# Small RNA expression pattern in multiply inbred lines and their hybrids of maize embryo

# Abstract

**Background:** Heterosis, also known as hybrid vigor or transgression, is the phenomenon wherein an F1 hybrid produced from crossing two cultivars of the same or different species shows superior phenotypes. Heterosis in maize has been found and applied in breeding for more than one hundred years. However, its underlying molecular mechanisms are still poorly understood. To investigate whether small RNAs (sRNAs) participate in the regulation of heterosis, we profiled the sRNA expression patterns in the germ seeds of five inbred lines and theirs three F1 hybrids using high-throughput sequencing technology.

**Results:** The sequencing result show maize sRNAs are enriched in 22-nt length. Nearly 90% of 22-nt small RNA dominated clusters (22-nt SRCs) are in repeat regions, which mainly originated from Gypsy and Copia in retrotransposon. About 25% differentially expressed SRCs exist between parents, and hybrid gain almost all differentially expressed 22-nt SRCs. Twenty-four-nt sRNA also enriched in maize, which showed great sequence diversity and overall low expression across the genome. More than half of 24-nt SRCs originate from repeat, and 80% of them come from DNA transposon. Nearly 30% of 24-nt SRCs located in genes or their flanking (±2kb) regions, especially in flanking regions of “lipid metabolic process” and “macromolecule modification” related genes. Several thousands 24-nt SRCs are paternal or maternal specific expressed, and hybrids gain only half of differentially expressed 24-nt SRCs. Hundreds of 24-nt SRCs show high parent (HP) or above high parent (AHP) expression pattern in different hybrids, and them mainly from Tourist, haT, and CACTA in DNA transposon and enrich nearby “tRNA aminoacylation for protein translation” related genes. Also some 21-nt SRCs show conserved expression pattern in low parent (LP) or below low parent (BLP). They were enriched in anti-sense region of some genes, which functions are about “oxidative phosphorylation” and “translation”. MicroRNA (miRNAs) have a global downregulated tendency in hybrids. Three miRNAs (zma-miR408-5p, zma-miR1432-5p and zma-miR528-5p) are significant (P-value < 0.001) downregulated in each hybrid, and this phenomenon may cause their target genes more stable and contribute to hybrid vigor.

**Conclusion:** Taken together, our results illustrated that sRNAs may contribute to heterosis at the very early stage of seed germination through repressing of retrotransposon activity, regulation gene activity at gene and genic flanking regions, and promotion some gene expressions by downregulated miRNAs.

**Keywords:** maize, heterosis, sRNA-seq, small RNA cluster, retrotransposon

# Introduction

Heterosis, also known as hybrid vigor or transgression, is the phenomenon wherein an F1 hybrid produced from crossing two cultivars of the same or different species shows superior phenotypes [[1](#_ENREF_1)]. The phenomenon of heterosis, depends on genetic variation between parents and altered genetic states in their offspring [[2](#_ENREF_2)]. Maize is a suitable model for research the genetic mechanism of heterosis, because it includes highly diversified phenotypes, allele, and the high-quality genome [[3](#_ENREF_3)]. In maize, heterosis has been found and applied in breeding for more than one hundred years [[4](#_ENREF_4)]. Many heterosis-related QTLs have been mapped on maize genome, and many potential heterosis-related genes have been reported [[5](#_ENREF_5)]. However, due to its so complex character, the molecular mechanism of heterosis in maize is still poorly understood.

Small RNAs (sRNAs) are mainly 20 ~ 24 nucleotides (nt) length RNA, which mainly functions are regulating gene expression and maintaining genome integrity [[6-9](#_ENREF_6)]. SRNAs mainly included two groups, small interfering RNAs (siRNAs), and microRNAs (miRNAs) in plant [[10](#_ENREF_10), [11](#_ENREF_11)]. The 24-nt siRNA are mainly derived from TEs and repeats region, and interact with ARGONAUTE4 (AGO4) and lead to gene silencing and/or RNA-directed DNA methylation (RdDM) at target loci [[12](#_ENREF_12)]. In *Arabidopsis*, 24-nt siRNA great downregulated in hybrids suggest an epigenetic contribution to hybrid vigor [[13](#_ENREF_13)]. RdDM is important in maintaining genome steady, due to 85% of maize genome is transposable elements [[3](#_ENREF_3), [12](#_ENREF_12)].

