Marine Biology

September 2011 Vol.56 No.27: 2916–2921
doi: 10.1007/s11434-011-4678-6

Characterization of small RNAs from *Ulva prolifera* by high-throughput sequencing and bioinformatics analysis

HUANG AiYou[1,2], WANG GuangCe[1*], HE LinWen[1,2], NIU JianFeng[1] & ZHANG BaoYu[1]

[1] *Institute of Oceanology, the Chinese Academy of Sciences (IOCAS), Qingdao 266071, China;*
[2] *Graduate University of the Chinese Academy of Sciences, Beijing 100049, China*

Received April 8, 2011; accepted July 4, 2011

The seaweed *Ulva prolifera* exists in 2 different states; attached to rocks or free-floating. However, there is little difference between the structures of the 2 states. *U. prolifera* thalli show significant differences in growth rate, with the attached thalli growing at a normal rate and free-floating thalli growing at a much faster rate. This raised the possibility that the growth of the two states may be regulated differently. miRNAs are important post-transcriptional regulators. In higher plants and animals, miRNAs have been extensively studied but they have been rarely studied in algae. To identify *U. prolifera* miRNAs and to investigate their possible roles in proliferation, we constructed and sequenced small RNA (sRNA) libraries from *U. prolifera.* Our results show that *U. prolifera* has a complex small RNA system that might play important roles in various processes.

***Ulva prolifera*, small RNA, microRNA, Solexa, bioinformatics**

Citation: Huang A Y, Wang G C, He L W, et al. Characterization of small RNAs from *Ulva prolifera* by high-throughput sequencing and bioinformatics analysis. Chinese Sci Bull, 2011, 56: 2916–2921, doi: 10.1007/s11434-011-4678-6

Ulva prolifera is a green alga (phylum: *Chlorophyta*; class: *Chlorophyceae*; order: *Ulvales*; family: *Ulvaceae*; genus: *Ulva*) [1,2]. *U. prolifera* thalli exist in 2 states; attached and free-floating. Attached *U. prolifera* lives on intertidal rocks while the floating state drifts in seawater. Although there is little difference in structure, the two states of *U. prolifera* thalli have significantly different growth rates; the attached *U. prolifera* has a normal growth rate while floating *U. prolifera* has a much higher rate [3] and can cause the "green tide" phenomenon [4], which can be extremely harmful to mariculture and the tourist industry. This prompted us to hypothesize that *U. prolifera* might possess specific regulators of gene expression, such as miRNAs, that regulate its development and proliferation.

microRNAs (miRNAs) are endogenous ~21 nt non-coding small RNAs that play important regulatory roles at the post-transcriptional level [5–7]. It is believed that miRNAs exist extensively in animals, plants and viruses with high levels of conservation in each kingdom [8,9]. miRNAs can regulate gene expression by targeting untranslated regions (UTR) or coding sequences (CDS) of mRNAs for cleavage (this is most often the case in plants) or by translational repression (most often the case in animals) [8,10]. miRNAs are involved in various processes, such as developmental patterning, cell differentiation and proliferation, and stress resistance [8].

The traditional methods of miRNA identification include experimental approaches, such as cDNA cloning, and bioinformatics approaches, such as computational prediction. The greatest weakness of traditional small scale cloning is that it is inefficient at discovering miRNAs expressed at low levels [11,12]. The computational prediction method solves this problem well; however, it introduces a high level of false-positive results, which require extensive experimental validation [13,14]. The high-throughput sequencing and bioinformatics analysis approach is a developed recently method that efficiently identifies low-level expressed or non-conserved miRNAs in various organisms. Very briefly,

*Corresponding author (email: gcwang@qdio.ac.cn) fragments of 18–28 nt are gel-purified from total RNA and ligated to a 5′-adaptor and a 3′-adaptor and then RT-PCR-amplified. RT-PCR products are then sequenced directly using high-throughput sequencing technology [15–18]. There have been many miRNA studies in higher plants and animals; however relatively little information is available regarding algae miRNAs. To identify miRNAs and their probable roles in *U. prolifera* development and proliferation, we constructed and sequenced a small RNA library from *U. prolifera* thalli. This initial study provided insights into the expression of small silencing RNAs in *U. prolifera*.

