SHORT COMMUNICATION

Identification and verification of microRNA in wheat (*Triticum aestivum*)

Weibo Jin · Nannan Li · Bin Zhang · Fangli Wu · Wuju Li · Aiguang Guo · Zhiyong Deng

Abstract MicroRNAs (miRNAs) are small, endogenous RNAs that regulate gene expression in both plants and animals. A large number of miRNAs has been identified from various animals and model plant species such as *Arabidopsis thaliana* and rice (*Oryza sativa*); however, characteristics of wheat (*Triticum aestivum*) miRNAs are poorly understood. Here, computational identification of miRNAs from wheat EST sequences was preformed by using the in-house program GenomicSVM, a prediction model for miRNAs. This study resulted in the discovery of 79 miRNA candidates. Nine out of 22 miRNA representatives randomly selected from the 79 candidates were experimentally validated with Northern blotting, indicating that prediction accuracy is about 40%. For the 9 validated miRNAs, 59 wheat ESTs were predicted as their putative targets.

Keywords GenomicSVM · miRNA · Wheat ESTs

Weibo Jin and Nannan Li contributed equally to the work.

W. Jin · N. Li · B. Zhang · F. Wu · A. Guo (✉)
College of Life Science, Northwest A&F University,
Yangling, Shaanxi 712100, China
e-mail: guoaiguang@yahoo.com.cn

W. Li
Center of Computational Biology,
Beijing Institute of Basic Medical Sciences,
Beijing 100850, China
e-mail: liwj@nic.bmi.ac.cn

W. Jin · A. Guo
Key Laboratory for Molecular Biology of Agriculture,
Shaanxi 712100, China

Z. Deng
Wake Forest University School of Medicine,
Winston-Salem, NC 27157, USA

Introduction

MicroRNAs (miRNAs) have been found in a wide range of eukaryotes, such as *Arabidopsis thaliana*, Caenorhabditis elegans, mice and human beings (Bartel 2004). They constitute a large family of non-coding RNAs and play many important roles in post-transcriptional gene regulation by degrading target mRNAs or by repressing target gene translation in plants and animals (Bartel and Bartel 2003; Carrington and Ambros 2003; Hunter and Poethig 2003; Bartel 2004; Mallory and Vaucheret 2004). The lengths of mature miRNAs reported so far vary from 17 to 29 nucleotides (nt), and the majority of miRNAs are about 21–25 nts in length. Most miRNAs have the typical hairpin structure.

Computational prediction is an effective way for large-scale discovery of miRNAs from different plants and animals. However, due to the distribution of miRNAs predominantly in the intergenic regions or introns of coding genes, prediction for miRNAs through computational approaches is generally limited to the model organisms, for which we have detailed genome sequence information.

For many agriculturally important plants such as common wheat (*Triticum aestivum* L.), because of the poor knowledge about their genome sequences, it seems to be a challenge for us to predict their miRNAs by computational methods. However, recently several groups succeeded in the identification of miRNAs from wheat in spite of its undefined huge genome (16,000 Mb). For example, 16

miRNAs were identified from wheat ESTs sequences by computational methods (Zhang et al. 2005). Yao identified 16 miRNAs from wheat based on a large-scale sequencing approach combined with computational biology (Yao et al. 2007). These studies represented important initial works aiming at systematically discovering wheat miRNAs and also encouraged us to identify more wheat imRNAs based on computational strategy. In this paper, the systematic discovery of wheat miRNAs has been studied. First, we identified 38,463 pre-miRNA-like sequences from 613,015 wheat ESTs. Second, we applied the in-house program GenomicSVM to predict each pre-miRNA-like sequence and obtained 79 pre-miRNA candidates. Third, from the above 79 pre-miRNA candidates, we randomly selected 22 candidates for experimental validation. The results indicated that there are nine candidates verified by Northern blotting in leaves and roots of wheat seedlings. Therefore, the prediction accuracy is about 40%.