The magnitude of hybrid vigor in maize is relatively high and observed throughout its whole life cycle [[12](#_ENREF_12)]. Seed germination plays a pivotal role during the life cycle of plants, and many papers reported hybrid show obvious different from their parents in miRNAs and proteins [[5](#_ENREF_5), [14](#_ENREF_14), [15](#_ENREF_15)]. Barber *et al*. reported repeat associated sRNAs vary among parents and following hybridization in maize of B73 and Mol17 [[2](#_ENREF_2)]. Ding *et al*. reported miRNAs prefer downregulated in hybrid of maize seed germination [[5](#_ENREF_5)]. These works were only used one hybrid and their parents (inbred lines), and their finding about differences of sRNAs between hybrid and their parents is hardly to say a conserved mechanism related to heterosis.

In this study, to learn more about how sRNAs change in hybridization of maize and whether a conserved expression pattern exist between hybrid and their parents, we sequenced sRNAs from the germinated seed under imbibition 24 hours from three hybrids and their inbred lines parents, which are main varieties in China. We identified small RNA cluster (SRCs) across maize genome, and found SRCs of 22-nt sRNAs dominant expressed (22-nt SRCs) account for 40% and 24-nt SRCs account for 48%. Twenty-two-nt SRCs mainly originated from superfamily of Gypsy and Copia in retrotransposon. Twenty-four-nt SRCs not only enriched in DNA transposon region, but also 30% of them located in genes and their flanking (±2kb) regions, especially in flanking regions of “lipid metabolic process” and “macromolecule modification” related genes. Maize hybrids have more SRCs with expression levels of high parent (HP) and above high parent (AHP) than low parent (LP) and below low parent (BLP). SRCs with expression levels of AHP and HP tend to be 24-nt SRCs, and SRCs with expression levels of BLP and LP enrich in 21-nt SRCs. Between parents have 25% differentially expressed SRCs, and hybrid gain almost all in 22-nt SRCs and half in 24-nt SRCs. As for the miRNAs, compared with middle parents value (MPV), eighteen miRNA families were common downregulated in hybrids, but no one upregulated. Zma-miR528-5p, zma-miR408-5p, and zma-miR1432-5p were significant downregulated in hybrid, which may upregulate their target genes. Our results illustrated that sRNAs may contribute to heterosis at the very early stage of seed germination through repressing of retrotransposon activity, regulation gene activity in genic flanking regions, and promotion some gene expressions by downregulated miRNAs. Our data also can serve as a useful resource to better understand the sRNAs potential role in heterosis of maize.

# Result

## Characterization of sRNAs and SRCs in Maize

More than 120 million reads of sRNA from eight samples were sequenced, and about 80% reads were perfect match on the reference genome of maize B73 (Table S1). Maize have great abundance of sRNA in 22-nt length in embryo, even more than the abundance of 21-nt sRNA, which mainly consist of miRNAs (Figure S1D). In some varieties, the proportion of 22-nt sRNA is the most abundance class (Figure S1B/C). The length distributions of sRNA are different in each inbred lines and their hybrids, no conserved pattern between parent and hybrid (Figure S1A/B/C). Maize sRNAs also prefer have multiple loci on genome, nearly 50% of total reads have more than 20 loci (Figure S2A), and this ratio is less than 3% in *Arabidopsis thaliana* (Figure S2B).

Small RNA prefer to cluster distributed in genome, so sRNAs are grouped into small RNA clusters (SRCs), based on the mapped sRNAs. We found 84689 SRCs across maize genome, which content 75.41% of total selected reads and coverage 0.56% of genome.