1 Materials and methods

1.1 Sample preparation

U. prolifera thalli were collected from the seashore of Qingdao, immersed in distilled water, patted dry with filter paper, instantly frozen in liquid nitrogen and then stored at −80°C to await RNA extraction.

1.2 Total RNA extraction

U. prolifera thalli were homogenized in liquid nitrogen and total RNA was extracted using the Trizol (Invitrogen, Carlsbad, CA, USA) method according to the manufacturer's protocol.

1.3 Small RNA library construction and sequencing

Total RNA was separated on denaturing polyacrylamide gels and fragments of 18–28 nt were recovered by gel purification. The small RNAs were ligated to a 3′ adaptor and a 5′ adaptor with T_4 RNA ligase, RT-PCR amplified and then sequenced using a Solexa 1G Genome Analyzer according to the manufacturer's protocols.

1.4 Initial processing of sequences

Total reads were processed as follows (Figure 1): (1) After removing adaptors and filtering low quality reads, the length distribution of the remaining reads was analyzed; (2) tags smaller than 18 nt were removed; (3) the clean reads were mapped on to EST sequences of *U. prolifera* and onto the genomes of *Chlamydomonas reinhardtii*, *Arabidopsis thaliana* and *Phaeodactylum tricornutum*, using SOAP; (4) non-coding RNA (rRNA, tRNA, snRNA and snoRNA) degradation fragments were identified by comparing the

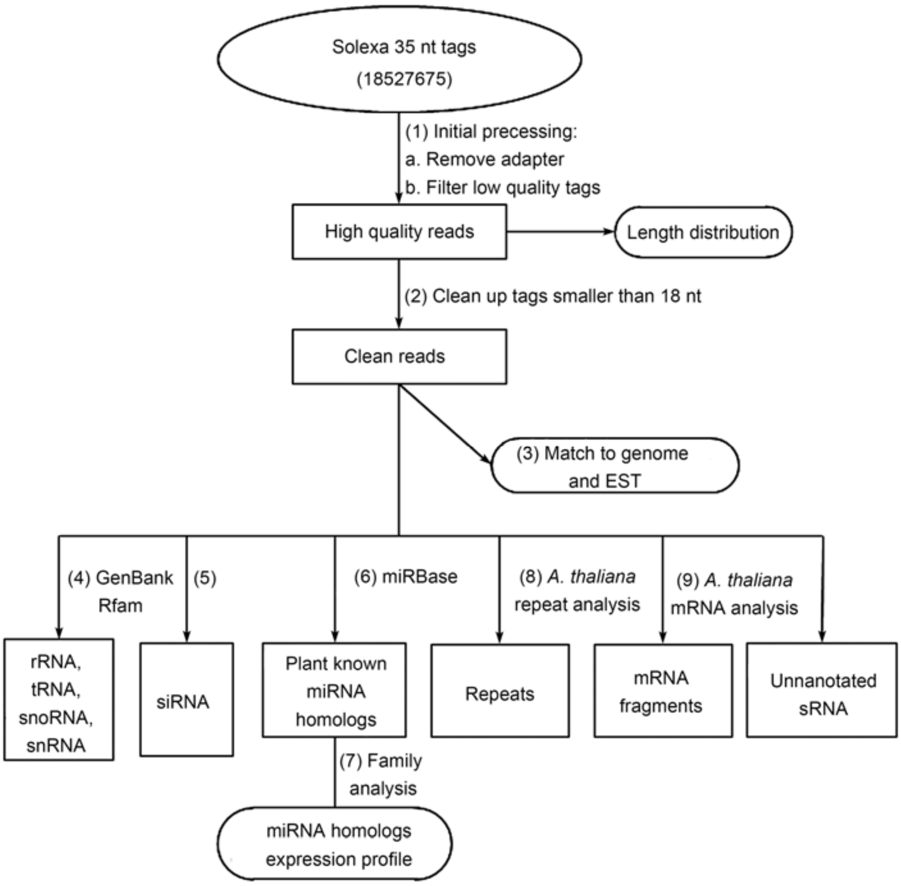
Solexa 35 nt tags
(18527675)
(1) Initial precessing:
a. Remove adapter
b. Filter low quality tags
High quality reads
Length distribution
(2) Clean up tags smaller than 18 nt
Clean reads
(3) Match to genome and EST
(4) GenBank Rfam
(5)
(6) miRBase
(8) *A. thaliana* repeat analysis
(9) *A. thaliana* mRNA analysis
rRNA, tRNA, snoRNA, snRNA
siRNA
Plant known miRNA homologs
Repeats
mRNA fragments
Unnanotated sRNA
(7) Family analysis
miRNA homologs expression profile

