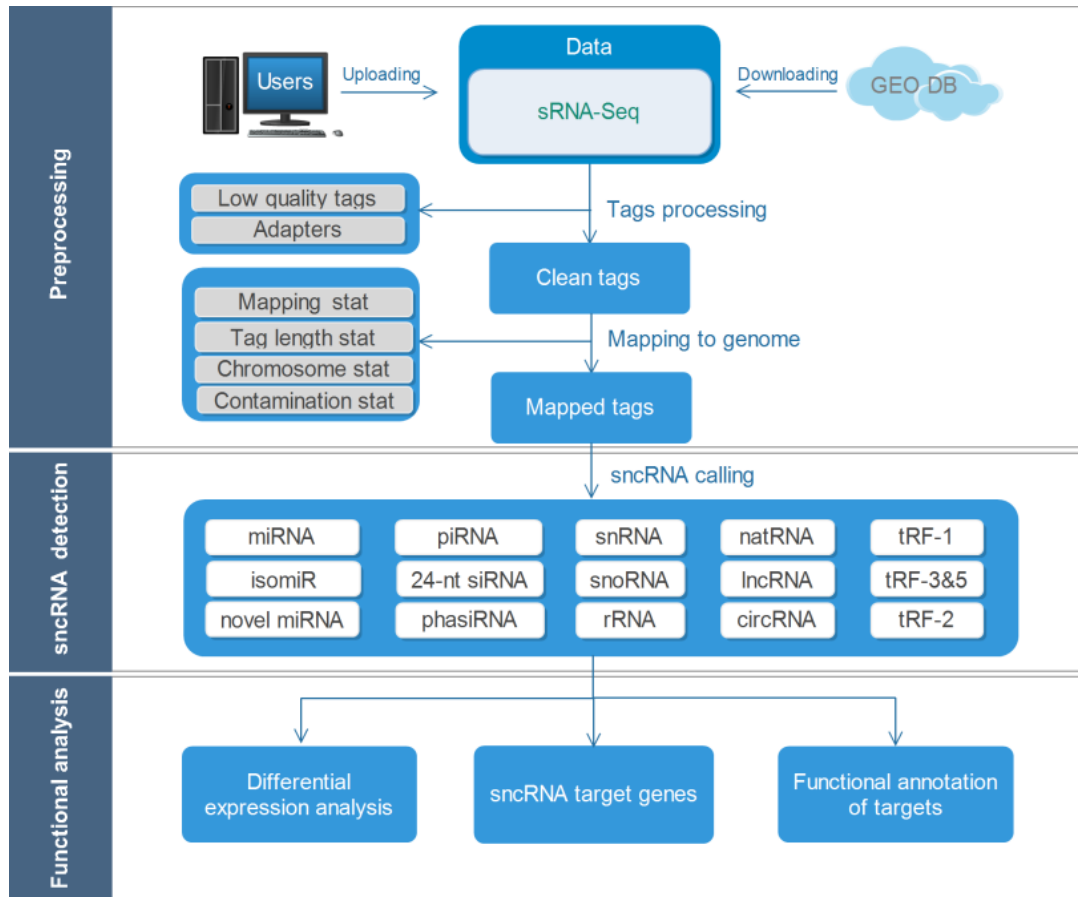


sRNAtools Tutorial

Last updated: 19/09/2019



Developed and maintained by:

Affiliations

State key laboratory of tree genetics and Breeding, Xiangshan Road, Haidian District, Beijing,

P.R.China 100091

Precision Medicine Research Center, West China Hospital, Sichuan University, Chengdu, Sichuan,

China 610041

Questions and feedback Email:

biolq668@gmail.com

langxiaoqiang@wchscu.cn

sRNAtools Tutorial	1
<i>1. Introduction.....</i>	<i>3</i>
<i>2. Getting Start.....</i>	<i>4</i>
2.1 Webserver.....	4
2.2 Standalone Versions: Encapsulate In A Docker.....	4
2.2.1 Docker Installation.....	4
2.2.2 Dependencies.....	6
2.2.3 sRNAtools Docker Image Installation.....	6
2.2.3.1 Docker image installation.....	7
2.2.3.2 Construct image from Dockerfile.....	7
2.2.4 Share Directory and Quickly start.....	7
2.3 VirtualBox version.....	8
2.3.1 VirtualBox installation.....	8
2.3.2 sRNAtools VirtualBox Download.....	8
2.3.3 VirtualBox Configure.....	9
2.3.4 Share Directory.....	10
2.3.5 Start VirtualBox.....	10
2.3.6 Change keyboard layout and time zone (Optionally).....	11
2.4 Local program package.....	11
2.4.1 Minimum Requirements.....	11
2.4.1.1 Required Software Packages:.....	11
2.4.1.2 Other program included in the distribution:.....	13
2.4.2 Installation Packages.....	13
2.4.3 Run test.....	14
2.5 Data sources.....	14
<i>3. sRNAtools.....</i>	<i>15</i>
3.1 Launch Analysis Locally.....	16
3.1.2 Prepare the inputs.....	17
3.1.3 Analysis Start.....	18
3.1.4 Analysis Result Outputs.....	19
3.1.4.1 Mapping Results.....	20
3.1.4.2 Detecting tRF Results.....	23
3.1.4.3 isomir Results.....	24
3.1.4.4 PhasiRNA Results.....	25
3.1.4.5 Novel miRNAs Results.....	26
3.1.4.6 Expression Results.....	28
3.1.4.7 24-nt siRNA Results.....	29
3.2 Launch Analysis Use Website.....	31
3.2.1 Prepare the inputs.....	31
3.2.3 Retrieve the results.....	37
3.2.4 Explain the results.....	38
3.3 Local server VS webserver analysis.....	58
<i>4. References.....</i>	<i>59</i>

1.Introduction

Small non-coding RNAs (sncRNAs) are generated from different genome locus and have revealed playing important roles in biological process, such as cell proliferation, priming of viral reverse transcriptases, regulation of gene expression, RNA processing. Next-generation sequencing (NGS) has provided an unprecedented opportunity to discover and quantify sncRNAs and there are many tools have been developed to analysis sncRNAs. However, to the best of our knowledge, most of available tools pay the main attention to study miRNAs. The comprehensive web-based tool to bench analysis other diverse kind of sncRNAs and their potential function is still lacking. Besides, many small RNAs, such as tRF, rRNA defived sncRNA, and even miRNA, contain diverse kind of RNA modifications. The highly repetitive elements content in these small RNAs impede the identification of sncRNA modification based on conventional immunoprecipitation methods, such as MeRIP-Seq. In this study, a novel web server sRNAtools is presented, which can be used to discovery, profile and functional annotate diverse kind of sncRNAs (including miRNA, piRNA, tRFs, siRNA, snRNA, snoRNA, rRNA, and plant specific 24-nt siRNA, phasiRNA and natsiRNA) which can not only be used in sRNA-Seq, but it is can also be used in single cell sRNA-seq. Beside, sncRNA modifications can be identified with site resolution based on cleavage based technology.

sRNAtools have these main features:

- sRNA can be detected and profiled for as many as 21 species.
- Unbiased classify the sncRNA into different categories.
- Differential expression analysis of sncRNA with paired cases small RNAs transcriptome.
- Differential expression analysis with group case small RNAs transcriptome.
- Gene targets and function prediction for sncRNA.
- sncRNA modification calling using cleavage-based method
- Very user-friendly web interfaces and convenient data analysis queue system.






2. Getting Start

There are four ways to use sRNAtools (or parts of it).















2.1 Webservice

The webservice can be accessed: <https://bioinformatics.caf.ac.cn/sRNAtools/>.

And in the Download Page, all required data sets, pipeline scripts and programs can be downloaded such as sRNAtools Docker image, Dockerfile, virtualbox version, Local version, sRNAtools database and Test_data to help the users to analyze their data locally.

 [DOWNLOAD LOCAL VERSIONS](#)    

Download local versions of sRNAtools

 sRNAtools local package	
 sRNAtools Dockerfile	
 sRNAtools docker image	
 sRNAtools VirtualBox version	
 Local version help document	
 sRNAtools database	
 Test data	

2.2 Docker versions

We integrated all the software and dependencies into a Docker image. Users can download the image and install it on their own local servers by using the command ‘docker load’. And we also provided a Dockerfile which contain a series of commands and parameters that are applied to create a new Docker image by using the command ‘docker build’. The Docker Image and Dockerfile can download from our website.

2.2.1 Docker Installation

- Official installation documentation:

<http://docs.docker.com/engine/installation/#installation>

Here we provide two methods on ubuntu operating system as example:

```
1. $ sudo apt-get install -y docker.io
```

```
2. $ sudo apt-get update
```

```
$ sudo apt-get install curl
```

```
curl -k -sSl https://get.docker.com | sudo sh
```

Note: Different operation systems may have some differences. To refer the official documentation if commands above do not work.

- Verify if Docker is installed correctly

Once Docker installation is completed, we can run **hello-world** image to verify if Docker is installed correctly.

Type the following command:

```
$ docker run hello-world
```

Installed successfully if the screen displays the following information:

```
Hello from Docker!
This message shows that your installation appears to be working correctly.

To generate this message, Docker took the following steps:
1. The Docker client contacted the Docker daemon.
2. The Docker daemon pulled the "hello-world" image from the Docker Hub.
   (amd64)
3. The Docker daemon created a new container from that image which runs the
   executable that produces the output you are currently reading.
4. The Docker daemon streamed that output to the Docker client, which sent it
   to your terminal.

To try something more ambitious, you can run an Ubuntu container with:
$ docker run -it ubuntu bash

Share images, automate workflows, and more with a free Docker ID:
https://hub.docker.com/

For more examples and ideas, visit:
https://docs.docker.com/engine/userguide/
```

Note: The Docker daemon binds to a Unix socket instead of a TCP port. By default that Unix socket is owned by the user root and other users can access it with sudo. For this reason, Docker daemon always runs as the root user.

To avoid having to use sudo when you use the Docker command, create a Unix group called Docker and add users to it. When the Docker daemon starts, it makes the ownership of the Unix socket read/writable by the Docker group.

2.2.2 Dependencies

- Perl 5.14 or above, bioperl with multithread support, and following Perl (CPAN) modules:

Scalar::Util	Data::Dumper	Parallel::ForkManager	Getopt::Long
experimental	SVG	File::Spec	List::Util
Math::CDF	Try::Tiny	JSON	

- python3 (>=V3.0)
- Bowtie, An ultra-fast short read aligner, version (v1.0).
- Samtools, provide various utilities for manipulating alignments in the SAM format, including sorting, merging, indexing and generating alignments in a per-position format (V1.9).
- Bedtools, a flexible suite of utilities for comparing genomic features
Biochemistry of multidrug resistance mediated by the multidrug transporter.
- ViennaRNA, for RNA Comparison and Secondary Structure Prediction (>=V2)
- R and R Packages

seqinr	XML	RCurl	data.table
ggpubr	Jsonlite	GOstats	GenomicAlignments
GenomicFeatures			

2.2.3 sRNAtools Docker Image Installation

2.2.3.1 Docker image installation

Download Docker image from sRNAtools website

(<https://bioinformatics.caf.ac.cn/sRNAtools/download.php>):

```
$ wget https://bioinformatics.caf.ac.cn/sRNAtools/data/download/sRNAtools.image.tar.gz
```

Image uncompressing by the following command:

```
$ tar -zxvf sRNAtools.image.tar.gz
```

Image load by the following command:

```
$ docker load < sRNAtoolsimage
```

View the loaded image by the following command:

```
$ docker images
```

Rename the image by the following command (optional):

```
$ docker tag IMAGE_ID REPOSITORY: TAG
```

2.2.3.2 Construct image from Dockerfile

Download Dockerfile from our website:

```
$ wget https://bioinformatics.caf.ac.cn/sRNAtools/data/download/Dockerfile.zip
```

Create a new image by the following command:

```
$ docker build -t REPOSITORY:TAG .
```

View the new image created by Dockerfile:

```
$ docker images
```

Rename the image by the following command (optional):

```
$ docker tag IMAGE_ID REPOSITORY: TAG
```

2.2.4 Share Directory and Quickly start

Once image installed successfully, start a new container by the following command:

```
$ docker run -it --name=your_own_test_name -v /host_dataset_directory :/dir_of_container/  
REPOSITORY:TAG
```

Note: *host_dataset_directory* is user's private dataset, eg:/home/test; directory of container is container's data directory, eg:/data/; REPOSITORY:TAG is the image's REPOSITORY and TAG, eg:'sRNAtools:v1.0'

Note: The parameter '-v' of Docker can mount host directory of dataset to sRNAtools Docker image, then users can conveniently analysis their own dataset.

2.3 VirtualBox version

VirtualBox is a general-purpose full virtualizer for x86 hardware, targeted at server, desktop and embedded use. Here we also provided a VirtualBox platform for sRNAtools, which has integrated all the software and dependencies. The users can perform the analysis through different operating systems (Windows hosts, OS X hosts, Linux distributions and Solaris hosts). Here, we use sRNAtools VirtualBox run the analysis on Windows operating systems.

2.3.1 VirtualBox installation

Install VirtualBox (<https://www.virtualbox.org/wiki/Downloads>)

Note: Different operating systems select the appropriate version (Windows, OS X, Linux) and Extension Pack. Once installed successfully, restart VirtualBox.

2.3.2 sRNAtools VirtualBox Download

Download the OVA file from our website

```
$wget https://bioinformatics.caf.ac.cn/sRNAtools/data/download/sRNAtools.ova
```

Then open sRNAtools VirtualBox

File → *Import Appliance*

Import the OVA file, in which all required software and dependencies are already installed.



2.3.3 VirtualBox Configure

Once imported successfully, you will find the tool in the left menu of VirtualBox.

Click on it to start the configuration:

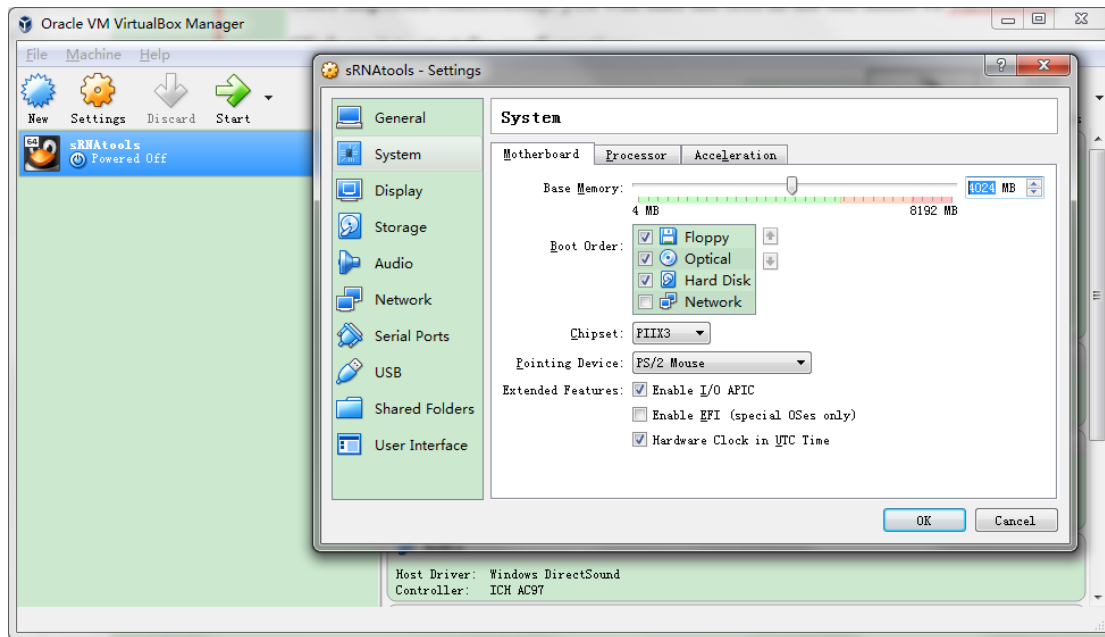
Change the memory allocation

Settings (orange gear) → System → Motherboard → Base Memory

Change the number of CPUs

Settings (orange gear) → System → Processor → Processor(s)

Note: Four or more CPUs are suggested.



2.3.4 Share Directory

With the analysis going on, VirtualBox disk will grow maddeningly. To facilitate users to store datasets, scripts, assemblies and results, it is strongly recommended that users share host folders.

Select the shared folder:

Settings (orange gear) → Shared Folders

To ensure that shared files are accessible, two types of information are needed:

- 1) The shared folder's name in the virtual machine
- 2) The folder's path and name in the host machine

Note: it is important to select auto load function

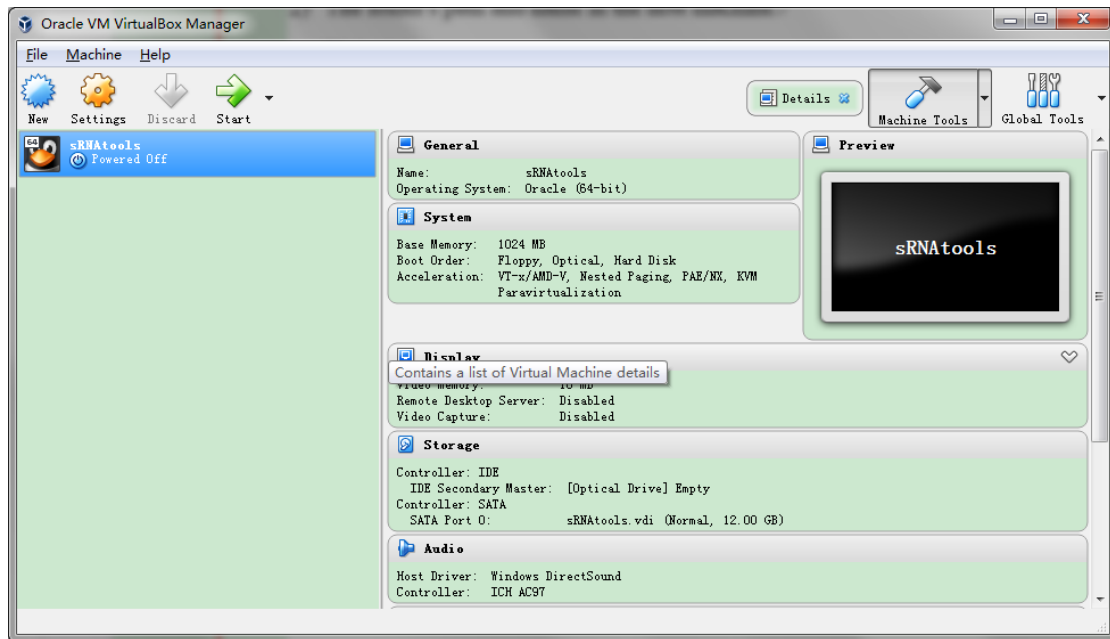
2.3.5 Start VirtualBox

Once the configuration is completed, select the VirtualBox interface and click the start button (green arrow) and the tool will start in a new window. Use the following account to login in and enter your credentials:

User: sncrnabench

Password: sRNAtools

If the users want to change the password, type the command: *passwd*



2.3.6 Change keyboard layout and time zone (Optionally)

Typing the following command in a terminal to change the keyboard layout:

```
sudo dpkg-reconfigure keyboard-configuration
```

Enter your password and select layout

Typing the following command in a terminal to change your time zone:

```
sudo tzselect
```

Enter your password and select your time zone

2.4 Local program package

2.4.1 Minimum Requirements

2.4.1.1 Required Software Packages:

* Perl 5.16 or above, installed with multithread support, and following Perl (CPAN) modules:

Scalar::Util
Data::Dumper
Parallel::ForkManager
Getopt::Long
Experimental
SVG
File::Spec
List::Util
Math::CDF
Try::Tiny
JSON

* Python3

* Bowtie and Bowtie2: An ultra-fast short read aligner, both version 1.0 and version 2.0.

* Samtools: Provide various utilities for manipulating alignments in the SAM format, including sorting, merging, indexing and generating alignments in a per-position format.

* Bedtools: A flexible suite of utilities for comparing genomic features. Biochemistry of multidrug resistance mediated by the multidrug transporter.

* ViennaRNA: For RNA Comparison and Secondary Structure Prediction ($\geq V2$)

* R and R Packages:

seqinr
XML
RCurl
data.table
ggpubr
jsonlite
GOstats
GenomicAlignments
GenomicFeatures

2.4.1.2 Other program included in the distribution:

- * sRNAtools/program/external/megablast
- * sRNAtools/program/external/ShortStack
- * sRNAtools/program/mireap
- * sRNAtools/program/mirtrace
- * sRNAtools/program/phasiRNAClassifier
- * sRNAtools/program/PHASIS
- * sRNAtools/program/TargetFinder_1.6

2.4.2 Installation Packages

- * Perl (<http://www.perl.org/>)
- * Python3 (<https://www.python.org/>)
- * Bowtie2
(https://sourceforge.net/projects/bowtie-bio/files/bowtie2/2.3.4.3/bowtie2-2.3.4.3-linux-x86_64.zip)
- * Bowtie
(https://sourceforge.net/projects/bowtie-bio/files/bowtie/1.2.2/bowtie-1.2.2-linux-x86_64.zip)
- * Samtools
(<https://github.com/samtools/samtools/releases/download/1.9/samtools-1.9.tar.bz2>)
- * Bedtools
(<https://github.com/arq5x/bedtools2/releases/download/v2.25.0/bedtools-2.25.0.tar.gz>)
- * ViennaRNA
(https://www.tbi.univie.ac.at/RNA/download/sourcecode/2_4_x/ViennaRNA-2.4.10.tar.gz)
- * R (<https://www.r-project.org/>)

Note: Different computers may exist some differences, to access the official documentation.

2.4.3 Run test

The detailed steps can be found in the next chapter (3.1 Launch Analysis In Local).