We found the distribution of SRCs is similar with the distribution of genes (Figure 1A), and the SRC and gene number have positive correlation in each chromosome (Figure 1B). Classified the SRCs by their dominant length sRNA (e.g. If one cluster contain more than half 21-nt sRNA, then it classified as 21-nt SRCs), and found nearly 90% SRCs were 22/24-nt SRCs (Figure 1C). We distributed them on genome, and found 22-nt SRCs especially enriched in repeat region (Figure 1D) and 24-nt SRCs enriched in repeat, genic flank (upstream and downstream 2 kb of genes) (Figure 1D). Twenty-nt and Twenty-one-nt SRCs usually have short length and high expression in genome, and most of the miRNAs and ta-siRNAs are in these groups. Other SRCs usually contain abundance reads, because these SRCs commonly originate from highly expressed gene regions, such as tRNA and rRNA, maybe theirs degraded products.

## Twenty-two-nt SRCs enrichment in retrotransposon and twenty-four-nt SRCs enrichment in DNA transposon

Here, we annotated these defined SRCs by using annotation from maize transposon element consortium (MTEC), and found the relationship between SRCs and repeat sequences. First, repeat sequences were separated into two types: retrotransposon, and DNA transposon. Then, we counted each type repeat and SRCs in repeat region, and found the retrotransposon was account for more than 80% of total repeat, but only included less than 70% SRCs of repeat region, which represented DNA transposons were more activity than retrotransposons in originating SRCs (Figure 2A). We also checked the distribution of SRCs on retrotransposon and DNA transposon regions. Nearly 70% of 22-nt SRCs are located in retrotransposon regions, and more than 80% of 24-nt SRCs are in DNA transposon regions (Figure 2B). Twenty-four-nt sRNA were widely studied [[13](#_ENREF_13), [16](#_ENREF_16)], however, little known about potential functions of 22-nt sRNA in plant. We further focused on the 22-nt SRCs part, and found the 22-nt SRCs mainly from three superfamilies of retrotransposon: Gypsy, Copia and Unknown, which all have less than 5% activity (Activity = SRCs numbers / repeat members \* 100%) (Figure 2C). Twenty-two-nt SRC number and activity in each family were calculated, and we found gyma and ji families were the major contributors for 22-nt SRCs originating in Gypsy and Copia superfamily, respectively, but their activity vary from 3.1% to 14.6%. However, in some repeat families, such as raider, wiwa, and dagaf were had more than 30% activity in originating 22-nt SRCs (Figure 2D).

Above mentioned 24-nt SRCs enriched in genic flank (Figure 1D), so we further discussed these 22/24-nt SRCs relationship with genes, and found 24-nt SRCs great enriched in 1 kb region of genic flank (Figure 3A). These genes, which include 24-nt SRCs in their flank (1 kb) regions, were functional enrichment in “cellular process” and “metabolic process” of biological process by AgriGO [[17](#_ENREF_17)] (Figure 3B).

## Expression of SRCs among parents and hybrid

To show expression patterns of SRCs in hybrid, we classify the SRCs into 7 groups according to the expression among hybrid and parents (detail in Methods). We showed expression pattern of each type of SRCs and found most of SRCs were middle parent (MP) expression in hybrid (Figure 4). Another apparently expression pattern of SRCs in high parent (HP) and above high parent (AHP) were more than low parent (LP) and below low parent (BLP) (Figure 4). This expression pattern conserved in three cross combinations, especially in 24-nt SRCs region (Figure 4, Figure S3A/B). However, the 22-nt SRCs have different patterns in each cross combination (Figure 4, Figure S3A/B). We further compare three cross combinations, and found some SRCs had conserved expression patterns in different hybrids. An interesting result showed nearly all common HP and AHP SRCs were 24-nt SRCs, and common LP and BLP SRCs were enriched in 21-nt SRCs (Figure 5A). These conserved 24-nt SRCs in expression pattern of AHP were mainly (85.6%) from Tourist, haT, CACTA, and PIF families of DNA transposon, some of them enriched nearby “tRNA aminoacylation for protein translation” related genes (Figure 5B). These conserved 24-nt SRCs in expression pattern of HP were enriched in genic regions that functions about “metabolic process” and “gene expression”. As for these conserved 21-nt SRCs expression pattern of BLP were significant enriched in anti-sense region of some genes, which functions are about “oxidative phosphorylation” (*P*-value = 2.25e-11) and “translation” (*P*-value = 7.88e-8) (Figure 5C). Only twelve 21-nt SRCs have conserved expression pattern in LP, and them also exist in anti-sense regions of some important genes, such as photosystem II reaction center protein C, DNA-directed RNA polymerase family protein, lysine histidine transporter, and so on.