Figure 1 Flow chart of initial processing of reads.

clean reads with sequences of non-coding RNAs available in Rfam (http://www.sanger.ac.uk/software/Rfam) and in the GenBank non-coding RNA database (http://www.ncbi.nlm.nih.gov/) using Blastn; (5) all the clean reads were compared with each other to identify potential siRNAs (two perfectly complementary sRNAs with a 2 nt 3′-end overhang); (6) the clean reads were compared with all plant miRNAs in miRBase (http://www.mirbase.org/) to identify homologs of known miRNAs (sequences that exhibited homology with other known miRNAs with ≤2 mismatches); (7) the expression pattern of known miRNA homologs was analyzed; (8) the clean reads were compared with *A. thaliana* repeat sequences; and (9) the clean reads were compared with *A. thaliana* mRNA sequences. All clean reads were then annotated according to their similarities with the small RNA categories mentioned above. If a small RNA was mapped to more than one category, the following rule was adopted: rRNA, tRNA, snRNA and snoRNA (in which GenBank > Rfam) > homologs of known miRNA > siRNA [19]. That is to say, if one sRNA was mapped to rRNA and miRNA, it was annotated as rRNA not homologs of miRNA. Sequences that mapped to none of the small RNA categories were named as non-annotated small RNAs.

1.5 Novel miRNA prediction

We used *A. thaliana* genome sequences and *U. prolifera* EST sequences as a reference dataset to identify potential novel miRNA precursors in *U. prolifera*. Non-annotated small RNAs were mapped onto the reference dataset and complementarity with 300 nt of upstream and downstream flanking sequences was determined to examine if hairpin secondary structures of a potential pre-miRNA could be formed. Criteria adopted were as follows: (1) ≤20 perfect matches on reference sequences; (2) minimum free energy (MFE) of precursors should be ≤−18 kcal/mol, checked by Mfold; (3) the complementary between miRNA and miRNA* should ≥16 bp and ≤4 bulges or asymmetries; (4) the space between miRNA and miRNA* should be smaller than 300 nt; and (5) mature sequence length should range from 18 to 25 nt.

1.6 Novel miRNA target prediction

We used the miRanda method to identify potential targets of novel miRNAs. The parameters adopted were as follows: match score ≥90, target duplex free energy ≤−20 kcal/mol and scaling parameter = 2. Then we checked the duplex manually according to criteria suggested for plant miRNA target prediction: (1) no more than 4 mismatches between the small RNA and the target in positions 2–21; (2) no adjacent mismatches in positions 2–12; (3) no more than 2 adjacent mismatches in all positions; (4) no mismatches in positions 10–11; and (5) no more than 2.5 mismatches in positions 1–12 (counting from the 5′-end of the miRNAs and G-U bases as 0.5 mismatches). In addition, the MFE of the miRNA/target duplex should ≥ 74% of their perfect complement.

