

UPTEC X 04 039  
SEP 2004

ISSN 1401-2138

JOHAN REIMEGÅRD

Bioinformatical and  
experimental approaches  
to miRNA:s in  
*Arabidopsis thaliana*

Master's degree project



UPPSALA  
UNIVERSITET

## Molecular Biotechnology Programme

Uppsala University School of Engineering

<b>UPTEC X 04 039</b>	<b>Date of issue 2004-09</b>	
Author	<b>Johan Reimegård</b>	
Title (English)	<b>Bioinformatical and experimental approaches to miRNA:s in <i>Arabidopsis thaliana</i></b>	
Title (Swedish)		
Abstract	<p>miRNA are small non-coding RNAs that are important in the development of plants and animals. Through an anti-sense mechanism the miRNA turns of the expression of one or many specific mRNA. The aim of this study was to create a database containing all known information about miRNA in plants, design a bioinformatical approach to predict the targets of the miRNA:s in plants and test the predicted interactions between one of the miRNA and its targets.</p>	
Keywords	miRNA, ncRNA, RNA, miR169, Arabidopsis, rice, plant	
Supervisors	<b>Sandra Kuusk</b> <b>Department of cell and molecular biology, Uppsala University</b>	
Scientific reviewer	<b>Gerhart Wagner</b> <b>Department of cell and molecular biology, Uppsala University</b>	
Project name	Sponsors	
Language	Security	
<b>English</b>		
<b>ISSN 1401-2138</b>	Classification	
Supplementary bibliographical information	Pages	
	<b>27</b>	
<b>Biology Education Centre</b> Box 592 S-75124 Uppsala	<b>Biomedical Center</b> Tel +46 (0)18 4710000	<b>Husargatan 3 Uppsala</b> Fax +46 (0)18 555217

## **Samanfattning**

I samband med att flera genomprojekt blev klara under början av 2000-talet blev det möjligt att leta efter gener på ett systematiskt sätt. Speciellt ickekodande gener, det vill säga gener som inte kodar för något protein, som tidigare bara hade hittats på ett slumpmässigt sätt, blev möjliga att hitta med hjälp av bioinformatik. En sorts ickekodande RNA, så kallade mikroRNA, har visat sig vara relativt lätta att identifiera med hjälp av sökalgoritmer eftersom de har en väldefinierad struktur. mikroRNA finns i både växter och djur men verkar rent mekanistiskt fungera något olika i respektive organismklass. I växter binder varje mikroRNA med nästan perfekt komplementaritet till ett messengerRNA, vilket i de flesta fall leder till att detta messengerRNA degraderas. På grund av denna höga grad av komplementaritet så är det möjligt att hitta de flesta messengerRNA som ett givet mikroRNA binder till genom att göra en blast-sökning. Två grupper har publicerat artiklar där de med hjälp av sådana sökningar har hittat potentiella mikroRNA-messengerRNA-interaktioner i modellväxten *Arabidopsis thaliana*. Gruppernas resultat skiljer sig dock något åt och för ett flertal mikroRNA har man inte lyckats finna några troliga interaktions-messengerRNA.

Projektet inleddes med skapandet av en databas för mikroRNA i växter. Ett webbaserat gränssnitt finns och möjlighet att lägga till och ta bort information i databasen via interfacet har gjorts. En sökalgoritm för att hitta troliga mikroRNA-messengerRNA-interaktioner har skapats. Alla redan kända interaktion blev funna och några nya har förutspått men måste testas innan deras funktion kan säkerställas. Ett mikroRNA i *Arabidopsis thaliana*, miR169, har transinfekterats in i en *Arabidopsis thaliana* bakom en konstitutiv promotor. Någon avvikelse från den normala fenotypen kunde inte konstateras men den transgena växten är ändå viktig för framtida studier av miR169.

**Johan Reimegård**

**Uppsala universitet  
augusti 2004**

1. Background	5
1.1. A new type of regulatory element	5
1.1.1. Non-coding regulatory RNAs	5
1.1.2. Non-coding regulatory RNAs in Eukaryotes	6
1.1.3. What's the big fuzz?	7
1.2. miRNA	8
1.2.1. History	8
1.2.2. miRNA transcripts	8
1.2.3. Going from a transcript to a mature miRNA	9
1.2.4. Target regulation	9
1.3. Plants	10
1.3.1. miRNA in plants	10
1.3.2. Arabidopsis	11
1.3.3. Introducing a new gene in Arabidopsis using <i>Agrobacterium tumefaciens</i> mediated transformation	11
1.3.4. Rice	12
1.4. Bioinformatical tools	12
2. Material and Methods	13
2.1. Plant material and growth conditions	13
2.2. DNA preparation	13
2.3. PCR amplification	13
2.4. Electrophoresis	13
2.5. Cloning procedures	14
2.6. The making of a transgenic Arabidopsis plant	14
2.7. <i>In silico</i> material	15
2.8. <i>In silico</i> methods	15
3. Results	16
3.1. Database	16
3.2. Website	16
3.3. Predicting targets for miRNA	17
3.3.1. Sliding window	18
3.3.2. Target Finder	18
3.3.3. Homology	19
3.4. Prediction of new miRNA homologs in rice	19
3.5. A transgenic Arabidopsis	20
4. Discussion	23
4.1. miRNA target finder	23
4.2. miRNA finder	23
4.3. miR169s transformed Arabidopsis	24
4.4. Database	24
4.5. Website	24
5. Acknowledgments	24
6. References	25

# 1 Background

## 1.1 A new type of regulatory element

### 1.1.1 non-coding regulatory RNAs

All living organisms keep the expression of genes under tight regulation in each cell. A large part of the genes in an organism codes for regulatory proteins. Until the beginning of the 21<sup>st</sup> Century only a few regulatory non-coding RNA (**ncRNA**) had been found and these were all considered to be exceptions from the rule that proteins serve as the only regulatory factor. In the book *BIOLOGY*<sup>1</sup> RNA is described as “Ribonucleic acid (**RNA**) (*RY-boh-noo-KLAY-ik*) a single-stranded nucleic acid molecule involved in protein synthesis, the structure of which is specified by DNA”. This definition includes messenger RNA (**mRNA**) that codes for the protein and the ncRNAs: the transfer RNA (**tRNA**) and the ribosomal RNA (**rRNA**), which are parts of the machinery that builds the protein from the mRNA template. In eukaryotic cells small nuclear RNA (**snRNA**) and small nucleolar RNA (**snoRNA**) are involved in the processing of mRNA (Table 1).

- |  |
|--|
| <ul style="list-style-type: none"> <li>• rRNA<br/>ribosomal RNA</li> <li>• tRNA<br/>transfer RNA</li> <li>• snRNA<br/>small nuclear RNA</li> <li>• snoRNA<br/>small nucleolar RNA</li> <li>• sRNA<br/>small RNA</li> <li>• siRNA<br/>small interfering RNA</li> <li>• miRNA<br/>micro RNA</li> </ul> |
|--|

**Table 1 different classes of non-coding RNAs.**

Different kinds of ncRNAs. miRNA and siRNA in eukaryotes and sRNA in bacteria as important factors in the development of species. The majority of the regulatory ncRNA has been discovered during the last five years.

Recent discoveries of new classes of ncRNA have made this description of RNA in *BIOLOGY* obsolete. These new ncRNA are regulatory factors and are not rare exceptions but found in all kingdoms and in great numbers. In prokaryotic cells the non-coding regulatory RNAs are called small RNA (**sRNA**)<sup>2-4</sup> and in eukaryotic cells there are two new classes called micro RNA (**miRNA**)<sup>5-19</sup> and small interfering RNA (**siRNA**)<sup>20-24</sup>.

RNA is generally considered to be single stranded but can form secondary structures and bind to complementary sequences. RNA binds complementarily to its canonical base pair like DNA e.g. Adenine (**A**) binds to Uracil (**U**) and Cytosine (**C**) binds to Guanine (**G**). It has also been shown that G and U binds strongly to each other. A single stranded RNA will find its most relaxed state either by forming its own structure base pairing with itself, causing parts of the RNA to become double stranded or, like DNA, bind to another complementary sequence.