## Parents exist great difference in SRC regions and uneven inherited by hybrid

The percentage of differentially expressed SRCs between parents was very high, nearly a quarter of SRCs have more than two fold significant difference (*P*-value < 0.01) in expression (Figure 6A). We found 3258 SRCs had two fold difference, and common exist in 3 cross combinations (Figure 6B), but no enrichment or bias distribution of these SRCs.

To investigate the differential SRCs between parents and hybrid, reads per million (RPM) more than 5 was used to define expressed SRCs in each samples. We found many 22/24-nt SRCs are paternal or maternal specific expressed (Figure 7 A/B). About 21% of 22-nt highly expressed SRCs were different between paternal and maternal, and hybrid gain almost all different SRCs (Figure 7A). This pattern were also conserved in other two cross combinations (Figure S4 A/E). The expression tendency between hybrid and parents were not found in these 22-nt SRCs (Figure 7C). As for the 24-nt highly expressed SRCs, only a half of parents specific SRCs were gained by their hybrid (Figure 7B). The similar pattern also show in B and C combinations (Figure S4 B/F). More SRCs tendency downregulated in hybrids, especially at higher threshold of fold changed level (Figure S4C/D/G/H). However, this phenomenon were not show in A cross combination (Figure 7 C/D).

## Downregulated miRNAs in hybrids

Compare with middle parents values (MPVs), percentage of miRNA show a tendency of downregulated in hybrids. This pattern was conserved in three cross combinations (Figure 8A). Percentage of miRNA even lower than their both parents in B and C combinations.

To investigate which miRNAs changed in hybrids, we counted the RPM values of each miRNA family. Compared with MPV, the results showed that most of miRNA families were downregulated in hybrid (Figure 8B). B and C combinations were also show the similar pattern (Figure S5). As for 32 miRNA families, which expression more than five RPM, we found 18 of them were common downregulated in three combinations. Among of them, two miRNA families (zma-miR408-5p and zma-miR1432-5p) are two fold downregulated, and six miRNA families (zma-miR528-5p, zma-miR408-3p, zma-miR168-3p, zma-miR156-3p, zma-miR529-5p, and zma-miR398-5p) are 1.5 fold downregulated. As well as, none miRNA family was common upregulated in three combinations. The results of three cross combinations show a conserved pattern that miRNAs global downregulated in hybrids.

# Discussion

Maize have a big part of 22-nt sRNA originate from retrotransposon, and even higher expressed than 21-nt or 24-nt sRNA in some varieties (Figure S1A/B/C). We also compared maize sRNA with sRNAs of *Arabidopsis thaliana*, and found maize sRNA of 21 and 24-nt length decreased was due to dilution of increased 22-nt sRNAs (Data download from GEO, Figure S1D).

Twenty-two-nt sRNAs are great enriched in maize (Figure S1D), and 22-nt SRC is the second large (about 40%) group in maize (Figure 1C). However, the function of 22-nt SRC is poorly understood in maize. Many papers reported 24-nt siRNAs suppress DNA transposon expression through RNA directed DNA methylation (RdDM) pathway [[12](#_ENREF_12), [13](#_ENREF_13), [18](#_ENREF_18), [19](#_ENREF_19)]. In this study, we found nearly 70% 22-nt SRCs enriched in retrotransposon regions (Figure 2B). Does these 22-nt SRCs have relationship with retrotransposon, and have conserved pattern in hybrids? Above result showed the expression pattern of 22-nt SRCs do not show a bias of upregulated or downregulated in hybrid (Figure 7C, Figure S4C/G). But the kinds of expressed 22-nt SRCs were great abundance in hybrid, which inherited nearly all expressed 22-nt SRCs from parents (Figure 7A, Figure S4A/E). This result give a clue, hybrid gain diversity 22-nt SRCs may better control activity of retrotransposons, reduce waste energy and materials on transposon, maintain integrity of genome. All of these potential result will benefit to hybrids and show strong phenotype in hybrid vigor.