1.7 Experimental verification of the expression of known miRNA homologs

RT-PCR was used to detect the expression of the 10 most abundant homologs of known plant miRNAs. miRNA cDNAs were synthesized using the NCode™ VILO™ miRNA cDNA Synthesis Kit (Invitrogen) according to the manufacturer's protocol. The cDNA was used for PCR amplification of known plant miRNAs homologs. The sense primers were designed according to each known plant miRNA homolog and the antisense primer was the universal primer supplied in the cDNA synthesis kit.

2 Results

2.1 A diverse set of small RNAs in *U. prolifera*

To identify *U. prolifera* miRNAs, we constructed and sequenced a small RNA library from *U. prolifera* thalli. After removing adaptors and filtering low quality sequences, we obtained abundant small RNAs ranging from 10 to 43 nt, with most sequences being 25 nt (Figure 2). After removing sequences smaller than 18 nt, we obtained 16052321 total reads, representing 3460994 unique clean reads (Table 1). Of these small RNAs, 1338946 (8.34%) total reads and

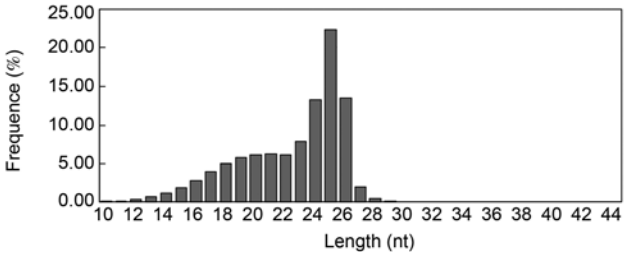

Figure 2 Length distributions of unique small RNAs in *U. prolifera*.

Table 1 Summary of data cleaning and length distribution of tags produced by *U. prolifera* small RNA sequencing

Type	Count	Percent (%)
Total reads	18527675	
High quality	18197647	100
3′ adaptor null	55009	0.30
Insert null	13993	0.08
5′ adaptor contaminants	77201	0.42
Smaller than 18 nt	1998394	10.98
PolyA	729	0.00
Clean reads	16052321	88.21

72493 (2.09%) unique reads had at least one perfect match with *U. prolifera* EST sequences or *C. reinhardtii*, *A. thaliana* or *P. tricornutum* genomic sequences, although most hits were to *U. prolifera* EST sequences (Figure 3, Table 2). Some small RNAs were mapped onto *A. thaliana* repeat sequences and a greater number of these sequences were mapped onto LSU-rRNA_Ath:1and SSU-rRNA_Ath:1 than onto other repeat sequences (Table S1).

All clean reads were annotated according to their similarities with noncoding RNA, such as rRNA, known miRNAs, and siRNA with a priority rule of rRNA, tRNA, snRNA and snoRNA (in which GenBank > Rfam) > homologs of known miRNA > siRNA. Due to the lack of *U. prolifera* genomic sequences, approximately 92% of small RNAs were not annotated. Of the small RNAs annotated, rRNAs accounted for 3.64%; siRNAs accounted for 1.96% and tRNAs accounted for 1.90%. Other kinds of small RNAs represented only a small percentage (Figure 4, Table 3).

2.2 Identification of homologs of known miRNAs and evolutionary analysis

To identify homologs of known miRNAs, we compared all clean reads with all plant miRNAs in miRBase. A total of 20657 sequences were identified as homologs of known miRNAs. Allowing up to 2 mismatches, they were classified into 285 miRNA families (Table S2), whose expression levels ranged from one to more than 10000 copies (Figure 5). For example, miR1520 was represented in the library by 15489 copies, while miR2678 and miR1874 were also present with over 10000 copies.

