



小麦种子发育过程中的 miRNA及其表达

G07 张静文 桂月靖 刘浩 刘声传

主要内容

- 一、背景
- 二、方法
- 三、结果
- 四、总结

一、背景

MicroRNAs (miRNAs) 是在真核生物中发现的一类内源性的具有调控功能的非编码RNA，其大小长约20~25个核苷酸。成熟的miRNAs是由较长的初级转录物经过一系列核酸酶的剪切加工而产生的，随后组装进RNA诱导的沉默复合体，通过碱基互补配对的方式识别靶mRNA，并根据互补程度的不同指导沉默复合体降解靶mRNA或者阻遏靶mRNA的翻译。最近的研究表明miRNA参与各种各样的调节途径，包括发育、病毒防御、器官形成、细胞增殖和凋亡等。

对于禾谷类作物来说，产量直接与种子相关。Xue等人对水稻种子进行了研究发现，miRNA在水稻种子发育过程中起到了重要作用。小麦是一种重要的粮食作物，平常所指的小麦是禾本科植物小麦 (*Triticum aestivum*) 的种子。小麦种子的好坏与粮食产量息息相关，因此提高小麦种子产量和质量都具有重要的意义。miRNA在植物的各个生长发育阶段中都发挥着重要作用，但是在小麦中还没有人研究种子发育过程中不同阶段的miRNA。

为了探讨miRNA在小麦种子发育过程中的作用，本研究对豫麦18开花后5、10、15、25天四个时期的种子进行了高通量测序，通过生物信息学手段对测序结果进行发掘，寻找已知的保守miRNA的表达情况以及预测新的miRNA；并通过对miRNA的靶基因的预测，分析miRNA在小麦种子发育过程中的表达模式及作用，以期发现与小麦种子发育有关的miRNA。

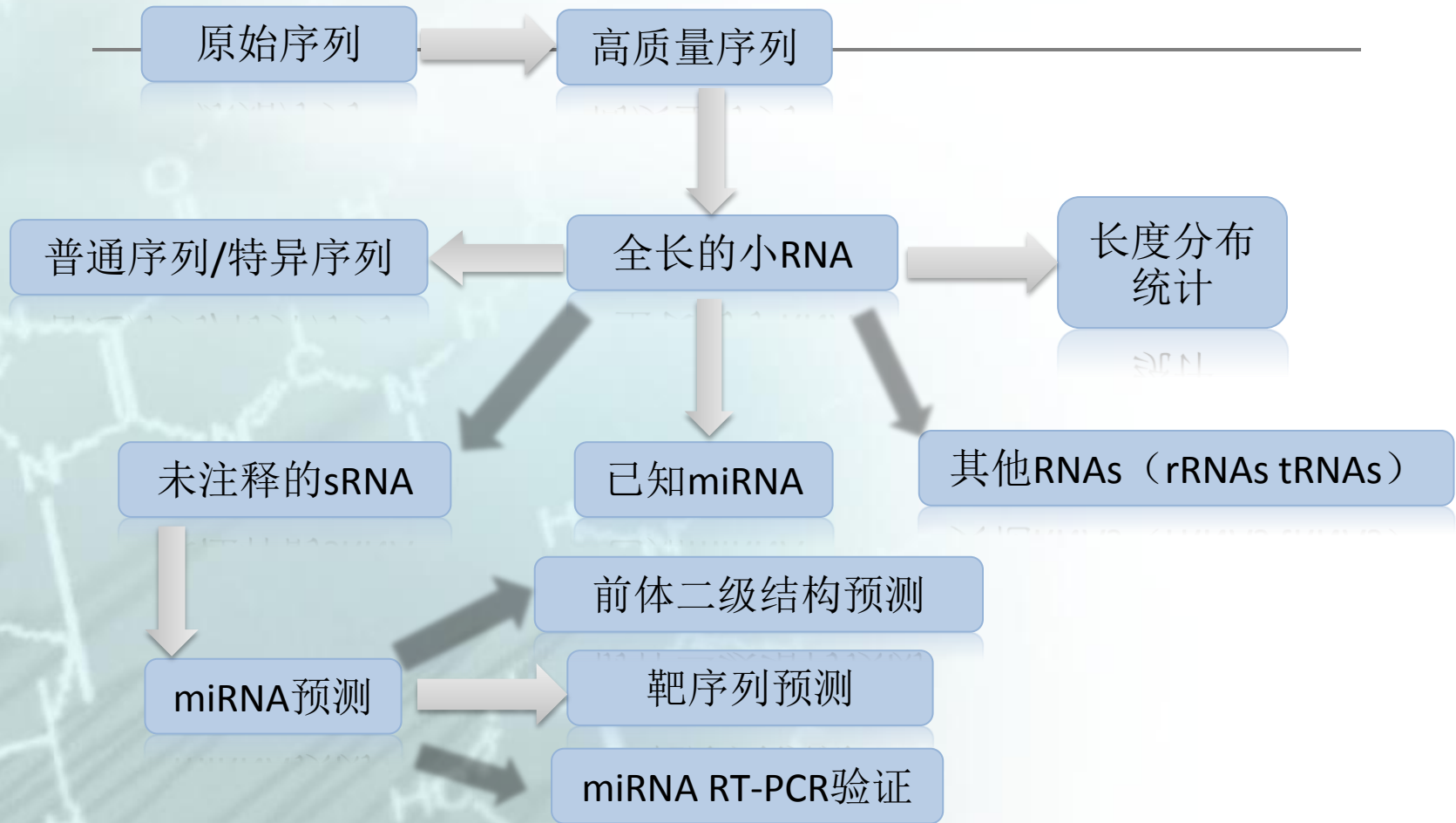
这对于研究小麦种子在发育过程中基因表达调控模式，探究miRNA与小麦种子形的关系，发现与产量性状有关的miRNA，有着重要的理论与现实意义。

