Available online at www.sciencedirect.com
ScienceDirect
Agricultural Sciences in China
2011, 10(9): 1336-1345
September 2011

In silico Detection of Novel MicroRNAs Genes in Soybean Genome

LIU Yong-xin[1*], CHANG Wei[1*], HAN Ying-peng[1], ZOU Quan[2], GUO Mao-zu[3] and LI Wen-bin[1]

[1] *Key Laboratory of Soybean Biology, Chinese Ministry of Education/Soybean Research Institute, Northeast Agricultural University, Harbin 150030, P.R.China*
[2] *School of Information Science and Technology, Xiamen University, Xiamen 361005, P.R.China*
[3] *Department of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, P.R.China*

Abstract

The importance of microRNAs (miRNAs) at the post-transcriptional regulation level has recently been recognized in both animals and plants. In this study, the simple and most effective method of comparative genomic approach was used. First known plants miRNAs BLAST against the soybean genome, and then the located candidates were searched for novel miRNAs by RNA folding method in the vicinity (±400 nt) of the candidates. The results showed that a total of 521 novel soybean miRNA genes, including 236 mature miRNAs, were identified. All these mature miRNAs were grouped into 58 families, of which 21 of them were novel family in soybean. The upstream 2 000 nt of potential pre-miRNAs was used for promoter prediction, in order to investigate prediction of miRNAs and detect transcript unit and clustering. In this study, miRNA genes less tend to be present as clusters in soybean. Only 9 clusters, containing 21 miRNA genes (accounted for 4.0% of the total), were observed as part of polycistronic transcripts. Detailed analysis of sequence characteristics of novel miRNAs in soybean and all previous known plants miRNAs, were carried out. These results of this study provide a reference point for further study on miRNAs identification in plants, and improve the understanding of genome in soybean.

Key words: soybean genome, microRNA, *in sillico*, comparative genomic approach, promoters prediction, cluster

INTRODUCTION

Soybean (*Glycine max*) is one of the most important agricultural crops around the world. As soybean seeds contain a high percentage of protein (40%) and oil (20%), soybean is considered the most nutritious crop and its seeds are processed into a variety of food products, such as soybean milk and tofu. Recently, soybean has been adopted as a potential source of biofuels (Zhang *et al*. 2008). The widespread agricultural use of soybeans and the demand for increased production will require the development of cultivars with higher yields with improved resistance to environmental stressors. Thus, there are growing needs to modify the soybean to increase its yield and resistance to different environmental stresses. Although progress has been made, several critical problems still exist, such as the disease resistance, the need for increased yield and so on. Newly discovered microRNAs (miRNAs) may play important roles in soybean development, nitrogen fixation, and the response to abiotic and biotic stresses (Subramanian *et al*. 2008; Zhang *et al*. 2008; Wang *et al*. 2009; Liu *et al*. 2010).

miRNAs, approximately 21 nucleotides (nt) long, are a large class of small non-coding RNAs (ncRNAs), which play important roles in post-transcriptional gene expression control. miRNAs can negatively regulate

Received 25 October, 2010 Accepted 12 August, 2011
LIU Yong-xin, MSc, E-mail: woodcorpse@163.com; Correspondence LI Wen-bin, Professor, Tel: +86-451-55190778, Fax: +86-451-55103336, E-mail: wenbinli@neau.edu.cn
[*] These authors contributed equally to this work.

doi:10.1016/S1671-2927(11)60126-0

gene expression through recognizing completely and partially complementary sequences in target mRNAs for mRNAs cleavage or inhibition of mRNAs translation (Voinnet 2009). The known plant miRNAs have a remarkable penchant for targeting transcription factor gene families, particularly those with known or suspected roles in developmental patterning or cell differentiation (Rhoades *et al*. 2002; Bartel 2004). Several studies have demonstrated that many miRNAs regulate various plant development processes, including leaf morphogenesis and polarity (Palatnik *et al*. 2003; Juarez *et al*. 2004), floral differentiation and development (Aukerman and Sakai 2003; Chen 2004), root initiation and development (Guo *et al*. 2005), vascular development (Kim *et al*. 2005), and transition of plant growth from vegetative growth to reproductive growth (Lauter *et al*. 2005). At present, three methods were used for identifying miRNAs including the cloning, NextGen sequencing, and computational approach. However, cloning and NextGen sequencing are not highly efficient approaches to find miRNAs, so computational approach has been widely applied to identify miRNA in plant such as *Arabidopsis thaliana*, *Oryza sativa*, *Brassica napus*, *Medicago truncatula*, *Glycine max*, *Solanum tuberosum*, *Populus euphratic*, and in diverse plant species (Bonnet *et al*. 2004a; Adai *et al*. 2005; Zhang *et al*. 2006a, 2008; Archak and Nagaraju 2007; Xie *et al*. 2007; Zhou *et al*. 2008, 2009; Li *et al*. 2009; Yang *et al*. 2009).

Although more and more miRNAs have been identified in diverse animal and plant species during recent years, there are no reports about miRNAs prediction in soybean genome so far and fewer records existed in current several miRNAs databases (Griffiths-Jones *et al*. 2008; Zhang *et al*. 2009). To the early in 2010, the 85 soybean miRNAs registered in the miRBase 14.0 (http://www.mirbase.org) were obtained from data of four published studies (Zhang *et al*. 2006a, 2008; Subramanian *et al*. 2008; Wang *et al*. 2009). Among those studies, two of those studies found that the expression of a set of miRNAs may be regulated by various biotic stresses, for instance, in nodulation-regulation, and the rest of studies found the potential miRNAs from soybean expressed sequence tags (ESTs). Although some studies have been done in miRNAs of soybean, the number of soybean miRNAs is limited. For larger genome of soybean, which reaches 950 mega bases (Mb), a large number miRNAs need to be identified for systematic research. In this study, miRNAs in genome of soybean were identified through comparative genomic approach. To further investigate the potential miRNAs loci in genome of soybean, prediction of promoters were done in upstream of potential pre-miRNAs.