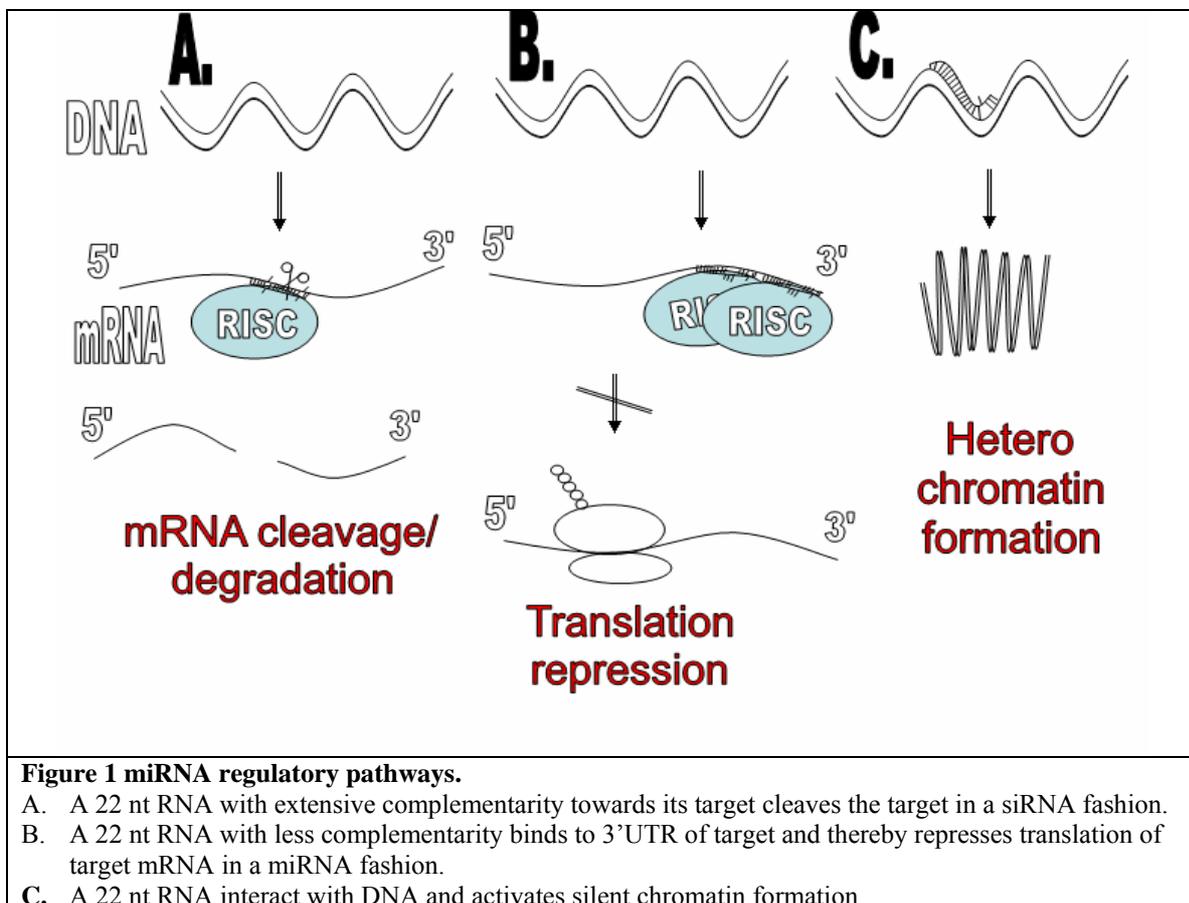
Despite different modes of action all regulatory RNAs have some common features, the structure and part of the sequence of the regulatory RNA play important roles. When comparing the same functional RNAs in different but closely related species one can see that the sequence has been altered during evolution but the structure preserved implying that there is a higher biological pressure towards keeping the structure than keeping the sequence. In most of the cases a certain part, or parts, of the sequence are important for ncRNAs to fulfil its function. It has been shown that because of complementarities between a regulatory RNA and an mRNA the regulatory RNA alters the expression of that mRNA.

### 1.1.2 non-coding regulatory RNAs in eukaryotes

During the 21<sup>st</sup> Century two new kinds of regulatory RNAs were recognised in eukaryotic systems, the siRNA and the miRNA. siRNA is the guiding RNA in the RNA interfering (**RNAi**) pathway<sup>20-24</sup> and is suggested to protect the cell against double stranded (**ds**) RNA viruses and transposable elements. If a ds RNA exist or is introduced into the cell, it will be cleaved into 22 nucleotides (**nt**) long ds RNA fragments by an enzyme known as Dicer. A protein complex called Ribonuclease induced silencing complex (**RISC**) associates with one of the strands of the 22 nt long sequence referred to as the siRNA. The siRNA will bind with perfect complementary to its target and recruit a protein of unknown identity which cuts the target RNA where siRNA has bound. This terminates the expression of the target RNA (Figure 1A).

miRNAs are thought to play important roles in the acquisition of diverse cell types in multicellular organisms. In its premature state, miRNA is produced in the cell as a single stranded (**ss**) RNA forming a hairpin structure. One unique part of the stem loop of the hairpin is cut out forming a 22 nt long ds RNA. One specific strand of the ds RNA, the miRNA, becomes part of the RISC. The miRNA regulates the gene expression by binding in the 3' untranslated region (**UTR**) of a target mRNA with imperfect complementarity, thereby hindering translation (Figure 1B).

It has also been shown that a 22 nt single stranded (**ss**) RNA, like a miRNA or a siRNA, with high enough complementarities towards DNA can initiate heterochromatin formation and thereby silence that part of the DNA in an epigenetic fashion (Figure 1C)<sup>25</sup>.



Even if their premature structures are different and although they serve two different functions in the cell miRNA and siRNA share some common features. siRNA and miRNA use the same pathway to get from an unmaturing state to a mature state. They both get processed by the same enzyme Dicer and they act as the guiding RNA of the RISC. However any part of the dsRNA can become a siRNA whereas only one specific part of the premature miRNA sequence becomes a miRNA.

When miRNA was discovered in plants experiments showed that most of the miRNA in plants work in a siRNA-like manner<sup>10</sup> and binds with perfect or almost perfect complementarities to the coding sequence of its target, thereby turning off the expression of the target. Recent studies have shown that when inserting siRNA in a cell, it will not only affect the perfectly complementary target, but will also affect other targets in a miRNA-like fashion<sup>26-28</sup>. The similarities between the two classes of RNA have speeded up the understanding of both classes.

### 1.1.3 What's the big fuzz?

miRNAs play an important role as a regulatory factors in multicellular organisms. But still for four years ago no miRNA had been found in mammals. Bioinformatical approaches to find miRNA estimate that there are about 250 different miRNA in humans<sup>18</sup>, which is approximately one percent of all genes. This is in the same amount of genes that other important regulatory gene families contain. Some of the miRNA exist in more than 50 000 copies per cell<sup>16</sup> which, apart from ribosomal RNA, is among the most abundant RNA in the cell. The important roles of miRNA in gene regulation have questioned the central dogma of microbiology and revised the role of RNA.

The potential use of RNAi as a tool in microbiology is also something new. Knocking-out a specific gene is expensive and tedious. By using RNAi, that is introducing ds RNA to the cell, any gene of interest can be silenced. The ds RNA will produce siRNA against the mRNA and turn off the expression. This method is a powerful tool in discovering the function of a particular gene.

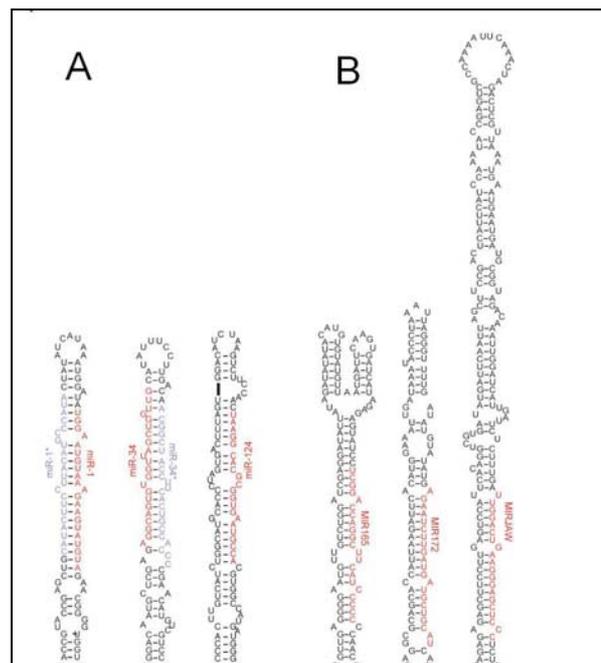
## 1.2 miRNA

### 1.2.1 History

The first miRNA, called lin-4, was discovered in the worm *Caenorhabditis elegans* (*C. elegans*) the year 1994 through genetic screens for mutants that lacked the ability to control the timing of specific cell fate switches during development<sup>5</sup>. Seven years later the same research group reported the discovery of another miRNA called let-7, which also is important for timing of development<sup>6</sup>. Since these miRNA were important for timing of cell-development and showed up during specific parts of development they were called small temporal RNA (**stRNA**). The precursor of the stRNAs formed a hairpin-like structure and the mature functional RNA were approximately 22 nt in length. The same year, more systematic approaches showed the existence of a large number of new small regulatory RNAs in *C. elegans*<sup>7-9</sup>. All of them folded in a hairpin precursor and a length of 22 nt. Some of them were found at a certain time of the development like lin-4 and let-7 but some were found in specific tissues. Since they all did not fill the criteria of stRNA a new name for these RNA was chosen, the miRNA<sup>29</sup>. To date several hundreds of miRNA have been discovered in many different multicellular organisms using cloning techniques and/or bioinformatical approaches<sup>13-19</sup>.

### 1.2.2 miRNA transcripts

Even though the length of the hairpin, called the pre-miRNA, is approximately 70 nt (Figure 2A), the transcript is assumed to be more than 200 nt. The transcript is called pri-miRNA. It is not known if the rest of the transcript, that do not form the hairpin, has a function. Many miRNA are believed to be transcribed by themselves or in miRNA clusters. Some miRNA in animals have been found in intron sequences of mRNAs, thus having the same expression pattern as that mRNA. In animals there is often one single copy for each miRNA but in plants some of the miRNAs reside in up to seven copies in the genome. In animals the length and the structure of the pre-miRNA is well conserved. Of all the hundreds of miRNAs found almost all fold into an approximately 70 nt long hairpin pre-miRNA structure. In plants, the length of the stem varies much more. The smallest predicted hairpin identified is of 70 nt like the miRNA in animals whereas the largest one found is assumed to be 313 nt<sup>11</sup> (Figure 2B).