二、方法

实验方法

1. 小RNA的高通量测序
2. 测序结果的生物信息学分析

高通量测序数据处理流程



三、结果与分析

- 高通量测序结果分析
 - 原始测序结果质量分析
 - 小RNA长度分布
 - 小RNA测序结果与参考基因组比对及小RNA分类注释
 - 小麦种子发育过程中已知miRNA的分析
 - 小麦种子发育过程中新发现的miRNA及靶基因预测

小RNA测序结果与参考基因组比对

- 通过SOAP将sRNA定位到基因组，分析sRNA在基因组上的表达和分布情况。
- 对于某些物种来说，miRNA基因也是成簇存在的，这样根据分布状况，就能方便的预测新miRNA

程序与参数

使用的为华大基因的SOAP软件，除此之外还有BOWTIE等软件。

```
./soap -v 0 -r 2 -s 7 -p 7 -a /home/lhac/clean.fa -d /home/lhac/TGA/split/5.fa -o 5.soap
```

SOAP2使用需要先建立索引文件：

```
./2bwt-builder /home/lhac/TGA/split/5.fa
```

然后使用SOAP2

```
./soap -a /home/lhac/clean.fa -D /home/lhac/TGA/split/5.fa.index -o 5.soap
```

id	seq	qual	#_of_hits	a/b[belonging to query a or b if pair-end]	length	+/-	ref	ref_location	type
t0000021	GATTCGTTTCGACTACGCTT	hhhhhhhhhhhhhhhhhhhh	1	a	19	-	contig00115	9640	0
t0000154	TCCGTCGTAGTCTAGGTGGTTA	hhhhhhhhhhhhhhhhhhhh	1872	a	22	+	contig00106	3868	0
t0000154	TCCGTCGTAGTCTAGGTGGTTA	hhhhhhhhhhhhhhhhhhhh	1872	a	22	+	contig00106	997	0
t0000318	TCCGTCGTAGTCTAGGTGGTT	hhhhhhhhhhhhhhhhhhhh	1872	a	21	+	contig00106	3868	0
t0000318	TCCGTCGTAGTCTAGGTGGTT	hhhhhhhhhhhhhhhhhhhh	1872	a	21	+	contig00106	997	0
t0000681	TTGTCTTCTTTGCCACAAGTACT	hhhhhhhhhhhhhhhhhhhh	14	a	24	-	contig00172	5216	0
t0001002	CGTGGCCTGAAAAAAGTAGATGAG	hhhhhhhhhhhhhhhhhhhh	704	a	24	+	contig00202	1990	0
t0001131	CATCTATTTTGGAACGGAGGG	hhhhhhhhhhhhhhhhhhhh	1846	a	21	+	contig00002	3	0
t0001131	CATCTATTTTGGAACGGAGGG	hhhhhhhhhhhhhhhhhhhh	1846	a	21	+	contig00045	4251	0
t0001234	CAAAGTGGAGAATCTGCATGTCAC	hhhhhhhhhhhhhhhhhhhh	103	a	24	+	contig00135	7493	0
t0001234	GTGACATGCAGATTCTCCACTTTG	hhhhhhhhhhhhhhhhhhhh	103	a	24	-	contig00046	3408	0
t0001322	ATTCGTTTCGACTACGCTT	hhhhhhhhhhhhhhhhhhhh	1	a	18	-	contig00115	9641	0
t0002169	GTGGCCTGAAAAAAGTAGATGAGT	hhhhhhhhhhhhhhhhhhhh	704	a	24	+	contig00202	1991	0
t0002651	TATTTGTCTTCTTTGCCACAAGT	hhhhhhhhhhhhhhhhhhhh	14	a	24	-	contig00172	5213	0
t0002870	AAAGTGGAGAATCTGCATGTCACA	hhhhhhhhhhhhhhhhhhhh	103	a	24	+	contig00135	7494	0
t0002870	TGTGACATGCAGATTCTCCACTTT	hhhhhhhhhhhhhhhhhhhh	103	a	24	-	contig00046	3407	0
t0003147	TCCGTCGTAGTCTAGGTG	hhhhhhhhhhhhhhhhhhhh	1873	a	18	+	contig00058	10260	0
t0003147	TCCGTCGTAGTCTAGGTG	hhhhhhhhhhhhhhhhhhhh	1873	a	18	+	contig00106	3868	0
t0003147	TCCGTCGTAGTCTAGGTG	hhhhhhhhhhhhhhhhhhhh	1873	a	18	+	contig00106	997	0

小RNA分类注释

我们将小麦小RNA与Rfam , miRBase 进行了比对 , 最后进行了注释。

Rfam数据库下载于 :

<ftp://ftp.sanger.ac.uk/pub/databases/Rfam/CURRENT/Rfam.fasta.gz>

miRBase数据库下载于 :

<ftp://mirbase.org/pub/mirbase/CURRENT/hairpin.fa.gz>

程序与参数

- BLAST数据库建立(BLAST版本:2.2.27+) :

- (linux) makeblastdb -in /home/lhac/hairpin.fa -dbtype nucl -title HAIR -out HAIR

- (Windows)makeblastdb -in D:\blast\Rfam.fasta -dbtype nucl -title rfam -out D:\blast\Rfam

- 使用BLASTN程序将测序结果与数据库比对 :

- (linux)blastn -task blastn-short -query 2.fa -db HAIR -out 2tomir -evaluate 0.01 -dust no -
num_threads 4 -outfmt 6

- (windows) blastn -task blastn-short -query D:\blast\1.fa -db d:\blast\rfam -out
D:\blast\1torfam.out -evaluate 0.01 -dust no -outfmt 6

