Non-coding RNA for ZM401, a Pollen-specific Gene of Zea mays

DAI Xiao-Yan, YU Jing-Juan*, ZHAO Qian, ZHU Deng-Yun, AO Guang-Ming
(State Key Laboratory for Agricultural Biotechnology, College of Biology, China Agricultural University, Beijing 100094, China)

Abstract: In our previous study, a cDNA library to poly(A) RNA isolated from mature pollen of Zea mays L. was constructed. One cDNA fragment, designated ZM401 (Z. mays), was obtained from maize mature pollen cDNA library by differential screening and cold-plaque screening method. It was specifically or preferentially expressed in mature pollen. According to the ZM401 cDNA fragment sequence, full length of ZM401 cDNA was generated by 5' and 3' RACE methods. ZM401 cDNA was 1,149 bp in length. Within the full-length cDNA sequence, a clear open reading frame (ORF) was undetectable by OMEGA 2.0 and DNAMAN softwares. The longest potential ORF was 269 nucleotides (791-1,060), coding 89 AA, had a poor consensus sequence for translation initiation. But it had a poly(A) tail. All these results suggested that ZM401 was one of a growing number of non-coding mRNA-like RNA transcripts that exerted their cellular functions directly as RNA. RT-PCR and Northern blotting analyses showed the ZM401 was transcribed from tetrad stage of microspore development and increased in concentration up to mature pollen, which suggested that ZM401 belongs to late gene in pollen development. Two transcripts of the ZM401 gene were detected by Northern blotting analysis (1.2 kb, 2.0 kb). Southern hybridisation showed that the ZM401 in corn was present in one or a very few copies in the maize genome.

Key words: Zea mays; pollen; non-coding RNA; 5' RACE; 3' RACE; overlapping PCR; developmental expression pattern

There were several reports of transcripts without a long open reading frame (ORF) in various eucaryotes (Brannan et al., 1990; Brockdorff et al., 1992; Brown et al., 1992; Askew et al., 1994; Crespiet al., 1994; Velic et al., 1994; Watanabe and Yamamoto, 1994; Yoshida et al., 1994), and it has been suggested that they function without being translated into proteins. Some genes encode RNAs, rather than proteins, as their final products. tRNA, rRNA, and the small nuclear RNAs and nucleolar RNAs have been studied extensively, and were relatively straightforward to identify by homology searches or with specialized algorithms (Lowe and Eddy, 1997; 1999). It has become apparent recently that in addition to these structural RNAs, other mRNA-like non-coding RNAs (ncRNAs) exist, which lack protein-coding capacity and exert their action mainly or exclusively at the RNA level (Eddy, 1999; Erdmann et al., 1999; 2000; 2001; Capnara and Nilsen, 2000; Stonz, 2002). Analyses of the properties and functions of ncRNAs indicated that they could act as gene regulators, as part of biotic and abiotic stress signals, or as part of RNA-protein complexes with various enzymatic and structural activities. A number of ncRNAs were processed in an mRNA-like manner. Consequently, they undergo splicing and have poly(A) tails and, presumably, caps (Erdmann et al., 2001). The presence of ncRNAs has been described in several systems, for example, in prokaryotic and eukaryotic systems (Willard and Salz, 1997; Panning and Janeisch, 1998; Akhtar et al., 2000).

Only a few ncRNAs from plants have been reported. One of the first transcripts identified as an ncRNA in plants was CR20 (Teramoto et al., 1996), a gene isolated from cucumber (Cucumis sativus) that repressed by cytokinins and by stress or developmental conditions (Teramoto et al., 1995). This gene was part of a family of ncRNA with members in several plant species (Taylor and Green, 1995; Teramoto et al., 1996; van Hoof et al., 1997). GUT15 was another characterized member of this family (Taylor and Green, 1995). The fact that transcripts of this family were hormonally regulated and had unstable transcripts suggested that they might play a role in regulatory processes, although their true functions were unknown. Another interesting family of ncRNAs present in plants was typified by Mt4 in Medicago truncatula (Burleigh and Harrison, 1998) and TPS11 in tomato (Lycopersicon esculentum) (Liu et al., 1997). As with the GUT15/CR20 family, these genes were regulated by biotic (cytokinins) and abiotic (phosphate starvation) signals. Several short non-conserved ORFs were present in Mt4/TPS11 family, and all of the transcripts showed regions of absolute identity at the nucleotide level.
(Martin et al., 2000). The high degree of nucleotide sequence conservation and low level of ORF conservation suggested that the final product of these genes was RNA and not protein.