Previous study reported global downregulated of miRNAs in hybrid [[5](#_ENREF_5)]. In this study, we also found the similar pattern in three hybrids. To elusive the potential functions of conserved and downregulated miRNAs, the miRNA targets were predicted by psRobot online [[20](#_ENREF_20)] . *Zma-miR408-3p/528-5p* are predicted co-targeted on LAC2 genes (GRMZM2G367668 and GRMZM2G169033), which have conserved target site in some dicots (include *Sorghum bicolor*，*Oryza sativa*, *Brachypodium distachyon*), and show shorten elongation of root in knockout mutation (Figure 9A) [[21](#_ENREF_21)]. Downregulated miRNAs in each group were used to predict target genes. The functions of these target genes are mainly related with “regulation of gene expression”, “metabolic process” and “regulation of gene expression” in different groups (Figure 9B). These result show different miRNA families may regulate similar functional genes to get the same result. In conclusion, miRNAs downregulated may cause some genes increased and contributed to hybrid vigor.

# Methods

## Plant materials, RNA extraction and deep sequencing

To investigate whether small RNAs involved in heterosis, and a conserved expression pattern of sRNAs exist in different cross combinations, next generation deep sequencing technique was performed. In this study, three hybrid varieties of maize, which were elite hybrids and widely cultivated in China that exhibited high heterosis for grain yield, and theirs five parental inbred lines were utilized (all samples relations were show in the Figure 10). Fifteen seeds of each hybrid and parental line were surface-sterilized in 70% (v/v) ethanol for 10 min, and then rinsed several times with sterile distilled water [[22](#_ENREF_22)]. Then the seed were soaked in sterile distilled water and incubated at 26°C in the dark for 24 hours. The embryo tissues were peeled from the seeds, and each embryo tissue material from same variety was mixed for RNA extraction. Total RNA from the all samples was extracted with TRIzol reagent (Invitrogen, 15596-026) following the manufacturer’s protocol. Each sample of 15 ng RNA was used for small RNA sequencing libraries prepared and on Illumina Genome Analyzer II system by BGI [[23](#_ENREF_23)].

## Identification small RNA clusters and analysis difference

Maize genome were downloaded from website of Maizesequence, and filtered genes set (FGS) and TE Consortium repeat annotations (MTEC) were used [[3](#_ENREF_3)]. Small RNAs (sRNAs) were mapping on the reference genome by psRobot toolkits [[20](#_ENREF_20)]. Maize gene annotations of homolog were downloaded from Phytozome [[24](#_ENREF_24)]. Maize tRNA and rRNA were identified by RepeatMasker 4.01 ([www.repeatmasker.org](http://www.repeatmasker.org)), based on RepBase version 20130422 [[25](#_ENREF_25)]. Based on all the sRNA mapping result, high and consecutive expressed sRNA region were identified as sRNA clusters (SRCs) by the following criteria (Figure 11): (1) sequence need to perfect match reference genome; (2) loci (hits) < 500 mapped sequences for exclude high repeat sequences; (3) reads counts >1 sequences in order to exclude occasion transcription or sequencing noise; (4) sRNA consecutive detected in one cluster, any cluster distance >1 nt as independence cluster; (5) in each cluster contain loci number > 3, read per million (RPM) > 1, and reads per kilobase per million (RPKM) > 5.

Statistics of SRCs basic information were procedure by Perl script, and histogram were draw by Microsoft Excel 2013. mapped sRNAs compare with SRCs were done by intersectBed from BEDtools 2.16.2 [[26](#_ENREF_26)]. Distribution of SRCs and genes were showed by IGV 2.2.4 [[27](#_ENREF_27)].

SRCs have highly correlation with repeat, and two types of repeat annotations exist in maize genome: maize transposon element consortium (MTEC) and MIPS/Recat. We selected MTEC for further study, due to it has more TE members, detail annotation and classification than MIPS. As for the 1971471 annotated repeat in MTEC, we found some information is redundancy, such as same loci have duplication annotation, similar repeat members of same family overlap each other. After remove redundancy and merge overlapped repeat in same family, we got 1472084 repeat annotations. Compare annotations with SRCs by BEDtools [[26](#_ENREF_26)], and summarize results by Perl script.