MATERIALS AND METHODS

Sequences data

To search potential miRNAs, all previous known 2 043 miRNAs and their precursor sequences from *A. thaliana*, *O. sativa*, *P. trichocarpa*, *G. max*, *M. truncatula*, *Sorghum bicolor*, *Zea mays*, and so on were obtained from the miRBase (Release 14.0, September 2009, http://www.mirbase.org, Griffiths-Jones *et al*. 2008). Most of these were identified or verified by experiments, and others were computationally predicted as their close homolog. The genomic sequence of soybean (Glyma 1.0) was downloaded from Phytozome (http://www.phytozome.net/, Schmutz *et al*. 2010).

In silico detection of novel soybean miRNA genes

The overview of *in silico* detection of novel soybean miRNA genes is presented in Fig. 1. First, repeat and similar sequences (less than five mismatches) of all known plant mature miRNAs were removed in order to exclude redundant; second, remaining known plant miRNAs were compared with the soybean genome using BLASTN (setting at wordlength 7 and *E*-value cutoff 10), which were used to identify potential miRNA candidates (Wang *et al*. 2005), and then, miRNA candidate sequences with less five mismatches variations against known miRNAs, were adopted; third, sequences of 400 nt upstream and 400 nt downstream of the potential miRNA candidates were extracted from soybean genome; fourth, secondary structure of pre-miRNAs was predicted on the web-based software RNAfold of Vienna RNA package ver. 1.8.2, which is publicly available at http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi (Gruber *et al*. 2008), and this software default parameters were used; finally, the following criteria were used for screening the candidates of potential miRNAs or

pre-miRNAs: (1) The miRNA and miRNA* are derived from opposite stem-arms such that they form a duplex with two nucleotide, 3′ overhangs; (2) base-pairing between the miRNA and the other arm of the hairpin, which includes the miRNA*, is extensive such that there are typically four or fewer mismatched miRNA bases; (3) asymmetric bulges are minimal in size (one or two bases) and frequency (typically one or less), especially within the miRNA/miRNA* duplex (Meyers *et al*. 2008). Through the above analysis, conservative novel miRNA genes of soybean can be identified.

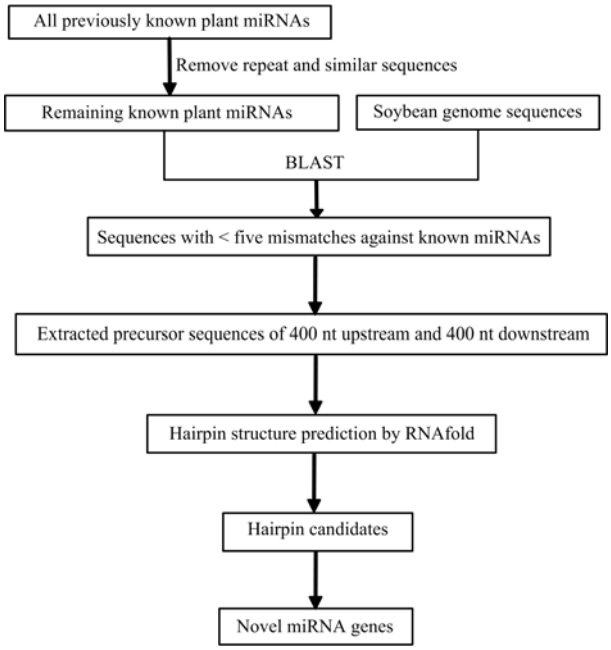

Fig. 1 Overview of *in silico* detection of novel conservative miRNA genes.

Analysis of characteristics of miRNA genes in the soybean genome

The minimal folding free energy (MFE) of the secondary structure for identified soybean miRNAs were obtained through web-based software RNAfold (Gruber *et al*. 2008). The adjusted minimal folding free energy (AMFE) was calculated by formula: MFE×Length of RNA sequence)/100, and the minimal folding free energy index (MFEI) was calculated by formula: AMFE×100/(G+C)% as described earlier by Zhang *et al*. (2006b). The sequences of upstream 2 000 nt of novel miRNAs were obtained for further promoter analysis. Sequences of TSS (transcript start site) and TATA-box were predicted using TSSP (http://mendel.cs.rhul.ac.uk/mendel.php?topic=gen). It is one of the best promoter prediction methods for plant miRNAs and the predictions were obtained at the default thresholds of TSSP. The program has been trained and tested on independent sets of well-known promoters (Shahmuradov *et al*. 2005). TSSP of Softberry was adopted for supplemental predictions (http://www.Softberry.com, Solovyev and Shahmuradov 2003). The distribution of novel miRNAs was analyzed by Mapchart 2.1 (Voorrips 2002). Furthermore, these novel miRNA genes were determined whether to cluster as previously described (Cui *et al*. 2009). These putative clusters used to combine the data of predicted promoters, were used for further analysis. All data on sequence characteristics in soybean and all previous known plant miRNAs, including the contents of A, C, G, U, A+U, and G+C, base composition of pre-miRNA sequences, and base composition at each position of mature miRNA sequences, were processed using Perl scripts.

