

Chapter 5 Small RNA analysis

Longjiang Fan Zhejiang University House-keeping noncoding RNAs:

- rRNA

- tRNA
- snRNA
- snoRNA

Small noncoding RNAs:

- siRNA
- miRNA

Long or large noncoding RNAs (>200nt):

- Precursor of small noncoding RNA
- Pol IV- and Pol V-dependent long noncoding RNA (plant-specific)
- mRNA-like long noncoding RNA

New types

- Circular RNA

http://rfam.xfam.org/

Complexity of transcription patterns of ncRNAs



1. Small non-coding RNAs in plants

Discovery of microRNA

- MicroRNAs were discovered in 1993 by Victor Ambros, Rosalind Lee and Rhonda Feinbaum during a study of the gene *lin-14* in *C. elegans* development.
- However, miRNAs were not recognized as a distinct class of biological regulators with conserved functions until the early 2000s. Since then, miRNA research has revealed multiple roles in negative regulation (transcript degradation and sequestering, translational suppression) and possible involvement in positive regulation (transcriptional and translational activation). By affecting gene regulation, miRNAs are likely to be involved in most biological processes

wikipedia

miRNA: miRNAs (microRNAs) are small regulatory RNAs of 20–24 nt that are encoded by endogenous MIR genes. Their primary transcripts form precursor RNAs, which have a partially doublestranded stem-loop structure and which are processed by DCL proteins to release mature miRNAs.

siRNA: siRNAs are generated from perfectly double-stranded RNAs that can originate from different sources such as RNAs transcribed from inverted repeats, natural cis-antisense transcript pairs, the action of RNA-dependent RNA polymerases (RDRs) which convert single stranded RNA into dsRNA, the replication of RNA viruses, and regions of the genome rich in retroelements.

A) miRNA biogenesis and function



B) siRNA biogenesis and function

(Khraiwesh et al. 2011)

miRNA

- Originate from capped & polyadenylated full length precursors; single chain (pri-miRNA)
- Hairpin precursor 60-400 nt (pre-miRNA)
- Mature miRNA 20-24 nt (miRNA)
- miRNA::miRNA* complex (non-perfect matching)
- Dicer-like RNase III endonuclease (DCL) and Argonaute protein (AGO)
- RISC (RNA-induced silencing complex)
- PTGS (posttranscriptional gene silencing)
- *trans-* or *cis-* acting
- First discovered in 1993 by Victor Ambros at Harvard (*lin-4*
- *Let-7* discovered in 2000 by Frank Slack as a postdoc at Harvard (Ruvkun lab)
- The first miRNA of plants was discovered in *Arabidopsis*



lin-4 and *let-7*



MIR827														
AACCCUUG	AAUGUGUUU	GUUGAUUG	AUAUCUAC.	CAUGUUG	AUCAUCC	UGUGUUG	AUCGAU	UGGUU	JUAGAU	GACCA	UCAAC	AAACUC	UUUCG	UGGUUUU
((.((.((((.(.((((((((.)	.(((((.	((((((((()))))))))).)))))).).)))))))))).)	.))))	.))
		UUGUUGAU	UGAUAUCU.	ACAC										
									. TTAG	ATGAC	CATCA	ACAAAC	T	
									TAG	ATGAC	CATCA	ACAAAC	T	

siRNA

- Originate from dsRNA (synthesis by RDR or bimolecular duplexes: NATs, natural antisense transcripts)
- Various types due to the divergence of biogenesis and mechanism processes
- Cleaved by miRNA, sometimes
- Dicer-like RNase III endonuclease (DCL) and Argonaute protein (AGO)
- RISC (RNA-induced silencing complex)
- PTGS and TGS (transcriptional gene silencing)
- Nearly perfect matching with target sites
- Cis- acting (also trans-)
- New tool for researchers



Comparison of miRNAs and siRNAs

	miRNAs	siRNAs
Dicer	DCL1	DCL1-4
5' end	Mostly U	Mostly A for 24-mers
Origin	Mostly intergenic region	Transposable-element, repeat sequence, heterochromatic sequence
Precursor	Hairpin pri-miRNA	Long dsRNA
Copy number	One pre-miRNA, one miRNA (one arm of the foldback structure)	One precursor, many siRNA (both strands of the precursor)
Conservation	Yes	No
Target gene	trans	cis and trans
Function	developmental regulation and stress response	Chromatin modification and stress response
Mechanism	mRNA cleavage or translational repression	DNA methylation and mRNA cleavage