For discussion expression pattern of SRCs between hybrid and parents, the SRCs were classified into 7 groups based on SRCs expression of hybrid relation with parents’, include high parent (HP), low parent (LP), average expression between parents called middle parent (MP), above high parent (AHP), below low parent (BLP), between HP and MP (BHM), and between HP and LP (BHL). The classifying procedure is in the Figure 12, and results are showed by Circos [[28](#_ENREF_28)].

## Calculation expression of miRNAs

All the sRNAs were mapped on known pre-miRNA of maize from miRBase 19.0 by local version of psRobot 1.01 [[20](#_ENREF_20), [29](#_ENREF_29)]. The percentage of miRNA in total sRNA was calculated by Perl scripts. All the expressions of miRNAs were normalized to reads per million (RPM), and grouped by family. The expression of each miRNA family were showed in heatmap and normalized to Z-score by R language. For the potential targets of miRNA, and conservation of target sites were predicted by psRobot online [[20](#_ENREF_20)]. Gene ontology (GO) enrichment analysis was proceeded by AgriGO and GOEAST [[17](#_ENREF_17), [30](#_ENREF_30)].

# Acknowledge

The authors thank Dr. Jihua Tang for kindly providing maize inbred line 87-1. The authors also thank Dr. Hua-Jun Wu and Meng Wang for the suggestion of data analysis.

# Reference

1. Shull GH, Gowen J: **Beginnings of the heterosis concept**. *Heterosis* 1952:14-48.

2. Barber WT, Zhang W, Win H, Varala KK, Dorweiler JE, Hudson ME, Moose SP: **Repeat associated small RNAs vary among parents and following hybridization in maize**. *Proc Natl Acad Sci USA* 2012, **109**(26):10444-10449.

3. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA *et al*: **The B73 maize genome: complexity, diversity, and dynamics**. *Science* 2009, **326**(5956):1112-1115.

4. Duvick DN: **Biotechnology in the 1930s: the development of hybrid maize**. *Nat Rev Genet* 2001, **2**(1):69-74.

5. Ding D, Wang Y, Han M, Fu Z, Li W, Liu Z, Hu Y, Tang J: **MicroRNA Transcriptomic Analysis of Heterosis during Maize Seed Germination**. *PLoS ONE* 2012, **7**(6):e39578.

6. Li S, Liu L, Zhuang X, Yu Y, Liu X, Cui X, Ji L, Pan Z, Cao X, Mo B *et al*: **MicroRNAs Inhibit the Translation of Target mRNAs on the Endoplasmic Reticulum in Arabidopsis**. *Cell* 2013, **153**(3):562-574.

7. Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AM, Strahl BD, Patel DJ, Jacobsen SE: **Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1**. *Nature* 2013, **498**(7454):385-389.

8. Li J, Liu Y-x, Han Y-p, Li Y-g, Guo M-z, Li W-b: **MicroRNA Primary Transcripts and Promoter Elements Analysis in Soybean (Glycine max L. Merril.)**. *Journal of Integrative Agriculture* 2013, **12**(9):1522-1529.

9. Liu YX, Han YP, Chang W, Zou QA, Guo MZ, Li WB: **Genomic Analysis of MicroRNA Promoters and Their Cis-Acting Elements in Soybean**. *Agricultural Sciences in China* 2010, **9**(11):1561-1570.

10. Liu YX, Wang M, Wang XJ: **Endogenous Small RNA Clusters in Plants**. *Genomics Proteomics Bioinformatics* 2014, **12**(2):64-71.

11. Liu YX, Chang W, Han YP, Zou Q, Guo MZ, Li WB: ***In silico* Detection of Novel MicroRNAs Genes in Soybean Genome**. *Agricultural Sciences in China* 2011, **10**(9):1336-1345.

12. Chen ZJ: **Genomic and epigenetic insights into the molecular bases of heterosis**. *Nat Rev Genet* 2013, **14**(7):471-482.