RESULTS AND DISCUSSION

Novel miRNA genes in soybean genome

The comparative genomic approach to scan the soybean genome sequences was used for identifying novel conservative miRNA genes. As a result, a total of 521 novel miRNA genes including 236 mature miRNAs, located in 3′ or 5′ arm of the stem-loop hairpin structures, were identified. The detailed information of novel soybean miRNA genes including sequence of the pre-miRNAs, sequence of the mature miRNAs, minimal folding free energy (MFE), and positions in chromosomes are also listed in Appendix A.

In this study, a total of 571 miRNAs of soybean were identified, including 50 known and 521 novel miRNAs of soybean. For all of 85 known miRNAs of soybean, nearly 60% of them were covered, and the other part of them weren't shown in the result due to the following reason: (1) Some known miRNAs of soybean could not located in genome that might be due to the incompleteness of genome sequences, such as gma-MIR1525 and gma-MIR1527; (2) part of the hairpin

structural features of pre-miRNA of known miRNAs was not obvious and not according to the earlier described rule (Meyers *et al*. 2008). For instance, miRNA and miRNA* of gma-MIR2107 had seven mismatches (threshold≤five mismatches), the size of asymmetric bulges of miRNA and miRNA* of gma-MIR2119 was five bases (threshold≤two bases), and frequency of asymmetric bulges of miRNA and miRNA* of gma-MIR2108 was third (threshold≤1). The number of the miRNA genes was greater than the number of mature miRNAs because mature miRNAs could be derived from two or more pre-miRNAs (Ambros *et al*. 2003).

Sequence characteristics of novel soybean and all previous known plant pre-miRNAs and mature miRNAs

Detail analysis of sequence characteristics of pre-miRNAs and mature miRNA of soybean and all previous known plant, and a comparison between them, was made. All major sequence characteristics of the newly identified soybean and all previous known plant pre-miRNAs are summarized in Table 1. In this study, the average of the length, the content of adenine (A%) , the content of uracil (U%), ratio of guanine to cytosine (G/C), MFE, and MFEI of these newly identified soybean miRNA precursors were more than that of all known plants. On the contrary, the average of the content of cytosine (C%), the content of guanine (G%), ratio of uracil to adenine (U/A), and AMFE were less than that of all known plants. The distribution of the length of pre-miRNAs of soybean and all previous known plant is shown in Fig. 2. The distribution of length of novel pre-miRNA in soybean was different from that of all known plants. The significant three peaks (80-99, 220-239, and 360-379 nt), together contain 205 of the total novel pre-miRNAs (39.3%) (Fig. 2). The distribution of the length mature miRNAs and base composition at each position of soybean and all plants was also analyzed in this study (Figs. 3 and 4). The length of mature miRNAs in plant ranged from 17 to 25 nt, and 21 nt long mature miRNAs were the majority (Fig. 3). U were dominated in the 5′ end of mature miRNA sequences. The content of every kinds of base was nearly equal in all known plants, but the (A+U)% was higher in novel miRNAs of soybean (Fig. 4).

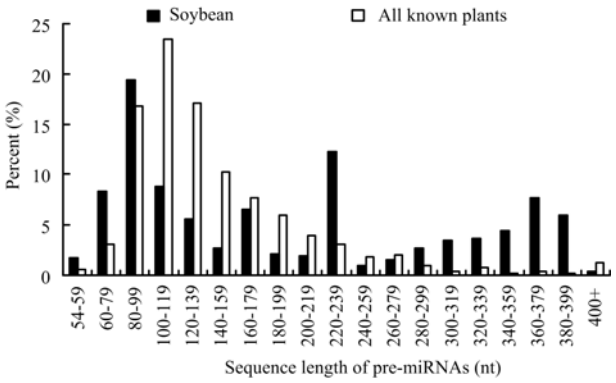

Fig. 2 The length distribution of novel soybean and all previously known plant pre-miRNAs.

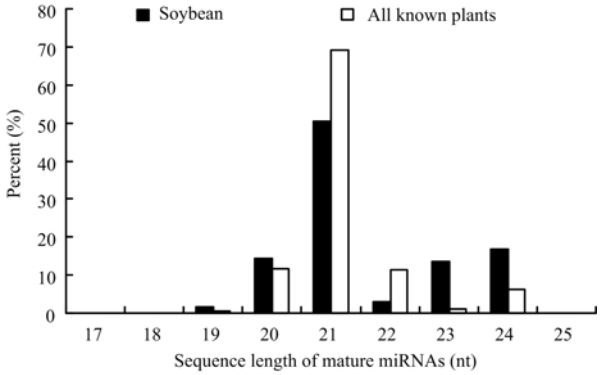

Fig. 3 The length distribution of novel soybean and all previously known plant mature miRNAs.

Sequence characteristics of pre-miRNAs and mature miRNAs in plants have been reported in recent studies (Zhang *et al*. 2006a, b, 2008). The secondary structure of pre-miRNA is less stable because of a higher (A+U) content, and easier to be processed into mature miRNA by the RISC complex (Zhang *et al*. 2006b). In this study, the (A+U) content of novel pre-miRNA of soybean was 11.47% higher than that of all known plants, which might be caused by the strictly threshold set in procedure of this study. The ratio of G/C and U/A in newly identified soybean pre-miRNAs was very similar to that of other previous plant pre-miRNAs, which suggested the ratio of G/C and U/A in plant pre-miRNA sequences may have a steady mean around 1.12 and 1.22, respectively (Table 1). The similar work has been carried out in animal (Zhou *et al*. 2009). Compared with them, the result demonstrated that base composition in animal and plant, pre-miRNA sequences had