2.5 Data sources

	<i>Homo sapiens</i> (hg38) ²	<i>Mus musculus</i> (mm10) ²	<i>Rattus norvegicus</i> (rn6) ²	<i>Danio rerio</i> (GRC-11) ²	<i>Pan troglodytes</i> (Pan_tro.3.0) ²	<i>Gorilla gorilla gorilla</i> (gor.Gor4) ²	<i>Gallus gallus</i> (GRC600) ²	<i>Sus scrofa</i> (susSer11) ²	<i>Macaca mulatta</i> (vieMac8) ²	<i>Drosophila melanogaster</i> (Dm6) ²	<i>Caenorhabditis elegans</i> (CE11) ²	<i>Arabidopsis thaliana</i> (TAIR10) ²	<i>Oryza sativa</i> (MSU7.0) ²	<i>Zea mays</i> (B73_RefGen_v4) ²	<i>Glycine max</i> (Wm82.a2) ²	<i>Sorghum bicolor</i> (NCBIv3) ²	<i>Solanum lycopersicum</i> (SL3.0) ²	<i>Brachypodium distachyon</i> (v3.0) ²	<i>Medicago truncatula</i> (v1.0) ²	<i>Populus trichocarpa</i> (Pop.tr1.v3) ²	<i>Vitis vinifera</i> (Grapevine 12X) ²
Reference genome sequences ²																					
Phytozome ³	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ensembl ⁴	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
RNA libraries ⁵																					
NONCODE ⁵	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CANTATAdb 2.0 ⁶	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CIRCpedia v2.0 ⁷	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
circAtlas 2.0 ⁸	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
PlantcircBase ⁹	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
RNAcentral NAT ¹⁰	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
PlantNATsDB ¹¹	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
piRBase ¹²	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
piRNAclusterDB ¹³	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pln24NT ¹⁴	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
miRBase ¹⁵	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
GtRNAdb ¹⁶	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Rfam ¹⁷	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

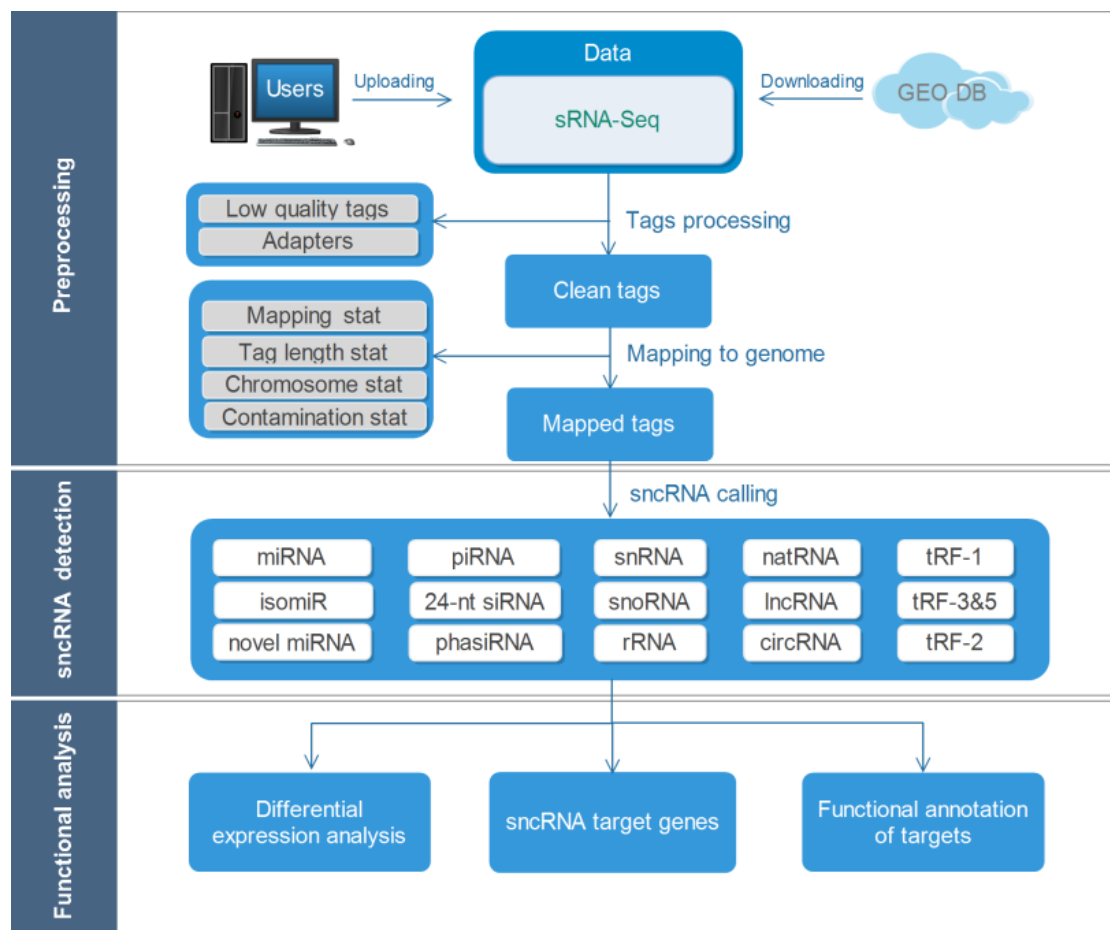
Dataset available for the 21 supported model species

Precursor and mature tRNA sequences were downloaded from the GtRNAdb database [1]. miRNA precursor sequences and other known small noncoding RNA sequences were retrieved from the latest miRBase database (release v22.0) (<http://www.mirbase.org/>) and latest RFAM database (release v14.0) [2], respectively. mRNA sequences and gene annotations were downloaded from the Phytozome database (v12.1) [3] for plant species and the Ensembl database [4] for other species. The lncRNA sequences were obtained from NONCODE [5] and CANTATAdb [6]. circRNA sequences were retrieved from CIRCpedia v2 [7], CircFunBase [8], and PlantcircBase [9]. Natural antisense transcripts (NATs) were downloaded from RNAcentral [10] and PlantNATsDB [11]. piRNAs and the loci that produce them were retrieved from piRBase [12] and piRNAclusterDB [13], respectively. Plant 24-nt siRNA-producing loci were obtained from Pln24NT [14]. Sequences containing tRNA genes and 100 bp downstream of the 3'-end of such genes were extracted as precursor

tRNA genes. For mature tRNA sequences, the introns were removed and “CCA” was added to the 3'-end of the tRNA gene sequences. The fully supported species for different RNA classes (miRNA, tRF, tRNA, piRNA/24-nt siRNA, piRNA/24-nt siRNA-producing sites, lncRNA, circRNA, NAT, snRNA, snoRNA, and rRNA) include human, mouse, rat, chicken, pig, Arabidopsis, rice, maize, soybean, and tomato. Other sncRNAs, including tRF, isomiR, novel miRNA, phasiRNA can be detected ab initio based on genomic sequences.

3. sRNAtools

The sRNAtools, which can be used to discovery, profile and functional annotate sncRNAs (including miRNA, piRNA, tRFs, siRNA, snRNA, snoRNA, rRNA, and plant specific 24-nt siRNA, phasiRNA and natsiRNA) based on high-throughput sequencing. The figure below shows the sRNAtools work flow.



Workflow of sRNAtools

3.1 Launch Analysis Locally

Assume that we have successfully installed sRNAtools Docker image, sRNAtools VirtualBox, and the locally_version_pipeline. Then we can launch analysis in our local server.

3.1.1 Database and test data download

The database download website: <https://bioinformatics.caf.ac.cn/sRNAtools/db/>, users can download required datasets. (As shown in the figure below):

Index of /sRNAtools/db

Name	Last modified	Size	Description
Parent Directory			-
24nt/	2019-09-13 19:57		-
circRNA/	2019-08-22 13:37		-
coordinate/	2019-09-18 13:14		-
dashr/	2019-08-22 17:14		-
functional/	2019-07-14 09:42		-
genome/	2019-09-14 13:09		-
idmap/	2019-07-11 10:49		-
lncRNA/	2019-09-05 13:02		-
mRNA/	2019-07-14 09:48		-
mirbase/	2019-08-12 13:12		-
mito_tRNA/	2019-08-14 05:58		-
nat/	2019-08-09 18:41		-
organism.txt	2019-08-13 12:29	603	
pirna/	2019-09-05 12:40		-
ppi/	2019-07-14 09:08		-
repeat/	2019-09-13 20:43		-
rfam/	2019-08-27 20:18		-
macenter/	2019-08-27 20:19		-
tRNA/	2019-09-05 11:41		-
targetdb/	2019-07-14 08:39		-

And test data download website:

<https://bioinformatics.caf.ac.cn/sRNAtools/data/testdata/>, where users can choose the data to run a test analysis. (As shown in the figure below):

Index of /sRNAtools/data/testdata

Name	Last modified	Size	Description
Parent Directory		-	
GSM632659_Group_C.fasta	2019-06-29 11:14	36M	
GSM632660_Control.fasta	2019-06-29 11:14	25M	
Mouse-INPUT-m7G.fa.gz	2019-06-29 11:20	17M	
Mouse-TREAT-m7G.fa.gz	2019-06-29 11:20	12M	
SRR6442216.clean.fa	2019-01-08 08:03	15M	
SRR6442216.clean.fa.gz	2019-06-29 11:21	3.9M	
SRR6442217.clean.fa	2019-01-08 08:03	29M	
SRR6442217.clean.fa.gz	2019-06-29 11:21	7.9M	
SRR6442218.clean.fa	2019-01-08 08:03	33M	
SRR6442218.clean.fa.gz	2019-06-29 11:22	9.0M	
SRR6442219.clean.fa	2019-01-08 08:03	28M	
SRR6442219.clean.fa.gz	2019-06-29 11:22	7.5M	
TGGAATTCTCGG_trimmed.>	2019-07-16 10:51	20	
TGGAATTCTCGG_trimmin.>	2019-07-16 10:51	1.3K	
arabidopsis_SRR08961.>	2019-06-29 11:07	180M	
arabidopsis_SRR08961.>	2019-07-16 11:14	145M	
arabidopsis_SRR61928.>	2019-06-29 11:13	398M	
arabidopsis_SRR61928.>	2019-07-16 11:18	263M	
formconver.pl	2019-06-29 11:13	193	
human_neural_stem_ce.>	2019-06-29 11:14	3.9M	
human_SRR6442216.fas.>	2019-06-29 11:19	325M	
human_SRR6442216_tri.>	2019-07-16 11:06	75M	
human_SRR7062776.fas.>	2019-07-16 09:08	148M	
human_SRR7062776_tri.>	2019-07-16 10:56	66M	
human_neural_stem_ce.>	2019-03-16 09:25	3.9M	
mESC.clean.fa	2019-01-09 06:10	95M	
mESC.clean.fa.gz	2019-03-16 09:25	2.8M	
mESC.clean.fake.fa	2019-03-22 09:00	91M	
mESC_v6.5.clean.fa.gz	2019-06-29 11:20	25M	
mouse.mESC.clean.fa.gz	2019-06-29 11:20	2.8M	
populus.control.clea.>	2019-06-29 11:20	9.1M	

3.1.2 Prepare the inputs

There have some configurations before we execute the analysis:

- Config ViennaRNA perl lib path in /program/mireap/bin/mireap.pl and /program/miRDeep_v2, such as "/software/ViennaRNA-2.4.10/interfaces/Perl/"
- Make 'megablast' and 'ShortStack' executable using command 'chmod a+x /program/external/*'
- Modify the script PHASIS /phastrigs.py: phaster_path="/sRNAtools/program/PHASIS/", make sure the path match your own server
- Modify the script mireap.pl, miRDeep_v2.pl and isomiR2Function_rev.pl, make sure the path or lib path match your own server

- Config db/ path in DBCONFIG.txt (as shown in the figure below)

```
dbbasepath=/sRNAtools/db
genomepath=/sRNAtools/db/genome
mirbasepath=/sRNAtools/db/mirbase
rfampath=/sRNAtools/db/rfam
mrnapath=/sRNAtools/db/mRNA
lncRNAPath=/sRNAtools/db/lncRNA
circRNAPath=/sRNAtools/db/circRNA
tRNAPath=/sRNAtools/db/tRNA
24ntpath=/sRNAtools/db/24nt
rebasepath=/sRNAtools/db/repeat
natrnpath=/sRNAtools/db/nat
```

3.1.3 Analysis Start

Launch the test analysis by typing the following command (Docker images have initiated above):

```
$ perl ./program/run.pl -infile test_data/D-MH-CM-1.fa -jobid 123456
```

Note: The jobid named by user himself and the absolute paths of perl scripts and test data are suggested, and if you want to perform animal analysis, you should take the perl script 'run_animal.pl', and there also have the script 'run_modify.pl' you can choose to perform sncRNA modification analysis.

Note: The test data and db data sets download from our website, if users take their own test data and data sets to perform analysis, please remember the index files of fasta sequence must be created by using blast or bowtie softwares.

We can find many files in the temporary folder of the result files during the analysis, and once the analysis completely, the temporary folder disappeared ultimately. And there some informations printed to the screen.

```

root root 109M Dec 27 21:14 123456.togenome.bam
root root 127M Dec 27 21:14 123456.togenome.bed
root root 17 Dec 27 21:14 all.mapping.stat.txt
root root 40M Dec 27 21:14 123456.togenome.read.txt
root root 152M Dec 27 21:14 123456.nonMir.fa
root root 991K Dec 27 21:14 123456.mirmapped.fa
root root 56M Dec 27 21:14 123456.tomir.bam
root root 352K Dec 27 21:15 123456.tomir.read.txt
root root 1.5M Dec 27 21:15 123456.rfammapped.fa
root root 152M Dec 27 21:15 123456.nonMirRfam.fa
root root 57M Dec 27 21:15 123456.toRfam.bam
root root 514K Dec 27 21:15 123456.toRfam.read.txt
root root 124M Dec 27 21:15 123456.nonMirRfamMRNA.fa
root root 70M Dec 27 21:15 123456.tomRNA.bam
root root 9.9M Dec 27 21:15 123456.tomRNA.read.txt
root root 62M Dec 27 21:16 123456.pritRNA.bam
root root 8.4M Dec 27 21:16 123456.topritRNA.read.txt
root root 61M Dec 27 21:16 123456.tRNA.bam
root root 966K Dec 27 21:17 123456.totRNA.read.txt

```

Temp intermediate file

```

#### Fn: Settings Reader #####
User Input runType      : G
User Input reference location : /mount/db/genome/ptc.fa
User Input Libs         : test_data/D-MH-CM-1.fa
user library format     : F
User Input for phase length : 21,24
User Input index location  : /mount/db/genome/ptc.fa
User Input for min. sRNA depth : 3
User Input distance b/w clusters : 300
Index status            : User specified
#### Converting FASTA format to counts #####
nprocPP                : 5

#### Fn: phaser #####
sRNA library located - Running phasing analysis
test_data/D-MH-CM-1.fa
Thu Dec 27 13:19:13 2018: Phasing analysis begins...
Thu Dec 27 13:19:13 2018: aligning...
Thu Dec 27 13:19:13 2018: script version is 1..
bowtie -f -a -v 0 -m 12 -p 3 /mount/db/genome/ptc.fa test_data/D-MH-CM-1.fa >././data/results/123456/phasedetect/D-MH-CM-1.txt.A4xp9YjTLz
# reads processed: 4266012
# reads with at least one reported alignment: 1053218 (24.69%)
# reads that failed to align: 3014060 (70.65%)
# reads with alignments suppressed due to -m: 198734 (4.66%)
Reported 2000380 alignments
Thu Dec 27 13:19:26 2018: aligning is done
././data/results/123456/phasedetect/D-MH-CM-1.txt.score
Thu Dec 27 13:19:36 2018: finished the loading of sRNA alignment data.
Thu Dec 27 13:19:37 2018: finished the sorting of presplitted sRNA clusters data
Thu Dec 27 13:19:37 2018: total 3565 presplitted sRNA clusters
Thu Dec 27 13:22:19 2018: Clusters have been scored...

#### Phasing Analysis finished successfully
#### Results are in folder: ./data/results/123456/phasedetect
#### 'phasmerge' can be run by command: python3 phasmerge -mode merge -dir ./data/results/123456/phasedetect

```

Phasing Analysis information in screen

3.1.4 Analysis Result Outputs

After the analysis completely, we will find many files in the folder where the results are stored, such as tRF detect results files (tRF-1.aln.txt, tRF-1.aln.web.txt, tRF-2.aln.txt, tRF-2.aln.web.txt, tRF-3.aln.txt, tRF-3.aln.web.txt, tRF-5.aln.txt,

tRF-5.aln.web.txt, RFstat.txt, RFs.txt), mapping results files (mapping_stat_by_tRFs_loci.txt, mapping_stat_by_tRNA.txt, mapping_stat_by_type.txt, mapping_stat.txt) and some analysis results in the subdirectory such as phasing analysis result directory `./phasedetect/`. Here we select some result files to explain the analysis result.

```

24nt.chr.stat.txt          novel_mirna_expression.list  torRNA.length.stat.pdf
24nt.cluster.length.stat.txt  novelpremature              torRNA.length.stat.png
24nt.loci.exp.read.txt       phasdetect                  torRNA.lengthStat.txt
24nt.loci.exp.txt           phasmerge                   tosnoRNA.length.stat.pdf
24nt.mapped.stat.txt        precursors.fa               tosnoRNA.length.stat.png
24nt.overlap.transposon.stat.txt  predictions                 tosnoRNA.lengthStat.txt
24nt.repeat.stat.txt        pritrRNA.sort.bam          tosnRNA.length.stat.pdf
24nt.transposons.tags.stat.txt  pritrRNA.sort.bam.bai     tosnRNA.length.stat.png
24nt.uniq.repeat.stat.txt     results_nr.txt             tosnRNA.lengthStat.txt
24nt.unmapped.stat.txt      results.txt                 total_mapping_stat_tRNA.txt
allmappingstat.txt          rfam_result_download       totRNA.length.stat.pdf
blast_dir_temp              rfam_sum_stat              totRNA.length.stat.png
chr.stat.txt                rRNA_stat_table.result    totRNA.lengthStat.txt
cluster_stat_for_plot.R     signatures                  tRF-1.aln.txt
hairpinFold                 snoRNA_stat_table.result  tRF-1.aln.web.txt
isomiRfinder.log            snRNA_stat_table.result   tRF-2.aln.txt
isomiR.read.txt             status.json                 tRF-2.aln.web.txt
length_stat.txt             sum_stat                   tRF-3.aln.txt
mapping_stat_by_tRFs_loci.txt  temp                       tRF-3.aln.web.txt
mapping_stat_by_tRNA.txt     templated                  tRF-5.aln.txt
mapping_stat_by_type.txt    templated_raw.quant1      tRF-5.aln.web.txt
mapping_stat.txt            togenome.length.stat.pdf  tRFstat.txt
miRNA_result_download       togenome.length.stat.png  tRFs.txt
miRNA_stat_table.result     togenome.lengthStat.txt  tRNA.sort.bam
miRNA_sum_stat              tomir.length.stat.pdf     tRNA.sort.bam.bai
non-templated               tomir.length.stat.png     tRNA_stat_table.result
non-templated_raw.quant1    tomir.lengthStat.txt     tRNA_sum_stat
novel_aln_mirdeep.txt       tootherRfam.length.stat.pdf  unique_mapping_stat_tRNA.txt
novel_hairpin.fa            tootherRfam.length.stat.png
novel_mature.fa             tootherRfam.lengthStat.txt

```

result files

3.1.4.1 Mapping Results

As the process begins to be analyzed, the first step is map the test sequence to sncRNAs sequence and genome, such as the step ‘mapping to genome’, ‘mapping to miRNA’, ‘mapping to rfam’, ‘mapping to the mRNA sequences’, ‘mapping to the pritrRNA sequences’ and ‘mapping to the mature tRNA sequences’. The mapping informations printed to the screen (in the figure below) and the result files (such as togenome.bam, tomir.bam, tomRNA.bam, toRfam.bam, tRNA.bam) generated in a temporary directory, the bam files contains the alignment result information of the sequences as shown in the figure below:

```

# reads processed: 4266012
# reads with at least one reported alignment: 3187906 (74.73%)
# reads that failed to align: 1078106 (25.27%)
Reported 3187906 alignments
# reads processed: 4266012
# reads with at least one reported alignment: 27480 (0.64%)
# reads that failed to align: 4238532 (99.36%)
Reported 27480 alignments
# reads processed: 4266012
# reads with at least one reported alignment: 39999 (0.94%)
# reads that failed to align: 4226013 (99.06%)
Reported 39999 alignments
mapping siRNA to the mRNA sequences
# reads processed: 4226013
# reads with at least one reported alignment: 789738 (18.69%)
# reads that failed to align: 3436275 (81.31%)
Reported 789738 alignments
mapping siRNA to the tRNA sequences
# reads processed: 4238532
# reads with at least one reported alignment: 75517 (1.78%)
# reads that failed to align: 4158585 (98.11%)
# reads with alignments suppressed due to -m: 4430 (0.10%)
Reported 665391 alignments
# reads processed: 4238532
# reads with at least one reported alignment: 52303 (1.23%)
# reads that failed to align: 4181630 (98.66%)
# reads with alignments suppressed due to -m: 4599 (0.11%)
Reported 588803 alignments

```

Alignments information in screen

100000153_x5937_0	scaffold_1.trna42-AlaAGC[1:44658543-44658715:++[1-73	1	255	18M	*	0	0	GGGGATGTAGCTCAGAC	IIIIIIIIIIIIIIIIIIII	XA:1:2	MD:Z:
100000153_x5937_0	scaffold_1.trna42-AlaAGC[1:44658543-44658715:++[1-73	1	255	20M	+	0	0	GGGGATGTAGCTCAGAGGG	IIIIIIIIIIIIIIIIIIII	XA:1:2	MD:Z:
100000153_x5937_0	scaffold_1.trna42-AlaAGC[1:44658543-44658715:++[1-73	1	255	18M	*	0	0	GGGGATGTAGCTCAGTGG	IIIIIIIIIIIIIIIIIIII	XA:1:2	MD:Z:
100000153_x5937_0	scaffold_1.trna42-AlaAGC[1:44658543-44658715:++[1-73	1	255	19M	*	0	0	GGGGATGTAGCTCAGACGT	IIIIIIIIIIIIIIIIIIII	XA:1:2	MD:Z:
100000153_x5937_0	scaffold_1.trna42-AlaAGC[1:44658543-44658715:++[1-73	1	255	19M	*	0	0	GGGGATGTAGCTCAGAGGA	IIIIIIIIIIIIIIIIIIII	XA:1:2	MD:Z:
100000153_x5937_0	scaffold_1.trna42-AlaAGC[1:44658543-44658715:++[1-73	1	255	19M	*	0	0	GGGGATGTAGCTCAGTGGG	IIIIIIIIIIIIIIIIIIII	XA:1:2	MD:Z:
100000153_x5937_0	scaffold_1.trna42-AlaAGC[1:44658543-44658715:++[1-73	1	255	23M	*	0	0	GGGGATATAGCTCAGTTGTTAGA	IIIIIIIIIIIIIIIIIIII	XA:1:2	MD:Z:
100000153_x5937_0	scaffold_1.trna42-AlaAGC[1:44658543-44658715:++[1-73	1	255	19M	*	0	0	GGGGATGTAGCTCAGGGGG	IIIIIIIIIIIIIIIIIIII	XA:1:2	MD:Z:
100000153_x5937_0	scaffold_1.trna42-AlaAGC[1:44658543-44658715:++[1-73	1	255	17M	*	0	0	GGGGATGTAGCTCAGTG	IIIIIIIIIIIIIIIIIIII	XA:1:2	MD:Z:
100000153_x5937_0	scaffold_1.trna42-AlaAGC[1:44658543-44658715:++[1-73	1	255	18M	*	0	0	GGGGATGTAGCTCAGGGG	IIIIIIIIIIIIIIIIIIII	XA:1:2	MD:Z:

Alignments result bam files

The result file allmappingstat.txt contains information of mapping statistics on different types of RNAs, including miRNA, pritrRNA, genome, other RFAM ncRNA, mRNA, lncRNA, and other mapping. Statistics are based on unique tags (second column) and total reads count (last column).

mir	461	955
pritrNA	1295	18232
genome	3187906	17549021
mRNA	16261	47359
Rfam	836	3553
other	1059253	1983874

allmappingstat.txt

The result file *mapping_stat_by_tRFs_loci.txt* contains information of mapping statistics of tRNA loci, and the result file *mapping_stat_by_type.txt* contains information of mapping statistics of tRNA Type, and there is also a summary mapping file *total_mapping_stat_tRNA.txt*. These three files as shown in the figure below:

#tRNA loci	Unique read	Total read
5'end	72854	43374223
Inner	330153	6294070
3'end	153742	4269830
3'trailer	31	41

mapping_stat_by_tRFs_loci.txt

#tRNA	Total read at 5'end	Total read at inner region	Total read at 3'end	Total read at 3'Trailer	Sum total read
scaffold_1.trna1-ProTGG	6	1279	15806	0	17091
scaffold_1.trna10-Undet???	1499	941	4576	3	7019
scaffold_1.trna11-ThrTGT	111081	5531	5903	1	122516
scaffold_1.trna12-SerGCT	3101	3622	7820	0	14543
scaffold_1.trna13-LeuAAG	0	1514	35	1	1550
scaffold_1.trna14-SerAGA	4252	2621	1045	0	7918
scaffold_1.trna15-ArgTCT	825	3280	3182	0	7287
scaffold_1.trna16-CysGCA	1656	5072	1454	0	8182
scaffold_1.trna17-LeuAAG	0	656	2	0	658
scaffold_1.trna18-SerGCT	3106	3626	7746	0	14478
scaffold_1.trna19-GlyTCC	3584	3974	4983	0	12541
scaffold_1.trna2-AlaAGC	7661	13479	2679	0	23819
scaffold_1.trna20-GlyTCC	3588	3992	5185	0	12765
scaffold_1.trna21-SerGCT	3101	3622	7820	0	14543
scaffold_1.trna22-AlaTGC	1396	3318	3194	0	7908
scaffold_1.trna23-GluTTC	13196	11700	7793	0	32689
scaffold_1.trna24-GlyGCC	7901	37808	30082	0	75791
scaffold_1.trna25-GluCTC	28603	20685	13329	0	62617
scaffold_1.trna26-ProTGG	6	1279	15806	0	17091
scaffold_1.trna27-ProTGG	6	1279	15806	0	17091
scaffold_1.trna28-ProAGG	6	1048	12035	0	13089
scaffold_1.trna29-ProTGG	6	1279	15806	0	17091
scaffold_1.trna3-MetCAT	120	2818	614	0	3552
scaffold_1.trna30-ProTGG	0	473	10160	0	10633

total_mapping_stat_tRNA.txt

#tRNA	Type	Unique read	Total read
AlaAGC	15839	476890	
AlaCGC	1709	23023	
AlaTGC	11441	339867	
ArgACG	7091	3552166	
ArgCCG	2205	224305	
ArgCCT	20310	35878584	
ArgTCG	6610	1191532	
ArgTCT	10538	83912	
AsnGTT	19063	330418	
AspGTC	1076	8881	
CysACA	650	6854	
CysGCA	9589	90867	
GlnCTG	20945	267823	
GlnTTG	14069	375661	
GluCTC	51169	1300542	
GluTTC	14980	457646	
GlyCCC	4907	105875	
GlyGCC	53619	1890742	
GlyTCC	17091	214407	
HisGTG	19019	2074007	
IleAAT	17662	249661	
IleGAT	73	374	
IleTAT	3536	197538	
LeuAAG	21771	168478	

mapping_stat_by_type.txt

3.1.4.2 Detecting tRF Results

There are about ten result files, tRF-1.aln.txt, tRF-1.aln.web.txt, tRF-2.aln.txt, tRF-2.aln.web.txt, tRF-3.aln.txt, tRF-3.aln.web.txt, tRF-5.aln.txt, tRF-5.aln.web.txt, tRFstat.txt and tRFs.txt. The file tRFstat.txt contains the information of tRFs type(first column), Unique read(second column) and Total read(last column), and the file tRFs.txt summarize the results of detect which contains the information of tRF type, source tRNA, RF length, locus on tRNA, read number, RPM, sequence and P-value. Two files shown in the figure below:

#tRFs	type	Unique read	Total read
tRF-5	112	9489335	
tRF-3	25	21869	
tRF-2	175	663114	
tRF-1	0	0	

tRFstat.txt

tRF-2	scaffold_8.trna6-HisGTG	18	37-54	61647	25017.62276	GCCGTGGAGACCTGGGCT	0.00	0.87497161348927	
tRF-2	scaffold_8.trna6-HisGTG	19	37-55	26249	10652.38503	GCCGTGGAGACCTGGGCTC	0.00	0.821230798110315	
tRF-2	scaffold_8.trna6-HisGTG	17	37-53	16786	6812.10465	GCCGTGGAGACCTGGGC	0.00	0.73558282208589	
tRF-2	scaffold_8.trna6-HisGTG	18	37-54	7718	3132.12342	GCCGTAGAGACCTGGGCT	0.00	0.876149392666591	
tRF-2	scaffold_8.trna6-HisGTG	17	39-55	5012	2033.97287	CTGGAGACCTGGGCTC	0.00	0.808256732785035	
tRF-2	scaffold_8.trna6-HisGTG	17	37-53	4943	2005.97124	GCCGTAGAGACCTGGGC	0.00	0.819191249585681	
tRF-2	scaffold_8.trna6-HisGTG	19	37-55	4623	1876.10865	GCCGTAGAGACCTGGGCTC	0.00	0.809065453272664	
tRF-2	scaffold_8.trna6-HisGTG	19	36-54	4139	1679.69148	GCCGTGGAGACCTGGGCT	0.00	0.791698546289212	
tRF-2	scaffold_8.trna6-HisGTG	20	36-55	3402	1380.60169	GCCGTGGAGACCTGGGCTC	0.00	0.75751503006012	
tRF-5	scaffold_1.trna66-LeuCAA		23	1-23	87	35.30639	GCCTTGGTGGTAAATGGTAGAC	0.00	0.813084112149533
tRF-5	scaffold_15.trna5-IleGAT		19	1-19	80	32.46565	CCGACCTTAGCTCAGTGGC	0.00	0.571428571420571
tRF-3	scaffold_15.trna5-IleGAT		17	59-75	186	75.48263	TCCGGCAGTCTGGACCA	0.00	0.978947368421053
tRF-3	scaffold_14.trna12-ArgTCT		17	60-76	376	152.58855	TCCCACGGCGTGCCA	0.00	0.397883597883598
tRF-2	scaffold_14.trna12-ArgTCT		21	36-56	133	53.97414	TAATCAAGCGATTGTGGGTTTC	0.00	0.685567010309278
tRF-2	scaffold_14.trna12-ArgTCT		17	40-56	112	45.45191	CAGCCGACTGTGGTTC	0.00	0.7
tRF-5	scaffold_4.trna32-GlnTTG		23	1-23	1092	443.15610	GGTTCATGGTGTAGTGGTTAGC	0.00	0.420161600615621
tRF-2	scaffold_16.trna16-MetCAT		19	38-56	50	20.29103	TTGGCGAATTCGTAGTTC	0.00	0.833333333333333
tRF-5	scaffold_15.trna21-AsnGTT		20	1-20	6201	2516.49356	GCTGGAATAGCTCAGTTGGC	0.00	0.676817288801572
tRF-5	scaffold_1.trna76-GlnCTG		18	1-18	490	198.85210	GGTTCATGGTCTAGCGG	0.00	0.306825297432686
tRF-2	scaffold_12.trna2-AlaCGC		21	36-56	1932	784.04541	CTTCCGAGAGGCACGGGTTTC	0.00	0.888038112923463
tRF-2	scaffold_12.trna2-AlaCGC		21	36-56	418	169.63301	CGTCCGAGAGGCACGGGTTTC	0.00	0.910675381263617
tRF-2	scaffold_12.trna2-AlaCGC		20	37-56	130	52.75668	TTCCGAGAGGCACGGGTTTC	0.00	0.474452554744526
tRF-2	scaffold_12.trna2-AlaCGC		20	36-55	123	49.91593	CTTCCGAGAGGCACGGGTTTC	0.00	0.75
tRF-2	scaffold_12.trna2-AlaCGC		20	37-56	104	42.20534	CTCCGAGAGGCACGGGTTTC	0.00	0.722222222222222
tRF-2	scaffold_14.trna8-HisGTG		18	32-49	61647	25017.62276	GCCGTGGAGACCTGGGCT	0.00	0.885376572643189
tRF-2	scaffold_14.trna8-HisGTG		19	32-50	26249	10652.38503	GCCGTGGAGACCTGGGCTC	0.00	0.843070499437932

tRFs.txt

3.1.4.3 isomir Results

The result file *isomiR.stat.txt* contains information of distribution of isomiR, including templated isomiR and non-templated isomiR (SS, MS, TS, CV and 3V) and the result file *isomiR.read.txt* contains the names of reads, and the log file *isomiRfinder.log* (shown below) have recorded the information of analysis date, the sequences source, the results storage location and analysis parameters. So we can quickly find the results by viewing this file

templated	templated	2520160	9022
non-templated	SS	209760	12692
non-templated	MS	24476	5516
non-templated	TS	139616	3550
non-templated	CV	22718	3150
non-templated	3V	42976	2228

isomiR.stat.txt


```
isomiRFinder start at Thu Mar  7 13:43:52 2019

Parameters:
  unique small RNA sequence file: test_data/D-MH-CM-1.fa
  pre-miRNA sequence file: /snrnabench/db/mirna/ptc_pri.fa
  miRNA sequence file: /snrnabench/db/mirna/ptc_mature.fa
  save results at: ./data/results/190307/
  species info:
  max internal snp allowed: 0
  max terminal substitution allowed: 0
  min isomiR abundance: 1
  min overlapping with canonical miRNA: 16
  min isomiR sequence length: 16
  max isomiR sequence length: 28
  max substitution or deletion in seed region: 0
  start of seed region: 2
  end of seed region: 8
  how many cores used for analysis: 2
```

isomiRFinder.log

3.1.4.4 PhasiRNA Results

This step we take PHASIS suite for phased clusters discovery, comparison, annotation and to identify miRNA triggers. PHASIS consists of the following three components:

1. Phasdetect: Identifies library-specific phased-loci, prepares and organizes the ancillary data.

2. Phasmerge: Summarizes the libraries-specific results to a non-overlapping genome- or transcriptome-level set, quantifies phased siRNAs, annotates the *PHAS* loci by matching to genome annotations or transcriptome GTF and facilitates comparison of *PHAS* summaries.

3. Phastrigs: Identifies miRNAs triggers for *PHAS* loci, if PARE data is provided then PARE-supported triggers are reported. It should be noted that the additional experimental data is optional and *phastrigs* can identify miRNA triggers with high accuracy even without such data.

The analysis will create three directories which named phasdetect, phasmerge and phastrigs respectively, the results files as shown in the figure below:

```
D-MH-CM-1.txt.cluster.boundary.without.PARE.validation.list
D-MH-CM-1.txt.output_all_sRNA_21_out.txt
D-MH-CM-1.txt.output_all_sRNA_24_out.txt
D-MH-CM-1.txt.score_p0.0001_sRNA_21_out.cluster
D-MH-CM-1.txt.score_p0.0001_sRNA_24_out.cluster
D-MH-CM-1.txt.score_p0.0005_sRNA_21_out.cluster
D-MH-CM-1.txt.score_p0.0005_sRNA_24_out.cluster
D-MH-CM-1.txt.score_p0.001_sRNA_21_out.cluster
D-MH-CM-1.txt.score_p0.001_sRNA_24_out.cluster
D-MH-CM-1.txt.score_p0.005_sRNA_21_out.cluster
D-MH-CM-1.txt.score_p0.005_sRNA_21_out.txt
D-MH-CM-1.txt.score_p0.005_sRNA_24_out.cluster
D-MH-CM-1.txt.score_p0.005_sRNA_24_out.txt
D-MH-CM-1.txt.score_p1e-05_sRNA_21_out.cluster
D-MH-CM-1.txt.score_p1e-06_sRNA_21_out.cluster
D-MH-CM-1.txt.score_p1e-07_sRNA_21_out.cluster
D-MH-CM-1.txt.score_p5e-05_sRNA_21_out.cluster
D-MH-CM-1.txt.score_p5e-06_sRNA_21_out.cluster
D-MH-CM-1.txt.score_p5e-07_sRNA_21_out.cluster
D-MH-CM-1.txt.summary.sRNA.v1.txt
phasis.set
```

Phasing analysis result files

3.1.4.5 Novel miRNAs Results

Mireap novel miRNA detect and expression analysis were performed in this step, and there have some intermediate files such as novel_hairpin.fa, novel_mature.fa, mireap-xxx.aln, mireap-xxx.gff, and mireap-xxx.log, the log file which contains the information of alignments result, the information of annotation and the information of parameters(as shown in the figure below)

```

mireap start at Tue Feb 26 02:23:05 2019

Parameters:
  minimal miRNA length: 18
  maximal miRNA length: 26
  minimal miRNA(reference) length: 20
  maximal miRNA(reference) length: 24
  uniqueness of miRNA: 20
  maximal energy: -18
  minimal space: 5
  maximal space: 35
  minimal mature pair: 14
  maximal mature bulge: 4
  maximal duplex asymmetry: 5
  flank sequence length: 10

Load short tag file: ./data/results/123456/query_sequence_for_mireap.fa
  small RNAs [18 - 26 nt]
  unique: 43253
  total: 820159

Load mapping file: ./data/results/123456/unclassified_reads_mapping_site.txt
  small RNAs mapped [18 - 26 nt]
  unique: 43253
  total: 820159

Load reference file: /smPlant/db/genome/ptc.fa
  total 1446 entries

miRNA discovery
  Chr10:+ 380 277(0) 14
  Chr10:- 371 234(0) 12
  Chr11:+ 701 398(0) 10
  Chr11:- 478 317(0) 13
  Chr12:+ 687 332(0) 7
  Chr12:- 617 260(0) 12
  Chr13:+ 583 290(0) 9
  Chr13:- 574 292(0) 10
  Chr14:+ 996 421(0) 21

```

mireap-xxx.log

The final result file which named novel_mirna_expression.list. All unclassified reads were considered for detecting candidate novel miRNA genes. Sequence of predicted putative miRNA and miRNA star along with the corresponding tag number, tag count and hairpin structure are provided. The file contains the information of miRNA names, strand, total tag numbers, total RPM, type, sequences, tag names, most abundant tags(shown below).

1	xxx-m48-3p	3	19024	542210.57	1	TCTTCCAACGCTCCCATACC	t00000071_x12767	12767	363877.33	TCT
2	xxx-m43-5p	5	4361	124294.59	1	AGGGATGTAGCTCAACT	t00000242_x3906_3906	111326.45		AGGGATGTAGC
3	xxx-m47-3p	3	2738	78036.82	1	TCTTGCTACTCTCCCATTC	t00000695_x1605_1605	45744.74		TCTTGCTACTCT
4	xxx-m20-5p	5	1703	40537.80	1	GGGATGTAGCTGAGTTGA	t00000712_x1571_1571	44775.69		GGGATGTAGC
5	xxx-m29-5p	5	1259	35883.26	1	ACTCCCCCTCAAGGGCTCCTGTT	t00006049_x262_262	7467.37		ACTCCCCCTCA
6	xxx-m14-5p	5	1188	33859.66	1	TGCTCAGGTTAAGGATTATGAAC	t00001316_x964_964	24745.35		TGCTCAGGTTA
7	xxx-m22-5p	5	802	22858.12	1	TAGTGTGTAGATATGTGAGGAC	t00007199_x219_219	6241.81		TAGTGTGTAGAT
8	xxx-m24-3p	3	724	20635.01	1	ATTTGGACTTCTAGAATCGAGAT	t00001935_x716_716	20407.00		ATTTGGACTTCT
9	xxx-m0-3p	3	309	11115.54	1	ATGGCATGATGATAGAACCA	t00018477_x74_74	2109.10		ATGGCATGATGAT
10	xxx-m7-5p	5	379	10802.03	1	TGTACTCAGCTTCTGTTGTGATA	t00004547_x347_347	9889.98		TGTACTCAGCTT
11	xxx-m35-3p	3	269	7666.88	1	ATTAAGAGAGATTGCCCGATCAGG	t00006313_x250_250	7125.35		ATTAAGAGAGATT
12	xxx-m40-5p	5	208	5928.29	1	ATTGGATGGAAACCTAACTTCACT	t00008139_x194_194	5529.27		ATTGGATGGAAAC
13	xxx-m13-3p	3	182	5187.25	1	ATGAATGGATTAACGAGAT	t00010089_x153_153	4360.71		ATGAATGGATTAAC
14	xxx-m18-5p	5	179	5101.75	1	AGGCTGTGATGATTTACCA	t00017569_x79_79	2251.61		AGGCTGTGATGAT
15	xxx-m44-3p	3	148	4218.21	1	CTGCATTGGCCTGGCCACCTCT	t00017453_x80_80	2280.11		CTGCATTGGCCTG
16	xxx-m41-5p	5	136	3876.19	1	ATTAGCCTTGATAACATTTGACCT	t00013503_x109_109	3106.65		ATTAGCCTTGATAA
17	xxx-m5-3p	3	116	3306.16	1	AACCAGATTGTCTGTTGAAGCTTT	t00024621_x51_51	51		AACCAGATTGTCT
18	xxx-m12-5p	5	83	2365.62	1	ATTAGTCTGGTACTGATAGCT	t00042488_x25_25	712.53		ATTAGTCTGGTACT
19	xxx-m4-3p	3	83	2365.62	1	AATATATAGTACTAGGACCT	t00050092_x20_20	570.03		AATATATAGTACT
20	xxx-m45-5p	5	80	2280.11	1	TCAGTTGGTTAGAGCACCCGTTT	t00018988_x71_71	2023.60		TCAGTTGGTTAGAG
21	xxx-m52-3p	3	75	2137.60	1	AAGAGAGATTGCCCGATCAGGAGG	t00024487_x51_51	51		AAGAGAGATTGCC
22	xxx-m1-3p	3	71	2023.60	1	AGGCTGTGCAGCAATAAGTGC	t00019031_x67_67	1909.59		AGGCTGTGCAGCA
23	xxx-m36-5p	5	64	1824.09	1	CCGAGCCCGGCTGGAGACAA	t00078893_x12_12	342.02		CCGAGCCCGGCTG
24	xxx-m8-5p	5	63	1795.59	1	GCAAAGATTGAAGCAATCT	t00027274_x45_45	1282.56		GCAAAGATTGAAGC
25	xxx-m11-5p	5	57	1624.58	1	TGTTCTTGAGGATTCATGTGCTC	t00022482_x57_57	57		TGTTCTTGAGGATT
26	xxx-m51-3p	3	52	1482.07	1	AAACCTGAACATTGCAAGAAA	t00028822_x41_41	1168.56		AAACCTGAACATT
27	xxx-m30-5p	5	50	1425.07	1	AAAGGATATGCTAGAAAGAAGGA	t00036827_x30_30	855.04		AAAGGATATGCTA
28	xxx-m34-3p	3	50	1425.07	1	TCCGGCACCGAAAGTATTCGC	t00030451_x39_39	1111.55		TCCGGCACCGAAAG
29	xxx-m39-5p	5	47	1339.57	1	GTCAGACTCAAAATCTTGGGTGC	t00050673_x20_20	570.03		GTCAGACTCAAAAT
30	xxx-m46-3p	3	43	1225.56	1	TGAATTGGCATGAACACCGATC	t00028249_x43_43	1225.56		TGAATTGGCATGAA
31	xxx-m49-3p	3	40	1140.06	1	TAAATACAAACAGGTTGGCTGACGT	t00029364_x40_40	1140.06		TAAATACAAACAG
32	xxx-m3-3p	3	39	1111.55	1	TGCAAAATGATCTGGGACAACT	t00034274_x33_33	940.55		TGCAAAATGATCT
33	xxx-m23-3p	3	37	1054.55	1	GAGATAAATTAATCAAGTGTGACT	t00071232_x13_13	379.52		GAGATAAATTAATC
34	xxx-m33-5p	5	29	826.54	1	TCTATGAGAGTGGAGTTGTTT	t00054536_x18_18	513.03		TCTATGAGAGTGGA
35	xxx-m16-5p	5	28	798.04	1	AACAAGCACTTATTCGGAAGT	t00047424_x22_22	627.03		AACAAGCACTTATT
36	xxx-m21-5p	5	26	741.04	1	TAGTGCACAAAATAAGATTTCTGAC	t00041510_x26_26	741.04		TAGTGCACAAAATA
37	xxx-m15-3p	3	25	712.53	1	AATTAACAATCTGATCAAGGGA	t00065684_x14_14	399.02		AATTAACAATCTGA
38	xxx-m27-3p	3	24	684.03	1	ATCGCCTGTGCTGAGCTACA	t00069605_x13_13	370.52		ATCGCCTGTGCTGA

novel_mirna_expression.list

3.1.4.6 Expression Results

This step we use megablast which is a tool of choice to identify a nucleotide sequence. And there have some intermediate files such as miresult.txt, miresult.txt.formated, mi.run.log, rfamesult.txt, rfamesult.txt.formated, rfam.run.log, trnaesult.txt, trnaesult.txt.formated, trna.run.log, and the final result files such as snRNA_stat_table.result, rRNA_stat_table.result, snoRNA_stat_table.result, miRNA_stat_table.result, tRNA_stat_table.result and some summary stat files such as rfam_sum_stat, miRNA_sum_stat and tRNA_sum_stat. The stat table results files contains the information of raw number(first column), tags names(second column), total expression(third column), total expression ratio(fourth column), rfam sequences(fifth column), uniq expression(sixth column) and other information such as most expression ratio and raw sequences.