We previously isolated some cDNAs for pollen-specific genes in Z. mays by differential screening (Li et al., 2001). One of cDNA fragments, named ZM401, was selected to be analyzed. The length of ZM401 cDNA fragment was 663 bp. Two transcripts (1.2 kb, 2.0 kb) of the ZM401 were detected by Northern blotting analysis. In this study, we began our molecular analysis of ZM401 by using 5'RACE and 3'RACE to clone a full-length cDNA for the 1.2 kb RNA. We discuss the possibility that ZM401 was a non-coding RNA. Using RT-PCR and Northern blotting analysis, we studied the developmental expression pattern of ZM401 in Z. mays pollen.

1 Materials and Methods

1.1 Plant material

Corn (Zea mays L.) pollen and anthers from various stages of microspore development were collected from field grown plants of the cultivar (Nongda 108). These were immediately frozen in liquid N2 and stored at −70°C until used for RNA isolation.

1.2 3' and 5' RACE-based cloning of ZM401 cDNA

To isolate a full-length cDNA from total RNA extracted from Z. mays pollen, we employed 3'RACE and 5'RACE. ZM401-specific primers were designed based on the ZM401 cDNA fragments. To amplify the 3' end of ZM401 mRNA, an anchored oligo(dT) primed single-strand cDNA was synthesized from 2 µg of total RNA using AMV reverse transcriptase followed by PCR with a ZM401-specific primer (NGSP2) 5'-GGTGAAGGGCTCA ATTATTAGG-3' and an anchor primer according to the manufacturer's protocol (3'RACE System For Rapid Amplification of cDNA Ends, Version 2.0, Gibco, BRL, No. 18373-019). 5'RACE was performed according to the manufacturer's protocol (5'RACE System For Rapid Amplification of cDNA Ends, Version 2.0, Gibco, BRL, No. 18374-058). In brief, single-strand cDNA was generated by reverse transcription with a primer GSP1: 5'-TAGCCACTATCCGCTGAC-3'. A homopolymeric dC tail was added to the 5'-end of the cDNA by terminal deoxynucleotidyl transferase followed by PCR with a nested primer, GSP4: 5'-CTCACCCGGCTGAACTCTGCA-3', and an abridged anchor primer was included in the kit. PCR was performed under conditions of 30 cycles of 1 min at 94°C, 1 min at 64°C and 1 min at 72°C followed by 10 min at 72°C in a thermal cycle. Amplified PCR fragments were cloned into pMD18-T and sequenced. ZCF1 (5'-TTGGCGAGCCTATGGAGCGGC-3') and ZCF2 (5'-TTTTTTTTTTTTTTGATCACGAACCTTATAATAAAGTTGAAGCTGG-3') primers were designed for splicing 5'RACE and 3'RACE amplified product.

1.3 Isolation of total RNA and reverse transcription-polymerase chain reaction (RT-PCR) analysis

Total RNA was isolated from mature pollen and anthers of various stages of microspore development. RNA was isolated using the hot phenol method and subsequently treated with RNase-free DNase I (Promega Co. Ltd.). Total RNAs (1 µg) were reverse-transcribed by AMV reverse transcriptase (Promega Co. Ltd.) using oligo(dT)12,18 as a primer followed by PCR with ZM401-specific internal primers (forward, P1 5'-ATCCGGAGCGAGTGGTGAC-3' and reverse P2 5'-GTCACTATGAAACCCAGAAG-3'). PCR reaction conditions were as follows: 94°C 1 min, 56°C 1 min, 72°C 1 min, 72°C extension 10 min, 30 cycle. PCR products were separated by electrophoresis in a 0.8% agarose gel followed by blotting to a nylon membrane (Amersham Pharmacia Biotech). The ZM401 products were detected by hybridization with a 32P-labeled ZM401 cDNA. In control experiments, RNA was treated as above but without reverse transcriptase. In all cases, these control reactions gave no signal after hybridization. All the RT-PCR experiments presented in this study were repeated using two or more independent RNA preparations.

1.4 Southern blot

Southern blotting analysis was performed as described by Church method with slight modifications. Twenty µg genomic DNA was digested with six restriction enzymes (Hind III, EcoRI, BamHI, KpnI, SacI, EcoRV), and electrophoresis on an agarose gel, and transferred to a nylon membrane (Amersham Pharmacia Biotech) (Sambrook et al., 1989). The membrane was prehybridized at 65°C for 6 h in a prehybridization solution consisting of 1% BSA, 1 mM EDTA, 0.5 mol/L Na2HPO4 (pH7.2), 7% SDS. Probes labelled with 32P were prepared from ZM401 cDNA using a random-primer DNA labelling kit (Promega). Hybridization was performed with 32P-labeled probes in the prehybridization solution for 18 h at 65°C. The membrane was washed at 65°C in 2× SSC plus 0.5% SDS for 30 min, 1× SSC plus 0.5% SDS for 30 min, 0.5× SSC plus 0.1% SDS for 30 min, and then, 0.1× SSC plus 0.1% SDS for 30 min.