**Figure 2 pre-miRNA structure in animals and plants.**

- A.** pre-miRNA in animals form single hairpin structure with an approximately length of 70 nt(miR-1, miR-35 and miR124)
- B.** pre-miRNA in plants have bigger diversity in structure and length(miR-165, miR-172 and miR319)

The structures are predicted by mfold using standard settings and the sequences are collected from the miRNA registry at the sanger institute.

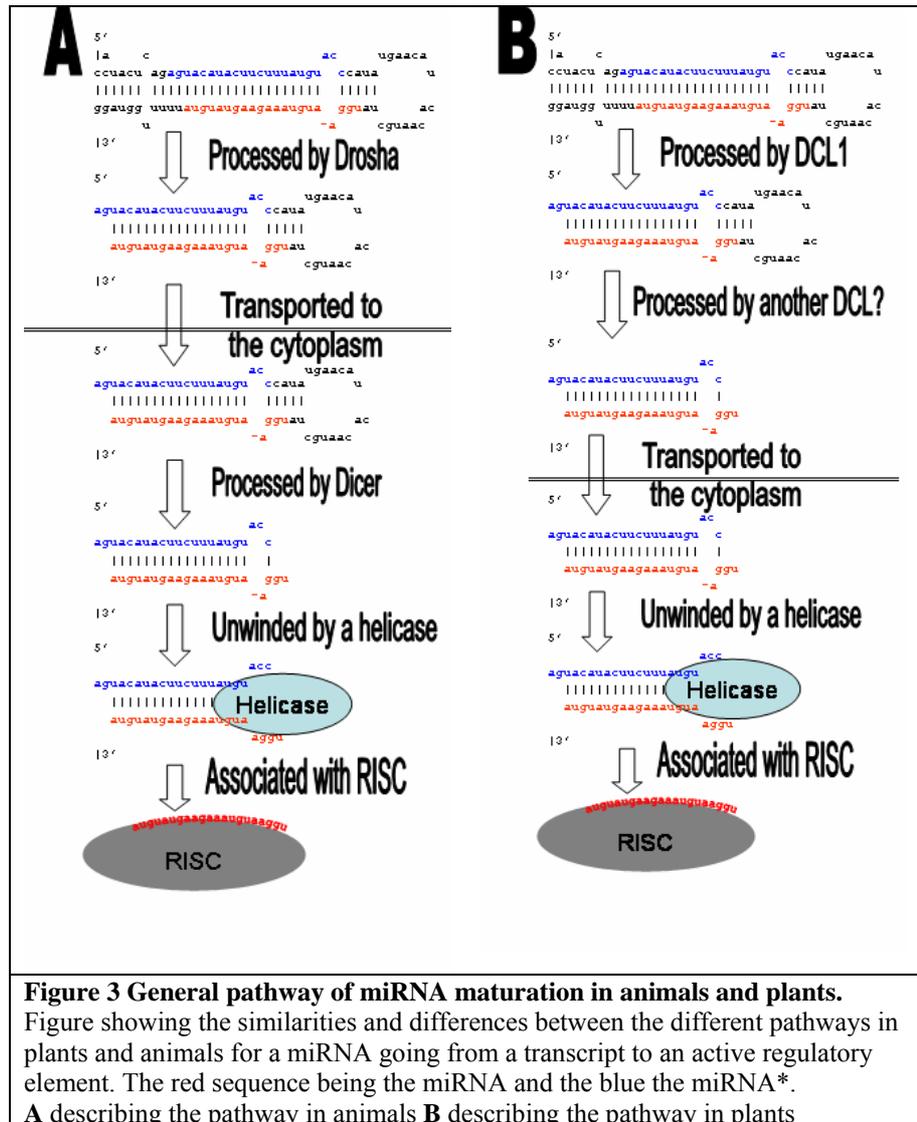
## From a transcript to a mature miRNA

The miRNA is transcribed in the nucleus and the hairpin structure is formed. In animals an enzyme called Drosha cuts the pre-miRNA at the end of the hairpin which gives rise to the pre-miRNA<sup>30</sup>. The pre-miRNA is then exported out of the nucleus and processed by the enzyme Dicer, which cuts out the miRNA and its complementary strand. This creates a 22 nt long dsRNA with a two nt 3' overhang consisting of the miRNA and its complement referred to as miRNA\*<sup>31</sup>. In plants the pre-miRNA is processed in the nucleus by an enzyme called DicerLike1 (DCL1), a Dicer homolog, and maybe another DCL

protein<sup>11</sup>. A helicase will unwind the miRNA:miRNA\* in the cytoplasm in both animals and plants and the miRNA will be associated in the RISC thus making the complex active. The difference between animals and plants is the fact that the pre-miRNA in plants does not have to pass through the nucleus membrane. This could be the reason why the pre-miRNA in plant does not have to restrict the size of their hairpin (Figure 3).

### 1.2.3 Target regulation

miRNAs regulate the expression of their mRNA target at the posttranscriptional level by using two different mechanisms. Either the miRNA binds to the 3'UTR of the target, thereby repressing the translation, or it binds to the target and mediates cleavage of the target mRNA (Figure 1). It appears as if the miRNA in animals regulate the expression of most of their targets by translation inhibition. The complementarity between the miRNA and the mRNA for translation inhibition to occur can be relatively low but some preliminary rules exist. When comparing the same miRNA in distantly related species, the sequence similarities were



low in many parts of the miRNA but the nt 2-8 from the 5' end of the miRNA were almost always conserved. Verified targets showed perfect complementarity between the heptamer 2-8 on the miRNA and its complement on the mRNA target. Also in searches to find new targets in animals, using a bioinformatics approach, the best ratio between true hits and noise was achieved when using the heptamer of the nt 2-8 as the strongest signal<sup>32,33</sup>. It is also important that there is more than one miRNA binding site in the 3' UTR. Criteria for multiple binding sites and high complementarity between the heptamer have been important factors when trying to find new miRNA targets in animals using a bioinformatical approach.

Silencing of mRNA targets by cleavage appears to be the most common pathway for miRNA action in plant. To be able to promote cleavage the sequence complementarity between the miRNA and the target has to be almost perfect, but as in animals, the nt heptamer (2-8) of the miRNA sequence is the most important. It has been suggested that the nt 2-8 in the miRNA sequence initiate the binding between the target and the miRNA. A protein called Argonaute, that is part of the RISC complex, contains a domain named PAZ. A hypothesis is that PAZ, which can bind ss and ds RNA, uses the nt 2-8 in the miRNA sequence for initiation of binding between the target and the miRNA. In order for the binding to take place the complementarity between the heptamer and its target has to be almost perfect or it will not fit in the groove of the PAZ domain where the miRNA is thought to be located. This could explain why the nt 2-8 in the miRNA sequence are the most conserved nt.

## 1.3 Plants

### 1.3.1 miRNA in plants

In *Arabidopsis thaliana* (**Arabidopsis**) 18 unique miRNAs have been found using cloning techniques<sup>10-12</sup>, and one using an activation tagging screen<sup>19</sup>. *Oryza sativa* (**Rice**) homologs have been found for eight of the 19 Arabidopsis miRNA where the miRNA sequence is totally conserved (Table 2). When allowing up to three mismatches between rice and Arabidopsis miRNA sequences, possible homologs can be found for almost all of the miRNA. Due to the fact that plant miRNA exhibit almost perfect complementary to their target, all of the miRNAs in plants have predicted targets. Only a few of them have been experimentally verified. A bioinformatics approach to find miRNA targets in plants have been presented, where no gaps and three mismatches were allowed. Out of the eight miRNA in rice, six had targets that could be related to the targets of the related miRNA (Table 2). So not only the miRNA had been conserved but also its targets. Even though Arabidopsis and rice are the only sequenced genomes that are available, other plants are being studied and large Expressed Sequence Tags (EST) libraries exist for many other plants. For one of the miRNA, miR165, where the target is verified, a HD-Zip transcription gene, there is evidence that this miRNA not only exists in flowering plants but in all land plants. Even if the miRNA in itself has not been found in all these plants, the region at the mRNA where

miRNA	Arabidopsis	Rice
156	X <sup>1,2</sup>	X <sup>2</sup>
157	X	
158	X	
159	X	
160	X <sup>1,2</sup>	X <sup>2</sup>
161	X	
162	X	X
163	X	
164	X <sup>1,2</sup>	X <sup>2</sup>
165	X <sup>1</sup>	
166	X	X
167	X <sup>1,2</sup>	X <sup>2</sup>
168	X	
169	X <sup>2</sup>	X <sup>2</sup>
170	X	
171	X <sup>2</sup>	X <sup>2</sup>
172	X <sup>1</sup>	
173	X	
319	X	

**Table 2 miRNA targets verified and homologs in rice.**  
Plant miRNA and where they are found.  
1. Targets verified  
2. Homolog targets found in rice

the miRNA will bind has been conserved and cleaved mRNA products at the target site were found<sup>34</sup>. This implies the existence of miRNA that date back more than 400 million years ago. No homolog miRNA present in both the two eukaryotic kingdoms have been found.