与miRBase比对结果

Query id	Subject id	% identity	alignment length	mismatches	gap openings	q. start	q. end	s. start	s. end	e-value	bit score
t0000001	ssp-MIR168a	100	21	0	0	1	21	18	38	4.00E-06	42.1
t0000001	bdi-MIR168	100	21	0	0	1	21	31	51	4.00E-06	42.1
t0000001	hvu-MIR168	100	21	0	0	1	21	23	43	4.00E-06	42.1
t0000001	zma-MIR168b	100	21	0	0	1	21	21	41	4.00E-06	42.1
t0000001	zma-MIR168a	100	21	0	0	1	21	21	41	4.00E-06	42.1
t0000001	sof-MIR168a	100	21	0	0	1	21	21	41	4.00E-06	42.1
t0000001	sbi-MIR168	100	21	0	0	1	21	21	41	4.00E-06	42.1
t0000001	osa-MIR168a	100	21	0	0	1	21	11	31	4.00E-06	42.1
t0000001	nta-MIR168c	95.24	21	1	0	1	21	11	31	9.00E-04	34.2
t0000001	nta-MIR168b	95.24	21	1	0	1	21	11	31	9.00E-04	34.2
t0000002	cme-MIR393a	100	12	0	0	6	17	82	93	0.67	24.3
t0000002	mdm-MIR393f	100	12	0	0	5	16	87	98	0.67	24.3
t0000002	mdm-MIR393a	100	12	0	0	5	16	87	98	0.67	24.3
t0000002	tcc-MIR393a	100	12	0	0	6	17	76	87	0.67	24.3
t0000002	aly-MIR393a	100	12	0	0	6	17	108	119	0.67	24.3
t0000002	vvi-MIR393a	100	12	0	0	6	17	92	103	0.67	24.3
t0000002	vvi-MIR393b	100	12	0	0	6	17	49	60	0.67	24.3
t0000002	ptc-MIR393c	100	12	0	0	5	16	76	87	0.67	24.3
t0000002	ptc-MIR393b	100	12	0	0	6	17	54	65	0.67	24.3
t0000002	ptc-MIR393a	100	12	0	0	6	17	55	66	0.67	24.3

与Rfam比对结果

Query id	Subject id	% identity	alignment length	mismatches	gap openings	q. start	q. end	s. start	s. end	e-value	bit score	
t0000001	RF00677;MIR168;ABXC01002411.1/32850-32747	100	21	0	0	0	1	21	19	39	7.00E-05	42.1
t0000001	RF00677;MIR168;ADDN01000763.1/307587-307484	100	21	0	0	0	1	21	19	39	7.00E-05	42.1
t0000001	RF00677;MIR168;AAAA02005008.1/24372-24269	100	21	0	0	0	1	21	19	39	7.00E-05	42.1
t0000003	RF00075;mir-166;GQ424377.1/265-178	100	21	0	0	0	1	21	63	83	7.00E-05	42.1
t0000003	RF00075;mir-166;CACCO1009518.1/3128-3041	100	21	0	0	0	1	21	63	83	7.00E-05	42.1
t0000003	RF00075;mir-166;CACCO1004759.1/68003-67914	100	21	0	0	0	1	21	65	85	7.00E-05	42.1
t0014080	RF00001;5S_rRNA;AEJG01010382.1/2790-2672	100	19	0	0	0	1	19	98	116	1.00E-03	38.2
t0014080	RF00001;5S_rRNA;AEJG01007543.1/132-14	100	19	0	0	0	1	19	98	116	1.00E-03	38.2
t0014080	RF00001;5S_rRNA;BABK01058006.1/962-843	100	19	0	0	0	1	19	99	117	1.00E-03	38.2
t0085064	RF00005;tRNA;BX640549.17/74037-73967	100	20	0	0	0	1	20	50	69	4.00E-04	40.1
t0085070	RF00002;5_8S_rRNA;EF418807.1/256-411	100	27	0	0	0	1	27	62	88	3.00E-08	54
t0085070	RF00002;5_8S_rRNA;AEC001044462.1/1713-1847	100	27	0	0	0	1	27	41	67	3.00E-08	54