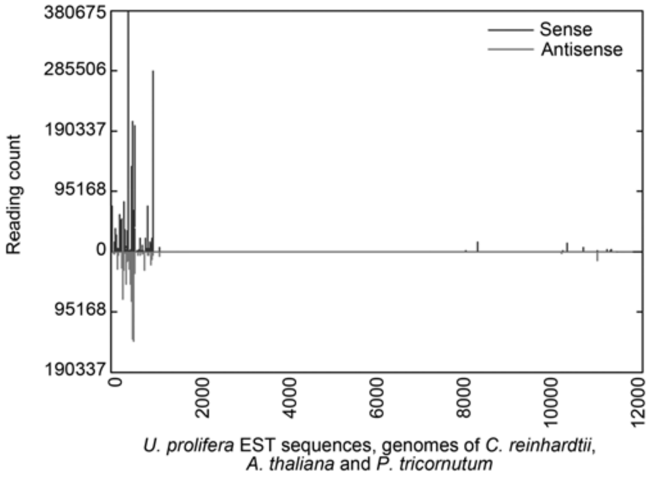

Figure 3 Small RNA (redundant sequences) distribution across *U. prolifera* EST sequences and genomes of *C. reinhardtii*, *A. thaliana* and *P. tricornutum*.

Table 2 Mapping statistics of U. prolifera small RNA sequences on U. prolifera EST sequences and genome sequences of three model organisms

	Unique sRNAs	Percent (%)	Total sRNAs	Percent (%)
Total sRNAs	3460994	100	16052321	100
Mapped to EST & genome	72493	2.09	1338946	8.34

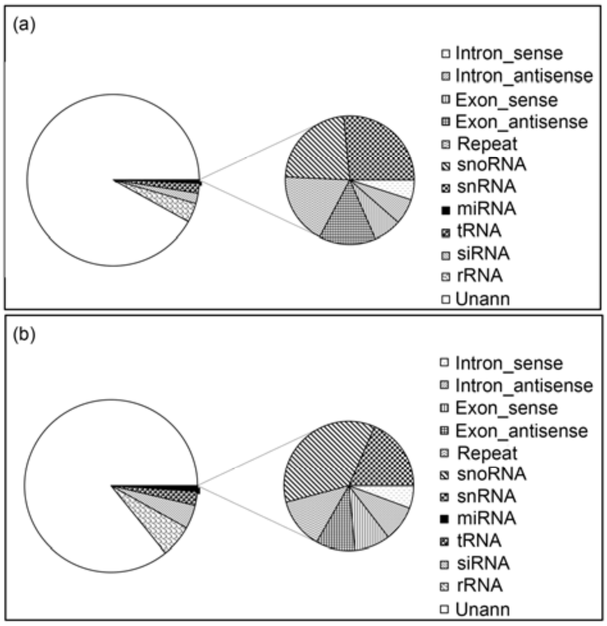

Figure 4 Categorization of *U. prolifera* small RNAs. (a) Categorization of unique small RNAs in *U. prolifera*. (b) Categorization of total small RNAs in *U. prolifera*.

Table 3 Annotation of *U. prolifera* small RNAs

Category	Unique sRNAs	Percent (%)	Total sRNAs	Percent (%)
Intron_sense	75	0.00	240	0.00
Intron_antisense	97	0.00	379	0.00
Exon_antisense	105	0.00	377	0.00
Exon_sense	213	0.01	415	0.00
repeat	270	0.01	536	0.00
snoRNA	341	0.01	1515	0.01
snRNA	398	0.01	784	0.00
miRNA	20657	0.60	170863	1.06
tRNA	65680	1.90	420839	2.62
siRNA	67866	1.96	710322	4.43
rRNA	125862	3.64	974059	6.07
Unann	3179430	91.86	13771992	85.79
Total	3460994	100	16052321	100

2.3 Identification of novel miRNAs

The hairpin structure of pre-miRNA made the prediction of miRNAs feasible. As the genome of *U. prolifera* has not been sequenced, we used *U. prolifera* EST sequences and *A. thaliana* genomic sequences to identify potential novel miRNAs. However, no novel miRNAs were indicated.