### 1.3.2 Arabidopsis

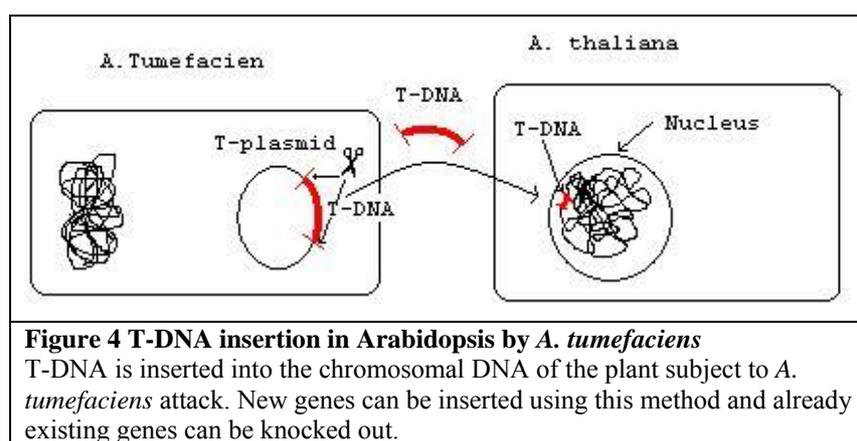
Arabidopsis is the model plant for flowering plants. Arabidopsis has a small genome (125 mega bases (MB)) and approximately 20 000 genes on five chromosomes. It lacks repetitive DNA, has a short generation time, is easy handled in small spaces and has abundant seed production. During the year of 2000, the entire genome of Arabidopsis was published<sup>35</sup>. Seed stock centers and databases of available mutants have increased the possibilities for efficient studies of Arabidopsis (Table 3).

### 1.3.3 Introducing a new gene in Arabidopsis using *Agrobacterium tumefaciens* Mediated transformation

The plant pathogen *Agrobacterium tumefaciens* (*A. tumefaciens*) is a bacterium that infects plants and induces tumor formation. The genes responsible for the tumour formation is located on a Tumor inducing (Ti)-plasmid<sup>36</sup>. *A. tumefaciens* has a mechanism for inserting part of the Ti-plasmid DNA, called the T-DNA, into the chromosomal DNA of a plant (Figure 4). In the wild type (wt) *A. tumefaciens* genes on the T-DNA code for hormone and opine biosynthesis enzymes.

Hormones encourage growth of the infected plant tissue, which induces the tumour

formation, and opines give bacteria a carbon and nitrogen source and generates a more favourable environment for the bacteria. By replacing the wt genes in the T-DNA with any other gene of interest it is possible to use *A. tumefaciens* to insert the gene of interest into the plants chromosomal DNA. The T-DNA is inserted randomly. If it is inserted into a gene, the gene will be silenced. Large mutant collections where T-DNA insert are localised are available (Table 3). *A. tumefaciens* does not only facilitate the insertion of new genes in the genome but is also used to generate knock-out plants



**Figure 4 T-DNA insertion in Arabidopsis by *A. tumefaciens***  
T-DNA is inserted into the chromosomal DNA of the plant subject to *A. tumefaciens* attack. New genes can be inserted using this method and already existing genes can be knocked out.

### 1.3.4 Rice

There are two large subgroups of flowering plants, the monocotyledons and the dicotyledons. The cotyledons are the "seed leaves" produced by the embryo. Arabidopsis is a dicotyledon. Rice is a monocotyledon. The rice genome is three times larger than the Arabidopsis genome but it still represents a relative small genome of approximately 420 MB and between 30 000 – 50 000 encoding genes. Two species of rice, *O. sativa* L. ssp Japonica and *O. sativa* L. ssp. Indica, are being sequenced and draft sequences are available<sup>37, 38</sup>. Monocotyledons and dicotyledons are supposed to have diverged about 140 million years ago.

Resource available	Internet address for information	Information provided
Arabidopsis database	<a href="http://www.arabidopsis.org">www.arabidopsis.org</a>	Primary source of information Links to relevant Internet sites Genomic sequence and tools
The Arabidopsis Biological Resource Center (ABRC)	<a href="http://arabidopsis.org/abrc/">http://arabidopsis.org/abrc/</a>	Collection, preservation and distribution of seeds DNA clone, library storage and distribution Data for all stocks and other information
The Nottingham Arabidopsis Stock Centre (NASC)	<a href="http://nasc.nott.ac.uk/">http://nasc.nott.ac.uk/</a>	provides seed and information resources to the International Arabidopsis Genome Programme

**Table 3 Public Arabidopsis resources.**  
Searchable databases and stock centers makes information and searchable mutant lines easy accessible for the plant scientists

## 1.4 Bioinformatical tools

Smith Waterman algorithm (S&W) was primarily designed for finding the best local alignment between two sequences<sup>39</sup>. The algorithm predicts all possible alignments between the two sequences and picks the local alignment with the highest score. S&W is made in two steps. First an Align Matrix (AM) that has the size  $l_1 \times l_2$ , where  $l_1$  are length of one of the sequences ( $s_1$ ) and  $l_2$  is the length of the other sequence ( $s_2$ ), are build. For each cell  $AM_{ij}$ , in the AM, where  $i$  is all values between 0 and  $l_1$  and  $j$  is all values between 0 and  $l_2$ , a score is assigned based on  $AM_{ij} = \max \{H_{j-1,i-1} + SM(s_1, s_2), \max \{H_{i-k,j} - G_k\}, \max \{H_{i,j-m} - G_m\}, 0\}$  where  $k$  is a value between 1 and  $i$  and  $m$  is a value between 1 and  $j$  and  $G$  is a gap penalty value. This means that for each cell the highest value, between aligning the two nt, inserting a gap or resetting the alignment, is chosen. When all cells are assigned in the AM the highest score is picked and then the best local alignment is build by backtracking from the highest back to where it started. S&W algorithm is still used to some extent because it tests all possibilities. A huge disadvantage is that it is a time consuming algorithm. Blast, which is based on an S&W algorithm is much faster but less accurate. For small regions, like the miRNA binding site, blast often miss to find similar regions. For comparison of two polypeptide sequences blast is a fast but crude method for assigning a common ancestor. For each hit blast gives an E-value. The E-value stands for how likely it is to find that kind of match by chance. If E-values are below  $10^{-4}$  the sequences can be concluded to originate from the same ancestor gene.

## 2 Material and Methods

### 2.1 Plant material and growth conditions

*Arabidopsis* seeds were surface-sterilized according to Fridborg *et al* (1999)<sup>40</sup>. Seeds from wild type *Arabidopsis*, ecotype Columbia (**Col**), were used and germinated on agar plates with Murashige&Skoog-medium. The plants were cultured in cool white fluorescent light at 20-22 °C under long day conditions. Samples were removed after two weeks of growth and used for further experiments or the seedlings were replanted in soil and grown at 20-22 °C under long day conditions.

### 2.2 DNA preparation

DNA was extracted from a Col plant according to Edward *et al* (1991)<sup>41</sup>. Small leaves were collected in a tube with extraction buffer, containing 250 mM NaCl and 25 mM EDTA in a 0,2 M TrisHCl buffer, pH 7.9, and 6-10 glass beads. The tubes were put in a FAST-prep machine where the cell walls are destroyed. 10 % SDS were added. The samples were centrifuged at 13000 rpm for 15 minutes. The supernatant was transferred to a new tube and an equal volume of isopropanol was added, in order to precipitate the DNA, and spun down, to gather the DNA from the sample. After centrifugation the supernatant was discarded and the pellet was left to air-dry and then dissolved in H<sub>2</sub>O. The samples were run on a 0.8% agarose-gel, asserting that DNA was present in each tube.

### 2.3 PCR amplification

The region of interested, the miRNA ath-miR169, was amplified using two ath-miR169 specific primers, **miR169-1b**(CCACtAtgaggatggagaagcatggagg) and **miR169-3**(agttacctcttctgcattgttc). Polymerase chain reaction (PCR) was carried out in a DNA Engine™, a PCR machine, in a total reaction mixture of 50 µl containing 0,3 mM dNTP(equal amount of each dNTP), 1 x PFU Buffer by Stratagene®, 1 U PFU polymerase and 0,4 µM each of **miR169-1b** and **miR169-3**. The temperature and time parameters were as follows: an initial denaturing period of 2 minutes at 95 °C, followed by 38 cycles of denaturation at 94 °C for 40 seconds, annealing at 58 °C for 1 minute and extension at 72 °C for 2 minutes. To ensure complete extension the PCR amplification was finished by a period of 10 minutes at 72°C followed by a period of 10 minutes at 4°C. Sterile water instead of DNA was used as a negative control. The PCR product was stored at 4 °C. In order to use the amplified product in a later step a nucleotide sequence of CCAC was put in the beginning at the 5' end of the forward primer and PFU enzyme was used asserting blunt end products.