小麦种子发育过程中新发现的miRNA

- miRNA预测的主要标准为含有茎环结构的pre-miRNA的“茎”的部位的精确切割。茎环前体结构是通过软件根据已知的ESTs序列进行预测的。由于植物产生的茎环结构前体在长度上变化范围比较大，所以预测主要有以下几个要点：（1）miRNA/miRNA*是从茎环结构上相反的两臂上产生的，并且有3'端有多于两个碱基。（2）miRNA与含有miRNA*的茎环结构上的臂的序列之间的配对，不能多于4个错配，（3）双链不对称的凸起里所包含的碱基不多于两个，并且这样的突起在配对区域不多于一个。
- miRNA预测使用的是华大基因开发的Mireap软件，预测过程为：（1）将比对上参考序列的sRNA两端截取一定长度的序列，（2）将该序列进行折叠，预测MFE以及按照上述的准则筛选折叠后的序列

小麦种子发育过程中新发现的miRNA

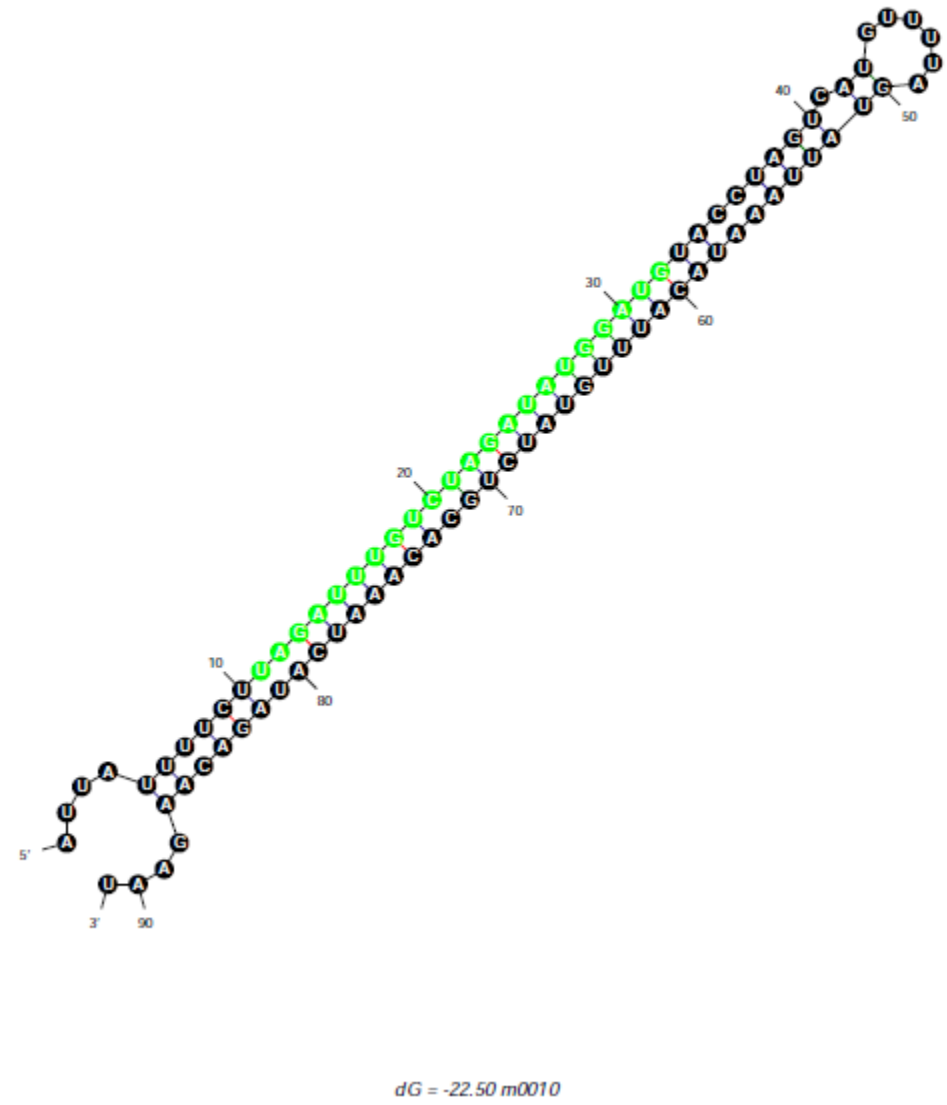
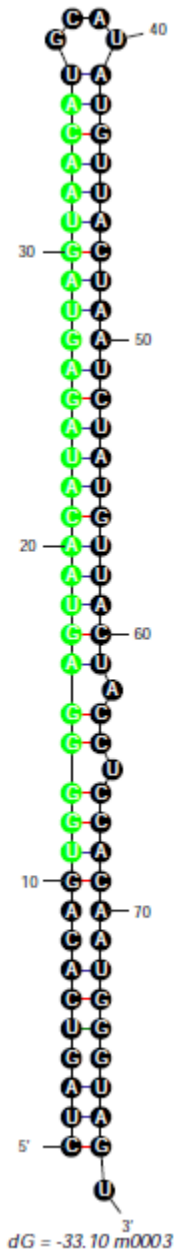
- 预测时候所使用的参数如下：
- miRNA最小长度 (18)
- miRNA最大长度 (25)
- miRNA最小参考序列长度 (20)
- miRNA最大参考序列长度 (23)
- miRNA在参考序列上的最大拷贝 (20)
- 最大自由能限制 (-18 kcal/mol)
- miRNA和miRNA*之前最大长度 (300)
- miRNA和miRNA*间最少配对碱基数 (16)
- miRNA和miRNA*间最大凸起 (4)
- miRNA/miRNA*复合体最大不对称数目 (4)
- miRNA在前体上侧翼序列截取长度 (20)
- 通过Mireap软件进行预测，共发现18个miRNA
- `mireap.pl -d 300 -p 16 -s 4 -f 10 -i '/home/lhac/clean.fa' -m '/home/lhac/match/map' -r /mnt/DOWN/genomic_Chinese_Spring_assembly.fasta -o /home/lhac/resultal/`