	1	10	20	30	40	50	60	70	80	90	100 110	1
ath-MTR15Ca				00202002	acsconeso	ACTORCES	nagegegegege	909- -099 6	CCORT-TT	COTOT-CO-T		į.
ath-MIR156c					CGCATAGA	ACTGACAG	AAGAGAGTGAGC	ACACAAAG	GCACTTT	GCATGTTCG-A	TGC	-
gna-HIR156a				CACACCAG	ATTGAGAGAG	GCTGACAG	AAGAGAGTGAGC	RCATGCTAGT	GGTATTT	GTATGAGGGCA	TAC	2
ath-MIR156d				GATGG	GGGAAAAGAA	AGTTGACAG	RAGAGAGTGAGCI	ACAC-AAAGG	GGAAGTT	GTATAAAAGTT	TTGTA	1
ath-MIR156b	GCTAG	AAGAGGGAGA	AGATGGTGAT	TGAGGAATGC	AACAGAGAAA	HAC TGACAGI	AAGAGAGTGAGC	ACAT GCA G-G	CACTG-TTA	TGTGTCTATAA	CTT1	5
osa-MIR156b	TTG	TCTTGAGAGO	GGGAAGAGAT	CTCTATGG	GTTTTGGAG	GTCTGACAG	RAGAGAGTGAGC	ACACACGGTG	CTTTC-TTA	GCATGCAAGAG	CCAT	
osa-filk156j	1166	ննեննենեն	GGHICGHGCH	IIHIGHHH66C	GHGH1 G1=-		HHGHGHGIGHGU	HUHUGGUGUG		GCCHICGGLGG	TCOTTOCCC TITCOT	£.
ath-AIK156e			COCT	сстерсерот	HUGH TCOT	1661 GHUHG	HAGHGHGIGHGU	ACACOTCCTC	CCTTT_CTT	CCOTOTTTCOO	TCC01	2
acn-nikijor			GAGA	BEAGAGAGEST	GAGAGGTGA	SETGRERG	RAGREGAGTERSC	ACACATGGTG	ACTIT-CIT	GCATACTAAA	GGACTCAT	2
zea-HTR156c			GGTGGB	IGGAGAGAGGTG-		GCTGBCBG	RAGRGRGTGRGC	BCACATGGTG	CCTTT-CTT	GCATGATGTAT	GATCEAGAGAGTTCAT	-
sof-HIR156			GGTGGR	GAAGAGGTG-	G80	GCTGACAG	RAGAGAGTGAGC	ACACATGGTG	CCTTT-CTT	GCATGATGAAC	GATCGAGAG-GTTCAT	6
zna-MIR156b				AGATGAGT	TTTTTGARGO	STT TGACAG	RAGAGAGTGAGC	ACACACGGTG	G-TTT-CTT	ACCATGAGT	GTCAT	-
sbi-MIR156c						TGACAG	RAGAGAGTGAGC	ACACATGGTG	CUTTT-CTT	GCATGATGAAT	GAGTGGATTCAT	5
osa-MIR156a					GGAC	GGTGACAG	AAGAGAGTGAGC	ACACGTGGTT	GTTTC-CTT	GCATAAATGAT	GCCTAT	3
ath-MIR156g					ATAACGAA	IGCCACAG	AAGAGAGTGAGC	ACACATGGCT	CTTTTTCTA	GCATGCT	CAT	2
SD1-RIK156D				COCODCCT_C	TCOTCOCO	TTCOCOC	HHUHUHUIUHUU	HUHUHUUUUU	DIC_C_CCC	CCOTOGO	TCT_01	2
JSG-HIKIJOU			10	ACCULT-C	TCC8C8C8	TTCOCOC	ADCOCOCTCOCC	ALALOOLOIO ALALOOLOIO	010-0-CC0	0CATACCA	TGT_91	-
zna-MTR156k			66	CTTAGGCC-C	CGAAGAGA	-TTGACAG	RAGEGEGEGEGE	ACCIGGCGCG	GCG-8-CCG	GCATGGAA	CGC-81	5
sbi-HIR156e			CG	ATTAGGCC-C	TGAAGAGA	-TTGACAG	RAGAGAGCGAGC	ACCC66C6C6	666-6-666	GCATGGAG	CGC-81	2
osa-MIR156e				G	GCGCGAGG	T-GACAG	AAGAGAGTGAGC	ACACGGCCGG	GCGTG-ACG	GCACCGGCGGG	CG1	1
zna-MIR156e			CC	CGCCGGGCG-C	GCGCGCGG	T-GACAG	AAGAGAGTGAGC	ACACGGCCGG	-TGTG-ACG	GCACCGGCGCG	GAT-GT	1
zna-MIR156h			CG	IGGCTCGCG-C	GCGCGCGG	T-GACAG	AAGAGAGTGAGC	ACA¢GGCCGG	GCGTG-ACG	GCA C	GT	1
zma-MIR156i				CGCG	GCGGGGGGG	T-GACAG	AAGAGAGTGAGC	ACACGGCCGG	GCGGA-ACG	GCACCGGCG-G	ATG-G1	2
OSA-ALK156h					GHGH116	I I bHLHb	HHAHAHAIAHAL	HEHEBBEBE	CCCCC OCC	LLHIL66L666	CO T CI	÷.
osa-filk1561					6	TCOCOC	HUHUHUHUIUHUU	HUHUGGUUGG	0160H-HL6	6C8 T88T 6	GH	-
osa-MTR156k				T	GAGAGTGA	TG8C8G	AAGAGAG A GAGC	ACA8CCCG	608-608	GCGBCGBCGGC	661060	-
sbi-MIR156d				TGGACCTC	TGGAGTGAGT	GA TGACAG	RAGAGAGAGAGAGC	ACAACCCA	GCA-CCA	GCGAGGAGAAG	CCTCGC	5
zma-MIR156,j			CG	AGTGGACCTC	GGGAGCGA	TGACAG	AAGAGAG <mark>A</mark> GAGCI	ACAACCCA	IGCA-CCA	GCGAGGAAAAAG	CCTCGC	2
Consensus					· · · · · · · · · · · · · · · · · · ·	TGACAG	AAGAGAGTGAGC	ACACg.G	gct.	GCAt	•• ••aT	1
								1.84			-	
	111	120	130	140	150	160	170	180	190	200 20	5	
ath-MTR156a	ACTTG	CITCICITS	CGTGCTCACT	arorrerona.	TCBGBTTCCC	GT	601681	CICITI	80	50		
ath-MTR156c	ATTTG	CITCICITE	CGTGCTCACT	GCTCTATCTG	TCBGATTCCC	i6	CT					
gna-HIR156a	AATTG	CGGGTGC	CGTGCTCRCT	TCTCTATCTG	TCAGCTTCCC	CAT	TETTTT	TTAC				
ath-MIR156d	TATEG	TTGCTTTTG	CGTGCTCACT	-CTCTTTTTG	TCRTARCI	TTC	TCCTTC	AT				
ath-MIR156b	6C G	TGT6(CGTGCTCACC	ICICITICIG	TCHGTTGCCT	RICICIGC	CTGCTTGACCTC	TETETETETE	TUTUTUTUT	CAAATTTGGCT		