2.4 Experimental validation of homologs of known miRNAs

We used RT-PCR to detect the expression of the ten most abundant homologs of known miRNAs. Six out of 10 were validated by PCR amplification. The expected sizes (60–80 bp) of PCR products were amplified (Figure 6), increasing confidence in their expression. Some larger PCR products might result from precursor RNAs.

3 Discussion

A total of 16052321 total sequence reads, representing 3460994 unique reads were obtained (Table 1), indicating a complex small RNA regulation mechanism in *U. prolifera*. Only 8.34% total and 2.09% unique sequences were mapped to the genomes of 3 model organisms, *C. reinhardtii*, *A. thaliana* and *P. tricornutum*, implying a different small RNA regulation mechanism from other organisms. Some small RNAs were mapped to repeat sequences of *A. thaliana*, which prompted us to think that small RNAs might play a role in silencing repeat sequences in *U. prolifera*. Small RNA annotation showed that siRNA represented the second highest read frequency of all annotated small RNAs, indicating a complex siRNA silencing mechanism in *U. prolifera*.

Small RNAs of 25 nt in size were the most abundant in *U. prolifera*, which is different from *A. thaliana* where the most abundant small RNA length was 24 nt [12,15]. *C. reinhardtii* and *Physcomitrella patens* are also reported to lack enrichment of 24 nt small RNAs [11,20,21]. The length of small RNAs depended on the enzymes that take part in their processing. For example, DCL2 derives small RNAs of 24 nt, while DCL1 derives small RNAs of 21 nt [20]. The lack of enrichment of 24 nt small RNAs in *U. prolifera*, *A. thaliana* and *P. patens* indicated that these lower photosynthetic organisms might have different RNA processing enzymes from *A. thaliana.* In fact, we designed primers according to protein sequences of Dicer-like enzymes from *A. thaliana*, *C. reinhardtii* and *P. patens*. We PCR amplified products of expected sizes from *U. prolifera* cDNA, which share some sequence similarities with Dicer-like enzymes in other organisms. This indicated that *U. prolifera* does have enzymes necessary for small RNA processing. More experiments are needed to determine their exact roles in *U. prolifera*.

There are many homologs of plant miRNAs in *U. prolifera*, indicating that the miRNA regulation pattern of *U. prolifera* might share some similarity to higher plants. Among these homologs, some were particularly highly expressed. For example, mir1520, mir2678 and mir1874 were present in the small RNA library at more than 10000 copies. They might play important roles in *U. prolifera*. To validate the expression of these homologs, we tested the 10 most abundant homologs using RT-PCR. Among them, 6 were verified, providing strong evidence for their expression. We tried to recover and sequence the products of the predicted size; however we failed, probably due to their low abundance and their small size, which lower the efficiency of gel-purification. We asked what role they played in *U. prolifera*. Due to the lack of genome sequences, we studied functions of these miRNA families in other organisms. However, they have only recently been identified and their functions are not clear. mir1520, mir2678 and miR1531 were identified in roots of *Soybean* and probably play a role in their mutualism with legume bacteria [22,23]. We

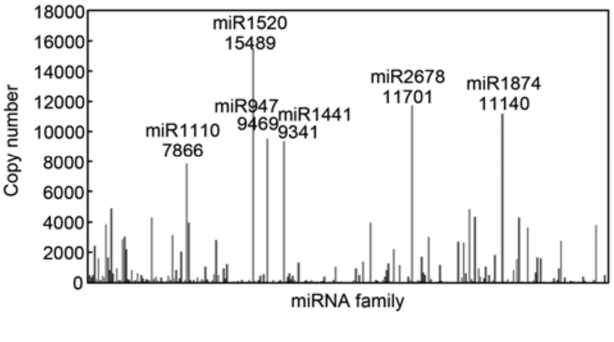

Figure 5 Expression levels of homologs of known miRNAs in *U. prolifera*.