1	T_RF00002;5_8S_rRNA;AARH01012340.1/5361-5516	122516	348048.91	GACTATCGGCAACAGATATCTCGGCTCTCGCATTGATG
2	GTTGCGCCCGAGGACTCTTGGTCGAGGGCACATTTGCTGGGTGTC	t00000241_x3918_3918	11130.43	ACCATCGAGTCTTTGAACGCCAA
2	T_RF00002;5_8S_rRNA;AARH01011946.1/6498-6359	88975	252764.14	GACTCTCGGCAACGGATATCTAGCTCTTGATCGATG
3	CTTGTTGAGGGCACTTCTGCTTGGGTGTC	t00000170_x5167_5167	14678.64	GTGAATTGCAGAAATCCCGTGAAC
3	T_RF00001;5S_rRNA;AARH01009280.1/1989-1871	49406	140354.76	GAGTGAATCATACCATCATTAAATGCACCGGATCCCAT
4	GTTGCACCCC_t00000145_x6173_6173	17536.53		CACCGGATCCCATCAGAAGTCC
4	T_RF00001;5S_rRNA;AARH01017907.1/5093-4975	40609	115363.86	TGGTGTGATCATACGAGCACTAATGCACCATATCCCAT
5	GTTGCACTCC_t00000106_x7839_7839	22269.38		CAGAAGCTCCGAAGTTAAACGT
5	T_RF00001;5S_rRNA;AARH01009280.1/3741-3629	20234	57481.65	GTGTCCGGTTCGCGAAGGCCAGGATCAAAATCAAAC
6	CCCC_t00000490_x2149_2149	6104.97		GCGAGAGTAGTACTAGGAT
6	Rf00001;5S_rRNA;AARH01009280.1/1989-1871	17056	48453.44	GAGUGCAUACUACCAUACUUAUAGCCCGGAUCCCAU
7	GUUGCACCCC_t00000050_x17056	17056	48453.44	CACCGGATCCCATCAGAAGTCCGA
7	T_RF00001;5S_rRNA;AC217035.1/58989-59087	8460	24033.54	GGGTGCAATCATATCAGCACTAAGTCTGTAGTTAAACA
1068	x1120_1120	3181.75		TACTAGGATGGGTGACCTCT

rRNA_stat_table.result

```

1 T_RF01847;Plant_U3;AARH0101655_1/42823-42810 17643 211049.67 ACCGACTTACTGAACAGATCTGCTCTATAGGATCGTACCTGTGTATCTTTAACTACGAGGAGACAGGAACTAACTGCTGGCTGATGATCATACCGCTGTGATCAT
2 T_RF01847;Plant_U3;AC209197_1/68384-68891 5162 61746.41 ACCGACTTACTGAACAGGATCTGCTCTATAGGATCGGACCTGTGATCTTTAATCACTAAGGAGACAGGAACTAACTGCTGGTGGAGGATCATGACCTGTGATCAT
3 T_RF00358;snz101;AARH0109456_1/205048-205138 2908 34784.69 TTGGGGGCTATGATGATCAATGTTTTAATGGGTAATTTGGCTGAAATTTTGTGTGCTTATCACCTGGAGACCTGATGCCCTTT t00083519_x44
4 T_RF00267;snrR64;AARH0109291_1/42426-42331 2660 34210.53 CTTCTGCAGTGAAGTCATGGATAGATCCACATGACACTTCAGAGATTTTAAATCGATAGAGAGACTTCCATACTTCAGGACAGAT t00080
5 T_RF01421;snrR114;AARH01023813_1/1261-1176 2532 38287.00 GAGCAATGATGAAGCGCCCTGGGATGGTGGGTCATCCGCAAAATGATGATACAGGAGACACTGGCTGTGTGGGCTG t00083789_x10_410
6 T_RF00149;snz103;AC208376_1/11244-11153 2463 29461.72 ATGGAAAGTATGATTTTTCCCAATGATCGAAGCTGAATAACTGAGCGACTTCTCTCGGAGAACCAAGCCAGATCTGTGATCAT t00084144_x37
7 T_RF00358;snz101;AARH0109456_1/205216-205304 2118 25334.93 ATCGGGCAGATGATGATCTGTGTTTTAATGGGTAATTTGATCTGAAITTCGTTGATGCTTCAACACTTGAACACTGATGCCCTTC t00084888_x32
8 T_RF01847;Plant_U3;AARH0106491_1/88074-88661 2695 25059.81 ACCGACTTACTGAACAGGATCTGCTCTATAGGATCGTACCTGTGATCTTTAATCACTAAGGAGACAGGAACTAACTGCTGGTGGAGAACCTAGCCGCTGATCAT
9 T_RF00134;snz106;AARH0109507_1/19451-19478 2654 24069.38 TGCATCTGTGATGATTAATCTCCGAGTAACTGATTTAAGAGTATAGCATTTTAACTACCATCTGCTGATGATGCA t00087957_x15
10 T_RF00295;snrR41;AARH0106692_1/284595-284677 2638 24377.99 TATGGCTCAGTATGATGAAGCACTACCTATTTTATGATCTTCTGTGATTAATAAACCAACTGATGATCTTGGAGACTCA t00085889_x268_268
11 T_RF00816;SMOR14;AARH0101274_1/10536-10645 2037 24366.03 TCAAAATGGATGCAAAAAGCGTTTTCTCCATGGTATCTAAAGAGGAAGTTTACATGACACTAACTCTATGTGATCTTCAATCTCTTGGATCTTGAAGTTG
12 T_RF00309;snosR60_215;AARH0108781_1/12889-12316 1942 23229.67 CTGGTCAATATGATATACAGCATGTTCACTGATTAATCTGTGATTAACAAACCTTTTATCTTCTGCTATCTGATGACTCG t00088
13 T_RF00955;SMOR96;AARH0109562_1/239727-239635 1863 22284.69 GGGCCATGATGATATATCAACAAACCTTTTGGGACTGAATGCTGGAATGATGCTTCTGACAGACTTCTGCAATCTGAGGCT t00082888_x51
14 T_RF00358;snz152;AARH01088776_1/2554-2445 1501 17954.55 GTTGCTTGAAGAGTTAAGCACTGAAATGCTGTAAGGCGAGTGGAGGACTTGAATCTTCCGCTGAGCTTTTAAATAACGACTTCTCTTCCACTGACCT
15 T_RF00337;snz112;AARH01001267_1/60959-61168 1172 14019.14 ATCCAAAGATAGGAAGTGGCTATGACACTTAAATCAGAGCGTCACTCATGATCCCAAGCTTTCCATCCACCACTCAGCCAAAGAGAGGCGGTTCTCTCTCC
16 T_RF01421;snrR114;AARH0106495_1/201033-501122 1158 13651.67 AAGCTGTGAGGAAATCTGTGCTCAATGTGATGATAGGTTATCACTACCTCCGATCCATCCGACCAAGAGTGCATCTGAGCT t00084650_x32

```

snoRNA_stat_table.result

```

1 T_RF00004;U2;AARH01005049_1/80644-80449 25900 318894.83 ATACCTTTCTCGGCCCTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATCAGTTTAAATATGATACGTTGGCCATAGG
TCCACCGCAGTACTGCTACTGGGATCTCTACCGGCTGCCCATGCGTTGCACTACTGCAAGGGCCAGGGCCACCTCC t000800149_x6054_6054 74540.13 CCAGTGAAGGGCCTGGCCACCTC
2 T_RF00004;U2;AARH01022937_1/317-121 22896 272857.92 ATACCTTTCTCGGCCCTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATCAGTTTAAATATGATACGTTGGCCATAGG
CCATCACAAGCAGTTTCTGCTGGGTTCTCGAGGCTGCGCTGTGCTGCACTACTGCAATGGCTGGCCGGCCCTCT t000800175_x4977_4977 61279.52 CTACTGAAGGGCCTGGCCGCTC
3 T_RF00004;U2;AARH01001231_1/52284-52089 14476 178162.48 ATACCTTTCTCGGCCCTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATCAGTTTAAATATGATACGTTGGCCATAGG
CTGTTCAAAGTACTGCTACTGAGGCTCTCGGGCTGCGCTATGCTTTGCACTGTTGCATGGCTGGCCAGCCCTC t000800118_x7163_7163 88194.74 CTGCAAGGGCCTGGCCACCTC
4 T_RF00007;U1;AARH01003540_1/648176-648022 8984 110615.87 TTGCCTTAAACTAATGAGTAAAGAAAACAAGCCGAGAGGTGGGAGCCCTCGCCGGCCGACACAGTTTGTG
GTTGACGCTGCTTACTTTCAGAACTCCGCTTCCGGCCGCTCCCTA t000800613_x1765_1765 21731.64 GCAGAACTCTTGGCCGGCCCTCC
5 T_RF00004;U2;AARH01005634_1/120594-120789 1883 23184.52 ATACCTTTCTCGGCCCTTTGGCTAAGATCAAGTGTAGTATCTGTTCTTATCAGTTTAAATATGATACGTTGG
GGTGAAGGGCCTGATACAGCAGCTGCTGCTGGCTTCTCGAGGCTGCGCTGCTCCACTGCTGCAAGTGGCTGGCCAGCCCTC t000807627_x207_207 2548.70 CTGCTGAAGGGCCTGGCCAGCCCTC
6 T_RF00003;U1;AARH01008492_1/271690-271530 1298 15981.68 ATACTTACTGGACGGGATCAATGGGTGATCTCGTAAGGCCCATGGCTAGGCTGTGACCTTCCATTGCACAC
AGCCTACGTCAAAATTTGGCAGAGGGGGCTGCGCGAGTCCGGCCCTAT t00020413_x65_65 800.32 CCTGGAGGGATCAATGGGTGATCA
7 T_RF00003;U1;AARH01016232_1/4482-4639 839 10330.22 ACAATATGGACGGGATCAATGGGTGATCAAGAGACCTATGGTCTAGGTCAGTGACCTCCATTCGACCCCGGAGGGGTG
AAAATTTGTGTTGAGGGGGCTGCCAAGTCCGGCCCTG t00011942_x126_126 1551.38 CCATGGTCTAGGTCAGTGACCTCCA

```

snoRNA_stat_table.result

3.1.4.7 24-nt siRNA Results

Firstly we use intermediate file nonMirRfam.fa and the Perl script ‘get24ntSeq.pl’ to generate 24-nt filtered sequences, and then mapping the filtered 24-nt siRNAs to the genome sequences to call 24-nt siRNA clusters by using the tools bowtie, samtools, ShortStack, bedtools and Perl script ‘parse_result_24nt.pl’ to get 24-nt sorted clusters, and next we take loci expression analysis and 24-nt siRNA statistics on repeats/transposons by using the software bedtools and our own Perl scripts such as ‘getExp.pl’, ‘getReadExp.pl’, and ‘getRepeatMapping.pl’, and finally we did the basic mapping statistics so that we can get information from the result files such as 24nt.mapped.stat.txt, 24nt.transposons.tags.stat.txt and 24nt.loci.exp.txt.

There are many 24nt named files in the result file, The result files 24nt.cluster.length.stat.txt, 24nt.clusters.exp.read.txt and 24nt.clusters.exp.txt are the results of the step which mapping the filtered 24-nt siRNAs to the genome sequences to call 24-nt siRNA clusters and loci expression. The file 24nt.clusters.exp.read.txt contains location information (first and second column), rpm value (third column), RPKM value (fourth column), reads number (fifth column), reads length (sixth column) and the last column which determines if it is a repeat; The file 24nt.clusters.exp.read.txt contains tags name and sequence (first and second column), reads numbers (third column), TPM value which means reads number multiply

1,000,000 before divide mapping total number (fourth column) and the cluster names in the last column. The two files as shown in the figure below:

Cluster_23528	Chr11:4581801-4582035	9.97	42.43	175	235	TRUE
Cluster_25952	Chr12:10727133-10727369	9.97	42.08	175	237	TRUE
Cluster_11142	Chr4:15615787-15616047	9.97	38.21	175	261	TRUE
Cluster_4875	Chr1:46156389-46159063	9.97	3.73	175	2675	TRUE
Cluster_14849	Chr6:3078762-3079031	9.97	36.93	175	270	TRUE
Cluster_5829	Chr2:5653202-5653508	9.97	32.48	175	307	TRUE
Cluster_11488	Chr4:18227691-18228005	9.97	31.66	175	315	FALSE
Cluster_37004	Chr19:10775820-10776164	9.97	28.90	175	345	TRUE
Cluster_19134	Chr8:8861092-8861450	9.97	27.78	175	359	TRUE
Cluster_19304	Chr8:11321414-11321790	9.97	26.45	175	377	TRUE
Cluster_39946	scaffold_131:4698-7206	99.72	39.75	1750	2509	FALSE
Cluster_22852	Chr10:20793560-20793991	9.97	23.08	175	432	TRUE
Cluster_10511	Chr4:9005103-9005534	9.97	23.08	175	432	TRUE
Cluster_12957	Chr5:9248881-9249360	9.97	20.78	175	480	TRUE

24nt.clusters.exp.txt

t00008951_x175	TGGAGAAGCGTCCTCAGCGGCGGA	175	9.97	NUM
t00008969_x175	ATCGCACCCATATGGACATTGTAG	175	9.97	Cluster_27174
t00000628_x1746	TAGAACCCACAAAGGGTGTGGTC	1746	99.49	NUM
t00009025_x174	AAGTTGGGGCTCGAAGACGATCA	174	9.92	NUM
t00009002_x174	ACTGGTACGGACAAGGGGAATCCG	174	9.92	NUM
t00009012_x174	CCAGAATTTCAACAGATCACCTAC	174	9.92	Cluster_26276
t00009086_x173	TGGTCTAGTGGTAGAATAGTACCC	173	9.86	NUM
t00009063_x173	AGCTTGAACGTCTCTCTGGCAAC	173	9.86	Cluster_34327
t00009057_x173	CGTTCAAGATCATTGCTCATTGGA	173	9.86	Cluster_34326
t00009109_x172	TGGTCCTAAGAATTCTGAACAACA	172	9.80	NUM

24nt.clusters.exp.read.txt

The result file 24nt.mapped.stat.txt, 24nt.unmapped.stat.txt and 24nt.chr.stat.txt contains information of mapping statistics with the total stats reads and unique stats:

Chr10	395844	152370
Chr1	1370942	483100
Chr11	438963	176414
Chr12	420589	137211
Chr13	299553	139876
Chr14	2863809	246722
Chr15	268759	118797
Chr16	379450	126393
Chr17	428474	162282
Chr18	256796	135046
Chr19	295356	151680
Chr2	405525	184278
Chr3	329330	165182
Chr4	422702	214359
Chr5	5706535	408035
Chr6	450382	212079
Chr7	269083	127688
Chr8	2250497	215537
Chr9	167512	76096

24nt.chr.stat.txt

22637163

4232059

24nt.mapped.stat.txt

748625 439398

24nt.unmapped.stat.txt

3.2 Launch Analysis Use Website

3.2.1 Prepare the inputs

Data Preprocessing: We can do the following steps which below the figure in the red border to preprocess the data.

Raw fastq sequences

```

1 @SRR392136.1 HwUST-EAS455_12_FC30L1KAAXX:5:1:1159:298 length=36
2 GAAGCGATAAAGTGTACACCTGAGGCACCATCAA
3 +SRR392136.1 HwUST-EAS455_12_FC30L1KAAXX:5:1:1159:298 length=36
4 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
5 @SRR392136.2 HwUST-EAS455_12_FC30L1KAAXX:5:1:263:1626 length=36
6 GTTTATTGTCATCCCAAAAATGACTGTAGGCACCA
7 +SRR392136.2 HwUST-EAS455_12_FC30L1KAAXX:5:1:263:1626 length=36
8 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
9 @SRR392136.3 HwUST-EAS455_12_FC30L1KAAXX:5:1:427:888 length=36
10 GGATAAGTGGGAGCTTCGGCGCAAGTGTAGGCA
11 +SRR392136.3 HwUST-EAS455_12_FC30L1KAAXX:5:1:427:888 length=36
12 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
13 @SRR392136.4 HwUST-EAS455_12_FC30L1KAAXX:5:1:1068:2017 length=36
14 GAAAAGGACTTTGAAAAGAGAGCTCAGTGGGCATC
15 +SRR392136.4 HwUST-EAS455_12_FC30L1KAAXX:5:1:1068:2017 length=36
16 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
17 @SRR392136.5 HwUST-EAS455_12_FC30L1KAAXX:5:1:429:811 length=36
18 GATAACATCATAGGATTTCCGATCCTGTAGGCACC
19 +SRR392136.5 HwUST-EAS455_12_FC30L1KAAXX:5:1:429:811 length=36
20 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

```

The reads/tags file should not contain any adapters. Users can use **cutadapt** (<http://cutadapt.readthedocs.io/en/stable/>) or **trim_galore** (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to remove the adapters.

Clean fastq sequences

```

1 @SRR392136.1 HwUST-EAS455_12_FC30L1KAAXX:5:1:1159:298 length=36
2 GAAGCGATAAAGTGTACA
3 +SRR392136.1 HwUST-EAS455_12_FC30L1KAAXX:5:1:1159:298 length=36
4 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!
5 @SRR392136.2 HwUST-EAS455_12_FC30L1KAAXX:5:1:263:1626 length=36
6 GTTTATTGTCATCCCAAAAATGA
7 +SRR392136.2 HwUST-EAS455_12_FC30L1KAAXX:5:1:263:1626 length=36
8 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!
9 @SRR392136.3 HwUST-EAS455_12_FC30L1KAAXX:5:1:427:888 length=36
10 GGATAAGTGGGAGCTTCGGCGCAAGTG
11 +SRR392136.3 HwUST-EAS455_12_FC30L1KAAXX:5:1:427:888 length=36
12 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!
13 @SRR392136.4 HwUST-EAS455_12_FC30L1KAAXX:5:1:1068:2017 length=36
14 GAAAAGGACTTTGAAAAGAGAGTCA
15 +SRR392136.4 HwUST-EAS455_12_FC30L1KAAXX:5:1:1068:2017 length=36
16 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!
17 @SRR392136.5 HwUST-EAS455_12_FC30L1KAAXX:5:1:429:811 length=36
18 GATAACATCATAGGATTTCCGATCCT
19 +SRR392136.5 HwUST-EAS455_12_FC30L1KAAXX:5:1:429:811 length=36
20 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!

```

Users can use "**fq2collapsedFa.pl**" (<http://bioinformatics.caf.ac.cn/tRFtools/fq2collapsedFa.pl>) to convert their data in FASTQ to **collapsed FASTA format**.

Besides, the above clean step and format conversion step could also be conduct at one stream by tool provided by mirtools (http://centre.bioinformatics.zj.cn/mr2_dev/start_filter.php)

Collapsed fasta sequences

```

1 >seq1_x132449
2 TGAAGTGCAGCATGATCTA
3 >seq2_x63297
4 TCCCAATGTAGCAAGCA
5 >seq3_x40868
6 TGACAGAAGAGAGTGAGCAC
7 >seq4_x27666
8 AGAATCTTGATGATGCTGCAT
9 >seq5_x19740
10 TTGACAGAAGATAGAGAGCAC
11 >seq6_x14945
12 AGTTACTAATTCATGATGGCAT
13 >seq7_x12741
14 ATGATATTTGTAGTAATGGCG
15 >seq8_x10596
16 TTGACAGAAGAGAGTGAGCAC
17 >seq9_x9158
18 TTGAAAGTGACTACATCGGGG
19 >seq10_x7979
20 TCAATGCATTGAAAAGTGACTA

```

Reads files in **collapsed FASTA format** are required as inputs, that combines the identical reads into unique tags (e.g. seq1_x290, where "seq1" = unique tag ID, "290" = reads count. The ID and read count number is separated by "_x").

To further minimize the size of uploading files, users can compress the fasta file into **.zip** file or **.gz** file.