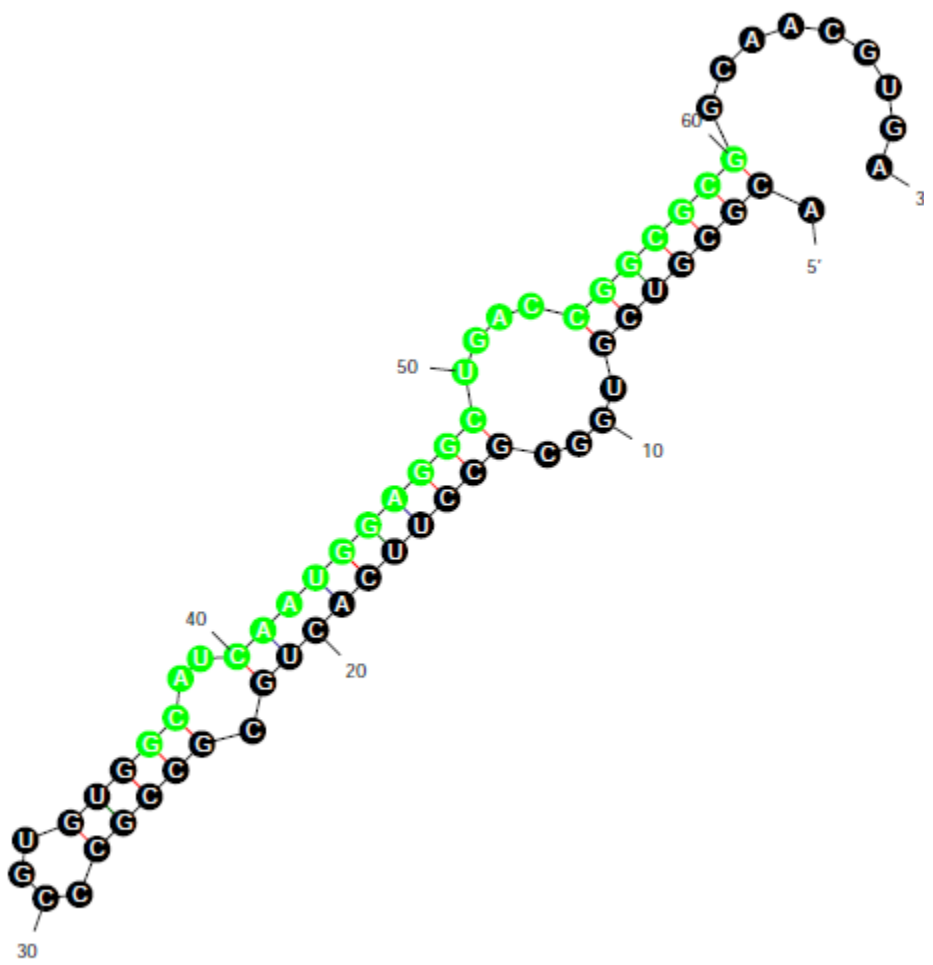
Mireap软件预测结果

```

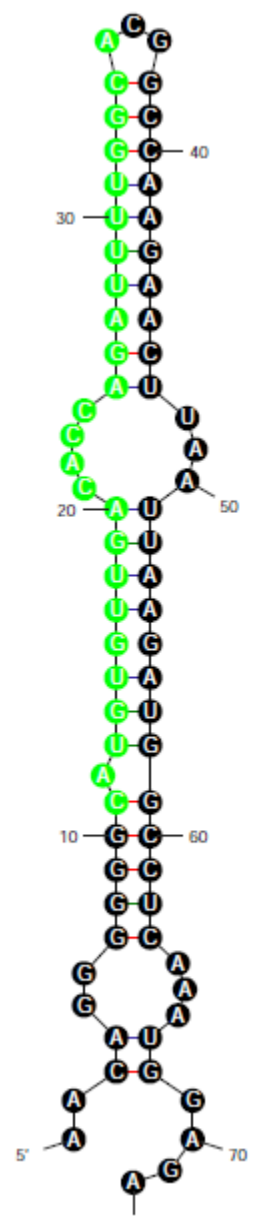
■ mireap
■ xxx-m0004 contig00045:4195:4281:+ 87(nt) -44.40(kcal/mol)
■ ATGACATACTCCCTCTGTTCCAAAACAGATGACCCAAATTTATACTAACTTTGGGCCATCTATTTTGG AACGGAGGGAGCTCGGCGT xxx-m0004 949
■ ..... (((((((((((((((((((((( (((((((((((((((((((((( (((((((((((((((((((((( .....
■ *****CATCTATTTTGG AACGGAGGG***** xxx-m0004-3p 943
■ -----CATCTATTTTGG AACGGAGG----- t0186797 8
■ -----CATCTATTTTGG AACGGAGGG----- t0001131 623
■ -----CATCTATTTTGG AACGGAGGGA----- t0590723 2
■ -----CATCTATTTTGG AACGGAGGGAG----- t0289695 5
■ -----ATCTATTTTGG AACGGAG----- t0508956 3
■ -----ATCTATTTTGG AACGGAGG----- t0241908 6
■ -----ATCTATTTTGG AACGGAGGG----- t0101637 16
■ -----ATCTATTTTGG AACGGAGGGA----- t0007082 176
■ -----ATCTATTTTGG AACGGAGGGAG----- t0628336 2
■ -----TCTATTTTGG AACGGAGG----- t0499336 3
■ -----TCTATTTTGG AACGGAGGGA----- t0476372 3
■ -----TCTATTTTGG AACGGAGGGAG----- t0017469 84
■ -----CTATTTTGG AACGGAGGG----- t1015852 1
■ -----CTATTTTGG AACGGAGGGA----- t0446718 3
■ -----CTATTTTGG AACGGAGGGAG----- t0189990 8
■ -----TATTTTGG AACGGAGGGA----- t0482194 3
■ -----TATTTTGG AACGGAGGGAG----- t0409713 3
■ //