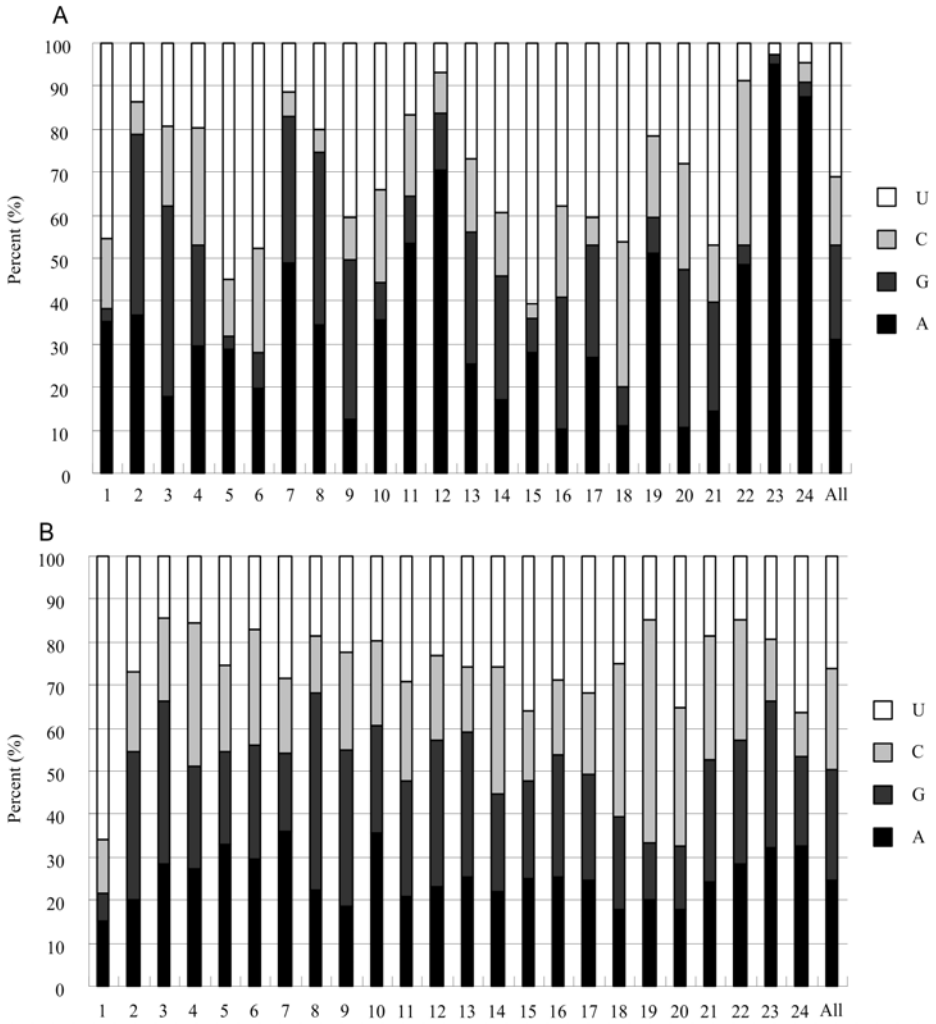

Fig. 4 The percentage distribution of base composition at each position in soybean (A) and all previously known plant (B) mature miRNA sequences.

similar pattern. However, more differences have existed; one of the most significant differences is the length of pre-miRNA between plants and animals.

The miRNA precursors have lower MFE than other non-coding RNAs and random sequence (Bonnet *et al*. 2004b). The MFE has been considered as one of important features to identify new miRNA genes in previous prediction methods (Wang *et al*. 2005). In this study, these newly identified soybean miRNA precursors had negative MFE, ranging from -6.1 to -261 kcal mol^{-1}, with an average of -67.57 kcal mol^{-1}, which was slightly lower than that of all known plants. However, MFE was strongly and positively correlated with their sequence length. To normalize the potential effect of sequence length on MFE and to better distinguish miRNAs from other RNAs, AMFE and MFEI were put forward to measure difference between miRNAs and other non-coding RNAs. The candidate RNA sequence is more likely to be a miRNA when the MFEI is greater than 0.85. In this study, MFEI of soybean pre-miRNAs ranged from 0.55 to 5.88, with an average of 1.19, higher than that of all known plants, and a 88.9% of novel soybean pre-miRNAs had MFEI value larger than 0.85.

As for the length of pre-miRNAs, the change is greater in plants than that in animals. In plants the change is widely ranged from 54 to 932 nt, with an average of (146±73) nt. On the contrary, all previous known animal miRNA precursors vary from 47 to 177 nt, with an average of (87±29) nt (Zhou *et al*. 2009). The

Table 1 Major sequence characteristics of the newly identified soybean and all previous known plant pre-miRNAs

Characteristic	Soybean					All known plants				
	Minimal	Maximum	Median	Mean	SD	Minimal	Maximum	Median	Mean	SD
Sequence length (nt)	43.00	400.00	169.00	197.68	110.68	54.00	932.00	125.00	146.12	73.01
A (%)	15.89	55.29	30.65	31.19	7.27	4.76	45.22	24.86	25.09	5.61
C (%)	1.72	34.62	16.92	16.12	5.55	3.36	40.78	21.36	21.68	5.26
G (%)	1.18	36.36	18.70	17.61	6.10	4.70	44.32	23.53	23.55	5.16
U (%)	17.65	58.46	35.00	35.03	5.70	10.14	50.34	30.00	29.66	5.61
G+C (%)	3.53	62.50	36.48	33.73	10.83	8.05	82.54	45.19	45.22	9.25
A+U (%)	37.50	96.47	63.47	66.22	10.89	17.46	91.95	54.79	54.75	9.28
G/C	0.40	8.00	1.13	1.13	0.41	0.30	6.67	1.10	1.12	0.28
U/A	0.42	2.18	1.11	1.17	0.27	0.47	3.29	1.18	1.22	0.31
MFE (kcal mol^{-1})	6.10	261.00	60.80	67.57	34.67	9.30	901.50	56.70	66.13	44.27
AMFE (kcal mol^{-1})	10.00	80.47	37.53	37.60	11.96	11.41	103.80	44.35	45.38	11.63
MFEI	0.55	5.88	1.08	1.19	0.48	0.36	2.45	0.98	1.02	0.26

length of extracted precursor sequences is also need to be discussed in experimental procedure, because the precursor of the plant is more complicated. In this study, the 400 nt of length of precursor sequences was determined. This length could cover the length of 100% of all known soybean pre-miRNA and 98.8% of all known plants pre-miRNA.