13. Groszmann M, Greaves IK, Albertyn ZI, Scofield GN, Peacock WJ, Dennis ES: **Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor**. *Proc Natl Acad Sci USA* 2011, **108**(6):2617-2622.

14. Fu Z, Jin X, Ding D, Li Y, Tang J: **Proteomic analysis of heterosis during maize seed germination**. *Proteomics* 2011, **11**(8):1462-1472.

15. He G, Chen B, Wang X, Li X, Li J, He H, Yang M, Lu L, Qi Y, Wang X *et al*: **Conservation and divergence of transcriptomic and epigenomic variation in maize hybrids**. *Genome Biol* 2013, **14**(6):R57.

16. Li Y, Varala K, Moose SP, Hudson ME: **The Inheritance Pattern of 24 nt siRNA Clusters in Arabidopsis Hybrids Is Influenced by Proximity to Transposable Elements**. *PLoS ONE* 2012, **7**(10):e47043.

17. Du Z, Zhou X, Ling Y, Zhang Z, Su Z: **agriGO: a GO analysis toolkit for the agricultural community**. *Nucleic Acids Res* 2010, **38**(Web Server issue):W64-70.

18. Zhang H, He X, Zhu J-K: **RNA-directed DNA methylation in plants: Where to start?** *RNA Biol* 2013, **10**(10):1593-1596.

19. Jia Y, Lisch DR, Ohtsu K, Scanlon MJ, Nettleton D, Schnable PS: **Loss of RNA-dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs**. *PLoS Genet* 2009, **5**(11):e1000737.

20. Wu HJ, Ma YK, Chen T, Wang M, Wang XJ: **PsRobot: a web-based plant small RNA meta-analysis toolbox**. *Nucleic Acids Res* 2012, **40**(Web Server issue):W22-28.

21. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M *et al*: **The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools**. *Nucleic Acids Res* 2012, **40**(D1):D1202-D1210.

22. Li D, Wang L, Liu X, Cui D, Chen T, Zhang H, Jiang C, Xu C, Li P, Li S *et al*: **Deep Sequencing of Maize Small RNAs Reveals a Diverse Set of MicroRNA in Dry and Imbibed Seeds**. *PLoS ONE* 2013, **8**(1):e55107.

23. Wang L, Liu H, Li D, Chen H: **Identification and characterization of maize microRNAs involved in the very early stage of seed germination**. *BMC Genomics* 2011, **12**:154.

24. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N: **Phytozome: a comparative platform for green plant genomics**. *Nucleic Acids Res* 2012, **40**(D1):D1178-D1186.

25. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J: **Repbase Update, a database of eukaryotic repetitive elements**. *Cytogenet Genome Res* 2005, **110**(1-4):462-467.

26. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features**. *Bioinformatics* 2010, **26**(6):841-842.

27. Thorvaldsdottir H, Robinson JT, Mesirov JP: **Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration**. *Brief Bioinform* 2013, **14**(2):178-192.

28. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics**. *Genome Res* 2009, **19**(9):1639-1645.

29. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: **miRBase: microRNA sequences, targets and gene nomenclature**. *Nucleic Acids Res* 2006, **34**(Database issue):D140-144.

30. Zheng Q, Wang XJ: **GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis**. *Nucleic Acids Res* 2008, **36**(Web Server issue):W358-363.