```





$dG = -28.10 \text{ m0007}$



$dG = -16.90 \text{ m0016}$

新预测的miRNA的靶基因的预测

靶基因的预测使用的是psRNA Target，这是一个基于web的miRNA靶基因预测软件，并且已经包含了许多物种的序列信息，使用方便。

<http://www.plantgrn.org/psRNA Target/>

miRNA_Acc	Target_Acc	Expectation	UPE	miRNA_start	miRNA_end	Target_start	Target_end	miRNA_aligned_fragment	Target_aligned_fragment	Inhibition	Target_Desc.	multiplicity
m0001-5p	CA694965	0	13.308	1	24	69	92CGUUGGUGUCAAGCUCAGAGAUGU	ACAUUCUGAGCUUGACACCAACG	Cleavage			1
m0001-5p	BJ274961	0.5	14.459	1	24	57	80CGUUGGUGUCAAGCUCAGAGAUGU	ACAUUCUGAGCUUGACAUCAACG	Cleavage		similar to UniRef100_Q754D0 Cluster: AFR140Cp; n=1; Eremothecium gossypii Rep: AFR140Cp - Ashbya gossypii (Yeast), partial (3%)	1
m0001-5p	GH729969	1	17.136	1	24	46	69CGUUGGUGUCAAGCUCAGAGAUGU	ACAUUCUGAGCUUGAUCCCAACG	Cleavage			1
m0001-5p	CJ887622	1.5	18.568	1	24	21	44CGUUGGUGUCAAGCUCAGAGAUGU	ACAUUCUUGAGUUUGCACCAACG	Cleavage			1
m0001-5p	TC441034	2	18.477	1	23	770	792CGUUGGUGUCAAGCUCAGAGAUG	CACCUUUGAGGUUGACGCCAACG	Cleavage		homologue to UniRef100_Q24581 Cluster: Luminal-binding protein 3 precursor; n=1; Zea mays Rep: Luminal-binding protein 3 precursor - Zea mays (Maize), partial (57%)	1
m0001-5p	TC422779	2.5	20.883	1	23	469	491CGUUGGUGUCAAGCUCAGAGAUG	CACCUCCGAGGUUGACGCCAACG	Cleavage		homologue to UniRef100_Q40058 Cluster: HSP70 precursor; n=1; Hordeum vulgare Rep: HSP70 precursor - Hordeum vulgare (Barley), partial (30%)	1
m0002-3p	EC907531	1	23.367	1	21	73	93AUACAUCUGUAUCUAGACAAA	UUUGUCUGGAUAUGAUGUAU	Cleavage			1
m0002-3p	CK200923	1.5	14.889	1	21	227	247AUACAUCUGUAUCUAGACAAA	UUUGUCUAGAUACGGAUGUAC	Cleavage		homologue to UniRef100_A1ZXQ5 Cluster: PK domain protein; n=1; Microscilla marina ATCC 23134 Rep: PK domain protein -, partial (0%)	1
m0002-3p	AL819672	1	9.632	1	21	265	285AUACAUCUGUAUCUAGACAAA	UUUGUCAAGAUACAGAUGUAU	Cleavage		homologue to UniRef100_Q2PP61 Cluster: Sodium/hydrogen exchanger; n=1; Hordeum vulgare Rep: Sodium/hydrogen exchanger - Hordeum vulgare (Barley), partial (13%)	1
m0002-3p	TC445283	1.5	11.082	1	21	341	361AUACAUCUGUAUCUAGACAAA	UUUGUCUAGAUACAAUGUAU	Cleavage			1
m0002-3p	TC432307	1	13.836	1	21	608	628AUACAUCUGUAUCUAGACAAA	UUUGUCUAGAUACUGAUGUAU	Cleavage		similar to UniRef100_Q0E2N4 Cluster: Os02g0225300 protein; n=1; Oryza sativa Japonica Group Rep: Os02g0225300 protein - Oryza sativa subsp. japonica (Rice), partial (43%)	1

总结

- 参数对于BLAST等程序结果及运行速度的影响
- 数据集（数据库）对于程序结果及运行速度的影响
- 不同算法的运行速度的差别以及结果的差别

■ 谢谢！