Methods

Screening for pre-miRNAs on wheat ESTs

In the prediction of wheat miRNAs, 613,015 wheat EST sequences were downloaded from the GenBank database (http://www.ncbi.nlm.nih.gov/, October 2005). The prediction included two major procedures: searching pre-miRNA-like sequences and identifying pre-miRNAs. First, the RNAfold program was used to find the hairpin structure from the EST sequences (Hofacker 2003). Moreover, the strict criteria were adopted in the identification of pre-miRNA-like sequences from hairpin structure sequences. The criteria are as follows: (1) the minimum length of pre-miRNA sequences is 60 nt; (2) the stem of the hairpin structure (included the GU wobble pairs) comprises at least 17 base pairings; (3) the maximum free energy of the pre-miRNA secondary structure is −15 kcal/mol; (4) multi-branch loops are not allowed in the predicted secondary structure; (5) the GC content of pre-miRNA is between 24 and 71%. These criteria ensure that the extracted sequences are similar to real pre-miRNAs according to the widely accepted characteristics. Second, the real pre-miRNAs were identified from a large number of pre-miRNA-like sequences using the program GenomicSVM that was developed by our Lab using the 'Support Vector Machine' model. The human pre-miRNA dataset was used to train the model. The model showed 86.3% sensitivity and 98.1% specificity, respectively, on the human test dataset, which contained 30 positive human pre-miRNAs and 1,000 negative pre-miRNAs. The model has been successfully applied to identify pre-miRNAs from *Arabidopsis thaliana*

(Jin et al. 2007). The model and related dataset can be accessed from the webpage (http://geneweb.go3.icpcn.com/genomicSVM/).

Validation of the predicted miRNA using the Northern blot method

Total RNA was extracted from leaves and roots of 4-week-old wheat seedlings using the trizol reagent (Invitrogen). A 100-μg amount of total RNA for each sample was resolved on a 15% polyacrylamide/1× TBE/8 M urea gel and subsequently transferred to a GeneScreen membrane (NIN). To generate high-specific probes, DNA oligonucleotides that are perfectly complementary to candidate miRNAs were end-labeled with $[\gamma\text{-}^{32}P]ATP$ by T4 polynucleotide kinase (New England Biolabs). Hybridization and washing procedures were performed as described (Sunkar and Zhu 2004). The membranes were briefly air dried and then exposed to a phosphor imager.

Prediction of miRNA targets

Previous study has shown that most known plant miRNAs bind to the protein-coding region of their mRNA targets with perfect or nearly perfect sequence complementarily and promote degradation for the target mRNAs in a way similar to RNA interference (Wang et al. 2004). In our prediction for miRNA targets, the number of allowed mismatchs at complementary sites between miRNA sequences and potential mRNA targets was less than four, and also no gap was allowed at the complementary sites.

Results

Computational identification of wheat miRNAs

The secondary structures of the wheat ESTs were predicted using the RNAfold program (Hofacker 2003). The first step was to search for potential hairpin structures in the wheat EST sequences, which yielded 551,129 qualified sequences. The second step was to further search miRNA precursor-like sequences according to their nucleotide composition and free energy of the secondary structure (more details are given in the methods section), which yielded 129,957 miRNA precursor-like sequences. The third step was to remove repeat elements and protein-coding sequences with the BLASTN and BLASTX programs, which yielded 5,834 sequences. The fourth step was to apply GenomicSVM to identify pre-miRNAs in the precursor-like sequences. Consequently, 79 pre-miRNA candidates were identified.

Experimental verification of novel candidate miRNAs

In order to validate the predicted miRNAs, we randomly selected 22 candidates from 79 predicted miRNAs for experimental verification. The expression patterns of these 22 candidates were examined in roots and leaves of 4-week-old wheat seedlings using the Northern blotting method. There are nine novel miRNA genes confirmed. Therefore, the prediction accuracy is about 40% (Fig. 1, Table 1, Figure S1 of Electronic Supplementary Material). During the experimental validation, we also included ten 21-mers that were rejected by GenomicSVM program criteria as negative controls, which were used to evaluate the specificity of Northern blot hybridization. As expected, none of them produced a positive signal. Among the nine confirmed miRNAs, three are close paralogs of known miRNAs; Ta-miR21 is a paralog of miR159, Ta-miR22 is a paralog of miR156, and Ta-miR24 is a paralog of miR164. The other six miRNAs are currently considered as novel wheat-specific miRNAs. Two of nine confirmed miRNAs were expressed in specific tissues: Ta-miR06 was only detected in wheat root, and Ta-miR22 was only found its expression in wheat leaf.

Evolution analysis and target prediction of nine validated miRNAs

To understand the evolutionary relationships among the validated wheat miRNAs and previously published plant miRNAs, phylogenetic analysis was performed for these miRNAs (Fig. 2). The results demonstrated that these nine
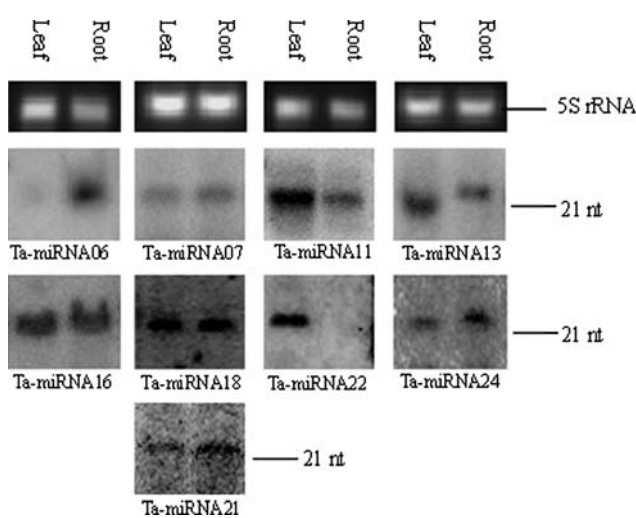