### 2.4 Electrophoresis

Electrophoresis in a 0.8% agarose gel with 0.2 µg/mL ethidium bromide was used to size separate the PCR-products. The electrophoresis was preformed at 100 V for 1.5 hour, in 0.5M TBE pH 7.5. The PCR-products were stained with ethidium bromide and visualized using Syngene Bioimaging system™. A  $\lambda$ -PstI DNA ladder was used to determine the size of the products.

## 2.5 Cloning procedures

The pENTRY Directional TOPO<sup>®</sup> Cloning Kit was used to insert the blunt end ds DNA into a TOPO<sup>®</sup> vector. The TOPO vector consists of one blunt end and one end with a GTGG overhang which assures the right directionality of the inserted product. Fresh miR169 PCR product was mixed gently with salt solution and TOPO vector provided in the kit and left to incubate for 5 minutes at room temperature and then placed on ice. 2 µl of the TOPO cloning reaction were added to a vial of One Shot TOP10 chemically competent *Escherichia coli* (*E. coli*). The solution was mixed gently and then placed on ice for 30 minutes. In order to get the vector into the *E. coli*, the cells were heat-shocked for 30 seconds at 42 °C. The tubes were then put on ice. 250 µl room tempered SOC was added to each tube and incubated for 1 hour with shaking and put on LB-plates containing 50 µg/ml Kanamycin (**Kan**), selecting for cells with the inserted vector. A pENTR TOPO vector with the right insert will be referred to as p\*169Entry. To confirm that there was an insert of right size and direction, the plasmids were cut with restriction enzymes and visualized in gel electrophoresis. To determine direction of the insert and ensure that no PCR-generated sequence errors had been introduced part of the plasmid and the insert was sequenced.

The miR169 insert was transferred from the pENTR TOPO vector to a GATEWAY<sup>™</sup> vector: pk7WG2 by homologous recombination<sup>42</sup> placing a constitutive promoter, CaMV35S, in front of the insert. Protocol and solutions were included in the Gateway<sup>®</sup> LR Clonase<sup>™</sup> Enzyme Mix. p\*169Entry was used as the entry vector and pk7WG2 as the LR vector. In order to get the pk7WG2 with the miR169 insert into the cells the solution was heat shocked and then put on ice. SOC Medium was added and the tubes were left for 1 hour of shaking at 37 °C for incubation. Cells with the pk7WG2 were selected for using Spectinomycin (**Sty**) and Streptomycin (**Str**). pk7WG2 with the miR169 insert will be referred to as pk7WG2m169. pk7WG2 m169 was purified and size and direction was confirmed with restriction enzyme cleavage of the insert.

## 2.6 The making of a transgenic *Arabidopsis* plant

*A. tumefaciens* was grown in YEP medium at 28 °C. YEP contains 100 µg/ml rifampicin (**Rif**) and 40µg/ml Gentamycin (**GM**) assuring that only *A. tumefaciens* with the two plasmids GV3101 and pMP90, that carries Rif and Gm resistance and are vital for successful transformation of artificially constructed T-plasmids, are grown. pk7WG2 m169 was transformed to *A. tumefaciens* using the freeze and thaw method. *A. tumefaciens* cells were suspended in 1 ml ice-cold 20 mM CaCl<sub>2</sub> solution. 0.1 ml aliquots were dispensed in pre-chilled eppendorf test tubes. 10 µg pk7WG2 m169 was added and frozen in liquid Nitrogen. The solution was then thawed at 37 °C in a water-bath for 5 minutes. 1 ml YEP was added and the tubes were left to incubate at 28 °C for 2 hours with gentle shaking. The tubes were centrifuged at 13 000 rpm for 30 seconds and the supernatant was discarded and the cells were suspended in 0.1 ml fresh YEP-medium. The cells were then spread on YEP agar plates containing Rif, GM, Str, Sty selecting for cells that have GV3101, pMP90 and pk7WG2 m169. The plates were left for 3 days in 28 °C to create visible colonies. Colonies from the plates were picked and re-stroken on fresh YEP plates and put in 37 °C and 28 °C. The absence of colonies on the plates grown in 37 °C and the presence at 28 °C verifying the colonies to be *A. tumefaciens*. Colony PCR verifying the presence of a plasmid with the right insert was performed. The PCR reaction was carried out as described above (3.2.2), but

instead of adding DNA a sample of a colony was added and mixed with the primer solution **mir169-1b** and **mir169-3**, which amplifies the miRNA insert in pk7WG2 m169. The colony PCR product was run on a 0.8% agarose gel, according to the procedure described above (2.4), verifying that the colonies carries the insert. *A.tumefaciens* carrying the pk7WG2 m169 plasmid was introduced into *Arabidopsis* according to Bechtold *et al* (1993)<sup>43</sup>

## 2.7 *In silico* material

A Dell Precision 360 Computer with a 2.8 GHz Pentium IV hyper threading processor and 1 GB RAM is used for all computer work and used as a local web server and database server. The operating system is Windows (Win) XP professional. The webb server is Internet Information Service (IIS) 5.1 which is included in Win XP. MySQL was downloaded from <http://www.mysql.se/>(2003-09-20) . PHP 4.3.5 is used for handling of the web interface and was downloaded at <http://www.php.net/>(2003-09-20) . The programming language is Java, J2SDK1.4.1\_06, and was downloaded at <http://java.sun.com/>(2003-09-20). T\_Coffee was used for multi-sequence alignment and found at [http://igs-server.cnrsmrs.fr/~cnotred/-Projects\\_home\\_page/t\\_coffee\\_home\\_page.html](http://igs-server.cnrsmrs.fr/~cnotred/-Projects_home_page/t_coffee_home_page.html) (2003-09-20). A stand alone blast server was downloaded from <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/>(2003-09-20) . *Arabidopsis* annotated genome was downloaded from <ftp.arabidopsis.org/home/tair/Genes/>(2004-01-31). Information about different domains in predicted genes were downloaded at <ftp.arabidopsis.org/home/tair/Proteins/Domains/>(2004-03-01). Rice pseudo molecules of the chromosomes were downloaded from <ftp.tigr.org> (2004-03-01). Known information about miRNA in plants was collected from <http://www.sanger.ac.uk/Software/Rfam/mirna/index.shtml> (2003-09-25).

## 2.8 *In silico* methods

Computers handle comparison of integers much faster than comparison of character. In order to speed up the computing time characters were transformed to integers. Instead of using the characters A, C, G and U the integers 1, 2, 3 and 4 were used. All other pseudo nucleotides, that still not are assigned in the genomes, like N or P were assigned as 5. Since the miRNA binds the mRNA in the cytosol binary loadable libraries of both *Arabidopsis* and rice processed transcripts, meaning the sequence with the UTR and the coding sequence but without the introns, were created. This post-processed sequence libraries will be referred to as cDNA-libraries.

Score matrixes was used to speed up comparison between two nt, avoiding if statements in the code. A 5x5 large matrix was created aligning all possible nt pairs (Figure 5). The values assigned in the score matrix depend on what algorithm it is used in. Algorithms that look for complementarity on RNA level will use a score matrix with positive score for the AU, GC but also for GU and negative score for the others.