The distribution of length of all known pre-miRNA of plants tends to standardized normal distribution. However, that of novel pre-miRNA of soybean had its own pattern in some region (Fig. 2). This kind of abnormal distribution might be due to the species specificity of soybean, which is an palaeopolyploid (Schmutz *et al*. 2010).

In plants, previous study reported that U is the predominant nucleotide at the 5′ end of mature miRNA sequences, and it has been proposed that U may play an important role in miRNA biogenesis through recognition of targeted miRNA precursors by the RNA-induced silencing complex (RISC), and C is the dominant nucleotide at position 19 (Zhang *et al*. 2006a). In this study, the same results were observed in all known plants. However, some differences existed in novel mature miRNA sequences, for example, C was not the dominant nucleotide at position 19, and the content of bases was insignificantly correlated with position. This phenomenon was due to the following reasons: (1) In soybean genome, 75% of the genes present in multiple copies, which may affect base composition (Schmutz *et al*. 2010); (2) some non-conserved miRNA (only found in soybean at present) families have many members that is the majority of soybean, such as gma-MIR1517, gma-MIR1520, and gma-MIR1525.

Analysis of miRNA genes clusters and transcriptional units

For further analysis of miRNA genes clusters and transcriptional units, 521 novel miRNAs were mapped onto 20 chromosomes or linkages of soybean genome. The distribution characteristics of miRNA genes were visualized using Mapchart 2.1 as shown in Fig. 5 (Voorrips 2002). To determine expression of these miRNAs, the sequences of upstream 2 000 nt of pre-miRNA were used for promoter analysis by TSSP (Solovyev and Shahmuradov 2003; Shahmuradov *et al*. 2005). The result showed that a total of 81.8% putative miRNAs contained promoters in the region of upstream 2 000 nt (Appendix A). A small part of pri-miRNA promoter could not be predicted due to the following reasons: (1) The exact pri-miRNA was unknown, their promoter maybe be located in more than 2 kb upstream of the pre-miRNAs; (2) most promoter prediction softwares are based on the homology searching, hence, some pri-miRNA, having non-conserved promoters, could not be predicted; (3) some of pre-miRNAs, having multiple copies, might pseudogenes and do not have the real promoters (Liu *et al*. 2010).

Furthermore, these novel miRNA genes were determined whether to cluster as described previously (Cui *et al*. 2009). These potential clusters were analyzed, and the information of predicted promoter was combined (Table 2). Overall, only 9 clusters containing 21 miRNA genes were found. The promoters of all these clusters could be found by TSSP, which further enhanced the authenticity of the cluster. The members of most clusters were only from the same family. Animal MIR genes are often genomically clustered and de-