osa-MIR156b	BUT 66	6H-6C16160	LGIGUIUHU-	-ILILIHILIG	ICHGCU	I ICHC	UHIGLUUHH	THI-GHITHH	TECCOCCCC	CICHGIIGHCH	6	
osa-filk156j	COTTO	DO-CCTOTCI	TOTOCTTOCT	CLICITITU	TC8CC60	101010	HICHLILLL	ILLUGLH	ILLLLLLLLL	LLUH		
ath-MTP156f	GETTG	BB-SCTRTS1	IGTGCTCACT	-CICICICICIC		(TT		TECETCI	TETE			
osa-MIR156c	GCTTG	AA-GCTATG	IGTGCTCACT	TCTCTCTCTG	TCBGCCF	ATTTGA	TETETE	TITCICICIT	TETEE			
zna-MIR156c	GCTCG	AA-GCTATGO	CGTGCTCACT	TCTCTCTTTG	TCAGCCF	TTAGA	ACTCC-	TETATC				
sof-MIR156	GCTCG	AA-GCTATGO	CGTGCTCACT	TCTCTCTCTG	TCAGCCC	STTRGA	ACTC	TCTCTC				
zna-MIR156b	GCTAG	GA-GCTGTGC	CGTGCTCACC	-CTCTATCTG	TCAGTCA	ACTC-8	TCAAG-	CCCATCT				
sbi-MIR156c	GCTCG	AA-GCTATGO	CGTGCTCACT	ICICICICIG	TCH							
osa-MIR156a	GETTE	GH-GUTHUGU	LGIGUIUHUI	CICICICIC		CTETET						
ach-nikibog	CTTCC	CO_CCTCTCC	CETECTEOFT	-CICIDICIIO	1010010 TCO	actorer						
osa-MTR156d	CCC-G	TECTEBEEG	CGTGCTCACT	CETETTETG	TCB C CC1	CTT	TCTC-T	C-8666CTC8	OTTOP .			
zna-HIR156d	GCC-G	TCCTCGCCG	CGTGCTCACT	TCTCTTTCTG	TCRGCC1	ICTT	TCTC-T	CGATGGCT				
zma-MIR156k	600-6	TECCEGECEGE	CGTGCTCGCT	TCTCTTTCTG	TCHG===CCT	ICTC	CCTCGT	CCTGAG				
sbi-MIR156e	GCC-G	TCCCCGCCGC	CGTGCTCGCT	TCTCTTTCTG	TCAGCC1	ICTC	TETEGT	CCT6GG	H	omology	Rotwoon	
osa-MIR156c	GCC-G	TCGCGGCCGG	CGTGCTCACT	GCTCTTTCTG	TCATCCC	GTG	CC		11	uniuugy	Detween	
zna-HIR156e	6CC-6	TCGCGGCCGG	CGTGCTCACT	GCTCTCTCTG	TCATCCC	GCTG	GCACAC	-CCTCRCG				
zna-mikibbh		Luthilittii	The Real Property of the Prope		And the second of the				m	ATILIKA MI	VNAGOND	
2Ma-01K1361	CCC C	TECCCCCCCCCC	CETECTERET	COTOTOTOTO	TCOT CCC	DOTO:	TETE	CCCCCCCT	111	alure mi	ININAS AILU	
asa-MTP15Ch	GCC-G	TEGEGGACGO	CGTGCTCACT	GETETATETE		ACTC	TCTC	-CCCCGCT	111	ature mi	INITAS allu	
osa-MIR156h osa-MIR156i	GCC-G GCC-G GCCTG	TCGCGGGACGO	CGTGCTCACT CGTGCTCGCT CGTGCTCGCT	CCTCTTTCTG	TCATCCA TCAGCAT TCAGCAT	СТС ГСТС	TCTC	-CCCCGCT	Ш Ж	ature III Vorgona	hotween	
osa-MIR156h osa-MIR156i shi-MIR156a	GCC-G GCC-G GCC-G GTC-G	TCGCGGGACGO CCCCCGCCGO TCGCGGCCGO TCCTCGCCGO	CGTGCTCACT CGTGCTCGCT CGTGCTCACT CGTGCTCACT	GCTCTATCTG CCTCTTTCTG GCTCTGTCTG GCTCTGTCTG	TCATCCA TCAGCAT TCATC TCA	істс гстс	TCTC	-CCCCGCT	di	vergence	e between	
osa-MIR156h osa-MIR156i sbi-MIR156a osa-MIR156k	GCC-G GCC-G GCC-G GTC-G TTCTG	TCGCGGACGO CCCCCGCCGO TCGCGGCCGO TCCTCGCCGO CCAGGGCCGT	CGTGCTCACT CGTGCTCGCT CGTGCTCGCT CGTGCTCACT CGTGCTCACT TGTGCTCTCT	GCTCTATCTG CCTCTTTCTG GCTCTGTCTG GCTCTGTCTG TCTCTTTCTG	TCATCC TCAGCAT TCATC TCA TCATT-C	йстс гстс БССБ	TCTC	-CCCCGCT	di	vergence	e between	
osa-MIR156h osa-MIR156i sbi-MIR156a osa-MIR156k sbi-MIR156d	GCC-G GCC-G GCC-G GTC-G TTCTG TTCTG	TCGCGGACGC CCCCCGCCGC TCGCGGCCGC TCCTCGCCGC CCAGGGCCGT CGAGGGCCGT	CGTGCTCACT CGTGCTCCGCT CGTGCTCACT CGTGCTCACT CGTGCTCACT TGTGCTCTCT TGTGCTCTCT	GCTCTATCTG CCTCTTTCTG GCTCTGTCTG TCTCTTTCTG GATCTATCTG GCTCTCACTG	TCATCCF TCAGCAT TCATC TCA TCATT-C TCATC-C	істс гстс БССБ БССБ	TCTC	-CCCCGCT	di pr	vergence e-miRN	e between As	
oso-MIR156h osa-MIR156i sbi-MIR156a osa-MIR156k sbi-MIR156d zma-MIR156j	GCC-G GCC-G GCC-G GTC-G TTCTG TTCTG TTCTG	TCGCGGACGC CCCCCGCCGC TCGCGGCCGC TCCTCGCCGC CCAGGGCCGT CGAGGGCCGT CGAGGGCCGT	CGTGCTCACT CGTGCTCGCT CGTGCTCACT CGTGCTCACT CGTGCTCACT TGTGCTCTCT TGTGCTCTCT TGTGCTCTCT	GCTCTTTCTG GCTCTTTCTG GCTCTGTCTG GCTCTGTCTG	TCATCC TCAGCAT TCATC TCA TCATT-C TCATC-C TCATC-C	йстс гстс бссб бссб бссб	TCCA TCCA ACAGGC	-CCCCGCT CACCCAA CACCCAA	di pr	vergence e-miRN	e between As	