	A	C	G	U	N
A	AA	CA	GA	UA	NA
C	AC	CC	GC	UC	NC
G	AG	CG	GG	UG	NG
U	AU	CU	GU	UU	NU
N	AN	CN	GN	UN	NN

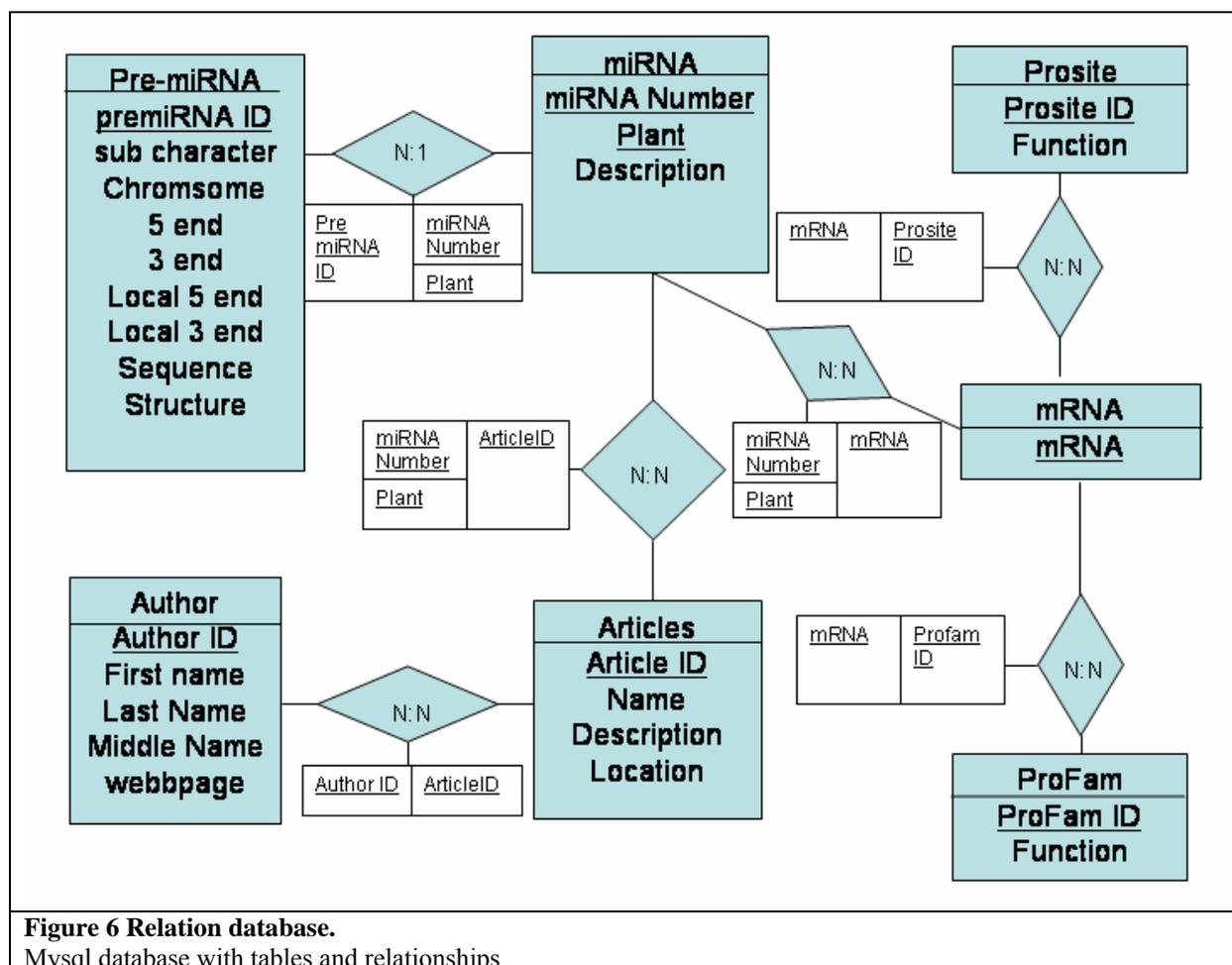
**Figure 5 Score matrix.**

Score matrices were build to speed up the algorithm processes. The scores in the score matrix is set to mimic binding between two RNA nt regarding to the algorithm.

### 3 Results

#### 3.1 Database

A database was set up for collection of new, relevant information on miRNA, and to make previously reported data on miRNA easily accessible. The database was built in mysql. In order to remove redundancy in the database the schema holds boyce/codd relationship. Relationships between Authors, Articles, miRNA, pre-miRNA, mRNA, proSite and proFam are build to establish links between the different parts (Figure 6). Because of the structure new relationships and tables are easy to add without disturbing the existing database. Information of coding genes is stored as xml documents.



#### 3.2 Website

The Website was build with PHP: Hypertext Preprocessor (**PHP**) for easy access of the mysql database and handling of information between different pages. Through the website everyone should be able to view the known miRNA in plants and have easy access to the different articles and authors that are related to miRNA research. Selected people at the lab that are assigned administrators of the website will be able to insert new articles, new miRNA and new pre-miRNA. Administrators will be able to start a new miRNA target search for new miRNA in the future and insert in the database. New administrators can be added or removed through the website. Most relationships in the mysql database are

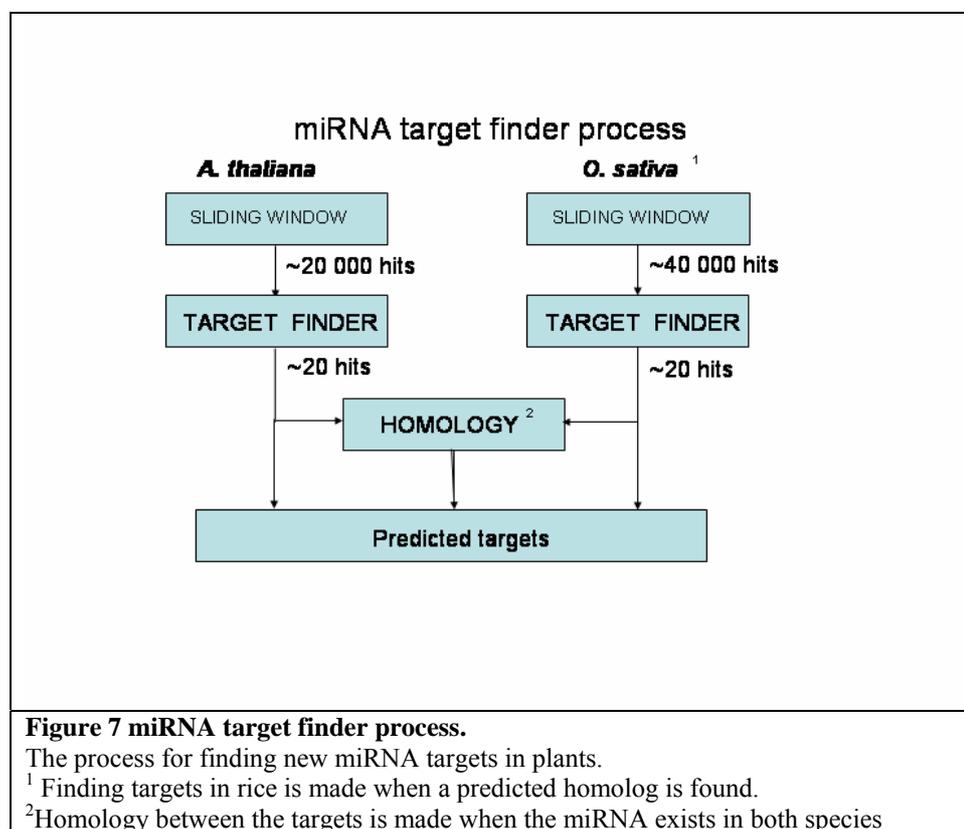
implemented on the website. A graphical interpretation of mRNA with domains and miRNA bindingsites has been created. A local blast server with website is accessible through the website.

### 3.3 Predicting targets for miRNA

Previous approaches to find miRNA targets that have been made have not been exhaustive where different groups have found different targets. Therefore a new approach of finding miRNA targets in plants was setup. The procedure for finding targets is divided into three steps (Figure 7). The first step is to find all alignments where there are seven or more nt complementarity in a row between a miRNA and a mRNA using a method referred to as the sliding window (see below). All hits from the first step are picked and run through the S&W based Target Finder program (see below) that allows all hits where the score is at least 75 percent of the maximum score. If the miRNA has a homolog in rice the same two steps where carried out in rice. As a third step the predicted targets for a miRNA in both plants are compared against each other on a peptide level. This step allowed for the findings of

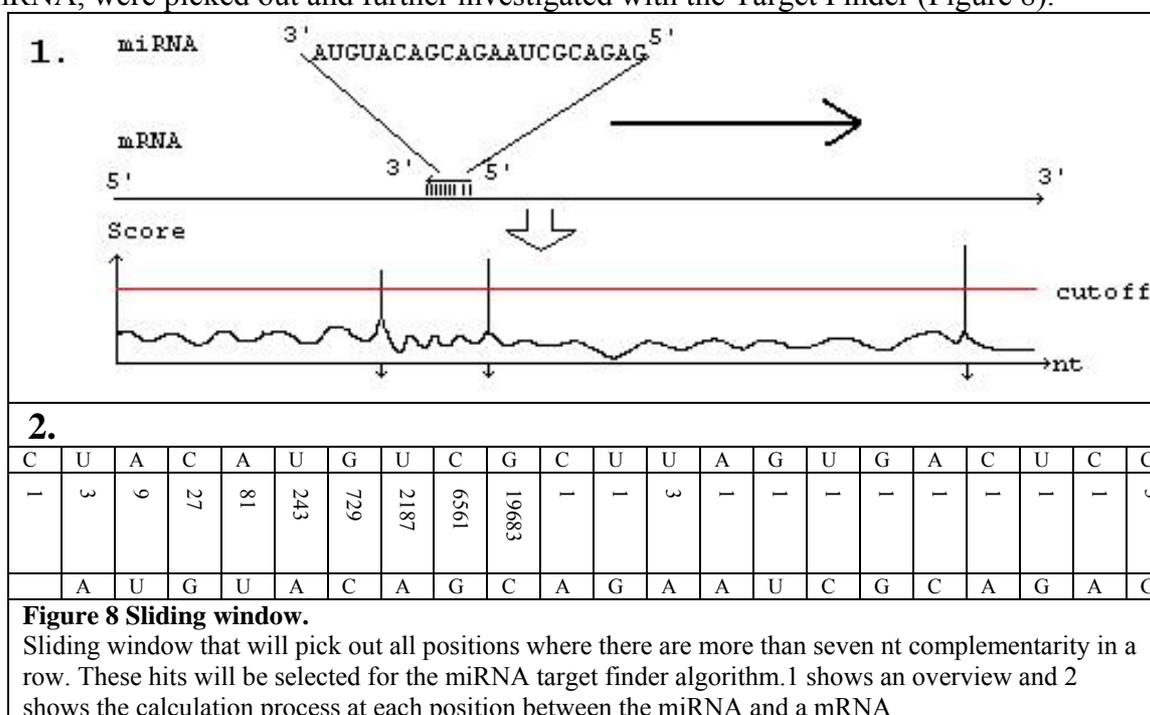
homologs among the predicted targets, not only between the plants but also within each plant to find if the miRNA regulate many genes in the same gene family. The search was divided in two steps because the sliding window algorithm computational time is linear in time to the size of the data and the Target Finder algorithm computational time is quadratic

to the size. A complete search with a miRNA that exists in both plants takes less than four minutes, making it possible to do searches online.