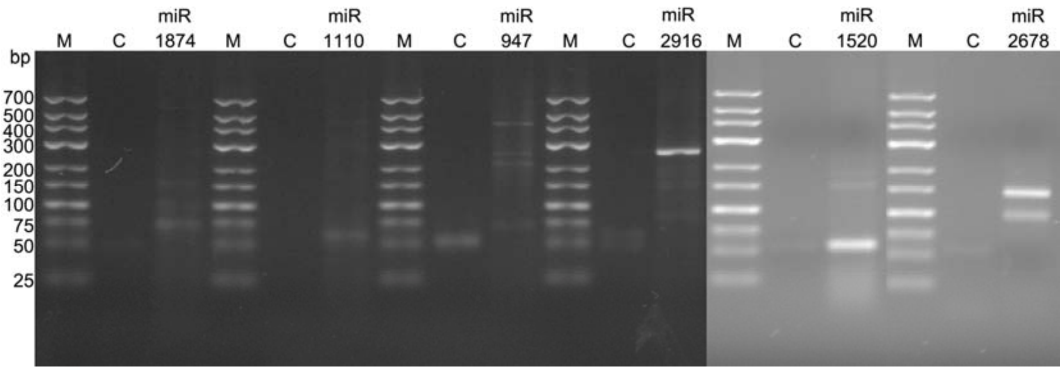

Figure 6 Experimental validation of some miRNAs homologs. M, marker; C, control.

proposed that they might regulate the rapid proliferation of roots. miR947 plays a role in the drought response in *Pinus taeda* [24], and the photosystem I in desiccated *U. prolifera* was still functional when the absolute water content of the thalli was close to 20%. In addition, photosynthesis activity was completely recovered after rehydration for 30 min, indicating that *U. prolifera* is resistant to desiccation [25]. We proposed that miR947 might play a role in the desiccation response in *U. prolifera*, although more experiments are needed to test this hypothesis. Ninety-two percent of small RNAs were not annotated, probably due to the distant evolutionary relationship of *U. prolifera* with other organisms, and thus *U. prolifera* has novel miRNAs that lack identifiable homologs in other organisms. As the genome of *U. prolifera* has not been sequenced, we could not verify the precursors of these miRNA homologs, nor could we identify novel miRNAs. To overcome this drawback, we used EST sequences to identify novel miRNAs, yet none were identified, probably due to the limit of our EST dataset, which contains only approximately one thousand sequences.

miRNA target prediction identified four potential complementary sites; however, more experiments are needed to verify what roles they play in *U. prolifera* and if they take part in regulation of the rapid proliferation of floating *U. prolifera*.

This work was supported by the Scientific and Technical Supporting Programs of China (2008BAC49B01) and the National Natural Science Foundation of China (30830015).

1 Zeng C K, Zhang D R, Zhang J P, et al. Flora of Chinese Economic Seaweeds. Beijing: Science Press, 1962

2 Hayden H, Blomster J, Maggs C, et al. Linnaeus was right all along: Ulva and enteromorpha are not distinct genera. Eur J Phycol, 2003, 38: 277–294

3 Gao S, Chen X Y, Yi Q Q, et al. A strategy for the proliferation of ulva prolifera, main causative species of green tides, with formation of sporangia by fragmentation. PLoS One, 2010, 5: e8571

4 Fletcher R. The Occurrence of “Green Tides” – A review. In: Schramm W, Nienhuis P H, eds. Ecological Studies, Vol. 123. Marine Benthic Vegetation. Recent Changes and the Effects of Eutrophication. Berlin: Springer, 1996. 7–43

5 Lau N, Lim L, Weinstein E, et al. An abundant class of tiny rnas with probable regulatory roles in caenorhabditis elegans. Science, 2001, 294: 858