Conserved Cotyledon miRNA Family miR156 miR160 miR166 miR171 miR319 miR390 miR408 miR159 miR162 miR164 miR169 miR395 miR397 miR167 miR168 miR172 miR393 miR394 miR398 miR399 miR403 miR2111 miR1507 miR1508 miR1509 miR1510 miR1514 miR2118 miR2119 miR3522 miR4414 miR530

The conservation of miRNA families in plants

Role of small RNAs in crops



(Khraiwesh et al. 2011)

Targets of conserved miRNAs



Regulatory Roles of Plant MicroRNAs

- miRNA and vegetative organ development
- miRNA and floral development and vegetative phase change
- miRNAs regulate miRNA and siRNA biogenesis and function
- miRNAs involved in signal transduction
- miRNAs involved in plant disease
- miRNAs involved in environmental stress responses
- Even more, miRNAs and DNA-methylation



Number of miRNAs

miRNAs (miRBase, release 22, March 2018)

Rice: 604 precursors, 738 mature
Soybean: 684 precursors, 756 mature
Maize: 174 precursors, 325 mature

Identification of miRNAs by us

- Rice: long miRNA-like (Zhu et al. 2008)
- Tobacco: >200 miRNAs (Tang et al. 2012)
- *O. rufipogon*: >300 miRNAs (Wang et al. 2012); >3000 phasiRNAs (Liu et al. 2013)
- *B. napus*: 360 conserved and 533 novel miRNAs (Shen *et al*, 2014; Shen *et al*., 2015)
- *N. benthamiana:* ~200 miRNA (Xiao et al. 2014)

Reads from miRNA loci

A GacgcagcaUgaga GUAAUAUACUAAUCCGUGCAUCCA (13) GUUGCACGGGUUUGUAUGUUGCAG (192) GUUGCACGGGUUUG (3) GUUGCACGGGUUUGUAUGUUGC (1) GUUGCACGGGUUUGUAUGUUGCA (1) GUUGCACGGGUUUGUAUGUUGCAGA (3) UUGCACGGGUUUGUAUGUUGCAG (3) 5' 5' GUARIAUA CAUCCOGC COLOCUUACACCO CAU-GUGARC CU G GUARIAUA CAUCCOGC COLOCUUACACCO CAU-GUGARC CU G GACGUUGUAUGUUGGCCA COUGCU-GUGAUGUG GUGUG GUGAU GUGAU GUGAU 3' A CU U A GUUGCACGGGUUUGUAUGUUGCAGC(1) B GUUGCACGGGUUUGUAUGUUGCAGG(1)

GUAÁ

2. Evolution of small RNAs: artificial selection

Role of small RNAs in crop domestication and genetic improvement



Transcript factor targets: protein-coding genes
 Non-coding small RNA targets: transcript factor
 Domestication targets: transcript factor

Question:

Targets of domestication selection: small RNAs?





в

miR397a	5'GUAGUUGCGACGUGAGUUACU 3'
	111111111111111111111111111111111111111
Os05g38420	CAUCAACGCUGCACUCAACGA(1)
Os01g62480	CAUCAACGCCGCGCUCAACGA(3)
Os05g38390	GAUCAACGCGGCGCUCAACGA(4)
Os01g62600	GAUCAACUCGGCGCUCAACGA (5)

The loss of miR397 binding site in its target family

(Guo et al. BMC Genomics, 2008)

Nucleotide mutations together with insertions/deletions are responsible for the gain/loss of miRNA binding sites during co-evolution of miRNAs and their target genes.