1.5 Northern blot

Total RNAs (20 µg) from different stages of microspore development and mature pollen grains were isolated and resolved by electrophoresis on a 1.2% agarose-formaldehyde gel. The RNAs were transferred to a nylon
membrane (Amersham Pharmacia Biotech), the membrane was prehybridized, hybridized, washed as described by Southern blots.

2 Results

2.1 Amplification of ZM401 full-length cDNA by 5' RACE and 3' RACE

Using the 5' RACE technique, we obtained a 651 bp single-band amplification product that extended the sequence of the 5' end of 403 bp (Fig.1). Using the 3' RACE technique, we obtained a 571 bp single-band amplification product that extended the sequence of the 3' end of 84 bp (Fig.2). Because ZM401 cDNA was a non-coding mRNA, we deduced the length of cDNA only by Northern blotting analysis. Two bands hybridizing with ZM401 cDNA fragment (663 bp) were detected by Northern blot. One was 1.2 kb, and the other was 2.0 kb. We obtained an 1 149 bp ZM401 cDNA by using 5' RACE and 3' RACE (Fig.3). To identify the 1 149 bp ZM401 cDNA was the transcript of 1.2 kb or 2.0 kb, we prepared three 32P-labelled probes with ZM401 cDNA of the 1 149 bp, 5' end of the 350 bp, 3' end of the 571 bp, respectively. Probes from 3' end of the 571 bp and 1 149 bp cDNA hybridized to both the 2.0- and 1.2-kb RNAs, while probes from 5' end of the 350 bp cDNA hybridized only to the 1.2-kb RNA (data not shown). We concluded that the 1 149 bp cDNA represented a full-length copy of 1.2 kb.

Because 5' RACE product and 3' RACE product had 70 bp overlapping, we obtained a 1 149 bp full-length cDNA clone by overlapping PCR (Fig.4) and cloned it into a TA vector pMD-18T, named pZFL-1.2, and sequenced. All the sequence presented in this study was sequenced using two or more independent clones.

2.2 The absence of a long ORF in the ZM401 transcript

Amino acid sequences were deduced from nucleotide sequence of ZM401 transcript. Figures 5, 6 showed the positions of initiation codons (ATG) and termination codons (TAA, TAG and TGA) in the six possible reading frames. There were many termination codons (TAA, TAG and TGA) in the six possible reading frames in ZM401 cDNA through each frame, and no long ORF was found (Figs.5, 6). The number of termination codons through three forward possible reading frames in the ZM401 cDNA was 13, 15, and 13, respectively. The longest potential ORF was 269 nucleotides long (791–1060), coding 89 AA, had a poor consensus sequence for translation initiation. But it had a poly(A) tail. Ecotopic expression of ZM401 cDNA in tobacco or overexpression in maize all resulted in pollen sterile (data not shown), suggesting that ZM401 functioned as a non-coding mRNA and was related to pollen development.