6 Lee R C, Ambros V. An extensive class of small rnas in caenorhabditis elegans. Science, 2001, 294: 862–864

7 Lagos-quintana M, Rauhut R, Lendeckel W, et al. Identification of novel genes coding for small expressed rnas. Science, 2001, 294: 853

8 Bartel D P. Micrornas genomics, biogenesis, mechanism, and function. Cell, 2004, 116: 281–297

9 Reinhart B J, Weinstein E G, Rhoades M W, et al. Micrornas in plants. United State: Patent Aplication Publication. 2004

10 Millar A, Waterhouse P. Plant and animal micrornas: Similarities and differences. Funct Integrat Genomics, 2005, 5: 129–135

11 Arazi T, Talmor-neiman M, Stav R, et al. Cloning and characterization of micro-rnas from moss. Plant J, 2005, 43: 837–848

12 Xie Z, Allen E, Fahlgren N, et al. Expression of *Arabidopsis* miRNA genes. Plant Physiol, 2005, 138: 2145

13 Jones-rhoades M W, Bartel D P. Computational identification of plant micrornas and their targets, including a stress-induced mirna. Mol Cell, 2004, 14: 787–799

14 Li L, Xu J, Yang D, et al. Computational approaches for microrna studies: A review. Mamm Genome, 2010, 21: 1–12

15 Rajagopalan R, Vaucheret H, Trejo J, et al. A diverse and evolutionarily fluid set of micrornas in *Arabidopsis thaliana*. Genes Dev, 2006, 20: 3407

16 Fahlgren N, Howell M, Kasschau K, et al. High-throughput sequencing of *Arabidopsis* microRNAs: Evidence for frequent birth and death of mirna genes. PLoS One, 2007, 2: e219

17 Sunkar R, Zhou X, Zheng Y, et al. Identification of novel and candidate mirnas in rice by high throughput sequencing. BMC Plant Biol, 2008, 8: 25

18 Wei B, Cai T, Zhang R, et al. Novel micrornas uncovered by deep sequencing of small rna transcriptomes in bread wheat (*Triticum aestivum* L.) and *Brachypodium distachyon* (L.) beauv. Funct Integrat Genomics, 2009, 9: 499–511

19 Calabrese J, Seila A, Yeo G, et al. Rna sequence analysis defines dicer’s role in mouse embryonic stem cells. Proc Natl Acad Sci USA, 2007, 104: 18097

20 Zhao T, Li G, Mi S, et al. A complex system of small rnas in the unicellular green alga *Chlamydomonas reinhardtii*. Genes Dev, 2007, 21: 1190

21 Axtell M, Snyder J, Bartel D. Common functions for diverse small RNAs of land plants. Plant Cell Online, 2007, 19: 1750

22 Subramanian S, Fu Y, Sunkar R, et al. Novel and nodulation-regulated micrornas in soybean roots. BMC Genomics, 2008, 9: 160

23 Lelandais-briere C, Naya L, Sallet E, et al. Genome-wide medicago truncatula small rna analysis revealed novel micrornas and isoforms differentially regulated in roots and nodules. Plant Cell Online, 2009, 21: 2780

24 Lu S, Sun Y H, Amerson H, et al. Micrornas in loblolly pine (*Pinus taeda* L.) and their association with fusiform rust gall development. Plant J, 2007, 51: 1077–1098

25 Gao S, Wang G C, Niu J F, et al. Psi-driven cyclic electron flow allows intertidal macro-algae *Ulva* sp. (Chlorophyta) to survive in desiccated conditions. Plant Cell Physiol, 2011, 10.1093/pcp/PCR038

Supporting Information

Table S1 Statistics of small RNAs mapped to repeat sequences of *A. thaliana.*

Table S2 Expression profile of homologs of known miRNA.

The supporting information is available online at csb.scichina.com and www.springerlink.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.