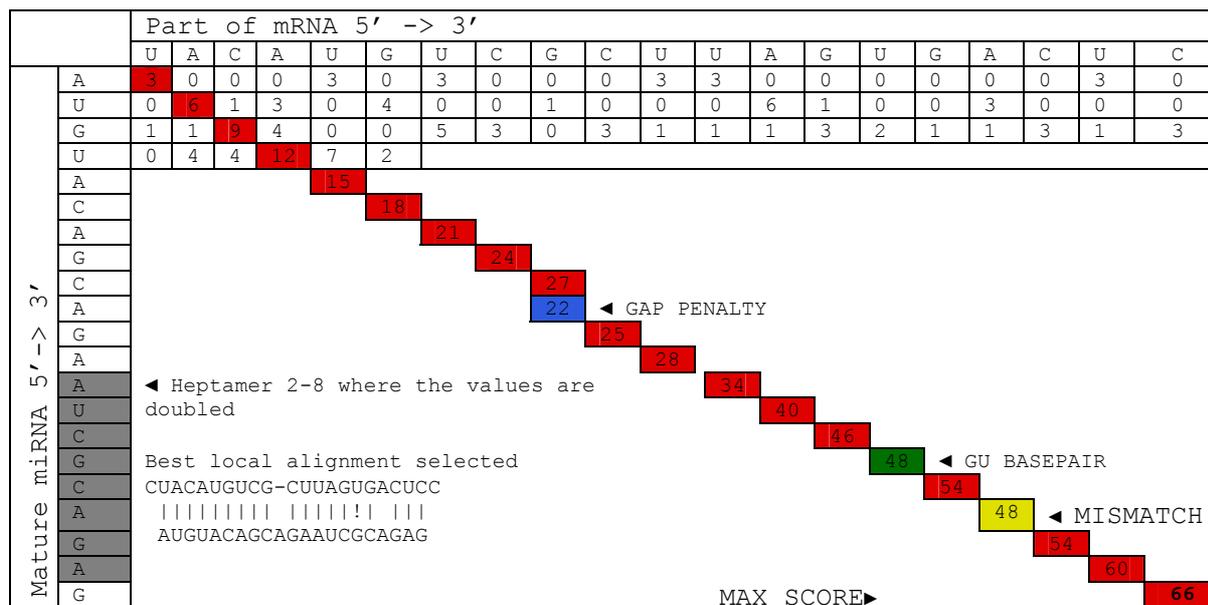
### 3.3.1 Sliding window

The method is referred to as the sliding window (**SW**) because of the action of sliding the miRNA over all position on a mRNA checking for short regions of complementarity. The miRNA, starting with the 3' end and ending with the 5' end, was compared against all the possible positions on all known mRNA in the plant. A score matrix was made for the local alignment (**LA**) in the SW, where a GC or an AU base pairing was assigned three points, GU wobbles were given one point and all the other were assigned zero points. On each mRNA the SW was applied on each position,  $SW_p = \max\{LA_{j,p+j}\}$ , where  $j$  is between 1 and the length of the miRNA and  $p$  is the current position on the mRNA. The LA algorithm goes from left to right assigning a value on each position in the putative miRNA:mRNA binding duplex based on the value to the left multiplied with the alignment in the score matrix. If the value turns 0, which it will each time there is a mismatch, the value was assigned to 1.  $LA_{j,p+j} = \max\{LA_{j-1,p+j-1} * LA_{sw}(miRNA_{length-j}, mRNA_{p+j}), 1\}$ . Positions where the  $SW_p > 2187$ , meaning that the miRNA has a stretch of seven or more nt with complementarity to the mRNA, were picked out and further investigated with the Target Finder (Figure 8).



### 3.3.2 Target Finder

For each position where the SW had a score above the cutoff (2187) a sequence of the length of the miRNA plus two adjacent nt of the mRNA were picked out and run against a S&W based algorithm that is referred to as the Target Finder (**TF**). Scores in the heptamer 2-8 of the miRNA were doubled giving a bias towards good alignments in this region. All alignments where the best local alignment score was above 75 percent of the maximum alignment score were picked out as potential miRNA targets (Figure 9, Table 4).



**Figure 9 Target Finder.** Target Finder work from left to right, top to bottom, deciding the score of a cell depending on the values of the previous cells. When the highest score is found the alignment is made through trace back. The blue cell shows a gap in the alignment, the green a GU basepair, the red a match and the yellow a mismatch. Only the best local alignment is highlighted. To be accepted as a miRNA target the score must be over 75 percent of max score, in this case 63.

### 3.3.3 Homology

If a given miRNA exists in both Arabidopsis and rice it is quite likely that it has homologous mRNA targets. When the Target Finder has produced a list of candidate genes that are above the threshold of 75 percent, the protein sequences of these targets are blasted against each other. Only hits with E-values below  $10^{-4}$ , using NCBI blastp, were assigned homologous miRNA to remove false positives. If homologs of predicted targets exist in Arabidopsis and rice at protein level the predicted binding sites are examined. If the miRNA binding site is conserved in both targets homologs, the TF candidate is picked out as a possible target (Table 4).

### 3.4 Prediction of new miRNA homologs in rice

Some of the known miRNA in Arabidopsis have a rice homolog with the same mature sequence in both plants (Table 2). A search for the homologs of the remaining miRNA was done. A number of criteria for assigning a sequence to be a rice miRNA homolog were setup. The miRNA sequences should not differ with more than three nt between the plants. The RNA sequence adjacent to the miRNA sequence in rice should be able to form a hairpin structure with good stability at the miRNA:miRNA\* site. The homolog in rice should have a predicted homologous target to the predicted targets of the miRNA in Arabidopsis. A predicted hairpin sequence and one hundred extra nt upstream of the sequence, that mimic a pri-miRNA, has to be in an intergenic region or in an intron of a gene. If all these criteria were satisfied the predicted miRNA in rice was assigned as a homologous miRNA. New homolog miRNA for miR157, miR159, miR160, miR168 (Figure 11) and miR319 were found using the criteria described above (Table 5). Two of the predicted miRNA, one of the



miRNA	TF Ara <sup>1</sup>	TF Rice <sup>2</sup>	Nr of Homologs <sup>3</sup>	Predicted targets <sup>4</sup>	Predicted targets function of classes
156	65	53	30(3)	19(1)	squamosa-promoter binding (SBP)-like proteins
157	47	35	32(4)	21(2)	SBP-like proteins(19) Putative DEAD-box RNA helicase(2)
158	27	No homologous miRNA found			No family found
159	29	20	17(3)	11(1)	myb family transcription factor
160	3	3	4(1)	4(1)	auxin response transcription factor
161	24				Pentatricopeptide repeat proteins <sup>5</sup>
162	5	8	2(1)	2(1)	Dicer Like Protein(DCL1)
163	13				SAM-dependent methyltransferases <sup>5</sup>
164	22	32	14(1)	14(1)	NAC domain proteins
165	9				HD-Zip transcription factors <sup>5</sup>
166	9	22	10(1)	10(1)	HD-Zip transcription factors
167	8	9	6(1)	6(1)	Auxin response factors
168	5	16	4(1)	4(1)	ARGONAUTE
169	12	9	10(1)	10(1)	CCAAT-binding factor
170	8	12	8(1)	8(1)	GRAS domain transcription factors
171	6	9	8(1)	8(1)	GRAS domain transcription factors
172	24				APETELA2-like transcription factors <sup>5</sup>
173	6	No homologous miRNA found			No family found
319	31	33	22(2)	22(2)	myb family transcription factor(9) TCP family transcription factor(13)

**Table 4 Predicted targets found using the miRNA target finder approach.**

Predicted targets using our approach based on TF and homology when the miRNA has been found in both species (3.3).

1 Number of putative targets found using the TF in Arabidopsis.

2 Number of putative targets found using the TF in rice.

3 The number of targets when doing homology search between targets of the two plants (number of different gene families).

4 The number of targets that, besides from being present in both Arabidopsis and rice, have a similar miRNA binding site (number of different gene families).

5 Predicted targets purely based on Target Finder and internal homology within Arabidopsis.



## 4 Discussion

miRNA is a hot topic and new information regarding new miRNA, which are their targets, how they regulate their targets and how they are processed is accumulating rapidly. An easy way to gather old and new information was needed. New evidence showed that the bioinformatical tools available for finding miRNA targets in plants would not give all the verified targets, therefore a new miRNA target finder was built.

### 4.1 miRNA target finder

A previously presented approach to find miRNA targets in plants had a cutoff at three mismatches and did not accept gaps between the miRNA and the mRNA. Some experimentally verified targets do however have more than three mismatches or gaps in the complementary region. Argonaute, which is a verified target for miR168 in Arabidopsis, has four mismatches and DCL-1, which is a verified target for miR162, has a gap in the complementary region. The approach to find miRNA targets in this study allows more mismatches and gaps in the alignment between the miRNA and the target. By using comparative genomics between and within the plants we can find more plausible target and still keep the same ratio between true and false negatives hits as in previous work. Our approach gives a more sensitive method in finding new targets without tampering on specificity. There are still much to find out regarding the binding of the RISC complex, miRNA and the target mRNA. When new information is published the TF could be more optimized to remove false positives and bring up false negatives. There are more plant genomes that are being sequenced right now and when they are published the homology step can play an even more important role.