Inputs data of collapsed FASTA

Upload samples and select the reference genome

Species * **Select the species for your study**
Please choose a species for the analysis

Input type * Collapsed FASTA FASTQ GSM ID / SRR ID Link **Select the collapse FASTA**

Collapsed Tags * **Upload your data here**
Please upload sequences files [One click test \(Hsa\)](#) [One click test \(Mmu\)](#) [One click test \(Populus\)](#)

Email **You can input your email to receive the job notification**
When the job is complete, you can get an notification (optional field)

Inputs data of FASTQ

Input type * Collapsed FASTA FASTQ GSM ID / SRR ID Link **Select FASTQ**

Adapter information: Raw fastq Adapter trimmed **Raw FASTQ or clean FASTQ**

Raw reads *
You can upload FASTQ file for sample (.fastq, .fq, .zip, .gz)
[hsa \(raw\)](#) [ath \(raw\)](#) **Upload your data here**

Adapter Trimming * smRNA-seq Truseq Illumina v1.0 small RNA Illumina v1.5 small RNA ENCODE microRNA-seq ENCODE short total RNA sequencing single-cell small RNA-seq (with UMIs) Customized protocol **The adaptor panel will be showed if "raw fastq" is selected.**

Inputs of GSM ID/SRR ID

Input type * Collapsed FASTA FASTQ GSM ID / SRR ID Link **Select to input GSM ID/SRR ID**

GSM ID *
GSM ID for sRNA-seq GEO sample (Example) **Input GSM ID from GEO or SRR ID from SRA**

Inputs of data link

Input type * Collapsed FASTA FASTQ GSM ID / SRR ID Link **Select to input link**

Link *
Provide link to download your collapsed fasta data **Input link to the collapsed FASTA data**

Adapter parameters

Adapter Trimming *

- smRNA-seq
 - Truseq
 - Illumina v1.0 small RNA
 - Illumina v1.5 small RNA
- ENCODE microRNA-seq
- ENCODE short total RNA sequencing
 - ENCODE3
 - A_Tailing_No_Barcode
 - A_Tailing_N3
 - A_Tailing_N4
- single-cell small RNA-seq (with UMIs)
- Customized protocol
 - Guess the adapter sequence (**Not recommended!**)

Select the adapter protocol or customize your adapter

Please **choose** the 3' adapter or input the adapter (**mandatory**)

Please **choose** the 5' adapter (optional)

1	v	6	v
---	---	---	---

Max. mismatches Min. adapter overlapping Length

0		0	
---	--	---	--

Remove 5' nucleotides Remove 3' nucleotides

3.2.2 Analysis how-to

Parameters for single case study:

Common parameters

Length of interval (bp) *	<input type="text" value="18"/>	←	Set the length range of tags used to perform the analysis
Shortest tag Length	<input type="text" value="45"/>		
Allowed mismatch in tags mapping *	<input type="text" value="1"/>	←	Maximum mismatch in tags mapping
Maximum of tag mapping times *	<input type="text" value="20"/>	←	Maximum times of tag multiple mapping

Default mapping time threshold for multiple mapping

miRNA analysis parameters

MEGABLAST Parameters	<input type="text" value="0.01"/>	<input type="text" value="5"/>	<input type="text" value="5"/>	←	Parameters of megablast used to align read to miRNA	
E value		B value	V value			
Novel miRNA prediction tool	<input checked="" type="radio"/> miRDeep <input type="radio"/> miReap				←	Choose from miReap and miRDeep to predict novel miRNA
miReap L value *	<input type="text" value="100"/>					
Maximum of tag mapping times *	<input type="text" value="10"/>				←	Maximum times of tag multiple mapping for miRNA

Flank sequence length of the query, Range 0 - 200
Maximum of mapping times for for miRNA

tRFs analysis parameters

Tag abundance cut-off *	<input type="text" value="20"/>	←	Set tag minimum abundance to call high confident tRFs
P-value cut-off *	<input type="text" value="0.01"/>	←	Set the p-value when call tRFs
Maximum of tag mapping times *	<input type="text" value="30"/>	←	Maximum times of tag multiple mapping for tRNA and tRF

Min. tag abundance (Read Rer Million)
P-value inferred based on Binomial statistics method to distinguish random fragments
m: Maximum of mapping times for tRNA&tRF

piRNA/24nt siRNA analysis parameters

Minimum depth *	<input type="text" value="10"/>		
Cluster padding distance *	<input type="text" value="150"/>	←	Parameter to call piRNA/24nt siRNAs producing loci
Maximum of tag mapping times *	<input type="text" value="50"/>	←	Maximum times of tag multiple mapping for piRNA/24nt-siRNA

Minimum read depth of 24-nt siRNAs to call clusters.
The neighbor clusters with distance smaller than this value will be merged. With step: 50
Maximum of mapping times for piRNA/24nt siRNA tags

PhasiRNA analysis parameters

Phased length *	<input type="text" value="21,24"/>		
Minimum depth *	<input type="text" value="3"/>	←	PhasiRNA and loci calling parameters

Periodic length of phased siRNAs (bp)
Minimum read depth for p-value computations.

Parameters for batch case study

Upload samples and select the reference genome

Species * Select the species
Please choose a species for the analysis

Input type * Collapsed FASTA FASTQ GSM IDs / SRR IDs LINKS Choose the type of data uploading

Sample1 * Click "+" to upload more samples
Upload your data here

Sample names * Sample name can be set here
separate multiple names by comma ,

Email When the job is complete, you can get a notification (optional field)

Parameters for group case study

Group 1 (control)

Replicate 1 * Example

Replicate 2 *

Group 2 (Case)

Replicate 1 * Example

Replicate 2 * Example

Input the single case "Job ID". The samples in each group could be added or reduced by click the right "+" and "-" buttons

Choose the object sncRNA type

sncRNA type * Choose the sncRNA type to perform differential expression analysis

Differential expression analysis method

Methods DESeq DESeq2 edgeR Student's t-test Wilcoxon signed-rank test Choose the method to perform differential expression analysis

Parameters for Statistics

P value * P-value cut-off for significant differential expression between two groups
Statistical significant P-value

F value * Expression change fold cut-off for significant differential expression between two groups
Fold change of normalized expression

Parameters for sncRNA targets gene analysis

Upload small RNA sequences to detect small RNA target genes

Species * ← **Choose your species from the supported species**
Please choose a species for the analysis.

small RNAs *
Please input or upload sequences in **fasta** format.
[Example1](#) [Example2](#) [Clear](#)
 Select file

Email
When the job is complete, you can get an notification (optional field)

Parameters for small RNA targets detection tools

Prediction tool miRanda RNAhybrid miRanda&RNAhybrid ← **Choose the tools for sncRNA target genes analysis. Tapirhybrid and Targetfinder for plants species. miRanda and RNAhybrid for other species**

MFE *
Minimal free energy cutoff ← **Set the parameters for according tools**

Score *
miRanda score cutoff

Parameters for sncRNA modification analysis

Upload samples and select the reference genome

Species * ← **Select the species**
Please choose a species for the analysis

Input type * Collapsed FASTA ← **Only collapse fasta is supported**

Upload sequences/GSE sample ID

INPUT *
You can upload raw reads file (.fastq, .fq, .zip, .gz) ← **Upload INPUT and chemical treated small RNA-seq data here**

Chemical treated *
You can upload raw reads file (.fastq, .fq, .zip, .gz)

Parameters for sncRNA modifications identification

Allowed mismatch in tags mapping * ← **Mismatch allowed in tags mapping**
Mismatch 0-3

Maximum of multiple-mapping * ← **Maximum times of tag multiple mapping**
multiple-mapping times

Minimum of cleavage score * ← **Minimum of cleavage score for both INPUT and treated samples**
Minimum of cleavage score

cleavage ratio threshold * ← **Threshold for cleavage ratio between INPUT and treated samples**
cleavage ratio threshold

P-value threshold * ← **Set the p-value when call cleavage**
P-value

Email
When the job is complete, you can get an notification (optional field)

3.2.3 Retrieve the results

Query the job status and retrieve the results

1 File uploaded 2 Tag mapping 3 sncRNA calling 4 Job complete

Job status progress bar. Four steps are included: File uploading, Tag mapping, sncRNA calling and Job complete. The finished steps are showed in blue circles

Dear user:
We have received your data. Your job ID number is assigned as:
1546140391 ← **Job ID which is used to retrieve the result**
The number of jobs ahead of yours is:
9 ← **Number of Jobs in the queue**
You can query the job status and get the results in [results page](#) with this job ID.

Query the job status

Retrieve the results by the job ID.

- Single case examples: 1234581 (*Homo sapiens*); 1234561 (*Arabidopsis thaliana*)
- Group case examples: 1473666115 (*Homo sapiens*); 1473687035 (*Arabidopsis thaliana*)
- Target analysis examples: 1474244807

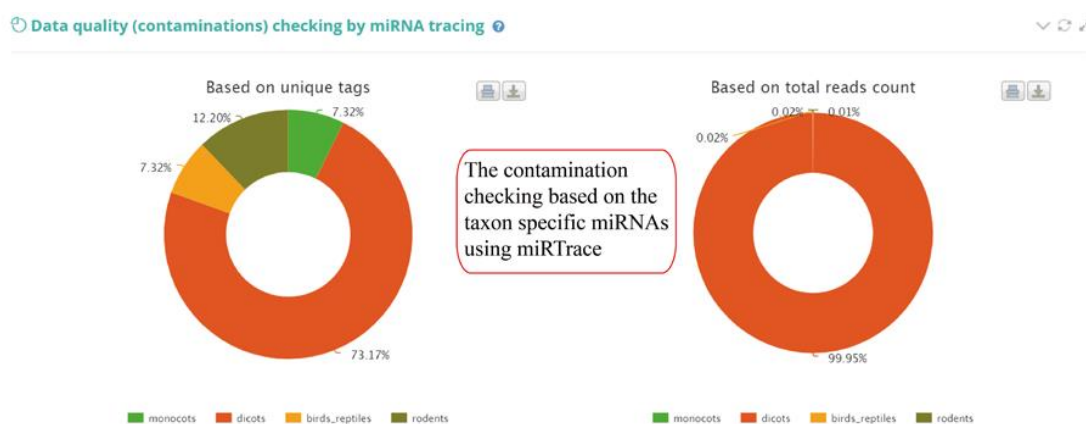
JobID

Type a jobID and check its status.

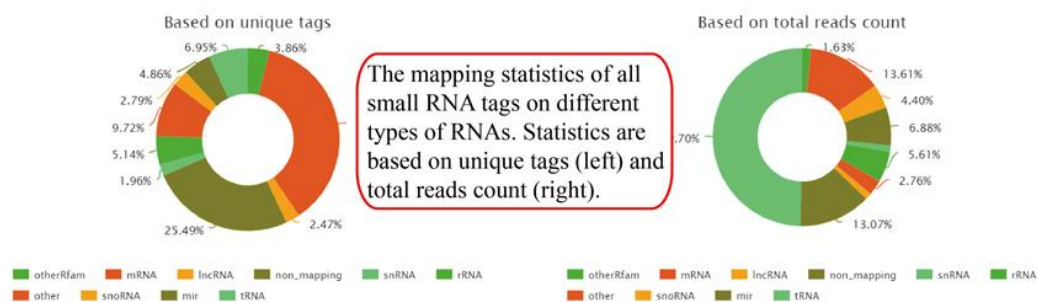
Query the job status and retrieve the results by inputting the Job ID that should be a ten digits number

3.2.4 Explain the results

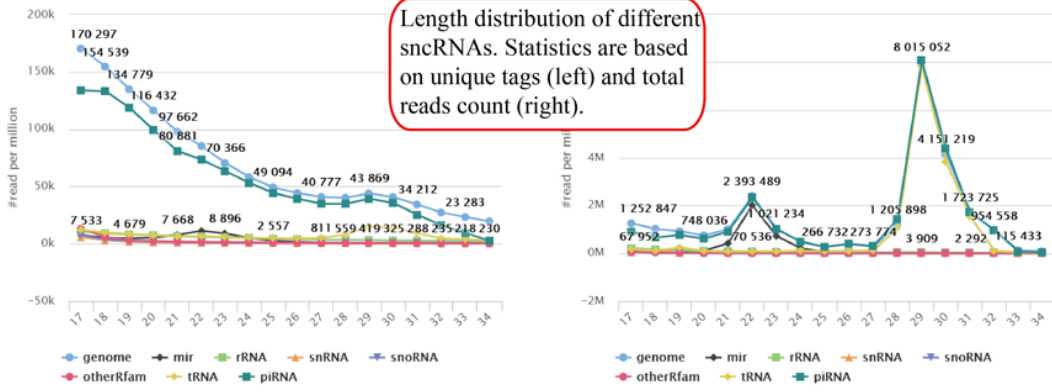
Sequence contamination evaluation by miRTrace



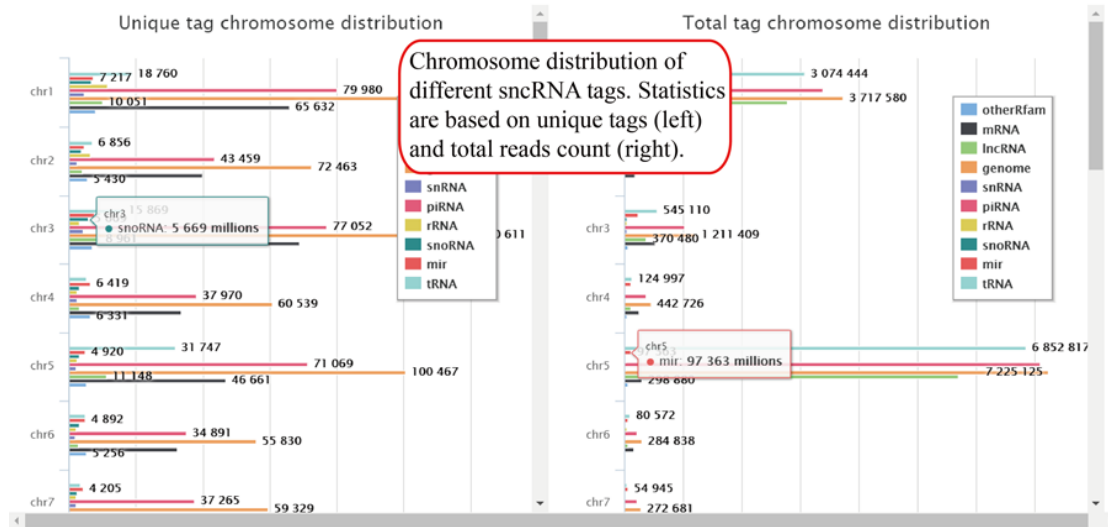
Tags mapping statistics on different types of RNAs, including miRNA, tRNA, rRNA, snRNA, snoRNA, other RFAM ncRNA, mRNA, lncRNA, other mapping



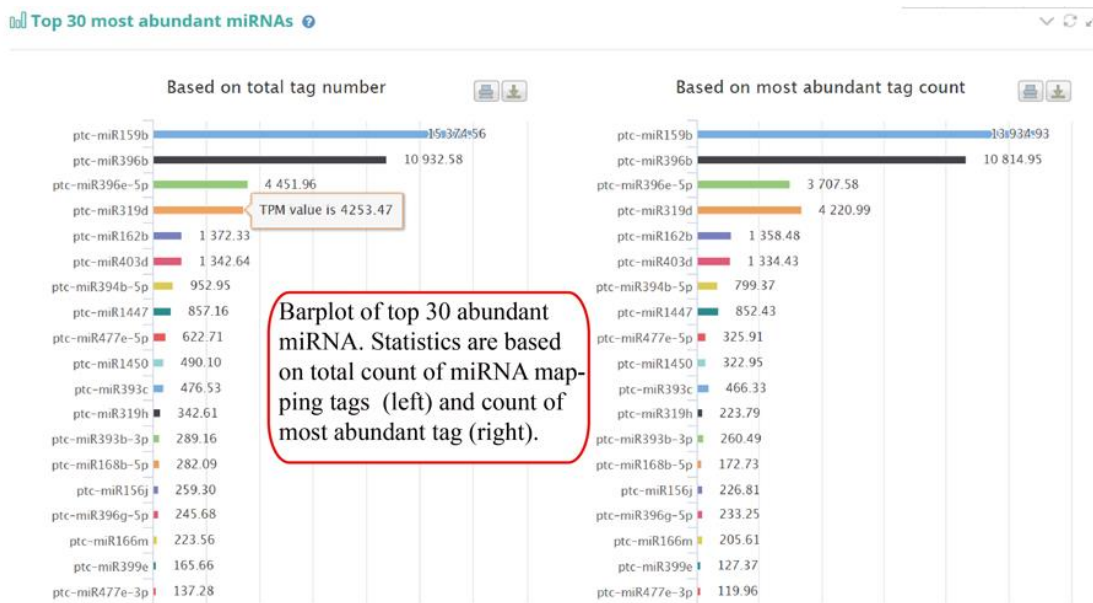
Tag length distribution of different sncRNAs. Statistics are based on unique tags and total tag counts



Tag length distribution of different sncRNAs. Statistics are based on unique tags and total tag counts



Top 30 most abundant miRNAs



miRNA expression list

miRNA expression profile

Download all tags expression list from this link

Download all records

Copy CSV Excel PDF Print

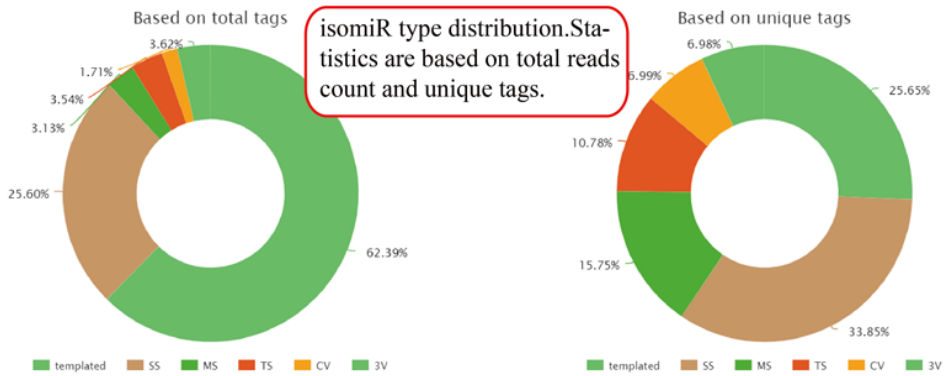
Search:

miRNA	Total tag raw number	Total TPM	Tag name	Most abundant tag	Most abundant TPM	Most abundant tag seq	Other tags	miRNA sequence	Genomic coordinate
ptc-miR1444a	60	3.42	t00044690_x23	23	1.31				
ptc-miR1446e	97	5.53	t00020162_x66	66	3.76				
ptc-miR1447	15041	857.16	t00000063_x14958	14958	852.43				
ptc-miR1448	1	0.06	t02895574_x1	1	0.06				
ptc-miR1450	8600				322.95				
ptc-miR156f	971				51.86				
ptc-miR156j	4550	259.3	t00000234_x3980	3980	226.81				
ptc-miR156l	3	0.17	t00558214_x2	2	0.11				
ptc-miR159b	269786	15374.56	t00000002_x244524	244524	13934.93				
ptc-miR159c	289	16.47	t00008252_x192	192	10.94				

Click the ID to show the miRNA detail in miRBase

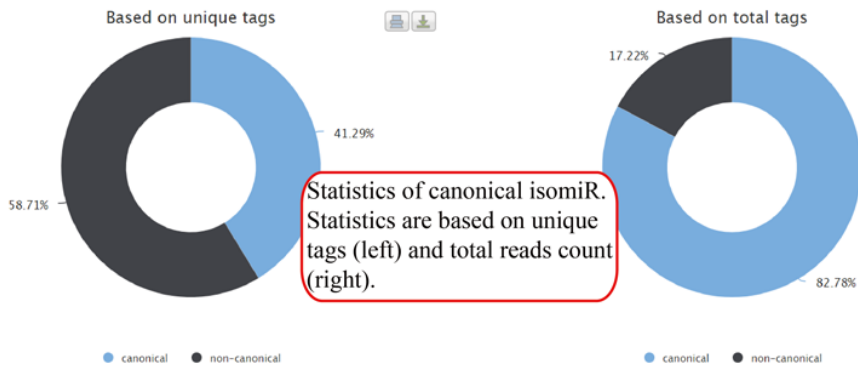
Click "Add" icon to show the detail

Tag distribution of isomiR, including templated isomiR and non-templated isomiR (SS, MS, TS, CV and 3V)



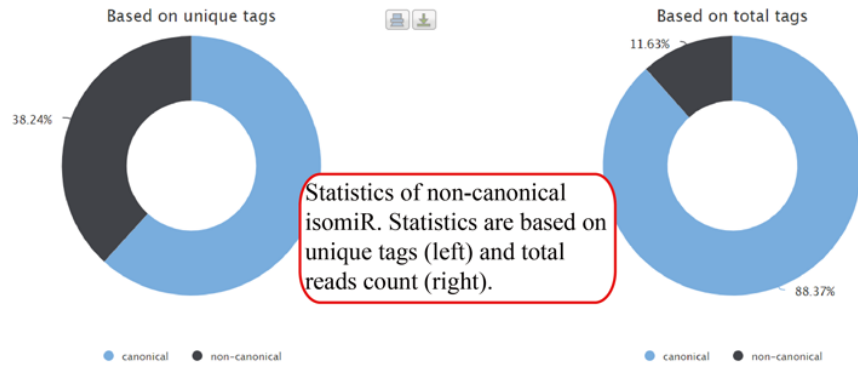
Statistics of canonical and non-canonical isomiR

Templated



Statistics of canonical isomiR. Statistics are based on unique tags (left) and total reads count (right).

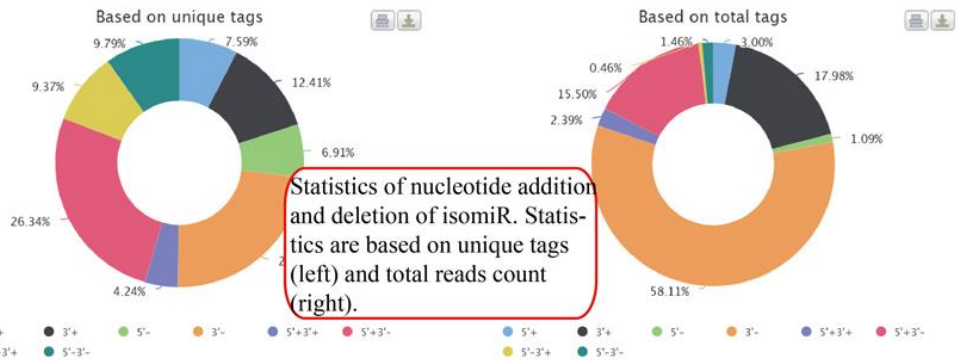
Non-templated



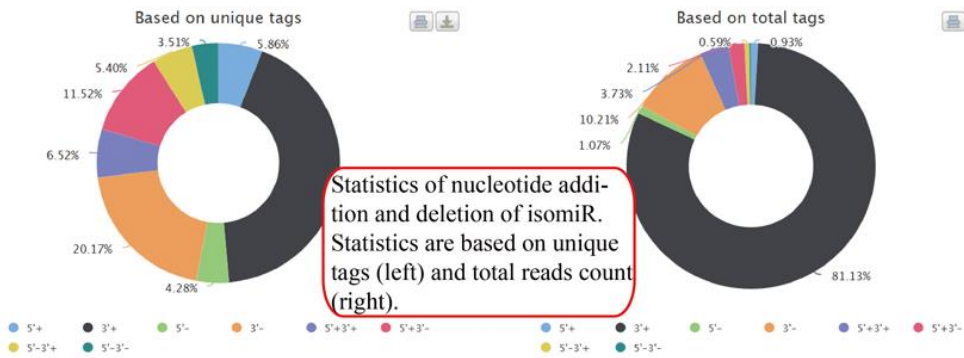
Statistics of non-canonical isomiR. Statistics are based on unique tags (left) and total reads count (right).

Statistics of isomiR nucleotides addition and deletion on both ends

Templated

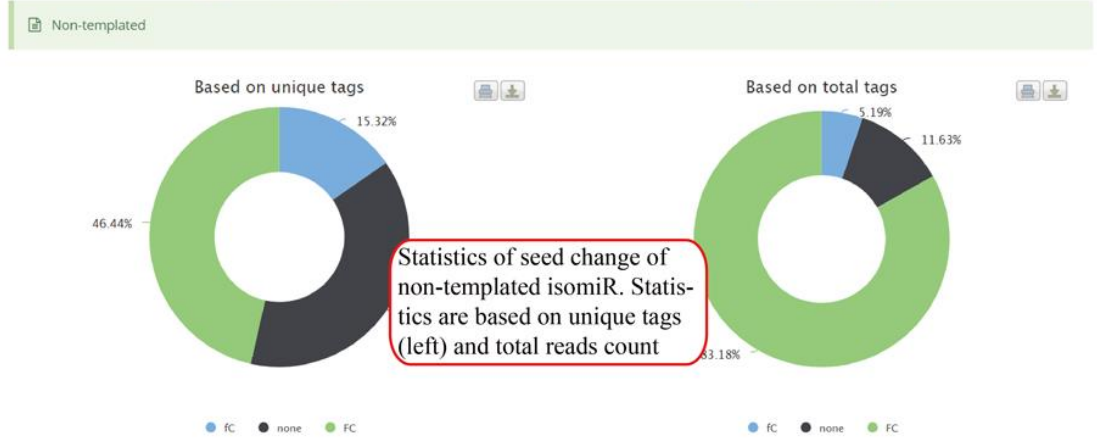
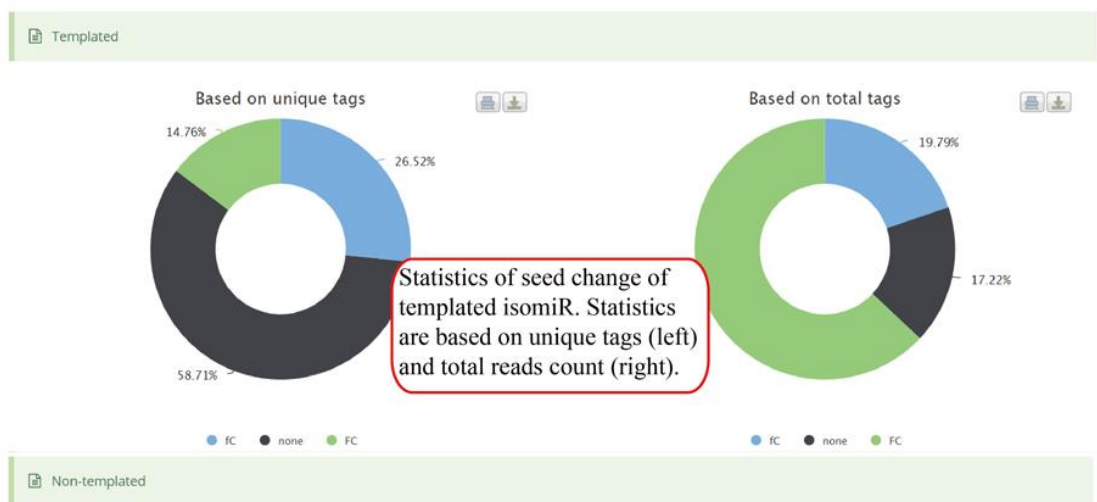


Non-templated



Statistics of isomiR seed changes

Statistics of isomiR seed changes



Statistics of mutations in isomiR



Non-templated isomiR list

Expressed List of non-templated isomiR

Show 10 entries

Download all tags expression list from this link

Download all records

Tag	PriMIRNA	Mature miRNA	IsoType	TagSeq	Variations	Tag raw number	TPM
t00000090_x9933	ptc-MIR393	ptc-miR393	3'+1	CGCATTGATT	TS: 21:C>T,22:C>T	9933	3115.827
t00000090_x9933	ptc-MIR393b	ptc-miR393b	3'+1	CGCATTGATT	TS: 21:C>T,22:C>T	9933	3115.827
t00000090_x9933	ptc-MIR393c	ptc-miR393c	3'+1	TCCAAGGGATCGCATTGATT	3V: 21:C>T	9933	3115.827
t00000115_x7274	ptc-MIR159a	ptc-miR159a	no drift	TTTGATTGAAGGGAGCTCT	SS: 21:A>T	7274	2281.74
t00000115_x7274	ptc-MIR159b	ptc-miR159b	no drift	TTTGATTGAAGGGAGCTCT	SS: 21:A>T	7274	2281.74
t00000193_x4629	ptc-MIR167a	ptc-miR167a	3'+1	TGAAGCTGCCAGCATGATCTGG	TS: 21:A>G,22:A>G	4629	1452.045
t00000193_x4629	ptc-MIR167b	ptc-miR167b	3'+1	TGAAGCTGCCAGCATGATCTGG	TS: 21:A>G,22:T>G	4629	1452.045
t00000193_x4629	ptc-MIR167c	ptc-miR167c	3'+1	TGAAGCTGCCAGCATGATCTGG	TS: 21:A>G,22:A>G	4629	1452.045
t00000193_x4629	ptc-MIR167d	ptc-miR167d	3'+1	TGAAGCTGCCAGCATGATCTGG	TS: 21:A>G,22:T>G	4629	1452.045
t00000193_x4629	ptc-MIR167e	ptc-miR167e	3'+1	TGAAGCTGCCAGCATGATCTGG	SS: 22:A>G	4629	1452.045

Click the ID to show the detail of pri-miRNA

Click the ID to show the detail of miRNA

Templated isomiR list

Expressed List of templated isomiR

Show 10 entries

Download all tags expression list from this link

Download all records

Tag	Pri_miRNA	mature miRNA	IsoType	TagSeq	Variations	Tag raw number	TPM
t00000002_x244524	ptc-MIR159a	ptc-miR159a	3'+1	AAGGGAGCTCT	templated	244524	76703.353
t00000002_x244524	ptc-MIR159b	ptc-miR159b	3'+1	AAGGGAGCTCT	templated	244524	76703.353
t00000014_x51703	ptc-MIR171c	none	other	TTGAGCCGCGCCAATCACT	templated	51703	16218.422
t00000014_x51703	ptc-MIR171d	none	other	TTGAGCCGCGCCAATCACT	templated	51703	16218.422
t00000020_x31267	ptc-MIR393c	ptc-miR393c	3'+1	TCCAAGGGATCGCATTGATT	templated	31267	9807.969
t00000029_x22794	ptc-MIR166g	ptc-miR166g	5'+2 3'-2	TCTCGGACCAGGCTTCATTCC	templated	22794	7150.121
t00000029_x22794	ptc-MIR166h	ptc-miR166h	5'+2 3'-2	TCTCGGACCAGGCTTCATTCC	templated	22794	7150.121
t00000029_x22794	ptc-MIR166m	ptc-miR166m	5'+2 3'-2	TCTCGGACCAGGCTTCATTCC	templated	22794	7150.121
t00000029_x22794	ptc-MIR166n	ptc-miR166n	5'+2 3'-2	TCTCGGACCAGGCTTCATTCC	templated	22794	7150.121
t00000029_x22794	ptc-MIR166o	ptc-miR166o	5'+2 3'-2	TCTCGGACCAGGCTTCATTCC	templated	22794	7150.121

Click the ID to show the detail of pri-miRNA

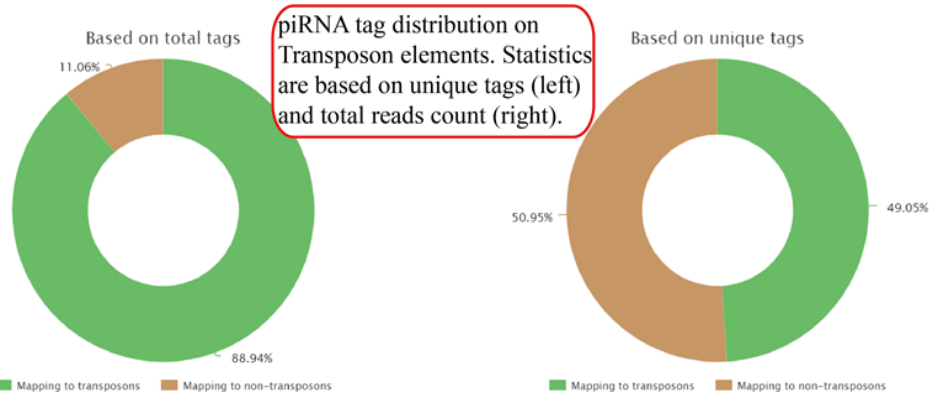
Click the ID to show the detail of miRNA

Novel miRNA list. All unclassified reads were considered for detecting candidate novel miRNA genes. Sequence of predicted putative miRNA and miRNA star along with the corresponding tag number, tag count and hairpin structure are provided:

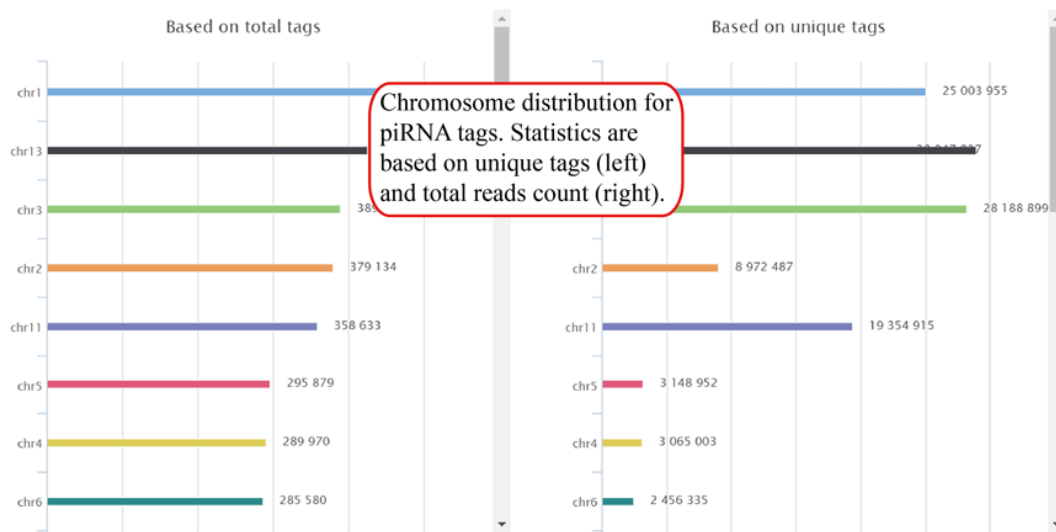
Show 20 entries

SN	miRNA	Strand	Total tag number	Total RPM	Type	Sequence	Tag name	Most abundant tag	Most abundance
2	xxx-m10-5p	5	485	349423.63	1	ACTATGAGCTAATGATTAG	seq14233_x105	105	75648.41
3	xxx-m8-5p	5	82	59077.81	1	TGCTTGTGTGTATGTGGGTGCTC	seq27452_x47	47	33861.67
4	xxx-m1-5p	5	49	35302.59	novel miRNA list		seq35611_x34	34	24495.68
5	xxx-m9-5p	5	26	18731.99	1	CGCTGGGACTGATTATTAG	seq44441_x26	26	18731.99
6	xxx-m7-5p	5	22	15850.14	1	CAGGGACAAGGGGAGG	seq70867_x15	15	10806.92
7	xxx-m5-5p	5	20	14409.22	1	GGCGGGGTCCCGGGCC	seq113913_x8	8	5763.69
8	xxx-m2-3p	3	18	12968.30	1	AAAGAATATCCCAGCATT	seq83424_x12	12	8645.53

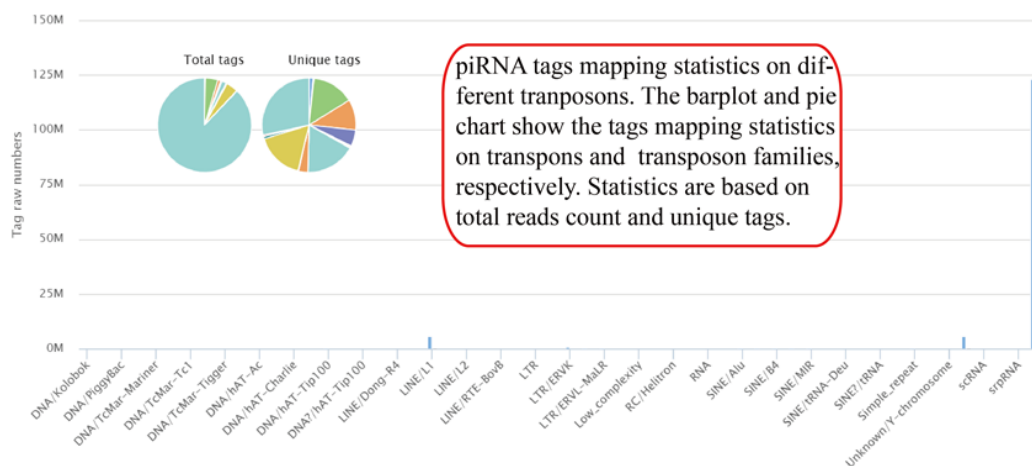
Tag distribution of piRNA on Transposon elements



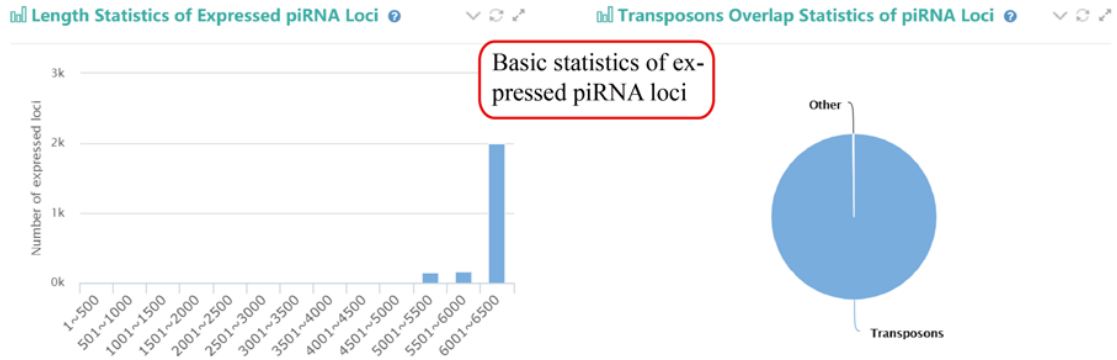
Tag distribution on different chromosomes for piRNA



piRNA tag distribution on different kind of transposon elements



piRNA producing loci basic statistics



Expressed piRNA producing loci list (top 200)

Show 20 entries

Locus	Chromosome:Start-End	RPM	RPKM	Tag number	Length	isTransposon
Cluster874	chr13:21167629-21172942	87602.90	16485.30	2576082	5314	TRUE
Cluster1279	chr3:97154435-97162672	66428.67		1953425	8238	TRUE
Cluster793	chr1:171042883-171052682	55380.61		1628542	9800	TRUE
Cluster1271	chr3:96309878-96315022	55090.47	10707.57	1620010	5145	TRUE
Cluster1274	chr3:96456798-96464903	53760.93	6632.24	1580913	8106	TRUE
Cluster1273	chr3:96409213-96414602	46507.43	8628.47	1367614	5390	TRUE
Cluster880	chr13:23425643-23438803	30304.88	2302.63	891156	13161	TRUE
Cluster877	chr13:21931006-21940996	27152.06	2717.65	798443	9991	TRUE

Top 200 list of expressed piRNA loci

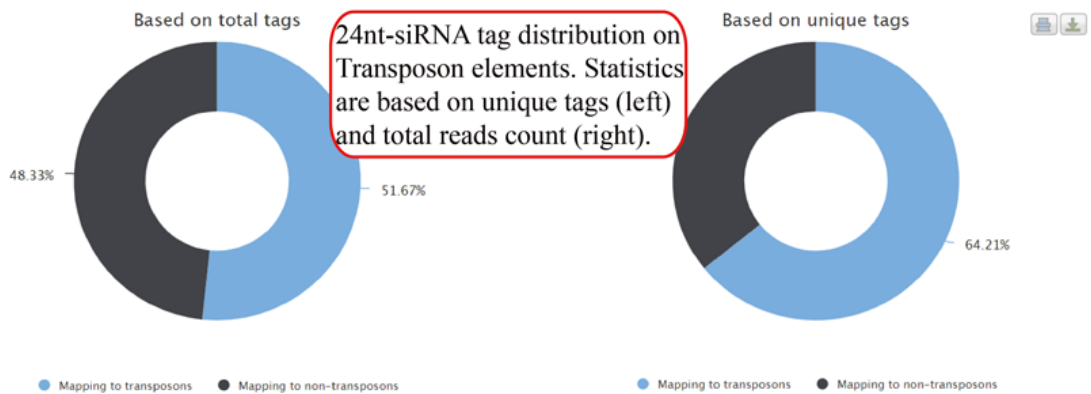
Expressed piRNA list (top 200)

Show 20 entries

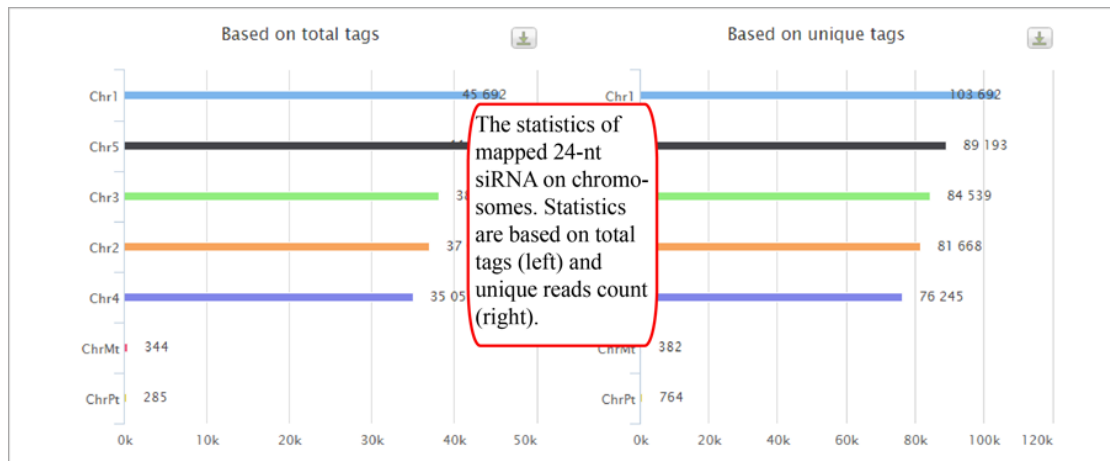
Tag name	Sequence	Tag raw number	RPM	Mapping locus
piR-mmu-64542836	GTTCCGTAGTGTAGTGGTTATCACGTTCCG	209861	7136.59	NA
piR-mmu-2004832	GCATGGGTGGTTCAGTGGTAGAATTCCTC	200174	7113.22	NA
piR-mmu-58055787	TCCCTGTGGTCTAGTGGTTAGG	171501	5399.10	NA
piR-mmu-55164778	AAGAAAGATTGCAAGA	71501	2431.48	NA
piR-mmu-61571449	GTTTCTGTAGTTGAATTACAACGATG	63230	2150.22	NA
piR-mmu-56655079	TCAACAAGTACCGTAAGGGAAAGTTG	61183	2080.60	NA
piR-mmu-23072140	TAGCTTATCAGACTGATGTTGAC	57990	1972.02	NA
piR-mmu-49097150	GCCCGGCTAGCTAGTCGGTAGAGCATGGGA	44288	1506.07	NA

Top 200 list of expressed piRNA

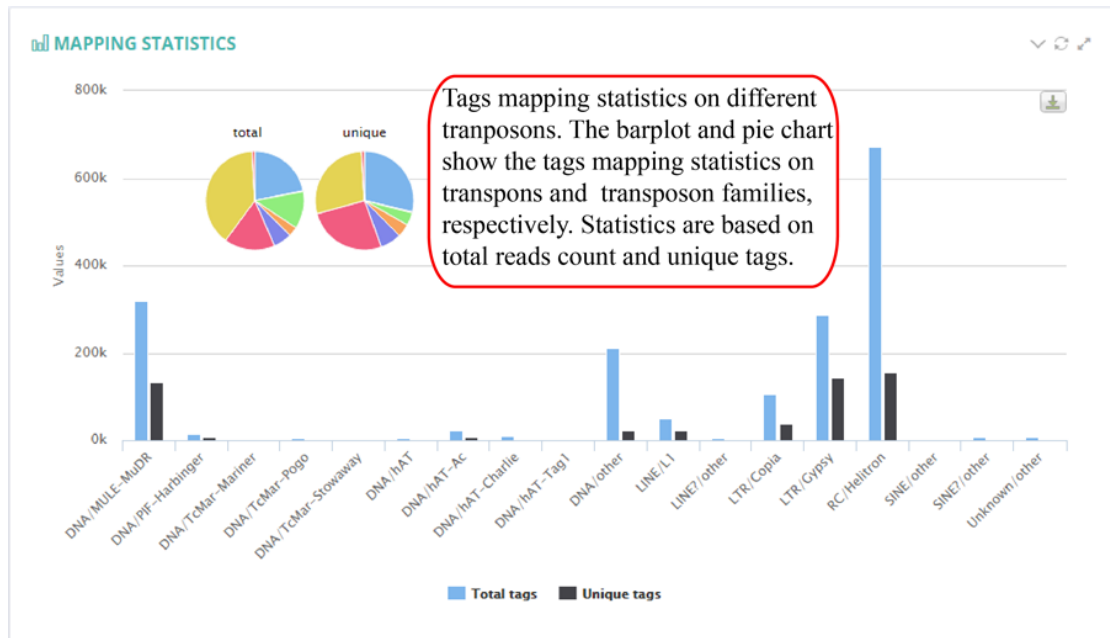
Tag distribution of 24nt-siRNA on Transposon elements



Distribution of 24-nt siRNA on different chromosomes



24nt siRNA tag distribution on different kind of transposon elements



24nt siRNA expression list

Tags Expression Profile (Top200)

RPM: normalized expression value, Reads Per Million mapped reads.
Mapping Locus: NUM (Non-Unique Mapping).

Show entries

Tag name	Sequence	Tag raw number	RPM	Mapping locus
seq1_7735	GCCGGCCGGGGGACGGACTGGGAA	7735	594.22	NUM
seq2_1610	AACGACTCTCGGCAACGGATATCT	1610	123.68	NUM
seq3_1488	AGTTACTAATTCATGATCTGGC	1488	114.37	NUM
seq4_1319	TAGCAACTGTTCTTTAGACGAC	1319	101.3	NUM
seq5_1273	AAGAACTTAAACCGCAACGGATC	1273	97.79	NUM
seq6_1101	AACTAGAAAAGACATTGGACATAT	1101	84.58	NUM
seq7_986	ATAAGAACTTAAACCGCAACGGGA	986	75.75	NUM
seq8_898	AAAACGACTCTCGGCAACGGATAT	898	68.99	NUM
seq9_748	ATCAGAACTCCGCAGTTAAGCGTG	748	57.46	NUM
seq10_742	TCTGACAGAAGACTCTCCATATAT	742	57.00	cluster7115

Download all tags expression list from this link

DOWNLOAD ALL RECORDES

Top200 list of 24-nt siRNA tags expression

Tag mapping locus. "NUM" indicates multiple mapping of the tag.

Expressed piRNA producing loci list (top 200)

Copy CSV Excel PDF Print Search:

Locus	Chromosome:Start-End	TPM	TPKM	Tag raw number	Length	isTransposon
cluster11299	chr10:15474036-15479525	117.91	21.49	2069	5490	TRUE
cluster11451	chr10:16937608-16943890	129.19	24.25	2267	6283	TRUE
cluster12160	chr11:1013364-1018568	1416.95	272.23	24864	5205	TRUE
cluster12164	chr11:1049947-1055458	266.76	48.40	4681	5512	TRUE
cluster12241	chr11:1455699-1456156	170.39	372.04	2990	458	TRUE
cluster12428	chr11:2488855-2490086	141.10	114.53	2476	1232	TRUE
cluster12513	chr11:3039113-3044439	1631.45	306.26	28628	5327	TRUE
cluster12541	chr11:3212632-3215290	113.01	42.50	1983	2659	TRUE
cluster12543	chr11:3235950-3243668	159.51	20.66	2799	7719	TRUE
cluster12712	chr11:4185118-4190419	297.59	56.13	5222	5302	TRUE

Top 200 list of expressed 24nt-siRNA loci

phasiRNA and loci list

PhasiRNA loci list

Download all tags expression list from this link

Download all records

Copy CSV Excel PDF Print

Search:

phase type	Name	Locus	Seq	Tag row count	TPM	phasiRNA genomic site	phasiRNA length	hts score	p-value
21	Phas-1	chr1:9950727-9951123	CGCGGCTCAATCAACAGCTCT	15	0.854820	9950727	21	2	2.923414930
21	Phas-1	chr1:9950727-9951123	AAGTAGAAAGTGCCCTGTGG	69	3.932172	9950787	21	1	2.923414930
21	Phas-1	chr1:9950727-9951123	CTAGGCGGTGGGGCAGTGACT	4	0.227952	9950850	21	1	1.937373648
21	Phas-1	chr1:9950727-9951123	AAGTCTTATCTGAGTTTCT	4	0.227952	9950895	20	2	1.937373648
21	Phas-1	chr1:9950727-9951123	GAA	98916	9950913	21	2	1.937373648	
21	Phas-1	chr1:9950727-9951123	CAC	862292	9950916	21	2	1.937373648	
21	Phas-1	chr1:9950727-9951123	TTGACAAATTGGATAAGTGA	4	0.227952	9950934	21	2	1.937373648
21	Phas-1	chr1:9950727-9951123	TCACTTGAATCAGGCTCTCC	16	0.911808	9950937	21	1	1.937373648
21	Phas-1	chr1:9950727-9951123	ATAGTATTCACATAGTGACT	198	11.283623	9950979	21	1	1.937373648
21	Phas-1	chr1:9950727-9951123	TTGATTTGGATTGGCGCTC	433	24.675801	9951000	21	1	1.937373648

Click to the tag mapping in IGV

PhasiRNA induced miRNA (trigger) list

Download all tags expression list from this link

Download all records

Copy CSV Excel PDF Print

Search:

phase type	miRNA	Target	Bind site	miRNA seq	target seq	Genomics bind site	Phase loci
21	ptc-miR171a-3p	Phas-1	116-136	GCACTATAACCGTGCCGAGTT	AGGGATATTGGCGGGCTCAA	chr1:9950716-9950736:w	chr1-9950727-99
21	ptc-miR171a-3p	Phas-4	116-136	GCACTATAACCGTGCCGAGTT	AGGGATATTGGCGGGCTCAA	chr14:4649950-4649970:c	chr14-4649773-4
21	ptc-miR171b	Phas-4	116-136	GCACTATAACCGTGCCGAGTT	AGGGATATTGGCGGGCTCAA	chr14:4649950-4649970:c	chr14-4649773-4
21	ptc-miR171b	Phas-1	116-136	GCACTATAACCGTGCCGAGTT	AGGGATATTGGCGGGCTCAA	chr1:9950716-9950736:w	chr1-9950727-99
21	ptc-miR171c	Phas-1	119-139	CTATAACCGCGCCGAGTTAGA	GATATTGGCGGGCTCAATCA	chr1:9950719-9950739:w	chr1-9950727-99
21	ptc-miR171d	Phas-1	119-139	CTATAACCGCGCCGAGTTAGA	GATATTGGCGGGCTCAATCA	chr1:9950719-9950739:w	chr1-9950727-99
21	ptc-miR171e	Phas-1	119-139	CTATAACCGCGCCGAGTTAGA	GATATTGGCGGGCTCAATCA	chr1:9950719-9950739:w	chr1-9950727-99
21	ptc-miR171f	Phas-1	119-139	CTATAACCGTGCCGAGTTAGT	GATATTGGCGGGCTCAATCA	chr1:9950719-9950739:w	chr1-9950727-99
21	ptc-miR171g-3p	Phas-1	119-139	CTATAACCGTGCCGAGTTAGT	GATATTGGCGGGCTCAATCA	chr1:9950719-9950739:w	chr1-9950727-99
21	ptc-miR393a-5p	Phas-6	116-136	CTAGTTACGCTAGGAAACCT	GATCAGAGCGATCCCTTGAG	chr14:11387358-11387378:w	chr14-11387369-

Click to the show the detail of miRNA in miR-Base

rRNA expression list

Download all tags expression list from this link

Search:

Copy CSV Excel PDF Print

tRNA	Total tag raw number	Total TPM	Tag name	Most abundant tag	Mo ablu TPM	tag seq	accun	Sequence	Genomic coordinate
tRNA-Ala-AGC-1-1	48	13.57	seq9730_x19	19	5.37				
tRNA-Ala-AGC-10-1	1	0.28	seq208625_x1	1	0.28				
tRNA-Ala-AGC-11-1	3	0.85	seq241073_x1	1	0.28				
tRNA-Ala-AGC-13-1	16	4.52	seq17827_x10	10	2.83				
tRNA-Ala-AGC-15-1	1	0.28	seq378895_x1	1	0.28				
tRNA-Ala-AGC-2-1	36				1.41				
tRNA-Ala-AGC-24-1	1				0.28				
tRNA-Ala-AGC-3-1	87				10.46				
tRNA-Ala-AGC-4-1	40938	11569.94	seq15_x34371	34371	9713.97				
tRNA-Ala-AGC-5-1	3	0.85	seq282127_x1	1	0.28				

Click the ID to link to show the detail of tRNA in GtRNadb

Click "Add" icon to show the detail

Showing 1 to 10 of 244 entries

Previous 1 2 3 4 5 ... 25 Next

snoRNA expression list

snoRNA expression profile

Download all tags expression list from this link Download all records

Copy CSV Excel PDF Print Search:

snoRNA	Total tag raw number	Total TPM	Tag name	Most abundant tag	Most abundant TPM	Most abundant tag seq	Other tags detail	Sequence	Genomic coordinate
URS000062A0D4	14	0.8	t00689891_x2	2	0.11				
URS0000638A41	65	3.7	t00057783_x17	17	0.97				
URS0000639EB2	223	12.71	t00035466_x32	32	1.82				
URS000063CAAE	5	0.28	t00540681_x2	2	0.11				
URS000063F064	4	0.06			0.06				
URS000065DC35	2	0.06			0.06				
URS000065DC93	2	0.11	t02639833_x1	1	0.06				
URS000065F996	4	0.23	t01124019_x1	1	0.06				
URS000066201B	1	0.06	t01949797_x1	1	0.06				
URS00006621C4	3	0.17	t04048396_x1	1	0.06				

Click the ID to link to show the detail of snoRNA Click "Add" icon to show the detail

snRNA expression list

sRNA expression profile Download all records

Copy CSV Excel PDF Print Search:

snRNA	Total tag raw number	Total TPM	Tag name	Most abundant tag	Most abundant TPM	Most abundant tag seq	Other tags detail	Sequence	Genomic coordinate
URS0000623EFD	10887	620.43	t00000501_x2096	2096	119.45				
URS0000664695	543	30.94	t00009664_x161	161	9.18				
URS00006ABC69	4744	270.35	t00001756_x773	773	44.05				
URS00006F20CA	9	0.51	t00533746_x2	2	0.11				
URS00006F4165	45	0.63			0.63				
URS00006FCDAC	1	0.06			0.06				
URS00007216E4	64	3.65	t00111028_x8	8	0.46				
URS0000904F67	305	17.38	t00036704_x30	30	1.71				
URS00009064BA	999	56.93	t00018186_x76	76	4.33				
URS0000906CDD	2	0.11	t03423140_x1	1	0.06				

Download all tags expression list from this link Download all records

Click the ID to link to show the detail of snRNA Click "Add" icon to show the detail

Other snRNA expression list

other snRNA expression profile Download all records

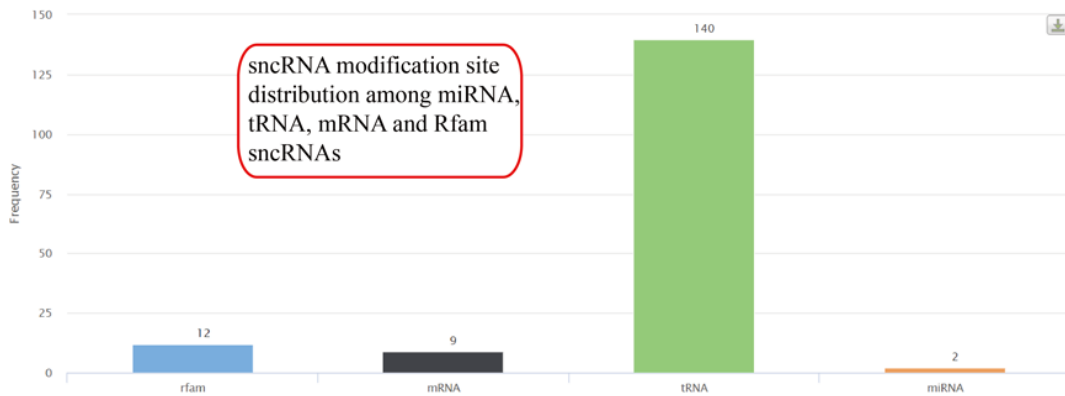
Copy CSV Excel PDF Print Search:

other snRNA	Total tag raw number	Total TPM	Tag name	Most abundant tag	Most abundant TPM	Most abundant tag seq	Other tags detail	Sequence
URS00006DF557_3694 antisense_RNA	14	0.8	t00217546_x4	4	0.23			
URS000072611E_3694 antisense_RNA	2	0.11	t01626175_x1	1	0.06			
URS000091FE87_3694 antisense_RNA	8	0.46	t00361015_x3	3	0.17			
URS000093688C_3694 SRP_RNA	342	19.49	t00057244_x17	17	0.97			
URS0000954804_3694 SRP_RNA	1	5.19			5.19			
URS0000977F1B_3694 antisense_RNA	1	0.11			0.11			
URS0000978C80_3694 SRP_RNA	4056	231.14	t00007294_x216	216	12.31			
URS00009A6A31_3694 RNase_MRP_RNA	9	0.51	t03884878_x1	1	0.06			
URS0000BECD8C_3694 SRP_RNA	223	12.71	t00026378_x47	47	2.68			
URS0000BEDBD6_3694 SRP_RNA	8641	492.43	t00003556_x435	435	24.79			

Download all tags expression list from this link Download all records

Click the ID to link to show the detail information Click "Add" icon to show the detail

Distribution of sncRNA with high confident modification



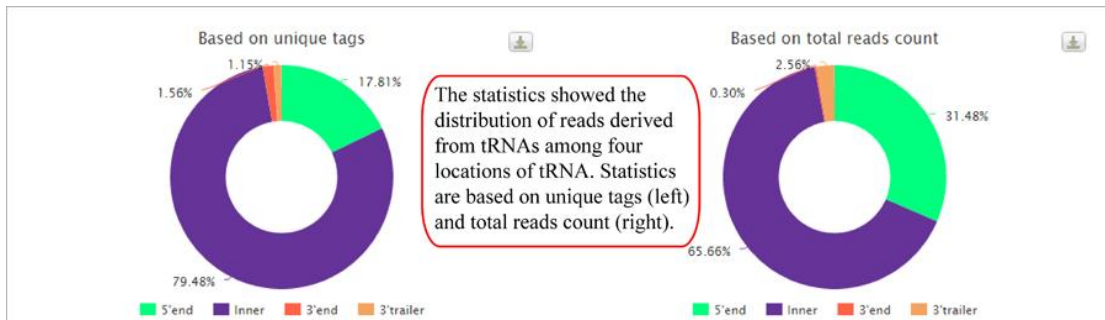
Modification site list of sncRNA

Download Table Search:

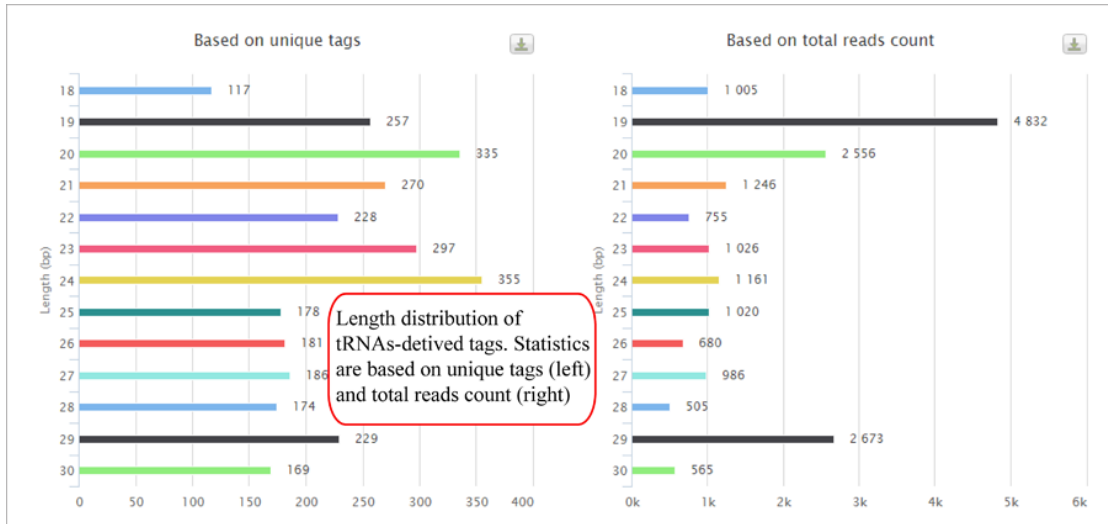
sncRNA	Modification site	coverage_input	count_input	faivalue_input	coverage_treated	count_treat
chr1.trna197-LysTTT chr1:133033962-133034034:+ 76	47	548	34	0.062043795620438	379	171
chr1.trna281-ProAGG chr1:165641634-165641705:+ 75	46			0.0658307210031348	189	92
chr1.trna451-AsnGTT chr1:171033942-171034015:- 77	48			0.0789473684210526	472	202
chr1.trna460-ProCGG chr1:165642271-165642342:- 75	46			0.0432989690721649	208	92
chr1.trna518-LysTTT chr1:133034327-133034399:- 76	47	548	34	0.062043795620438	379	171
chr10.trna167-AsnGTT chr10:80249159-80249232:+ 77	48	912	72	0.0789473684210526	472	202
chr10.trna301-SerCGA chr10:128458464-128458545:- 85	33	88	10	0.113636363636364	104	33
chr10.trna384-TrpCCA chr10:91181412-91181483:- 75	46	330	15	0.0454545454545455	380	189

A red-bordered text box is overlaid on the table with the text: 'sncRNA modification site list. P value < 0.05 is used as the threshold to define the modification sites'.

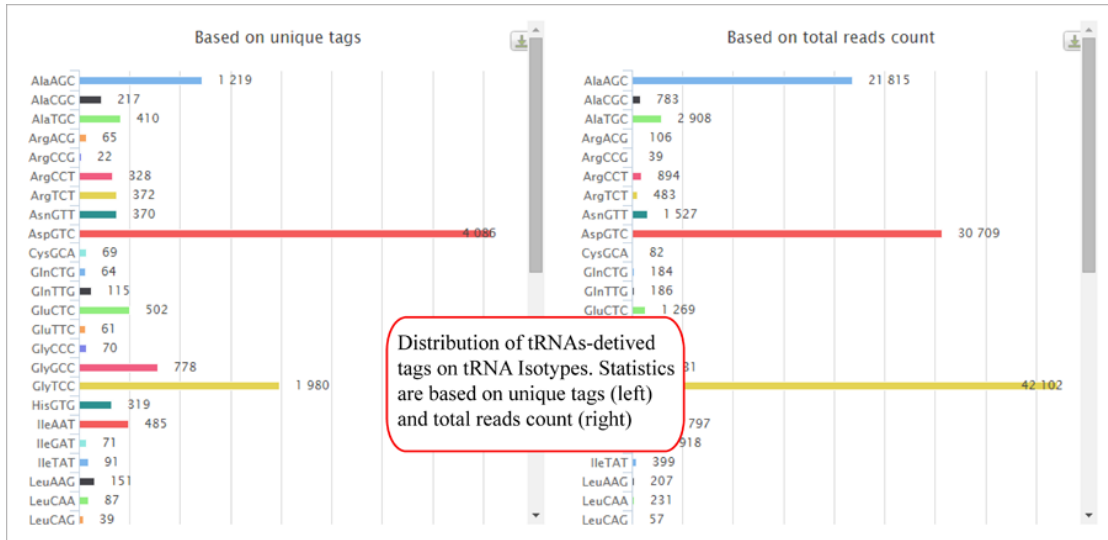
Tags mapping statistics on different regions



Statistics of tRNA family match results



Statistics of tRNA family match results



Statistics of tRNA match results



tRFs result list

tRF type: All Amino acid: All Anticodon: All Minimum read number: 10 P-value: 0.05

Re-filter the tRFs list by setting the criteria including tRF type, amino acid, anticodon, minimum read number and detection p-value.

Click to download 'tRFs list table' which could be used for ground case study

TSV Print Fasta

Show 20

#tRF Type	Source tRNA	RF Length	Locus on tRNA	Read number	RPM	Sequence	P-value
tRF-1	chr6.tma44-SerGCT	31	1-31	136	147.98372	ACCGTTTTCTTTAAGAGGGTAACCAGGTTTT	0.00
tRF-1	chr6.tma44-SerGCT	30	1-30	27	29.37912	ACCGTTTTCTTTAAGAGGGTAACCAGGTTT	0.00
	4-SerGCT	32				TCGTTTTCTTTAAGAGGGTAACCA	
	25-CysGCA	25				GTGCTCCGGAGTTACCTGTTTT	
	25-CysGCA	24				GTGCTCCGGAGTTACCTGTTT	
tRF-1	chr15.trna11-SerGCT	18	1-18	248	269.85267	ATGTGGTGCTTACTTTT	0.00
tRF-1	chr15.trna11-SerGCT	17	1-17	2155	2344.88912	ATGTGGTGCTTACTTT	0.00
tRF-1	chr1.tma22-GlnCTG	32			461.36101	TTCTGTTAATTAGGACGGCAATGTTGTT	0.00
tRF-1	chr16.trna3-ArgCCT	17			620.22589	AAGAAAGGCCGAATTTT	0.00
tRF-1	chr1.tma103-GlnCTG	32			82.69679	TTGTGTTTCATTGGCATGGTAAGCCGTGTT	0.01
tRF-1	chr6.tma36-ArgACG	19	1-19	59	64.19882	GGTGTAATCTTGGCTTTT	0.00
tRF-1	chr6.tma36-ArgACG	20	1-20	32	34.81970	GGTGTAATCTTGGCTTTT	0.00
tRF-1	chr1.tma90-ArgTCT	22	1-22	11	11.96927	GTCACCTGGCAGGTGCCTCTT	0.00
tRF-1	chr12.trna5-AspGTC	19			335.13961	GTGTGTAGCTGCATTTT	0.00
tRF-1	chr12.trna5-AspGTC	17			608.25662	GTGTGTAGCTGCATTT	0.00
tRF-1	chr12.trna5-AspGTC	18			866.14002	GTGTGTAGCTGCATTT	0.00
tRF-1	chr17.trna12-ThrCGT	20	1-20	146	158.86488	GAGACCCGAGGTAGGGCTTT	0.00
tRF-1	chr11.trna18-ValTAC	21	1-21	924	1005.41882	TGAGATGTACCTAGCGTTT	0.00
tRF-1	chr11.trna18-ValTAC	20	1-20	770	837.84901	TGAGATGTACCTAGCGTTT	0.00
tRF-1	chr5.tma6-LeuAAG	22	1-22	420	457.00855	GCTTGTGTGATTCCTCAATT	0.00

tRF types include tRF-1, tRF-2, tRF-3 and tRF-5

The location of tRFs on tRNA/3'trailer

Reads per million tRNA-derived reads

P-value calculated by binomial test

Click to view the detail annotation of the according tRNA on GtRNAdb

Click the sequences to view the detail annotation of the according tRF

Non-redundant tRFs list

Non-redundant tRF list

Copy CSV Excel PDF Print

Non-redundant list of tRF sequences

Search:

tRF sequence	tRF type	Tag raw number	TPM	#tRNA loci	tRNA locus details
AACCGGCGGAAACACCA	tRF-3	26	17.4454	4	⊕
ACATTGGTGGTTCAGTGGTAGAATCTCGCC	tRF-5	21	14.0905	1	⊕
ACATTGGTGGTTCAGTGGTAGAATCTCGCCT	tRF-5	23	15.4324	1	⊕
ACCAGGATGGCCGAGTGGTTAAGG	tRF-5	17	11.4066	2	⊕
AGCAGAGTGGCGCAGCGGAAGCGTGTGGGC	tRF-5	17	11.4066	2	⊕
AGCAGAGTGGCGCAGCGGAAGCGTGTGGGCC	tRF-5	18	12.0776	2	⊕
AGGATGGCCGAGCGGTCTAAG	tRF-2	73	48.9812	3	⊕
AGTCGGTAGAGCATCAGAC	tRF-2	85	57.0329	2	⊕
ATCCCACCGCTGCCACCA	tRF-3	38	25.4971	2	⊕
ATCCCACCTCTGACACCA	tRF-3	194	130.1693	4	⊕

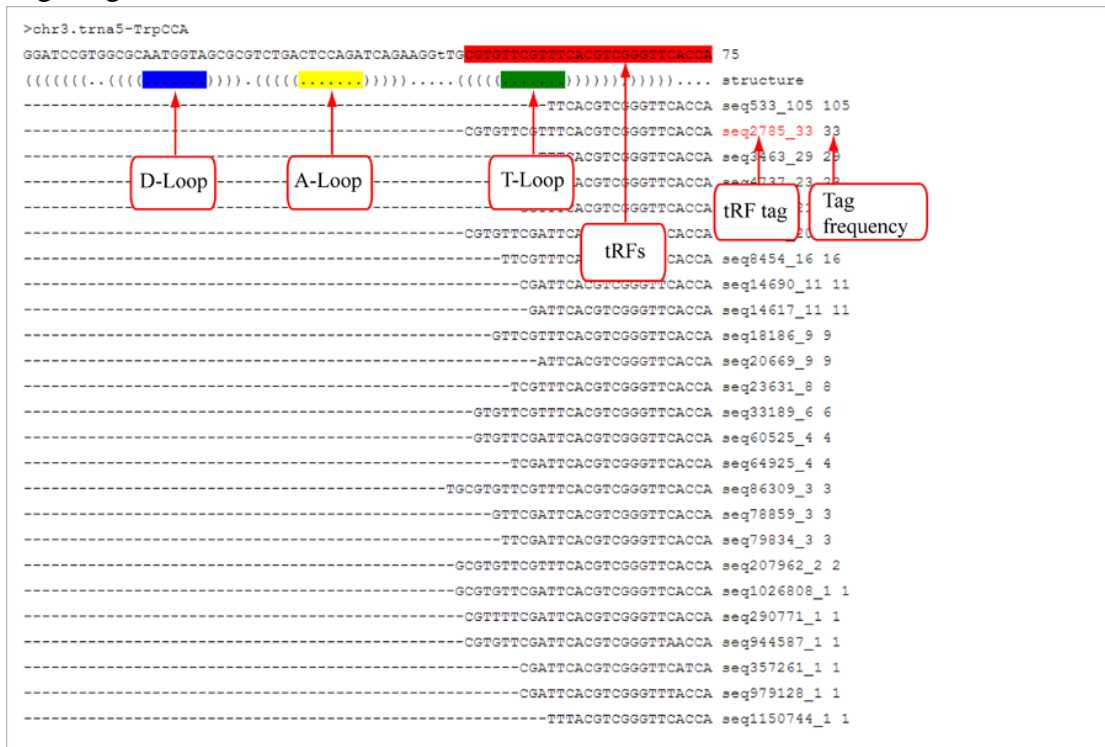
Click to show the tRNA producing according tRF

Basic information of tRFs

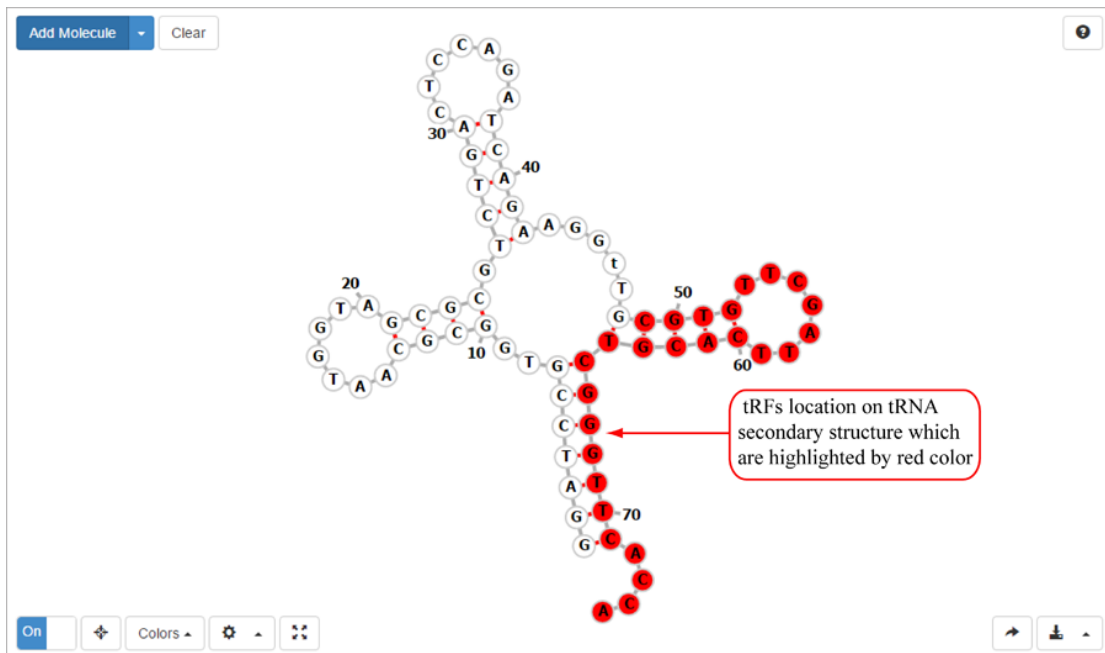
trf Type:	tRF-3
Source tRNA:	chr4.trna35-SerCGA
trf length:	23
Locus on tRNA:	63-85
Read number:	16
RPM:	841.66228
P-value:	0.00
Sequence:	TTCGAATCCTGCTGTTGACGCCA

← Basic information of tRFs

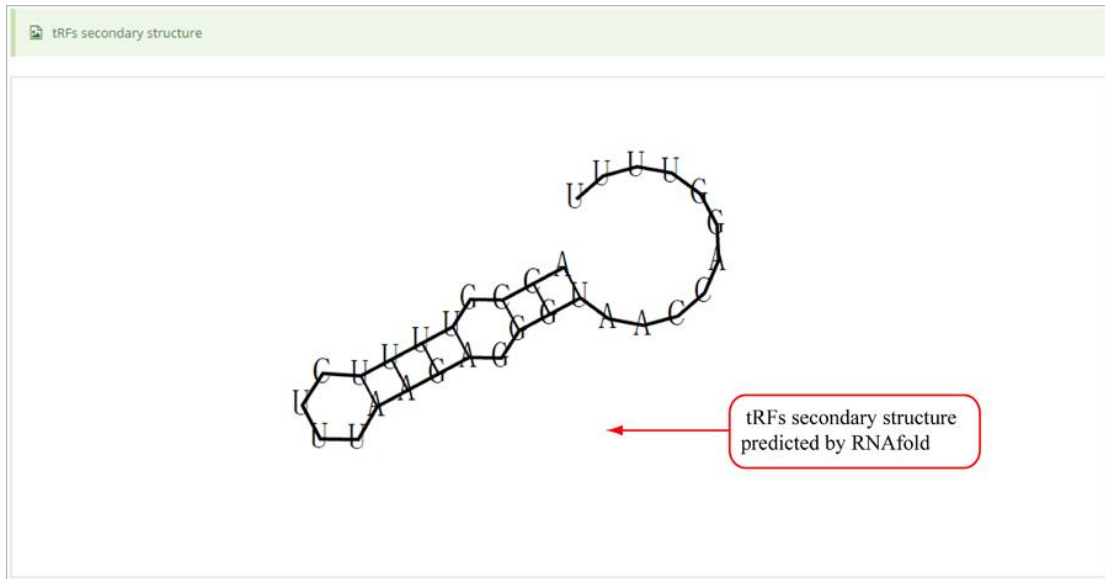
Tags alignments



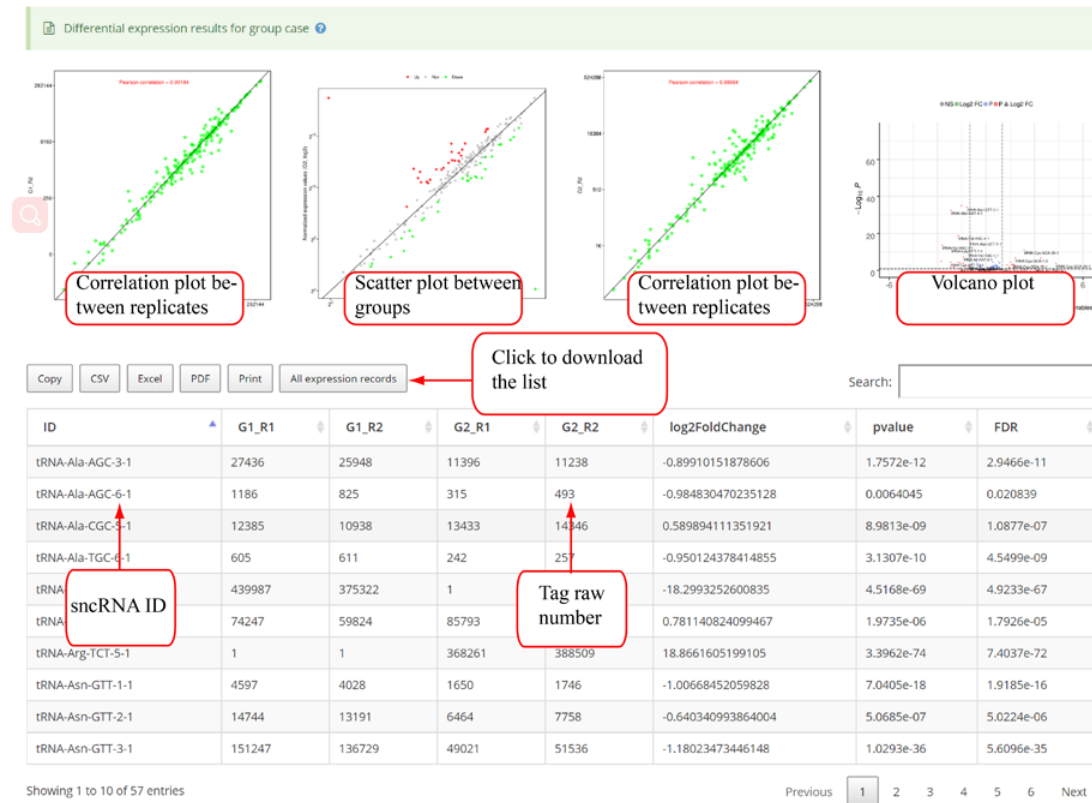
tRFs location on tRNA secondary structure



tRFs secondary structure



Differential expressed sncRNA in group case analysis



sncRNA target gene lists

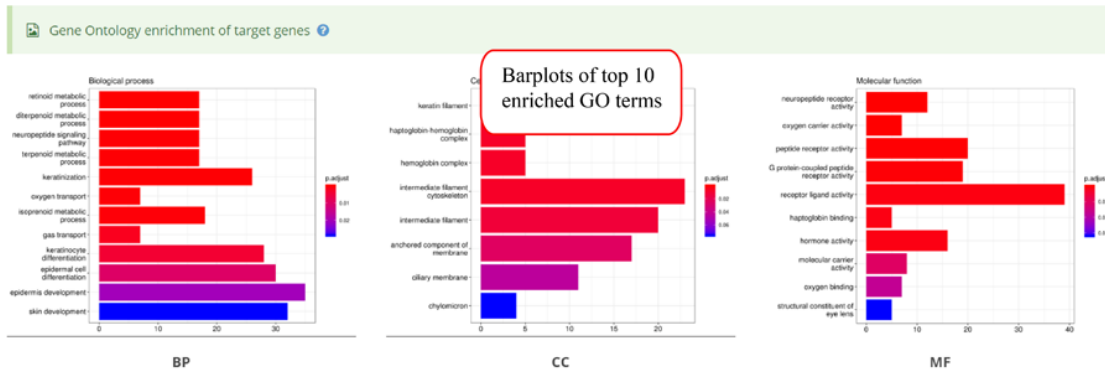
Click to download "Target list" → Excel Print

Show 20 entries Search:

Num	trf name	Number of targets	Targets
1	seq5_13954	2296	AT1G01400,AT1G01430,AT1G01450,AT1G01470,AT1G01610,AT1G01690,AT1G01790,AT1G02010,AT1G02030,AT1G02080,AT1G021
2	seq2_33665	3724	AT1G01040,AT1G01090,AT1G01100,AT1G01190,AT1G01305,AT1G01390,AT1G01450,AT1G01490,AT1G01610,AT1G01740,AT1G020
3	seq1_10000	3100	AT1G01080,AT1G01110,AT1G01220,AT1G01225,AT1G01390,AT1G01440,AT1G01460,AT1G015
4	seq4_10000	3100	AT1G01080,AT1G01110,AT1G01220,AT1G01230,AT1G01460,AT1G01560,AT1G01660,AT1G016
5	seq3_10000	3100	AT1G01070,AT1G01120,AT1G01200,AT1G01225,AT1G01320,AT1G01440,AT1G01590,AT1G016
6	seq7_11053	1472	AT1G01140,AT1G01500,AT1G01720,AT1G01740,AT1G01810,AT1G01880,AT1G01900,AT1G01950,AT1G02080,AT1G02300,AT1G029
7	seq6_11534	3594	AT1G01060,AT1G01150,AT1G01320,AT1G01390,AT1G01420,AT1G01560,AT1G01725,AT1G01740,AT1G01750,AT1G01780,AT1G017
8	seq8_9692	1036	AT1G01030,AT1G01190,AT1G01340,AT1G01390,AT1G01760,AT1G02360,AT1G02550,AT1G02690,AT1G02890,AT1G02940,AT1G029
9	seq9_9026	2296	AT1G01400,AT1G01430,AT1G01450,AT1G01470,AT1G01610,AT1G01690,AT1G01790,AT1G02010,AT1G02030,AT1G02080,AT1G021
10	seq3_25445	2790	AT1G01210,AT1G01220,AT1G01225,AT1G01305,AT1G01350,AT1G01400,AT1G01450,AT1G01550,AT1G01600,AT1G01610,AT1G016

Showing 1 to 10 of 10 entries

Gene Ontology enrichments of target genes



Gene Ontology enrichment for snCRNA targets

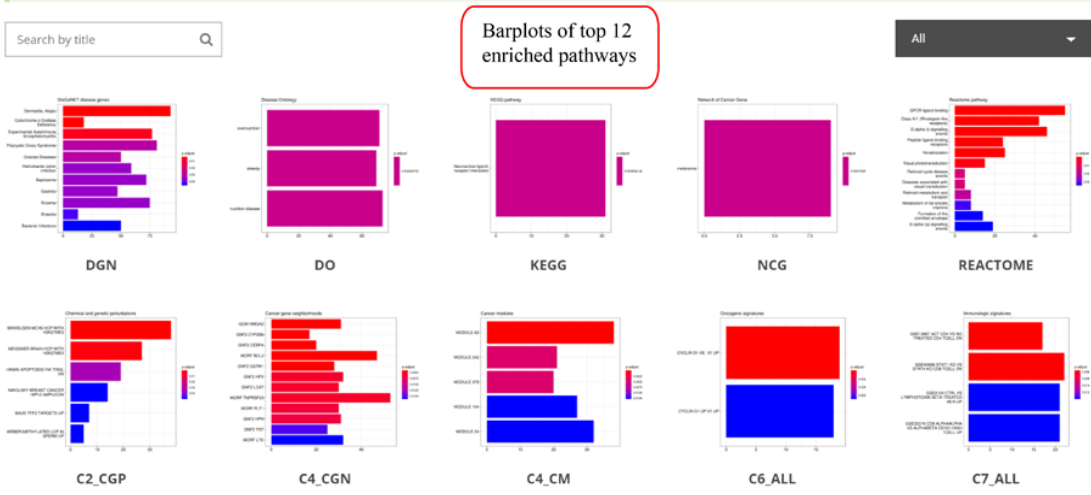
Copy CSV Excel PDF Print → Click to download all "GO-BP list" Search:

ID	Description	GeneRatio	BgRatio	p.value	p.adjust	GO
GO:0001523	retinoid metabolic process	17/690	86/18493	1.5545e-8	0.0001	BP
GO:0001653	peptide receptor activity	20/691	151/17632	0.0000018754	0.0005	MF
GO:0001653	hormone activity	12/691	121/17632	0	0.0005	MF
GO:0001653	structural constituent of eye lens	20/691	20/17632	0.0009	0.0009	MF
GO:0005344	oxygen carrier activity	7/691	14/17632	3.717e-7	0.0001	MF
GO:0005833	hemoglobin complex	5/755	12/19659	0.0001	0.0072	CC
GO:0005882	intermediate filament	20/755	203/19659	0.0001	0.0119	CC
GO:0006720	isoprenoid metabolic process	18/690	123/18493	7.0767e-7	0.0004	BP
GO:0006721	terpenoid metabolic process	17/690	104/18493	2.8678e-7	0.0003	BP
GO:0007200	phospholipase C-activating G protein-coupled...	12/690	97/18493	0.0003	0.0626	BP

Showing 1 to 10 of 42 entries

Previous 1 2 3 4 5 Next

Pathway enrichments of target genes



Enriched pathways

Copy CSV Excel PDF Print Search:

ID	Description	Link	GeneRatio	BgRatio	p.value	p.adjust	Pathway_type
hsa04080	hsa04080	Link	31/231	272/5894	5.2637e-8	0.00009001	KEGG pathway
R-HSA-500792	R-HSA-500792	Link	55/406	457/10554	2.3552e-14	2.0231e-11	Reactome
R-HSA-3	R-HSA-373076	Link	324/10554	3.0736e-1			
R-HSA-4	R-HSA-418594	Link	896/10554	1.3354e-1			
R-HSA-375276	R-HSA-375276	Link	24/406	190/10554	2.6995e-7	0.0001	Reactome
R-HSA-6805567	R-HSA-6805567	Link	25/406	214/10554	6.8302e-7	0.0001	Reactome
R-HSA-2187338	R-HSA-2187338	Link	15/406	103/10554	0.0000085025	0.0012	Reactome
R-HSA-2453864	R-HSA-2453864	Link	5/406	14/10554	0.0001	0.0132	Reactome
R-HSA-2474795	R-HSA-2474795	Link	5/406	14/10554	0.0001	0.0132	Reactome
R-HSA-975634	R-HSA-975634	Link	8/406	43/10554	0.0002	0.019	Reactome













Showing 1 to 10 of 31 entries Previous 1 2 3 4 Next

Functional gene network of target genes



Download local pipeline of sncRNAbench

DOWNLOAD LOCAL VERSIONS

Download local versions of sRNAtools		
 sRNAtools local package	Local program package	
 sRNAtools docker version (Image)	Docker image	
 sRNAtools VirtualBox version	VirtualBox version	
 Local version help document	Help document	
 sRNAtools database	RNA libraries required	
 Test data	Test data	

Browser tested

Browser name	Version
Internet Explorer	11
Chrome	70.0.3538.80
Firefox	66.0
Safari	5.1.7

3.3 Local server Vs. webserver analysis

All the result files are exactly similar between localization server analysis and webserver analysis (the figure shows below). Analysis on our website can get a quickly and visual results, and we also provide the URL where the results are stored, but if users want to take analysis more individuation and privately, download our sRNAtools Docker image, sRNAtools VMbox and install our process locally will be the best choice.

Localization server analysis VS webserver analysis

4. References

1. Chan PP, Lowe TM. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes, *Nucleic Acids Res* 2016;44:D184-189.
2. Kalvari I, Argasinska J, Quinones-Olvera N et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, *Nucleic Acids Res* 2018;46:D335-D342.
3. Goodstein DM, Shu S, Howson R et al. Phytozome: a comparative platform for green plant genomics, *Nucleic Acids Res* 2012;40:D1178-1186.
4. Hubbard T, Barker D, Birney E et al. The Ensembl genome database project, *Nucleic Acids Res* 2002;30:38-41.
5. Zhao Y, Li H, Fang S et al. NONCODE 2016: an informative and valuable data source of long non-coding RNAs, *Nucleic Acids Res* 2016;44:D203-208.
6. Szczesniak MW, Rosikiewicz W, Makalowska I. CANTATAdb: A Collection of Plant Long Non-Coding RNAs, *Plant Cell Physiol* 2016;57:e8.
7. Dong R, Ma XK, Li GW et al. CIRCpedia v2: An Updated Database for Comprehensive Circular RNA Annotation and Expression Comparison, *Genomics Proteomics Bioinformatics* 2018;16:226-233.
8. Meng X, Hu D, Zhang P et al. CircFunBase: a database for functional circular RNAs, *Database (Oxford)* 2019;2019.
9. Chu Q, Zhang X, Zhu X et al. PlantcircBase: A Database for Plant Circular RNAs, *Mol Plant* 2017;10:1126-1128.
10. Consortium TR. RNAcentral: a hub of information for non-coding RNA sequences, *Nucleic Acids Res* 2019;47:D1250-D1251.
11. Chen D, Yuan C, Zhang J et al. PlantNATsDB: a comprehensive database of plant natural antisense transcripts, *Nucleic Acids Res* 2012;40:D1187-1193.
12. Wang J, Zhang P, Lu Y et al. piRBase: a comprehensive database of piRNA sequences, *Nucleic Acids Res* 2019;47:D175-D180.
13. Rosenkranz D. piRNA cluster database: a web resource for piRNA producing loci, *Nucleic Acids Res* 2016;44:D223-230.

14. Liu Q, Ding C, Chu Y et al. PIn24NT: a web resource for plant 24-nt siRNA producing loci, *Bioinformatics* 2017;33:2065-2067.