# Figure and legend

1. **Summary and genomic distribution of small RNA clusters (SRCs)**. **A**. SRCs and genes distribution on ten chromosomes. **B**. SRCs and genes number in each chromosome. **C**. SRCs classified by length of dominant (>50%) expressed sRNA. **D**. Distribution of 22/24-nt SRCs on genome (Genome is classified into repeat, exon, intron, genic flank and intergenic, and “genic flank” means upstream and downstream 2 kb of genes.).
2. **Distribution of SRCs on repeat regions**. **A**. Distribution of SRCs on two types of repeat. **B**. Distribution of each SRC group on two types of repeat. **C**. Distribution of 22-nt SRCs on major superfamilies of retrotransposon (Activity = SRCs numbers / repeat members \* 100%). **D**. Distribution of 22-nt SRCs on major families of retrotransposon.
3. **Twenty-four-nt SRCs enriched in genic flank**. **A**. Distribution of 22/24-nt SRCs on gene body and genic flank (All the gene body scale to 0 to 100%, and genic flank include 2 kb of upstream and downstream of gene. Each part (upstream, gene body, and downstream) split into 50 bins, and density is SRCs number in each bin). **B**. Enrichment GO terms of genes, which include 24-nt SRCs in their flank 1 kb region.
4. **Classified SRCs based on expression hybrid compared with parents**. SRCs are group into 7 classes, from high to low are above high parent (AHP), high parent (HP), between MP and HP (BMH), middle parent (MP), between MP and LP (BML), low parent (LP), and below low parent (BLP).
5. **Conserved SRCs in each expression pattern and their related genes**. **A**. Proportion of each length type SRCs in conserved expression pattern. AHP & HP almost all 24-nt SRCs, and BLP and LP enrichment in 21-nt SRCs. **B**. Enriched GO terms of AHP SRCs related genes. **C**. Enriched GO terms of BLP SRCs related genes.
6. **Two fold different SRCs between parents**. **A**. Percentage in each cross combination. **B**. Overlapping among 3 cross combinations.
7. **Expressed SRCs among parents and hybrid in A cross combination**. **A**. Twenty-two-nt SRCs of parent specific were good inherited in hybrid. **B**. Twenty-four-nt SRCs of parent specific were half inherited in hybrid. **C**. Differentially expressed 22-nt SRCs under different threshold. **D**. Differentially expressed 24-nt SRCs under different threshold.
8. **Expression of miRNAs**. **A**. Percentage of miRNA in each samples and tendency downregulated in hybrid. **B**. Expression of known miRNA families in A cross (normalize to Z-Score).
9. **Predicted target genes of downregulated miRNAs**. **A**. Target genes of three conserved downregulated miRNAs in hybrids. **B**. Common GO terms of target genes of downregulated miRNAs in each group
10. **Used varieties of maize**. A, B, C represent three cross combinations.
11. **Definition and analysis small RNA clusters (SRCs)**. **A**. Model of define SRCs; **B**. Pipeline for analysis SRCs.
12. Flow chart for classifying SRCs based on hybrid expression compared to parents’.

# Supplement

Table S1. Statistics of sRNA sequencing and mapping

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Variety | Total reads | Mapped reads %1 | Selected reads %2 |
| A1 | Change 7-2 | 16966913 | 74.32% | 68.78% |
| A2 | Zheng 58 | 15810440 | 77.77% | 62.57% |
| A3 | Zhengdan 958 | 18399480 | 76.55% | 64.27% |
| B2 | 9058 | 18590068 | 80.71% | 67.91% |
| B3 | Lingdan 20 | 17249792 | 76.38% | 64.96% |
| C1 | 87-1 | 9735651 | 83.62% | 51.77% |
| C2 | Zong 3 | 13150801 | 83.55% | 50.58% |
| C3 | Yuyu 22 | 11068163 | 84.00% | 44.63% |
| Merge |  | 120971308 | 79.02% | 66.90% |

1 Mapped reads % = Perfect matched reads / Total reads; 2 Selected reads % = Selected reads / Perfect matched reads, selected reads need reads count > 1, reads loci <500 and not mapped on tRNA and rRNA region.

1. **Length distribution of total sRNA reads**. **A/B/C**. Length distribution of sRNA in A/B/C group. **D**. Length distribution of sRNA in *Zea may* (Zma) and *Arabidopsis thaliana* (Ath).
2. **Distribution of loci number in each length sRNA**. **A**. *Zea mays*. **B**. *Arabidopsis thaliana*.
3. **Compared with parents, length type SRCs expression changed in hybrid**. **A**. B cross combination. **B**. C cross combination.
4. **SRCs between parents and hybrid in B and C cross combination**. **A-D**. B combination. **A/C**. Expressed 22-nt SRCs. **B/D**. Expressed 24-nt SRCs. **E-H**. C combination. **E/G**. Expressed 22-nt SRCs. **F/H**. Expressed 24-nt SRCs.
5. **Expression of known miRNA families in B and C cross combinations**.