Fig. 1 Detecting miRNA with Northern blotting. RNA gel blots of total RNA isolated from different tissues were probed with labeled oligonucleotides. The 5S rRNA bands were visualized by ethidium bromide staining of agarose gels. Labeled RNA oligonucleotide was used as a size marker, and the position was indicated. *nt* nucleotides
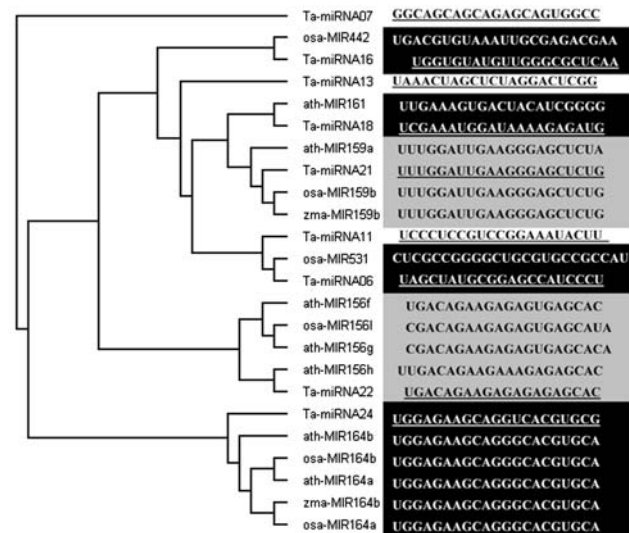

Fig. 2 Phylogenetic tree for the nine validated miRNAs and several published miRNAs. The tree was constructed using the neighbor-joining method. The sequences in the *black* and *grey boxes* indicate closely homologous (orthologs + paralogs) members. The *underlined sequences* are denoted as the new wheat miRNAs

Table 1 Nine wheat miRNAs were validated by Northern blotting

miRNA no.	miRNA sequence	EST accession no.	Precursor		Similar EST no.
			Start	End	
Ta-miR06	UAGCUAUGCGGAGCCAUCCCU	CA685345	1	146	251
Ta-miR07	GGCAGCAGCAGAGCAGUGGCC	CA738047	20	110	13
Ta-miR11	UCCCUCCGUCCGGAAAUACUU	CD886819	4	130	2
Ta-miR13	UAAACUAGCUCUAGGACUCGG	CD929884	4	129	35
Ta-miR16	UGGUGUAUGUUGGGCGCUCAA	BQ902027	135	214	1
Ta-miR18	UCGAAAUGGAUAAAAGAGAUG	CD910121	158	307	4
Ta-miR21	UUUGGAUUGAAGGGAGCUCUG	CA484819	419	485	1
Ta-miR22	UGACAGAAGAGAGAGAGCAC	BJ322639	411	478	2
Ta-miR24	UGGAGAAGCAGGUCACGUGCG	CD899685	601	690	3

Table 2 Predicted targets of the nine wheat miRNAs

miRNA no.	Target EST accession no.	Conserved miRNA	
Ta-miR06	BJ215281; BJ285284;	CDH1-D;	Wheat-specific
	BJ219790	Ubiquitin-conjugating enzyme;	
	BJ312089;	rRNA intron-encoded homing endonuclease;	
	BE515604; BJ231564;		
	BJ249472; BJ308743; BJ218544; BJ218512; BJ309229; BJ318141	Annotation not available	
Ta-miR07	CA696803	Ethylene-insensitive-3-like protein;	Wheat-specific
	TC214696	Q7YYP5;	
	CK215998	Annotation not available	
Ta-miR11	TC199924	ReMembR-H2 protein	Wheat-specific
	CD878370	Large Ala/Glu-rich protein	
	CD863874	CG14463-PA	
	BE516586	P0403C05.24	
	CK208408	H864 avr9 homolog F9L11.10	
	CD863281	P0038D11.16	
	TC219109	Q9XGV6	
	CK162908	Mla1	
	TC231203	Q6IGC5	
	CK211571	P0456A01.2	
	TC190166	Q7QPE7	
	CA730358	ZFP2	
	BQ744063; BI479952; CD896057; TC226920; TC193953; TC227706; BE585804; TC202580; CA611717; CK196395; TC190214; TC231385; TC220342; CA653869; BE419340; CN008790	Annotation not available	
Ta-miR13	TC190431; TC189250	Heat shock protein 80	Wheat-specific
Ta-miR16	TC190271; BJ259612;	Tubulin alpha chain	Wheat-specific
	CK152821; TC187691;	Annotation not available	
	TC188707; TC189447		
	BJ259548		
Ta-miR18	CD920072	Annotation not available	Wheat-specific
Ta-miR21	CA483944	Annotation not available	miR159
Ta-miR22	TC224574; TC210427	Squamosa promoter binding protein-like	miR156
	CK196549	PF6 protein	
	AL810223	Annotation not available	
Ta-miR24	TC224408; TC224409	OsNAC2 protein	miR164