An important factor that is not included in our approach is at what time and in which locations genes are expressed. So a miRNA with a perfect target complementarity will not affect expression unless they are expressed at the same time at the same place. One of the miRNA, miR319, gets two different predicted gene family classes of miRNA targets using our approach. One of the families is the MYB-genes and they are known to be regulated by miR159. Experimental data shows that the miR319 only affects the other class, the TCP family. This is thought to be due to differences in the expression pattern of the miRNA and the MYB-genes. Expression data for miRNA and mRNA could be another step in order to increase specificity. The SW and the TF are both written in an object oriented way so new features can easily be adopted.

Some targets that are not found in both Arabidopsis and rice are still interesting and will be tested. Since the plants diverged for more than 150 million years ago some diversity in targets is possible. To verify our new predicted targets 5' RACE experiments, which identify miRNA cleaved products, will be done.

### 4.2 miRNA finder

No bioinformatical approach to predict new miRNA in plants has been done. As the pre-miRNA structures in plants vary very much among miRNA duplicates, it is hard to define search criteria based on the length and structure of the pre-miRNA in plants. The fact that rice and Arabidopsis are the only plant genomes yet published has made comparative

genomics, which have been a key feature in finding new miRNAs in mammals, a very blunt tool. Since the miRNA and its target are well conserved in plants our approach to finding targets could be an important step in finding new miRNAs in plants.

### 4.3 miR169s transformed Arabidopsis

Even if some aberrant phenotypes could be found among the individuals of the first generation of transgenic plants expressing the introduced gene miR169, no general effect on phenotypic expression could be concluded at this stage. The transgenic plant will be used for further studies of miR169. Experiments involving the transgenic plants will not be carried out on the plant before the third generation in order to ensure a single homozygote inserts. Therefore no results could be presented at this point.

### 4.4 Database

The sequence and prediction of the rice genome is still under construction and must be updated relatively soon. Since the targets for a miRNA often are of the same gene family, relationships should be added for both Arabidopsis and rice using ProFam. Domains for the mRNA in rice should also be added, perhaps speeding up the search. Protocols and methods should be added making it easy accessible to get hold of methods and to know which miRNA different methods have been used.

### 4.5 Website

Since the website accesses many different processes on the computer a hostile attack against the website could cause severe damages not only to the computer which the website resides but also to other computers on the network. Therefore the website is not accessible from the web but only at the local network at ICM. Therefore the website is still under construction and some kind of security increase should be made before it can be accessible from the web. Improvement on stability and a more user-friendly interface is a pending factor when designing a web interface.

## 5 Acknowledgments

I would like to express my warmest gratitude to my supervisor Sandra Kuusk for sharing her knowledge and enthusiasm in the fields of plants and miRNA and for her patience and help in making this project possible. I would like to thank Pontus Larsson for his ideas and thoughts concerning the bioinformatical part of the project. Three cheers and one hooray to the people in the microbiology group at ICM for openly accepting me in the lab. Many thoughts to Lisa for doing more than her work so that I could do mine. Emil for giving me perspective to what I do. Finally, I thank Gerhart Wagner for his great wisdom and for reviewing my report.

## 6 References

- 1 Campbell, Biology (Fourth Edition)
- 2 Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EG, Margalit H, Altuvia S. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*, *Curr. Biol.* **11**, 941 (2001)
- 3 Rivas E, Klein RJ, Jones TA, Eddy SR. Computational identification of noncoding RNAs in *E. coli* by comparative genomics, *Curr. Biol.* **11**, 1369(2001)
- 4 Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S. Identification of novel small RNAs using comparative genomics and microarrays, *Genes Dev.* **15**, 1637 (2001)
- 5 Lee RC, Feinbaum RL, Ambros V. The heterochronic gene *lin-4* of *C. elegans* encodes small RNAs with antisense complementarity to *lin-14*, *Cell* **75**, 843 (1993)
- 6 Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G. The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*, *Nature* **403**, 901 (2000)
- 7 Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of Novel Genes Coding for Small Expressed RNAs. *Science* **294**, 853 (2001)
- 8 Lau NC, Lim LP, Weinstein EG, Bartel DP. An Abundant Class of Tiny RNAs with Probable Regulatory Roles in *Caenorhabditis elegans*, *Science* **294**, 858 (2001)
- 9 Lee RC, Ambros V. *et al* An Extensive Class of Small RNAs in *Caenorhabditis elegans*, *Science* **294**, 862 (2001)
- 10 Llave C, Kasschau KD, Rector MA, Carrington JC. Endogenous and silencing-Associated Small RNAs in Plants, *Plant Cell* **14**, 1605 (2002)
- 11 Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP. MicroRNAs in plants, *Genes Dev.* **16**, 1616 (2002)
- 12 Park W, Li J, Song R, Messing J, Chen X. CARPEL FACTORY, a Dicer Homolog, and HEN1, a Novel Protein, Act in microRNA Metabolism in *Arabidopsis thaliana*, *Curr. Biol.* **12**, 1484 (2002)
- 13 Ruvkun G. Glimpses of a Tiny RNA World, *Science* **294**, 797 (2001)
- 14 Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. Identification of Tissue-Specific MicroRNAs from Mouse, *Curr. Biol.* **12**, 735 (2002)
- 15 Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T. New microRNAs from mouse and human, *RNA* **9**, 175 (2003)
- 16 Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, Burge CB, Bartel DP. The microRNAs of *Caenorhabditis elegans*, *Genes Dev.* **17**, 991 (2003)
- 17 Mourelatos Z, Dostie J, Paushkin S, Sharma A, Charroux B, Abel L, Rappsilber J, Mann M, Dreyfuss G. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs, *Genes Dev.* **16**, 720 (2002)
- 18 Lim LP, Glasner ME, Yekta S, Burge CB, Bartel DP. Vertebrate MicroRNA Genes, *Science* **299**, 1540 (2003)
- 19 Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, Carrington JC, Weigel D. Control of leaf morphogenesis by microRNAs, *Nature* **425**, 257 (2003)
- 20 Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*, *Nature* **391**, 806 (1998)
- 21 Hunter C. P. Gene silencing: shrinking the black box of RNAi, *Curr. Biol.* **10**, 137 (2000)
- 22 Carthew R. W. Gene silencing by double-stranded RNA, *Curr. Opin. Cell Biol.* **13**, 244 (2001)
- 23 Sharp P. A. RNA interference, *Genes Dev.* **15**, 485–490 (2001)

- 24 Vance V, Vaucheret H. RNA silencing in plants - defense and counterdefense, *Science* **292**, 2277 (2001)
- 25 Jenuwein An RNA-Guided Pathway for the Epigenome, *Science* **297**, 2215 (2002)
- 26 Doench JG, Petersen CP, Sharp PA. siRNAs can function as miRNAs *Genes Dev.* **17**, 438 (2003)
- 27 Zeng Y, Yi R, Cullen BR. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms, *Proc. Natl. Acad. Sci.* **100**, 9779 (2003)
- 28 Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, Mao M, Li B, Cavet G, Linsley PS. Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* **21**, 635 (2003)
- 29 Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, Matzke M, Ruvkun G, Tuschl T. A uniform system for microRNA annotation, *RNA* **9**, 277 (2003)
- 30 Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN. The nuclear RNase III Drosha initiates microRNA processing, *Nature* **425**, 415 (2003)
- 31 Lee Y, Jeon K, Lee JT, Kim S, Kim VN. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.* **21**, 4663 (2002)
- 32 Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets, *Cell* **115**, 787 (2003)
- 33 Stark A, Brennecke J, Russell RB, Cohen SM. Identification of Drosophila microRNA targets, *PLOS Biol.* **1**, 60 (2003)
- 34 Floyd SK, Bowman JL. Ancient microRNA target sequences in plants, *Nature* **428** 485 (2004)
- 35 The Arabidopsis Genome Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, *Nature* **408**, 796 (2000)
- 36 Gelvin S. B. Agrobacterium and plant genes involved in T-DNA transfer and integration, *Annual Review of Plant Physiology and Plant Molecular Biology* **51**, 223 (2000)
- 37 Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S. A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica), *Science* **296** 92 (2002)
- 38 Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H. A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica), *Science* **296** 79 (2002)
- 39 Smith TF, Waterman MS. Identification of common molecular subsequences, *J Mol Biol.* **147**, 195 (1981)

- 40 Fridborg I, Kuusk S, Moritz T, Sundberg E. The Arabidopsis Dwarf Mutant shi Exhibits Reduced Gibberellin Responses Conferred by Overexpression of a New Putative Zinc Finger Protein, *Plant Cell* **11**, 1019 (1999)
- 41 Edwards K, Johnstone C, Thompson C. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis, *Nucleic Acids Res.* **19**, 1349 (1991)
- 42 Karimi M, Inze D, Depicker A. GATEWAY™ vectors for Agrobacterium-mediated plant transformation. *Trends in plant Science* **7**, 193 (2002)
- 43 Bechtold N, Pelletier G. In planta Agrobacterium mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants, *C.R. Acad. Sci. Paris Life Sci.* **316**, 1194 (1993)