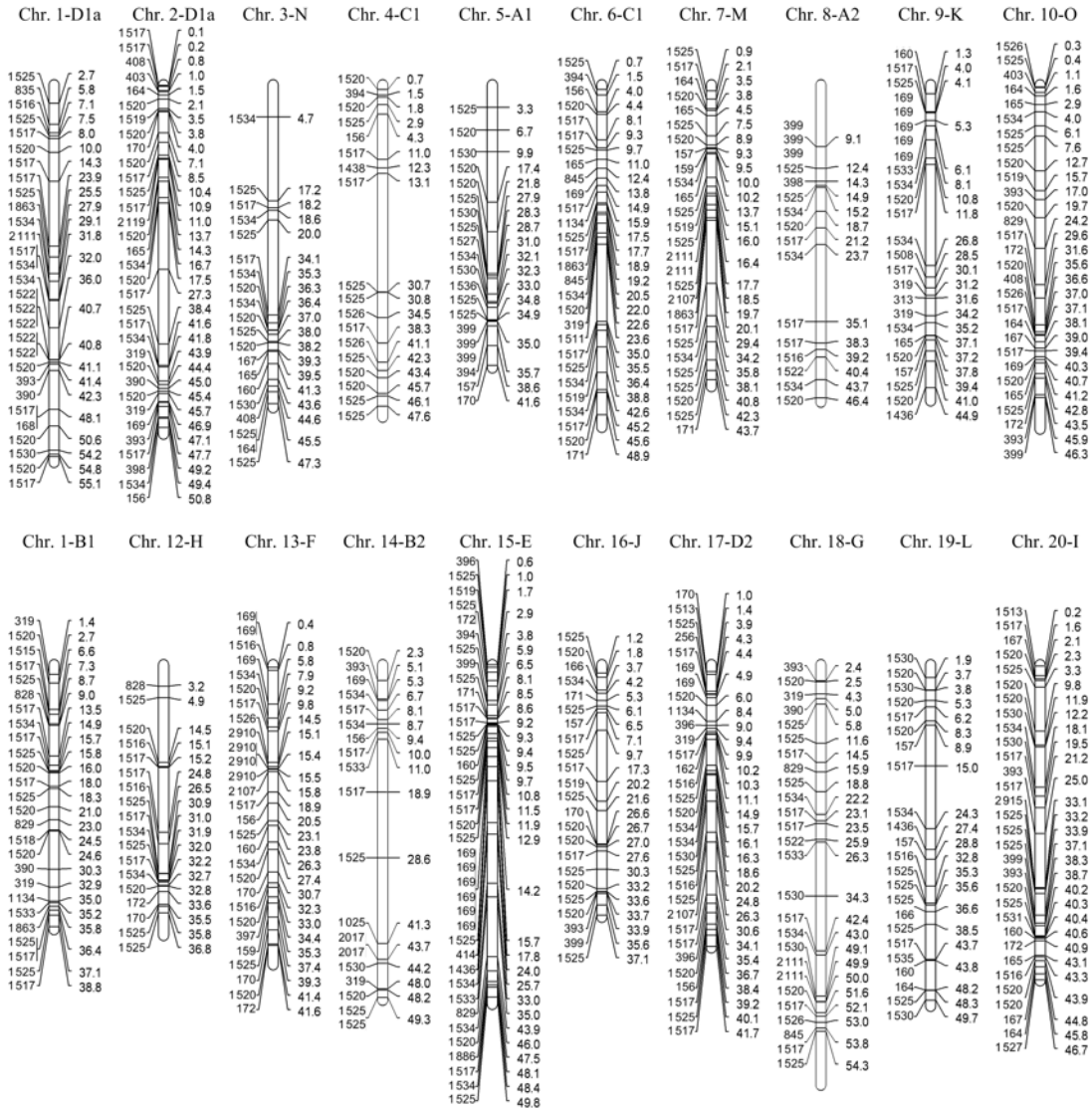

Fig. 5 The distribution characteristics of miRNA genes in soybean genome. The relative locations of 521 miRNA genes are shown across 20 soybean chromosomes. In every chromosome, the left number represents the miRNA name, the right number represents the position of miRNA (unit in Mb).

scribed as polycistronic RNAs (Kim 2005), for instance, 48% of miRNA genes appear as clusters within 10 kb in human (Altuvia *et al*. 2005). In contrast, plant MIR genes are rarely arranged in clustering (Voinnet 2009). For the known miRNAs of soybean, ten potential clusters were found in miRBase (Griffiths-Jones *et al*. 2008). Further analysis of these results showed that one cluster (included MIR396a and MIR396b) of these was not real, because the member of them distribute in different strand. Other nine clusters only contained one kind of pre-miRNA (MIR2107), which had 58 copies in soybean genome. The promoters of these nine potential clusters were also analyzed, however, few of them had promoters in the region of upstream 2 000 nt. The result suggested that these clusters might be pseudogenes and do not have the real promoters. In this study, potential cluster were identified by the following rules: (1) the inter-miRNA distance <10 000 nt; (2) all the members of potential cluster must distribute in the same strand; (3) the region of upstream 2 000 nt of all the clusters should have one more promoters predicted. Compared with rice, the soybean miRNAs are less tend to be present as cluster (Cui *et al*. 2009). In human, most miRNA clusters contain miRNAs from the different families and may consist of several different groups of non-homologous miRNAs (Yu *et al*. 2006).

Table 2 Features of predicted soybean miRNA clusters on the genome

No.	Members of cluster	Linkage	Chromosome	Strand	Location	Length (nt)	Number of promoter
1	miR399, miR399	A1	5	-	34 958 633-34 967 760	9 127	Multiple
2	miR399, miR399	A2	8	-	9 118 520-9 126 548	8 028	Multiple
3	miR169, miR169, miR169	K	9	+	5 282 127-5 287 983	5 856	Multiple
4	miR169, miR169	F	13	-	368 453-371 189	2 736	Multiple
5	miR1525, miR1525	E	15	-	9 345 019-9 351 157	6 138	Single
6	miR169, miR169	E	15	+	14 171 034-14 176 241	5 207	Single
7	miR169, miR169, miR169, miR169	E	15	+	14 191 199-14 202 558	11 359	Single
8	miR169, miR169	D2	17	-	4 861 835-4 864 274	2 439	Multiple
9	miR1535, miR160	L	19	-	43 786 823-43 796 035	9 212	Multiple

- and + mean negative and positive chains of reference geome.

However, in this study, the members of eight clusters of all were from the same family. The same phenomenon was also found in rice (Guddeti *et al*. 2005; Cui *et al*. 2009).

In addition, we studied the relationship between the number of miRNA genes and the density across 20 soybean chromosomes (Fig. 6). The result showed that the novel miRNAs of chromosome 15 had the highest density and number, and that of chromosome 8 had the lowest density and number. There was no sharp change in the distribution of number and density of miRNAs across 20 chromosomes in all studied cases.

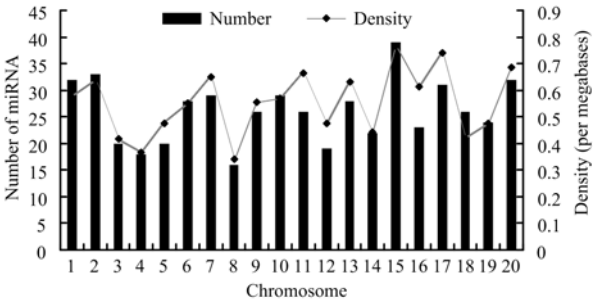

Fig. 6 The number of miRNA genes and the density across 20 chromosomes of soybean.

miRNA genes families in soybean genome

A total of 521 novel miRNAs were distributed into 58 families. The sizes of miRNA families in soybean are shown in Appendix B. Of them, 21 families were novel family. The members of novel family are shown in Fig. 7. The largest member number of novel family of soybean was miR399, which contained 11 members. This family had 78 known members, which distributed in 13 species of flora including *O. sativa*, *Vitis vinifera*, *M. truncatula*, and so on. It suggested that the family of miR399 is highly conserved family which is predicted to target mRNAs coding for a phosphatase transporter (Jones-Rhoades and Bartel 2004). Besides, miR408, 397, 497, and 157 are also highly conserved, their known members being over ten and across over four species. The families of miR403, 414, 845, and 828 are reported in two or three species. Other are found in only one species (Griffiths-Jones *et al*. 2008). The results of this study further proved the conservativeness of these families.