confirmed new wheat miRNAs are distantly homologous to each other, but some of them form subgroups in the phylogenetic tree with miRNAs from other plants, revealing a relatively close evolutionary relationship with them. Six of nine novel miRNAs, namely Ta-miR06, Ta-miR07, Ta-miR11, Ta-miR13, Ta-miR16 and Ta-miR18, for which their homologous miRNAs were not found, showed an unrelated evolutionary relationship with other miRNAs.

We used these miRNAs to search for wheat ESTs in order to determine potential regulatory targets for the nine miRNAs. In this prediction, the screening criteria were that there were no more than four mismatches in the complimentary region between miRNAs and their miRNA targets. Any gap and G:U as well as other non-canonical pairs were not allowed in the complimentary region and were considered as mismatches according to the description in the

methods section. Finally, we found 59 target ESTs for the 9 miRNAs (Table 2). There were 34 target ESTs encoding functional proteins and another 25 target ESTs coding unknown proteins. The detail binding information between miRNAs and their targets is provided in Fig. S1 of the Electronic Supplementary Material.

Discussion

In this study, 79 wheat miRNA candidates were identified by mining the wheat EST data with our improved computational strategy and technology. Of these, three miRNAs, Ta-miR21, Ta-miR22 and Ta-miR24, were predicted through homology-search methods by Zhang et al. (2005), and another three miRNAs, Ta-miR41, Ta-miR59 and Ta-miR73, were identified with experimental RNomics methods by Yao et al. (2007).

We randomly selected 22 miRNAs from these 79 candidates for experimental validation and found 9 authentic pre-miRNAs. This experimental evaluation showed an accuracy of 40% for our prediction. However, the total number of genes authenticated by this method is underestimated for the following two main reasons. First, mature miRNA can be derived from either arm of a given stem loop, and many predicted pre-miRNAs fold equally well on either strand. In some cases, the non-transcribed strand actually adopts a fold with longer continuous helices than the transcribed strand does. As we tested one or two probes for each candidate, a false-negative result will be obtained whenever either the incorrect arm or a strand of a putative miRNA is tested. Second, a significant fraction of miRNAs

is likely to be expressed at extremely low levels or in a highly tissue-specific manner, and miRNAs in that fraction may not be amenable to confirmation by these means.

References

Bartel B, Bartel DP (2003) MicroRNAs: at the root of plant development? Plant Physiol 132:709–717
Bartel DP (2004) MicroRNAs: Genomics, biogenesis, mechanism, and function. Cell 116:281–297
Carrington JC, Ambros V (2003) Role of microRNAs in plant and animal development. Science 301:336–338
Hunter C, Poethig RS (2003) Missing links: miRNAs and plant development. Curr Opin Genet Dev 13:372–378
Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31(13):3429–3431
Mallory AC, Vaucheret H (2004) MicroRNAs: something important between the genes. Curr Opin Plant Biol 7:120–125
Sunkar R, Zhu JK (2004) Novel and stress regulated micro-RNAs and other small RNAs from Arabidopsis. Plant Cell 16:2001–2019
Wang XJ, Reyes JL, Chua NH, Gaasterland T (2004) Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. Genome Biol 5:R65
Weibo J, Dong K, Xiaoming Y, Wuju L, Aiguang G (2007) Prediction and analysis of novel miRNA in *Arabidopsis thaliana*. Acta Biophysica Sinica 23:389–396
Yao YY, Guo GG, Ni ZF, Sunkar R, Du JK, Zhu JK, Sun QX (2007) Cloning and characterization of microRNAs from wheat (*Triticum aestivum* L.). Genome Biol 8:R96
Zhang BH, Pan XP, Wang QL, Cobb GP, Anderson TA (2005) Identification and characterization of new plant microRNAs using EST analysis. Cell Res 15:336–360