Question: How about domestication selection?



Distribution of Tajima's *D*, Fay and Wu's *H* and HKA test values across all small RNA sequencing fragments in cultivated rice populations. Significant (<0.05%) fragments under positive selection are colored red.



Expression divergence of the miR164 family in cultivated and wild rice. A: Log2 of the expression ratios of miR164 in six microarray chips including six cultivars and six *O. rufipogon* lines; **B:** The expression of miR164 family in developing grain and seedling as indicated by reads per million (RPM) in *O. sativa (japonica)* and *O. rufipogon* small RNA populations through Illumina sequencing.

miR164 family: selection evidence

indica					japonica		rufipogon			
Title	Seq_No.	Tajima'D	P (H test)	Seq_No.	Tajima' D	P (H test)	Seq_No.	Tajima'D	P (H test)	
miR164a	25	1.544	0.0772	22	-1.3099	0.7128	11	-0.8448	0.5446	
miR164b	23	-1.1361	0.2312	21	-0.0466	0.8038	12	-0.0143	0.3002	
miR164c	23	-0.5139	0.021	23	-0.6312	0.001	8	0.4691	0.105	
miR164d	25	-0.4764	0.0028	21	-1.2874	0.0966	12	-1.5419	0.2304	
miR164e	21	-1.9033	0.003	21	1.1834	0.6054	10	-1.0403	0.905	
miR164f	22	NA	NA	22	NA	NA	10	NA	NA	

Genomic variations in precursors of 38 previously identified conserved miRNA families in wild rice.

(Wang et al, 2012, New Phytologist)



miRNAs are targets of artificial selection

Our results illustrated *MIRNA* genes, like protein-coding genes, might have been significantly shaped during rice domestication and could be one of the driven forces contributed to rice domestication.

3. Identification of small RNAs

miRNAsphsiRNAs

Prediction of MIRNA gene

3 stages:

- □ 1. genetic screening technology; direct cloning of miRNAs → small RNA isolation
- □ 2. computational approaches
- □ 3. prediction based on high throughput sequencing data



Data mining based on high throughput sequencing technology

Small RNA sequencing:

- SE, < 50nt
- Millions of reads
- High sensitivity
- FASTQ & FASTA

Pipeline:

- Quality filtering
- Trim adaptors
- eliminate the redundancy
- Mapping to reference sequence
- Removing small RNA mapping to tRNA, rRNA, snRNA, snoRNA, mRNA et al.
- Prediction (sofware or webserver, based on structure and property of miRNA)

		Leaf		
Item	Topping	Wounding	Control	(wounding)
Total				
Raw reads	5,186,473	5,510,747	3,074,948	4,722,862
Unique raw reads	1,959,287	3,083,036	1,018,809	2,178,969
Mapped to GSS#				
Reads	2,632,084	2,700,532	1,277,575	2,469,534
Specific in each library	477,279	803,960	206,310	491,809
Conserved in all libraries	1,667,977	1,085,172	782,079	1,312,822
Unique	542,516	965,547	252,425	687,949
Specific in each libraries	385,469	702,522	148,981	442,337
Conserved in all libraries	25,780	25,780	25,780	25,780
Singleton	417,905	761,185	202,567	534,409
Mapped to EST#				
Reads	1,293,204	1,218,970	764,235	1,528,346
Specific in each libraries	99,553	137,509	43,758	107,459
Conserved in all libraries	985,513	811,452	563,565	1,165,580
Unique	129,710	186,601	72,750	149,454
Specific in each libraries	77,511	116,666	32,875	84,638
Conserved in all libraries	14,001	14,001	14,001	14,001
Singleton	88,566	131,515	48,486	102,406

Prediction of miRNA's targets

miRU (update to psRNATarget) http://bioinfo3.noble.org/miRNA/

<u>miRU.htm</u> (Dai and Zhao, 2011)

Principle
1. ≤ 4 mismatches (at most 7)
2. Nearly none of mismatch allowed
between 1st to 10th