2.3 The developmental expression pattern of ZM401 in Z. mays pollen

The analysis of RT-PCR products (Fig.7A) showed an expression pattern of ZM401 among different stages of pollen development in Z. mays. A strong amplified band was found at the tetrad stage of microspore development, the uninucleate stage of microspore development, the
<table>
<thead>
<tr>
<th></th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TTGGAGCCGAC ATGGACGGAG GCCGGATTCG ATCCGGGCGG TTACGTGCTA</td>
</tr>
<tr>
<td>51</td>
<td>CGGTGCTGTTT GAGCATGCGG AGACACCTAC GCGATGACCC CAAACACATT</td>
</tr>
<tr>
<td>101</td>
<td>CACGCTGCTCA ATTCGCGCAG AGGCCGGTTA CCACATACGT GCTGCCAGATA</td>
</tr>
<tr>
<td>151</td>
<td>GCGGGCAGAAC CGGATACGGC GCTGCCCCCT ACCGGCTAC TAAAGCGCTG</td>
</tr>
<tr>
<td>201</td>
<td>CGAGTTTATAT CAGGGATGTC GGAATGGCCT AGTGACTTAC TAAAGCGCTG</td>
</tr>
<tr>
<td>251</td>
<td>GTGCTCACGC AAATCTGGAC GAGACTACC AGGATGGCCG GCTGGCCGCG</td>
</tr>
<tr>
<td>301</td>
<td>AGATCTAGCT GCGGGCCGTC ATCCGGGCAA AACGCGCGG TTTCTGTCGCC</td>
</tr>
<tr>
<td>351</td>
<td>ATGCTTGACG TGGATCCCT GCGGCGCAAA CAAAGGCTTT TTCTCTGCT</td>
</tr>
<tr>
<td>401</td>
<td>AGAGGATGTG GAATGATGAG CCGGACAGAT ATGAAAGGAG TTTGGATGTC</td>
</tr>
<tr>
<td>451</td>
<td>CGTGGGAGCG GCGATGCTG CAGGGTCTGG TTCTTGCGGC GTTGGGAGG</td>
</tr>
<tr>
<td>501</td>
<td>AAAAGAAATT CGCGAGCTAT GGGATTATA CTGGGAGCCG GAGTGGGTGCT</td>
</tr>
<tr>
<td>551</td>
<td>CTGGGTGTCGT ACAAAGAGGG CGGGCAAGGG TGGAGCGTCA ATTTTATAGGG</td>
</tr>
<tr>
<td>601</td>
<td>ATGGACACAT CTTGCTACGA ATTTGGGCAA GATGAGCTGT GCGGCGGTTG</td>
</tr>
<tr>
<td>651</td>
<td>GTGGGTGAGT CACCGGAGTA TGGCTACGT GGGTGACAGT TGGAGCAAGGT</td>
</tr>
<tr>
<td>701</td>
<td>CTGGGGTGTG GAACTGGCTAT CACCAAGGAC TCATTGTGAC GGGTGAACAC</td>
</tr>
<tr>
<td>751</td>
<td>GGGTGATGGT GTGCCGGCAG GACGAGTGGT CAGGGATGC ACGGAGACG GACGACTTAC</td>
</tr>
<tr>
<td>801</td>
<td>TGCACCGGCG ACAGGGCGCG GCAATGCGGT AGCGAGACG AGGACGCGTC</td>
</tr>
<tr>
<td>851</td>
<td>GAACCGCCCT CGATTTTGT CAGCCGCTGG GGGGACTCT GCTGACAGCA</td>
</tr>
<tr>
<td>901</td>
<td>AGGAGCAGATC AGCGCCGCAA GCCGGATTC ATCATCGGAG ATTACGCGG</td>
</tr>
<tr>
<td>951</td>
<td>GCAAGTTGGT ATGCTCTGAC GTCCTCCTGA ATGAGCTCGT CATTGTTGAT</td>
</tr>
<tr>
<td>1001</td>
<td>AGCTCATTTA CCAGCTGAC AGAGAACATTT GAGAGCTTT GTCTCTGCT</td>
</tr>
<tr>
<td>1051</td>
<td>TTTCATGCTA CGAAGCTGCTA ATGCAGACTA ATCAAAGTG TTTGCGCTA</td>
</tr>
<tr>
<td>1 101</td>
<td>TACCAAGCTTC AACCTTTATTA TAAAGTTCGAT GATCAAAAAA AAAAAAAAA</td>
</tr>
</tbody>
</table>

**Fig. 3.** Sequence of ZM401 cDNA. The shadow indicates the primers used in this paper and arrows indicate the forward or reverse primer.

The ZM401 transcripts were first detectable at the tetrad stage, and they continued to accumulate thereafter, reaching a maximum concentration in the mature pollen grains (Fig. 8A). Thus the ZM401 gene seems to belong to “late gene”. The results provided a clue to the question of whether the two transcripts of ZM401 resulted from degradation during preparation of the RNA or not. Products of the degradation of rRNAs from each sample were nearly undetectable, and the patterns of bands of rRNAs were almost identical (Fig. 8B). However, two bands of ZM401-hybridizing RNA were detected by Northern blotting.

binucleate stage of microspore development and mature pollen, but no amplified band was detected at the pollen mother cell stage of microspore development. When they were probed with ZM401 cDNA, each PCR product showed a hybridisation signal (Fig. 7B), and the pattern was very similar to that in Fig. 7A. Identical expression patterns were also obtained in RNA gel blotting analysis.

Two transcripts of ZM401 were detected from the tetrad stage of microspore development to mature pollen, but not in the pollen mother cell stage of microspore development. We noticed that in RNA gel blotting analysis, the signal of
2.4 Was the ZM401 expressed in pollen, members of large or small families?