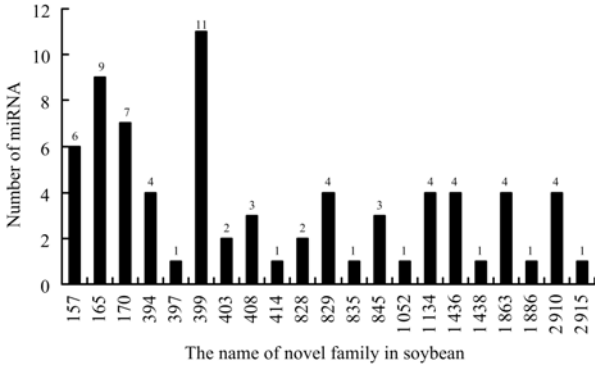

Fig. 7 Size of novel miRNA families in soybean.

CONCLUSION

Based on our improved comparative genomic integrated with promoters prediction approach, we identified 521 novel miRNA genes including 236 mature miRNAs. All these mature miRNAs were grouped into 58 families, of which 21 of them were novel family in soybean. All miRNA genes were distributed throughout 20 chromo-

somes in soybean genome. Furthermore, these miRNA genes were less tend to be present as clusters. All major sequence characteristics of novel soybean miRNAs including AMFE, MFEI, the contents of A, C, G, U, A+U, G+C, base composition at each position of pre-miRNA and mature miRNA sequences were examined in this study. Whereas, this paper does not include the experimental evidence, some comments on further studies will be needed for experimental proof in the future. The results of this study provide a reference point for further study on miRNAs identification in plants, and improve the understanding of genome in soybean.

Acknowledgements

This study was financially supported by the National High-Tech R&D Program of China (863 Program, 2006AA100104-4), the 948 Project, Ministry of Agriculture, China (2006-G5), the National Natural Science Foundation of China (30971810, 60932008), the National Basic Research Program of China (973 Program, 2009CB118400), the National Genetically Modified Organisms Breeding Major Projects of China (2009ZX08009-088B), the Postdoctoral Fund in Heilongjiang Province, China (LBH-Z07228), the Technology Project of Ministry of Education, Heilongjiang Province, China (11541025), and the Technology Project of Harbin, China (2009RFQXN085).

Appendix associated with this paper can be available on http://www.ChinaAgriSci.com/V2/En/appendix.htm

References

Adai A, Johnson C, Mlotshwa S, Archer-Evans S, Manocha V, Vance V, Sundaresan V. 2005. Computational prediction of miRNAs in *Arabidopsis thaliana*. *Genome Research*, **15**, 78-91.

Altuvia Y, Landgraf P, Lithwick G, Elefant N, Pfeffer S, Aravin A, Brownstein M J, Tuschl T, Margalit H. 2005. Clustering and conservation patterns of human microRNAs. *Nucleic Acids Research*, **33**, 2697-2706.

Ambros V, Bartel B, Bartel D P, Burge C B, Carrington J C, Chen X, Dreyfuss G, Eddy S R, Griffiths-Jones S, Marshall M, *et al*. 2003. A uniform system for microRNA annotation. *RNA*, **9**, 277-279.

Archak S, Nagaraju J. 2007. Computational prediction of rice (*Oryza sativa*) miRNA targets. *Genomics Proteomics Bioinformatics*, **5**, 196-206.

Aukerman M J, Sakai H. 2003. Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. *The Plant Cell*, **15**, 2730-2741.

Bartel D P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281-297.

Bonnet E, Wuyts J, Rouze P, van de Peer Y. 2004a. Detection of 91 potential conserved plant microRNAs in *Arabidopsis thaliana* and *Oryza sativa* identifies important target genes. *Proceedings of the National Academy of Sciences of the USA*, **101**, 11511-11516.

Bonnet E, Wuyts J, Rouze P, van de Peer Y. 2004b. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics*, **20**, 2911-2917.

Chen X. 2004. A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. *Science*, **303**, 2022-2025.

Cui X, Xu S M, Mu D S, Yang Z M. 2009. Genomic analysis of rice microRNA promoters and clusters. *Gene*, **431**, 61-66.

Griffiths-Jones S, Saini H K, van Dongen S, Enright A J. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Research*, **36**, D154-158.

Gruber A R, Lorenz R, Bernhart S H, Neubock R, Hofacker I L. 2008. The Vienna RNA websuite. *Nucleic Acids Research*, **36**, W70-W74.

Guddeti S, Zhang D C, Li A L, Leseberg C H, Kang H, Li X G, Zhai W X, Johns M A, Mao L. 2005. Molecular evolution of the rice miR395 gene family. *Cell Research*, **15**, 631-638.

Guo H S, Xie Q, Fei J F, Chua N H. 2005. MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for arabidopsis lateral root development. *The Plant Cell*, **17**, 1376-1386.

Jones-Rhoades M W, Bartel D P. 2004. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Molecular Cell*, **14**, 787-799.

Juarez M T, Kui J S, Thomas J, Heller B A, Timmermans M C. 2004. MicroRNA-mediated repression of rolled leaf1 specifies maize leaf polarity. *Nature*, **428**, 84-88.

Kim J, Jung J H, Reyes J L, Kim Y S, Kim S Y, Chung K S, Kim J A, Lee M, Lee Y, Narry Kim V, *et al*. 2005. MicroRNA-directed cleavage of ATHB15 mRNA regulates vascular development in *Arabidopsis* inflorescence stems. *Plant Journal*, **42**, 84-94.

Kim V N. 2005. MicroRNA biogenesis: coordinated cropping and dicing. *Nature Reviews Molecular Cell Biology*, **6**, 376-385.

Lauter N, Kampani A, Carlson S, Goebel M, Moose S P. 2005. microRNA172 down-regulates glossy15 to promote vegetative phase change in maize. *Proceedings of the National*

Academy of Sciences of the USA, **102**, 9412-9417.

Li B, Yin W, Xia X. 2009. Identification of microRNAs and their targets from *Populus euphratica*. *Biochemical and Biophysical Research Communications*, **388**, 272-277.

Liu Y X, Han Y P, Chang W, Zou Q, Guo M Z, Li W B. 2010. Genomic analysis of microRNA promoters and their cis-acting elements in soybean. *Agricultural Sciences in China*, **9**, 1561-1570.

Meyers B C, Axtell M J, Bartel B, Bartel D P, Baulcombe D, Bowman J L, Cao X, Carrington J C, Chen X, Green P J, *et al*. 2008. Criteria for annotation of plant microRNAs. *The Plant Cell*, **20**, 3186-3190.

Palatnik J F, Allen E, Wu X, Schommer C, Schwab R, Carrington J C, Weigel D. 2003. Control of leaf morphogenesis by microRNAs. *Nature*, **425**, 257-263.

Rhoades M W, Reinhart B J, Lim L P, Burge C B, Bartel B, Bartel D P. 2002. Prediction of plant microRNA targets. *Cell*, **110**, 513-520.

Schmutz J, Cannon S B, Schlueter J, Ma J, Mitros T, Nelson W, Hyten D L, Song Q, Thelen J J, Cheng J, *et al*. 2010. Genome sequence of the palaeopolyploid soybean. *Nature*, **463**, 178-183.

Shahmuradov I A, Solovyev V V, Gammerman A J. 2005. Plant promoter prediction with confidence estimation. *Nucleic Acids Research*, **33**, 1069-1076.

Solovyev V V, Shahmuradov I A. 2003. PromH: Promoters identification using orthologous genomic sequences. *Nucleic Acids Research*, **31**, 3540-3545.

Subramanian S, Fu Y, Sunkar R, Barbazuk W B, Zhu J K, Yu O. 2008. Novel and nodulation-regulated microRNAs in soybean roots. *BMC Genomics*, **9**, 160.

Voinnet O. 2009. Origin, biogenesis, and activity of plant microRNAs. *Cell*, **136**, 669-687.

Voorrips R E. 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. *The Journal of Heredity*, **93**, 77-78.

Wang X, Zhang J, Li F, Gu J, He T, Zhang X, Li Y. 2005. MicroRNA identification based on sequence and structure alignment. *Bioinformatics*, **21**, 3610-3614.

Wang Y, Li P, Cao X, Wang X, Zhang A, Li X. 2009. Identification and expression analysis of miRNAs from nitrogen-fixing soybean nodules. *Biochemical and Biophysical Research Communications*, **378**, 799-803.

Xie F L, Huang S Q, Guo K, Xiang A L, Zhu Y Y, Nie L, Yang Z M. 2007. Computational identification of novel microRNAs and targets in *Brassica napus*. *FEBS Letters*, **581**, 1464-1474.

Yang W Z, Liu X, Zhang J G, Feng J L, Li C, Chen J S. 2009. Prediction and validation of conservative microRNAs of *Solanum tuberosum* L. *Molecular Biology Reports*, **37**, 3081-3087.

Yu J, Wang F, Yang G H, Wang F L, Ma Y N, Du Z W, Zhang J W. 2006. Human microRNA clusters: genomic organization and expression profile in leukemia cell lines. *Biochemical and Biophysical Research Communications*, **349**, 59-68.

Zhang B H, Pan X P, Cannon C H, Cobb G P, Anderson T A. 2006a. Conservation and divergence of plant microRNA genes. *The Plant Journal*, **46**, 243-259.

Zhang B H, Pan X P, Cox S B, Cobb G P, Anderson T A. 2006b. Evidence that miRNAs are different from other RNAs. *Cellular and Molecular Life Sciences*, **63**, 246-254.

Zhang B H, Pan X P, Stellwag E J. 2008. Identification of soybean microRNAs and their targets. *Planta*, **229**, 161-182.

Zhang Z H, Yu J Y, Li D F, Zhang Z Y, Liu F X, Zhou X, Wang T, Ling Y, Su Z. 2010. PMRD: plant microRNA database. *Nucleic Acids Research*, **38**, D806-D813.

Zhou M, Wang Q, Sun J, Li X, Xu L, Yang H, Shi H, Ning S, Chen L, Li Y, *et al*. 2009. *In silico* detection and characteristics of novel microRNA genes in the *Equus caballus* genome using an integrated *ab initio* and comparative genomic approach. *Genomics*, **94**, 125-131.

Zhou Z S, Huang S Q, Yang Z M. 2008. Bioinformatic identification and expression analysis of new microRNAs from Medicago truncatula. *Biochemical and Biophysical Research Communications*, **374**, 538-542.

(Managing editor WANG Ning)