position from 5' end, espetially 9th to 11th position

3. One to many

NOBLE	matic: Laboratory		sRNATarg	et: A Pla	nt Small i	RNA Target .	Analysis Serve
					About	Citation	Analysis
cation: Analysis							
User-submitted small RNAs / preloaded transcripts	Prelesded small RNAs	/ user-submitted	tanacipta	Use	me hostimeture	all RNAs / user-st	bmitted transcripts
Upload small RNA sequence(s) in FASTA format: or paste sequences below:	[Load demo data] (238	ନ					
 tile input sequence size limit: 20M. 							
invalid small RNAs will be ignored during analysis	. 0						
- invalid small RNAs will be ignored during analysis	. 0						
- invalid small RNAs will be ignored during analysis Select a preloaded transcript genomic library for ta	s 😡		-				
 invalid small RNAs will be ignored during analysis Select a preloaded transcript genomic library for tax Allium_exps (Onion) DFCI Gene Index (ONG) Release Aphildraph (Addima TAIRQ) eDNA menod mPNA- 	s 🖗 rget search: 2 zms. oficiand 06/19/2000	1]				
invalid small RNAs will be ignored during analysis Select a preloaded transcript genomic library for ta Allium_exps (Onion) DFCI Gene Index (ONGI) Release Ambidepsis thaliana DFCI Gene Index (AOI) Release 1 Ambidepsis thaliana DFCI Gene Index (AOI) Release 1	s. 69 rget search: 2 genc, eleased 06/19/2009 5	1					
invalid small RNAs will be ignored during analysis belect a preloaded transcript genomic library for tas Allium_exps (Onise) DFCI Gene Index (ONGI) Release Ambidopsis thaliana DFCI Gene Index (AOI) Release I Ambidopsis thaliana DFCI Gene Index (AOI) Release I Ambidopsis thaliana, Genema(01/22/2004), 3.4% segm	 a. and a search: 2 2 gene, released 06 /19/2009 5 anna from +/- strand with 0.4K 	14 [] evenings	10				
 - invalid small RNAs will be ignored during analysis Select a preloaded transcript genomic library for tas Allium_exps (Onion) DFCI Gene hades (ONG) Release Ambidopsis thaliana TAIRO, cDNA, emered mRNA g Ambidopsis thaliana, Genome(01/22/2004), 3.4K segm Aquilagia DFCI Gene Index (AQQI) Release 2.1 Beta vulnatis (Bene) DFCI Gene Index (AQQI) Release 2.1 	 a. and a search: 2 2 gene, released 06 /19/2009 5 annus from +/- strand with 0.4K 	evalgp	- 100				
invalid small RNAs will be ignored during analysis islest a preloaded transcript genomic library for tas Allium_exps (Onion) DFCI Gene Index (ONGI) Release Ambidopsis thaliana TAIRO, eDNA, smarved mRNA g Ambidopsis thaliana RCI Gene Index (AdQI) Release 1: Ambidopsis thaliana, Geneme(01/22/2004), 3.4% segm Aquilegis DFCI Gene Index (AdQI) Release 2: Beta_vulgats (Ben) DFCI Gene Index (BVGI) Release 3 Bendypedium_distastyse (Uruple faits term), JOI v1	 a for the search: 2 gene, released 06/19/2009 5 genes +/- search with 0.4K 0 generate manacipes, fip-ip-ip 	eveningg	-				
 invalid amatl RNAs will be ignored during analysis Select a preloaded transcript genomic library for ta Allium_erps (Onion) DFCI Gene Index (OMG) Release 1 Ambidopsis thaliana TARS, 2DNA, meaved mRNAs q Ambidopsis thaliana DFCI Gene Index (AGI) Release 1 Ambidopsis thaliana DFCI Gene Index (AGI) Release 2.1 Beta_vulgais (Ben) DFCI Gene Index (AGI) Release 2.1 Benchypedium_distablyse (Purple_faire_brenc), IGI vi 	 A for a search: 2 geneselessed 06/19/2009 search with 0.4K 0 0 genemic measuript, fp-jgi-g Bone Index (BDGI) Release 1 	eveligg					
isvelid small RNAs will be ignored during analysis Select a preloaded transcript genomic library for its Allium_exps (Onion) DFCI Gene Index (ON00) Release 1 Ambidopsis thalians TARS, eDNA, merved mRNAs q Ambidopsis thalians DFCI Gene Index (AGI) Release 1 Ambidopsis thalians, Geneme(01/22/2004), 3.4K segm Aquilogis DFCI Gene Index (AGI) Release 2.1 Bets, vulganis (Bent) DFCI Gene Index (RVGI) Release 3 Banchypodium_distashyes (Purple_faire_brone), JGI v1 Benshypodium_distashyes (Purple_faire_brone), JGI v1 Benshypodium_distashyes (DFCI Gene Index (BNGI) Bensine negus (callend _mpc) DFCI Gene Index (BNGI)	 a. and a search: 2 2 pene, released 06/19/2009 5 ants from +/- stand with 0.4K 0 generate massaipts, fip-jgi-g Bene lades (BDGI) Release 1 Release 3 	evelep					
isvelid small RNAs will be ignored during analysis Select a preloaded transcript genomic library for its Allium_ceps (Onion) DFCI Gane Index (ONGI) Release Ambidopsis thalians TARK, 200A, smarved mRNAs q Ambidopsis thalians DFCI Gane Index (AGI) Release 1: Ambidopsis thalians, Geneme(01:22:2004), 3.4K segm Aquilagia DFCI Gane Index (AGG) Release 2: Benchypedium_dismahyee (Pupile_faire_bence), JGI v1 Benchypedium_dismahyee (Pupile_faire_bence), JGI v1 Benatypedium_dismahyee (Pupile_faire_bence), JGI v2 Benatypedium_dismahyee (Pupile_faire_bence), JGI v2 Benatypedium_dismahyee (Pupi	 a. If a search: 2 2 pene, released 06/19/2009 5 ants from +/- stand with 0.4K 6 6 genemic massedpts, fip-jgi-g Bolesse 1 Release 3 Release 4 	evalapp ef.ez					
isvalid amati RNAs will be ignored during analysis select a preloaded transcript genomic library for ta Allium_ceps (Onion) DFCI Gene Index (ONGI) Release Ambidepsis thalians TAR9, DNA, menved mRNAs g Ambidepsis thalians DFCI Gene Index (AGI) Release I Anbidepsis thalians DFCI Gene Index (AGI) Release I Beatypediam_dismayne (Pueple faits_berme), JGI vt. Beatypediam_dismayne (Pueple faits_berme), JGI vt. Beatypediam_dismayne (Pueple faits_berme), JGI vt. Beatsin segus (eilsted_mpc) DFCI Gene Index (BNGI) Beatsin segus (eilsted_mpc) DFCI Gene Index (BNGI) Beatsin eilste (eilsted_mpc) DFCI Gene Index (BNGI) Response to add/ update a transcript library. (Inviencem association (2 Station lows of the set (1990))	 a for search: 2 gene, edeased 06/19/2009 5 annts from +/- seand with 0.4K 0 0 genemic matacaipta, fip-ipi-gi-gi-gi-gi-gi-gi-gi-gi-gi-gi-gi-gi-gi	ovælepp ef ce					
isvelid amall RNAs will be ignored during analysis belect a preloaded transcript genomic library for its Allium_exps (Onion) DFCI Gane Index (ONGI) Release Ambidopsis thalians TARK, 200A, energed mRNAs q Ambidopsis thalians, DFCI Gane Index (AGI) Release 11 Ambidopsis thalians, Geneme(01/22/2004), 3.4K segm Aquilogis DFCI Gane Index (AGGI) Release 21 Beachypedium_distashyes (Pupils_fals_beam), JGI v1 Beachypedium_distashyes (Pupils_fals_beam), JGI v1 Beachyp	a. Orget search: 2 2 2 2 2 2 2 2 2 2 2 2 2	evenlagg si ez	a 0-50)	9			
 invalid amail RNAs will be ignored during analysis Select a preloaded transcript genomic library for ta Alium_exps (Onion) DFCI Gene Index (ONGI) Release Ambidopsis thalians AIRS, eDNA, merved mRNAs g Ambidopsis thalians, Genemo(01/22/2004), 3.4K regm Acultagis DFCI Gene Index (AGD) Release 1.1 Beta_vulgais DFCI Gene Index (AGD) Release 2.1 Beta_vulgais (Bent) DFCI Gene Index (AGD) Release 2.1 Benabypodium_distashyos (Purple_file_former) DFCI Gene Enabypodium_distashyos (Purple_file_former) DFCI Gene Benabypodium_distashyos (Purple_file_former) DFCI Gene Benabypodium_distashyos (Purple_file_former) DFCI Gene Benabypodium_distashyos (Purple_file_former) DFCI Gene Benation negus (eilened_mpc) DFCI Gene Index (BNGI) Request to add/ update a transcript Bbrary. Maximum expectation (* Profer Iowar false genith et nove atringent cale off threakeld (0-7.0); Profer Kingler Floate site mene relaxed cut-off threakeld (4-6.2) 	 a. O argot search: 2 argot search: 2 argot search: argot search:<	evelage si ce	e: 0-5.0)	9			
isvalid amati RNAs will be ignored during analysis select a preloaded transcript genomic library for its Allium_ceps (Onion) DFCI Ganc Index (ONGI) Release Ambidopsis thalians TARS, 200A, merved mRNAs q Ambidopsis thalians DFCI Ganc Index (AGI) Release 11 Ambidopsis thalians DFCI Ganc Index (AGI) Release 21 Bentypedium_diseadyses (AGGI) Release 2.1 Bentypedium_diseadyses (AGGI) Release 2.1 Bentypedium_diseadyses (Purple_faire_brene), IGI vt. Bentypedium_diseadyses (Purple_faire_brene), IGI vt. Bentypedium_diseadyses (Purple_faire_brene), IGI vt. Benation angus (eilteed_mpc) DFCI Ganc Index (BNGI) Benation angus (eilteed_mpc) DFCI Ganc Index (BNGI) Request to add/ update a transcript library. Uaximum expectation (* Prefer Iowar faire positive nore abriegent cas-6gf threaked [G-2.0], Prefer Ighe Please set a more released cut-off threaked [4.0-5.0] .ength for complementarity scoring (hspaire):	 a for search: 2 gene, eleased 06/19/2009 5 ants from +/- seand with 0.4K 6 0 genemic transcript, fop-jo-p 3ene Index (BDGI) Release 1 Release 3 Release 4 prediction rate? Please set a or prediction coverage? I): 	evelapp af ez 3.0 (rang 20 (range	e: 0-5.0) (15-30bp)	8			
invalid small RNAs will be ignored during analysis beloct a preloaded transcript genomic library for ta Allium espa (Onion) DFCI Gene Index (ONG) Release Ashidopsis thaliana AIRO, eDNA, mexeed mRNAs Analysis DFCI Gene Index (AGI) Release 1 Analysis DFCI Gene Index (AGI) Release 1 Beneypedium_distably on (Nuple, false, brome), 101 et Beneypedium_distably on (Nuple, false, brome), 101 et Beneypedium_distably on (Nuple, false, brome), 101 et Beneypedium_distably on (Puple, false, brome), 101 et Beneypedium_distable, 100 et Beneypediu	 a for the same of the same of	evalipp sf ec 3.0 (range 20 (range 25.0 (range	e: 0-5.0) 15-30bp) 0-100, less is	D D S better)	Ð		
 invalid amail RNAs will be ignored during analysis Select a preloaded transcript genomic library for ta Allium_expa (Onion) DFCI Gene Index (ONGI) Release Ambidoptis thaliana TARS, eDNA, menved mRNAs Ambidoptis thaliana DFCI Gene Index (AGI) Release 11 Ambidoptis thaliana, Cancen(01/22/2004), 3.4K segm Aquilagia DFCI Gene Index (AGI) Release 2.1 Benk-yoolam_distathyon (Pusple, faile, benen) DFCI G Benkhyoolam_distathyon (Pusple, faile, benen) DFCI G Benkhyoolam, agus (oilsend_mpc) DFCI Gene Index (BNGI) Benation angus (oilsend_mpc) DFCI Gene Index (BNGI) - Request to add / update a transcript fibrary. Maximum expectation (* Prefer lower faile positive more abriegent cut-off benational (0-2.0); Prefer leight Please set a more relaxed cut-off breakold [4.0-5.0] Length for complementarity scoring (hspasite): Target accessibility - allowed maximum energy to Planking length around target site for target accessib 	 a fight search: 2 2 prese, released 06/19/2009 5 ants from +/- stand with 0.4K 0 genemic tensoripts, fip.jg/rp none lades: (BDGI) Release 1 Release 3 Release 4 prediction: rate? Please set a or prediction: coverage? j): suspair the target site (UPE): sitity analysis	everlipp ef er 3.0 (range 20 (range 25.0 (range 17 bp in up	e: 0-5.0) (15-30bp) 0-100, tess is stream / [13	D B s better) (] bp is dow	D Nutream	Θ	



Degradome sequencing

The principle of degradome sequencing process of library construction





A(n) 3'OH

Prediction of phasiRNA

Howell et al., 2007

P = ln[(1 +
$$\sum_{i=1}^{8} k_i)^{n-2}$$
], P > 0,

n = number of phase cycle positions occupied by at least one small RNA read within an eight-cycle window,³² k = the total number of reads for all small RNAs with consolidated start coordinates in a given phase within an eight-cycle window. Phase cycle length was set at 21 nucleotides. A positive phase signal, therefore, was limited by the power function (n-2) to those loci with small RNAs occupying at least three cycle positions in a phased, 21-nucleotide register.





(B)



4. Related database

-miRBase

http://www.mirbase.org/



miRBase provides the following services:

- The miRBase database is a searchable database of published miRNA sequences and annotation. Each entry in the miRBase Sequence database represents a predicted hairpin portion of a miRNA transcript (termed mir in the database), with information on the location and sequence of the mature miRNA sequence (termed miR). Both hairpin and mature sequences are available for searching and browsing, and entries can also be retrieved by name, keyword, references and annotation. All sequence and annotation data are also available for download.
- The miRBase Registry provides miRNA gene hunters with unique names for novel miRNA genes prior to publication of results. Visit the help pages for more information about the naming service.

To receive email notification of data updates and feature changes please subscribe to the miRBase announcements mailing list. Any queries about the website or naming service should be directed at mirbase@manchester.ac.uk.

miRBase is managed by the Griffiths-Jones lab at the Faculty of Biology, Medicine and Health, University of Manchester with funding from the BBSRC. miRBase was previously hosted and supported by the Wellcome Trust Sanger Institute.

References

If you make use of the data presented here, please cite the following articles in addition to the primary data sources: miRBase: annotating high confidence microRNAs using deep sequencing data. Kozomara A, Griffiths-Jones S. NAR 2014 42:D68-D73 miRBase: integrating microRNA annotation and deep-sequencing data. Kozomara A, Griffiths-Jones S. NAR 2011 39:D152-D157 miRBase: tools for microRNA genomics. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. NAR 2008 36:D154-D158 miRBase: microRNA sequences, targets and gene nomenclature. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. NAR 2006 34:D140-D144 The microRNA Registry. Griffiths-Jones S.

miRBase Genome Browser

http://www.mirbase.org/cgi-bin/browse.pl

- monocotyledons
 - Asparagus officinalis (101 precursors, 101 mature) [Aspof.V1]
 - Aegilops tauschii (88 precursors, 173 mature) [ASM34733v2]
 - Brachypodium distachyon (317 precursors, 525 mature) [Bd21]
 - <u>Elaeis guineensis</u> (6 precursors, 6 mature)
 - Festuca arundinacea (15 precursors, 15 mature)
 - Hordeum vulgare (69 precursors, 71 mature) [Hvul_cvHarunaNijo3HBAC01]
 - Oryza sativa (604 precursors, 738 mature) [MSU7]
 - <u>Sorghum bicolor</u> (205 precursors, 241 mature) [Sorghum_bicolor_NCBIv3]
 - Saccharum officinarum (16 precursors, 16 mature)
 - <u>Saccharum sp.</u> (19 precursors, 20 mature)
 - Triticum aestivum (122 precursors, 125 mature)
 - <u>Triticum turgidum</u> (1 precursors, 1 mature)
 - Vriesea carinata (33 precursors, 65 mature)
 - Zea mays (174 precursors, 325 mature) [B73_RefGen_v4]

Version 22.0

The Increasing records in miRBase



Version

Stem-loop see	quence MI0000653
Accession	MI0000653
ID	osa-MIR156a
Description	Oryza sativa miR156a stem-loop
Stem-loop	g a uguu uu aau ggagg ugacaga agaga gugagcac cguggu ucc gcaua g a ccucc acugucu ucucu cacucgug gcaucg agg cguau u - c u c uu ccg Get sequence
Deep sequencing	<u>1111</u> reads, 3 experiments
Comments	The stem-loop sequence represented here is predicted based on homology to miRNAs rice.
Genome context	CoordinatesOverlapping transcripts01: 22523102-22523201 [-]intergenic
Gene family	MIPF000008; <u>MIR156</u>
Mature seque	ence MIMAT0000618
Accession	MIMAT0000618
ID	osa-miR156a
Sequence	7 - ugacagaagagagugagcac - 26 Get sequence
Deep sequencing	<u>1110</u> reads, 3 experiments
Evidence	by similarity; <u>MI0000183</u>
References	
1	PMID: <u>12101121</u> <u>"MicroRNAs in plants"</u> Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP Genes Dev. 16:1616-1626(2002).

Ŋ9