To determine if ZM401 expressed in corn pollen was present in the genome as members of large families of genes or, as single or very few genes, Southern hybridizations were carried out. Genomic DNA was prepared from leaves of maize. The DNA was digested with six restriction enzymes Hind III, EcoRI, BamHI, Kpn I, Sac I, EcoRV respectively, and each digestion was analysed by Southern hybridisation. 32P-labeled ZM401 cDNA was used as a probe, one hybridization band generated by EcoRI, KpnI, but there were two bands in the case of digestion with HindIII, EcoRV, BamHI and SacI (Fig.9). These results suggested that the ZM401 gene was present in one or a very few copies in the genome instead of a large and complex gene family.

2.5 Homology comparison of ZM401

The comparison of the nucleotide with sequences in GenBank, EMBL, and DDBJ data banks had not shown homology to any known sequence and not significant.
homology to any sequence of pollen-specific or anther-specific gene. Indeed, the sequence of ZM401 was unrelated to any in the database. The lack of a significant ORF together with the lack of homology to known coding sequence argued against any of the RNA species being translated. Therefore, These results suggested that ZM401 was a new category of non-coding mRNA-like RNA.

3 Discussion

In mammals, there were two well-studied non-coding RNAs, H19 (Brannan et al., 1990) and XIST (Brockdorff et al., 1992; Brown et al., 1992). The common feature of these RNAs was the absence of an extensive ORF in the transcripts, even though the transcripts were spliced and polyadenylated like other mRNAs. The expression of the H19 gene was activated in early embryogenesis of mouse (Pachnis et al., 1988), and the H19 RNA showed tumour-suppressor activity on two embryonic tumour cell lines (Hao et al., 1993). The XIST gene was mapped to the X chromosome inactivation center, and the gene was expressed exclusively from the inactive X chromosome, suggesting that the gene product played a role in X chromosome inactivation (Brown et al., 1992; Lee et al., 1999). Recently several other novel genes without extensive ORFs have been reported in fission yeast (Watanabe and Yamamoto, 1994), Dictyostelium discoideum (Yoshida et al., 1994), rat (Velleca et al., 1994), mouse (Aske et al., 1994), Medicago plants (Crespi et al., 1994), cucumber (Teramoto et al., 1996), Arabidopsis (Teramoto et al., 1996) and tomato (Liu et al.,
Although the transcripts of most of them, including H19, XIST, CR20, were thought to function as untranslatable RNAs, the molecular mechanisms of their function still remain unclear. Watanabe and Yamamoto (1994) demonstrated that mei RNA contained no long ORF and formed a complex with Mei2 protein and performed an essential role in the induction of meiosis in fission yeast. The amino acid sequences deduced from all possible reading frames of the ZM401 cDNA had many termination codons throughout each reading frame, and no long ORF was found, but it had poly(A) tail. The nucleotide sequence showed no significant homology to non-coding mRNA sequence isolated from plants. Therefore the ZM401 RNA could be a member of a new category of non-coding RNA.

Northern blotting analysis revealed ZM401 was pollen-specific gene (Li et al., 2001) and had two transcripts in Z mays mature pollen. RT-PCR and Northern blotting analysis showed ZM401 began to express from tetrad stage of microspore development to mature pollen. Northern blotting analysis revealed the presence of two transcripts (2.0 kb and 1.2 kb) of ZM401 in Z. mays. There would be little contribution of degradation products generated during preparation of RNA to the observed two RNA bands, as indicated by the fact that degradation of rRNAs was not detected, the results could be due to (1) different members of a gene family, (2) alternative splicing, (3) degradation of the RNAs in the intact plants, and (4) different transcription start site. Genomic Southern blotting analysis showed that ZM401 was encoded by a single or a very few copies (Fig.9). Because there were at least two bands of ZM401 transcripts on Northern blots, and different patterns of bands were obtained from different stages of microspore development, the hypothesis that a complex gene family existed and/or degradation of the RNAs in the intact plants occurred failed adequately to explain the results. Therefore, two transcripts of ZM401 might derive from alternative splicing of ZM401 gene or different transcription start site if a single copy gene encodes ZM401 cDNA. If a few genes encode ZM401, two transcripts might derive from homology gene of ZM401. Multiple transcripts, as in the case of ZM401, were generated by alternative splicing for human XIST (not in case of mouse) and cucumber CR20 (Brown et al., 1992; Hao et al., 1993), unstable mRNA generating multiple transcripts for GUT15 (MacIntosh et al., 2001). However, the significance of multiple transcripts for many non-coding genes remains unknown.

Ecotropic expression of ZM401 in tobacco or over-expression in maize all resulted in pollen sterile (data not shown), suggesting that ZM401 functions as a non-coding mRNA and plays an important role in pollen development of Z. mays.

References:


