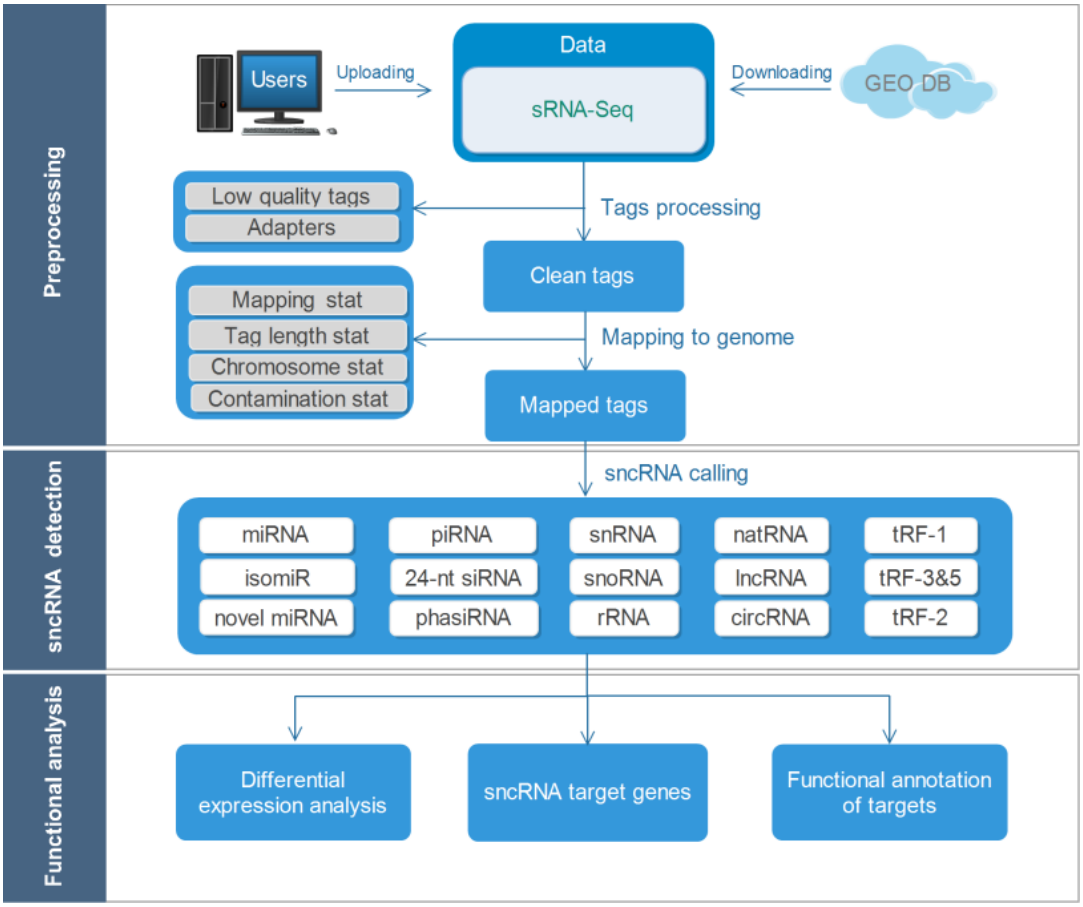
sRNAtools Tutorial

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**

1.Introduction *3*

2.Getting Start *4*

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

3. sRNAtools *15*

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

4. References *59*

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 Webserver

The webserver 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.

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.

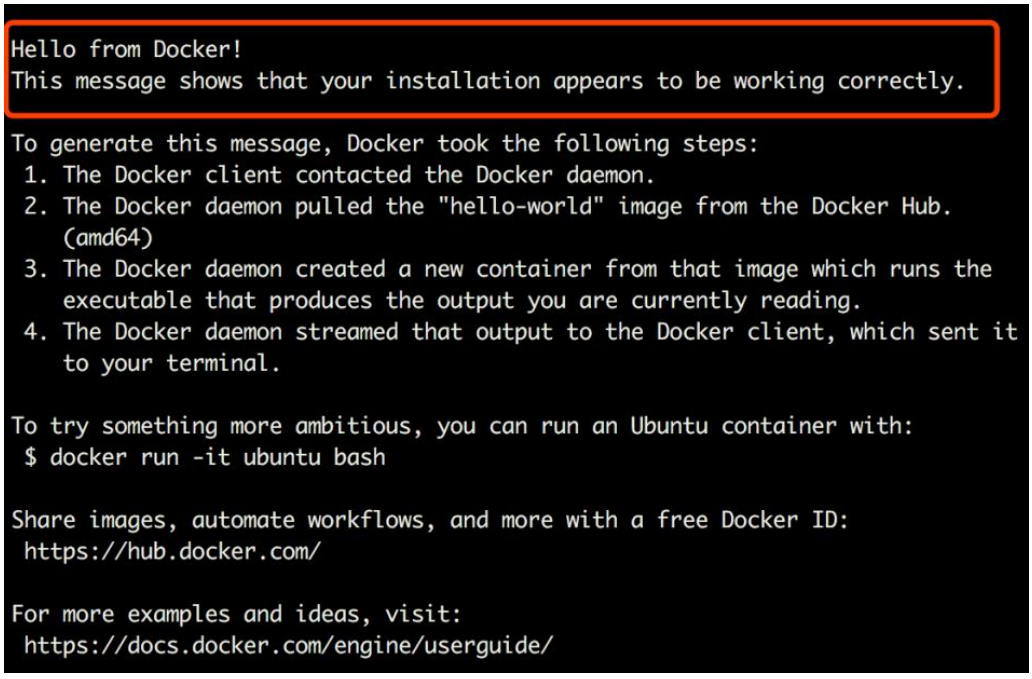
- Verify if Docker is installed correctly

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:



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

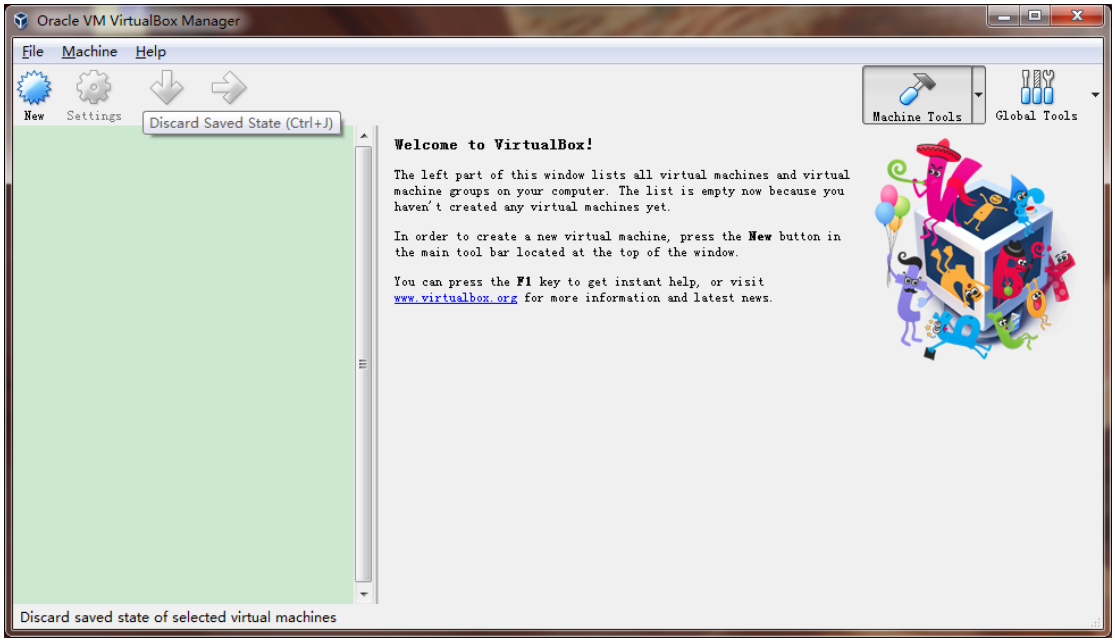
Download the OVA file from our website

$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:

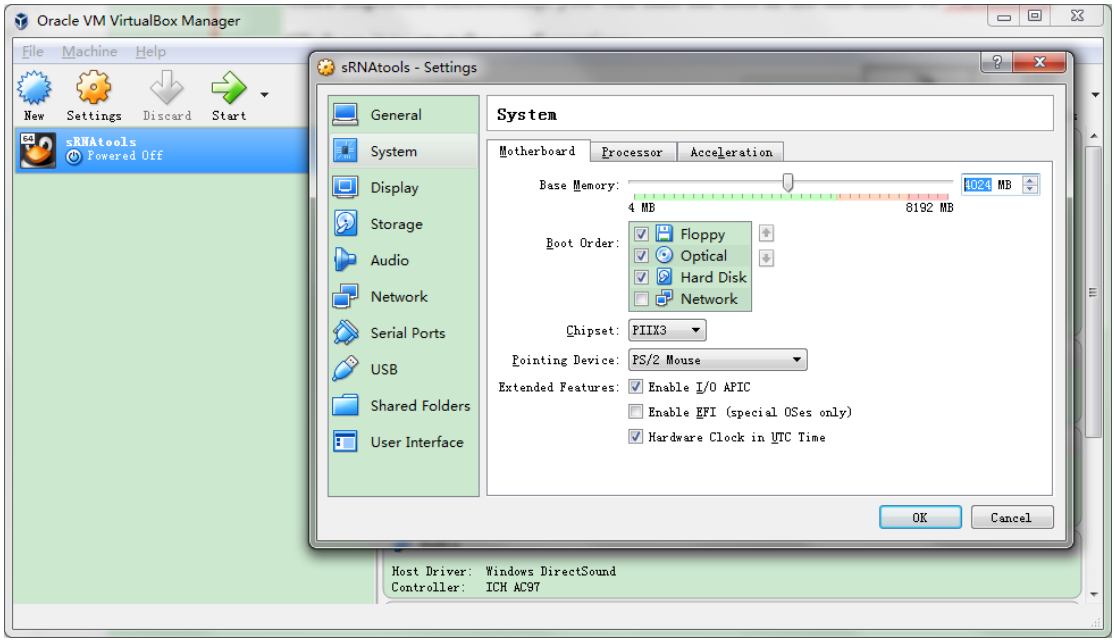
Change the memory allocation

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

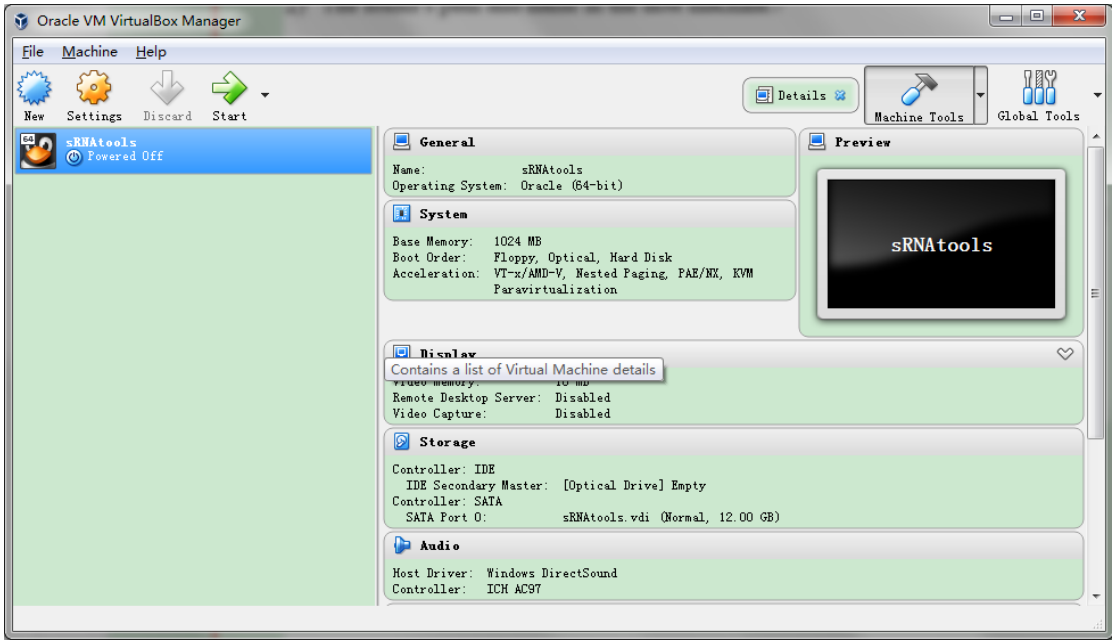
2.3.5 Start VirtualBox

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 version1.0 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 (>=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

	Homo sapiens (hg38)	*Mus musculus (mm10)*	*Rattus norvegicus (rn6)*	*Danio rerio (GRC11)*	*Pan troglodytes (Pan_tro 3.0)*	*Gorilla gorilla gorilla (gorGor4)*	*Gallus gallus (GRCg6a)*	*Sus scrofa (susScr11)*	*Macaca mulatta (rheMac8)*	*Drosophila melanogaster (Dm6)*	*Caenorhabditis elegans*	*Arabidopsis thaliana (TAIR10)*	*Oryza sativa (MSU7.0)*	*Zea mays (B73_RefGen_v4)*	*Glycine max (Wm82.a2)*	*Sorghum bicolor (NCBIv3)*	*Solanum lycopersicum (SL3.0)*	*Brachypodium distachyon (v3.0)*	*Medicago truncatula*	*Populus trichocarpa (Pop_tri_v3)*	*Vitis vinifera (Grapevine 12X)*
Reference genome sequences																					
Phytozome												√	√	√	√	√	√	√	√	√	√
Ensembl	√	√	√	√	√	√	√	√	√	√	√										
RNA libraries																					
NONCODE	√	√	√	√	√	√	√	√	√	√	√	√									
CANTATAdb 2.0												√	√	√	√	√	√	√	√	√	√
CIRCpedia v2.0	√	√	√	√						√	√										
circAtlas 2.0	√	√	√				√	√	√												
PlantcircBase												√	√	√	√		√				
RNAcentral NAT	√	√	√	√	√		√	√	√	√	√										
PlantNATsDB												√	√	√	√	√	√	√	√	√	√
piRBase	√	√	√	√			√	√	√	√	√										
piRNAclusterDB	√	√	√	√			√	√	√	√											
Pln24NT												√	√	√	√	√	√	√	√	√	√
miRBase	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
GtRNAdb	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Rfam	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√

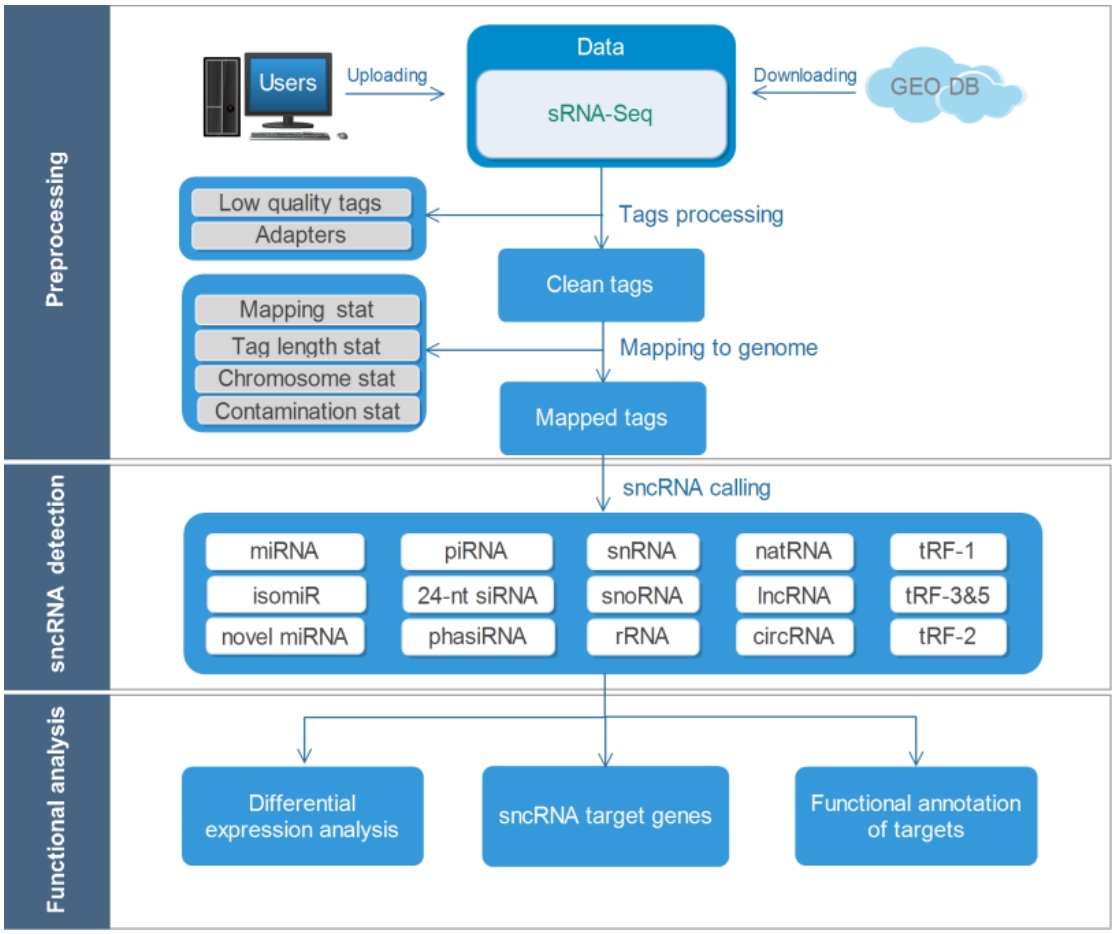
Dataset available for the 21 supported model species

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.



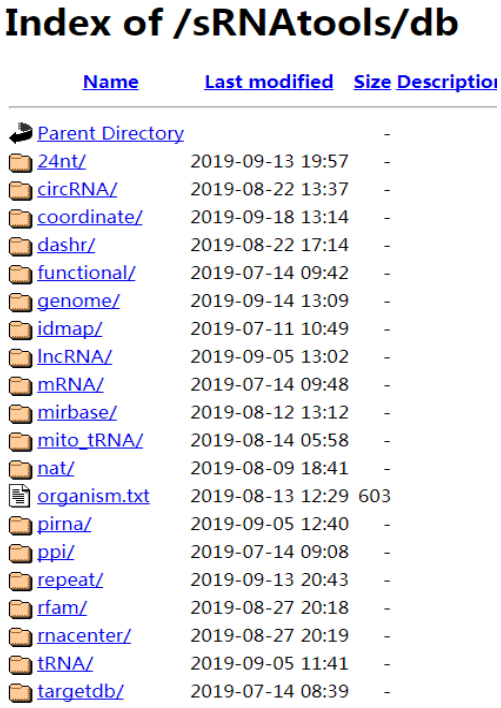
Workflow of sRNAtools

3.1 Launch Analysis Locally

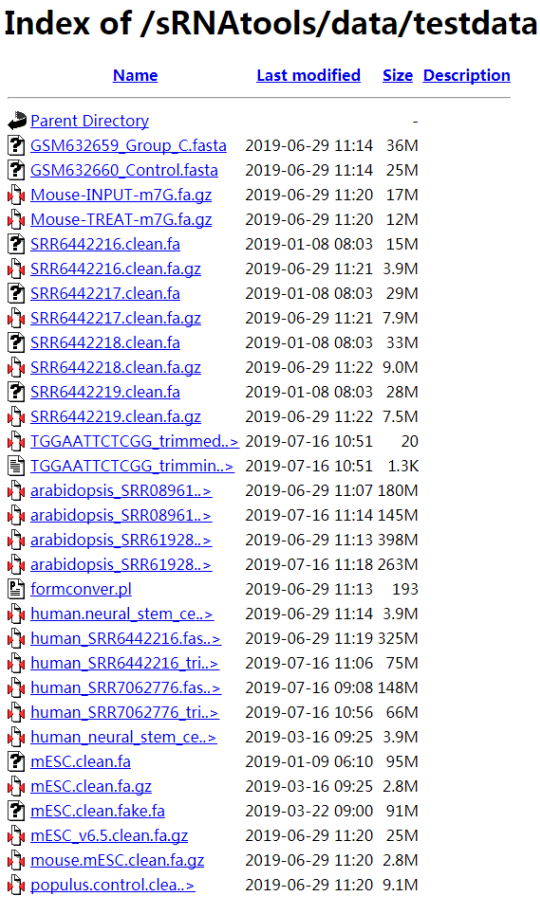
Assume that we have successfully installed sRNAtools Docker image, sRNAtools VirtualBox, and the locally_version_pipeline. Then we can launch analysis in our local server.

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):



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

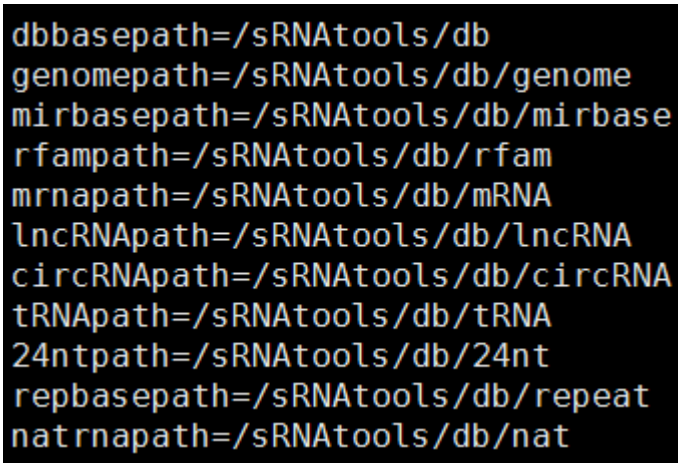
Name	Last modified	Size	Description
Parent Directory		-	
GSM632659_Group_C.fasta	2019-06-29 11:14	36M	
GSM632660_Control.fasta	2019-06-29 11:14	25M	
Mouse-INPUT-m7G.fa.gz	2019-06-29 11:20	17M	
Mouse-TREAT-m7G.fa.gz	2019-06-29 11:20	12M	
SRR6442216.clean.fa	2019-01-08 08:03	15M	
SRR6442216.clean.fa.gz	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' excutable 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)



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

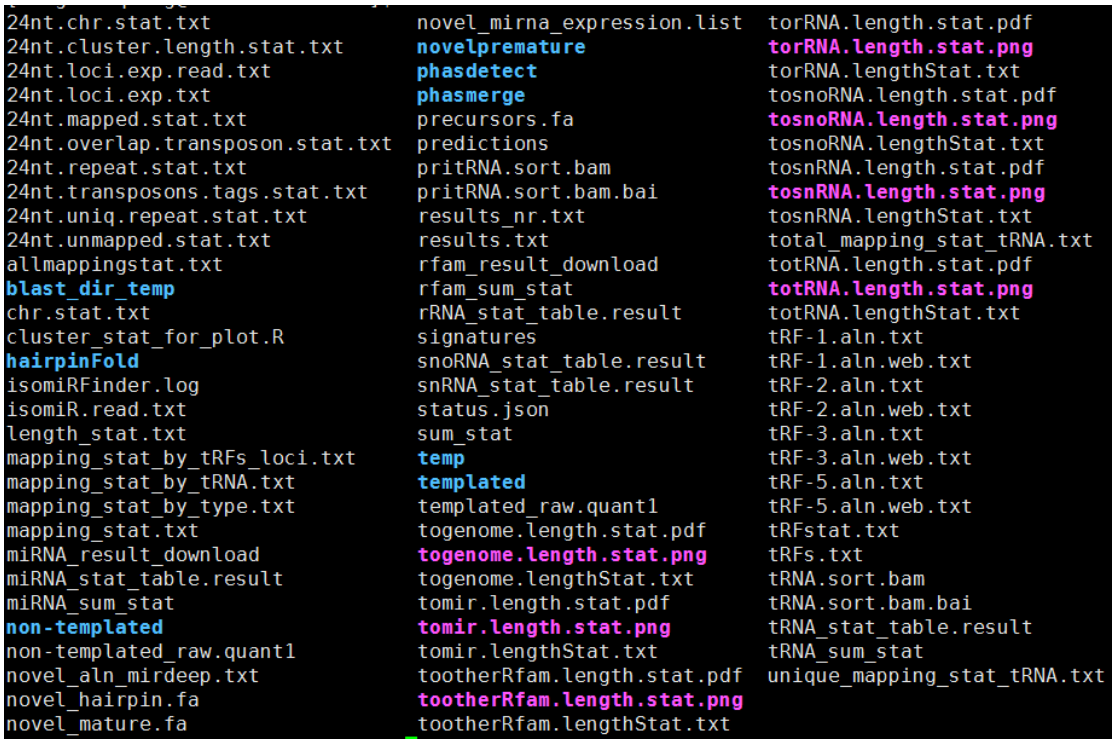
```
#### Fn: Settings Reader ####################
User Input runType              : G
User Input reference location   : /mount/db/genome/ptc.fa
User Input Libs                 : test_data/D-MH-CM-1.fa
user library format             : F
User Input for phase length     : 21,24
User Input index location       : /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/phasdetect/D-MH-CM-1.txt.A4xp9YjTlLz
# 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 2080380 alignments
Thu Dec 27 13:19:26 2018: aligning is done
./data/results/123456/phasdetect/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/phasdetect
#### 'phasmerge' can be run by command: python3 phasmerge -mode merge -dir ./data/results/123456/phasdetect
```

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, apping_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 ./phasdetect/. Here we select some result files to explain the analysis result.



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 pritRNA sequences’ and ‘mapping to the mature tRNA sequences’. The mapping informations printed to the screen (in the figure below) and the result files (sach 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:

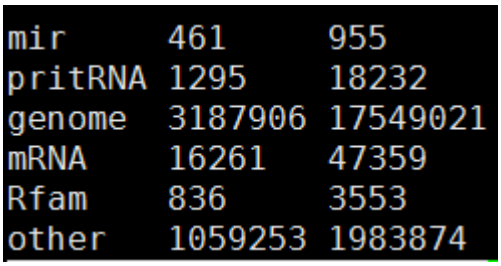
```
# reads processed: 4266012
# reads with at least one reported alignment: 3187906 (74.73%
# reads that failed to align: 1078106 (25.27%)
Reported 3187906 alignments
# reads processed: 4266012
# reads with at least one reported alignment: 27480 (0.64%)
# reads that failed to align: 4238532 (99.36%)
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

```
t00000153_x5937 0   scaffold_1.trna42-AlaAGC|1:44658543-44658715:+|1-73   1   255   18M   *   0   0   GGGGATGTAGCTCAGAGC   IIIIIIIIIIIIIIIII   XA:i:2  MD:Z:
16T0G0  NM:i:2
t00002075_x679  0   scaffold_1.trna42-AlaAGC|1:44658543-44658715:+|1-73   1   255   20M   *   0   0   GGGGATGTAGCTCAGAGGGG   IIIIIIIIIIIIIIIIIIII   XA:i:2  MD:Z:
16T2T0  NM:i:2
t00004202_x373  0   scaffold_1.trna42-AlaAGC|1:44658543-44658715:+|1-73   1   255   18M   *   0   0   GGGGATGTAGCTCAGTGG   IIIIIIIIIIIIIIIII   XA:i:2  MD:Z:
15A0T1  NM:i:2
t00012848_x116  0   scaffold_1.trna42-AlaAGC|1:44658543-44658715:+|1-73   1   255   19M   *   0   0   GGGGATGTAGCTCAGACGT   IIIIIIIIIIIIIIIIII   XA:i:2  MD:Z:
16T1G0  NM:i:2
t00014371_x101  0   scaffold_1.trna42-AlaAGC|1:44658543-44658715:+|1-73   1   255   19M   *   0   0   GGGGATGTAGCTCAGAGGA   IIIIIIIIIIIIIIIII   XA:i:2  MD:Z:
16T1G0  NM:i:2
t00037581_x29   0   scaffold_1.trna42-AlaAGC|1:44658543-44658715:+|1-73   1   255   19M   *   0   0   GGGGATGTAGCTCAGTGGG   IIIIIIIIIIIIIIIII   XA:i:2  MD:Z:
15A0T2  NM:i:2
t00041542_x26   0   scaffold_1.trna42-AlaAGC|1:44658543-44658715:+|1-73   1   255   23M   *   0   0   GGGGATATAGCTCAGTTGGTAGA   IIIIIIIIIIIIIIIIIIIIIII   XA:i:2  MD:Z:
6G8A7   NM:i:2
t00041940_x26   0   scaffold_1.trna42-AlaAGC|1:44658543-44658715:+|1-73   1   255   19M   *   0   0   GGGGATGTAGCTCAGGGGG   IIIIIIIIIIIIIIIII   XA:i:2  MD:Z:
15A0T2  NM:i:2
t00052273_x19   0   scaffold_1.trna42-AlaAGC|1:44658543-44658715:+|1-73   1   255   17M   *   0   0   GGGGATGTAGCTCAGTG   IIIIIIIIIIIIIIII   XA:i:2  MD:Z:
15A0T0  NM:i:2
t00052746_x19   0   scaffold_1.trna42-AlaAGC|1:44658543-44658715:+|1-73   1   255   18M   *   0   0   GGGGATGTAGCTCAGGGG   IIIIIIIIIIIIIIIII   XA:i:2  MD:Z:
15A0T1  NM:i:2
```

Alignments result bam files

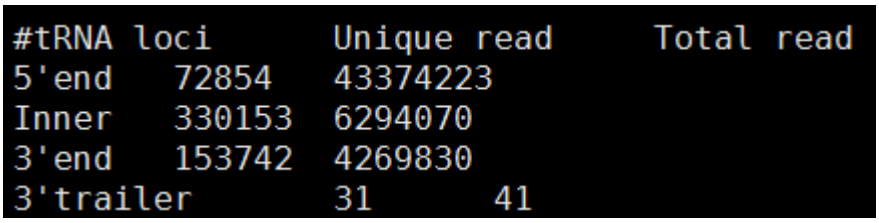
The result file allmappingstat.txt contains information of mapping statistics on different types of RNAs, including miRNA, pritRNA, genome, other RFAM ncRNA, mRNA, lncRNA,and other mapping. Statistics are based on unique tags (second column) and total reads count (last column).



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

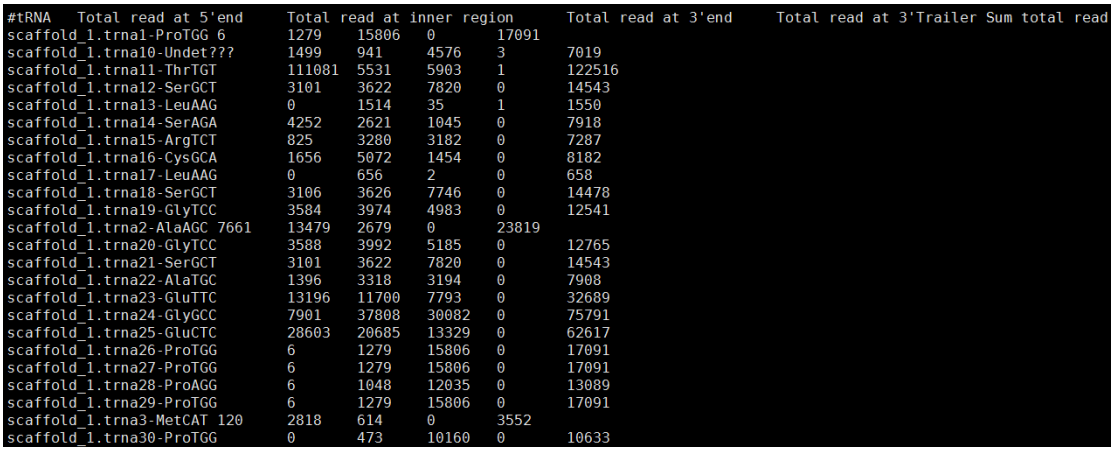
allmappingstat.txt

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:



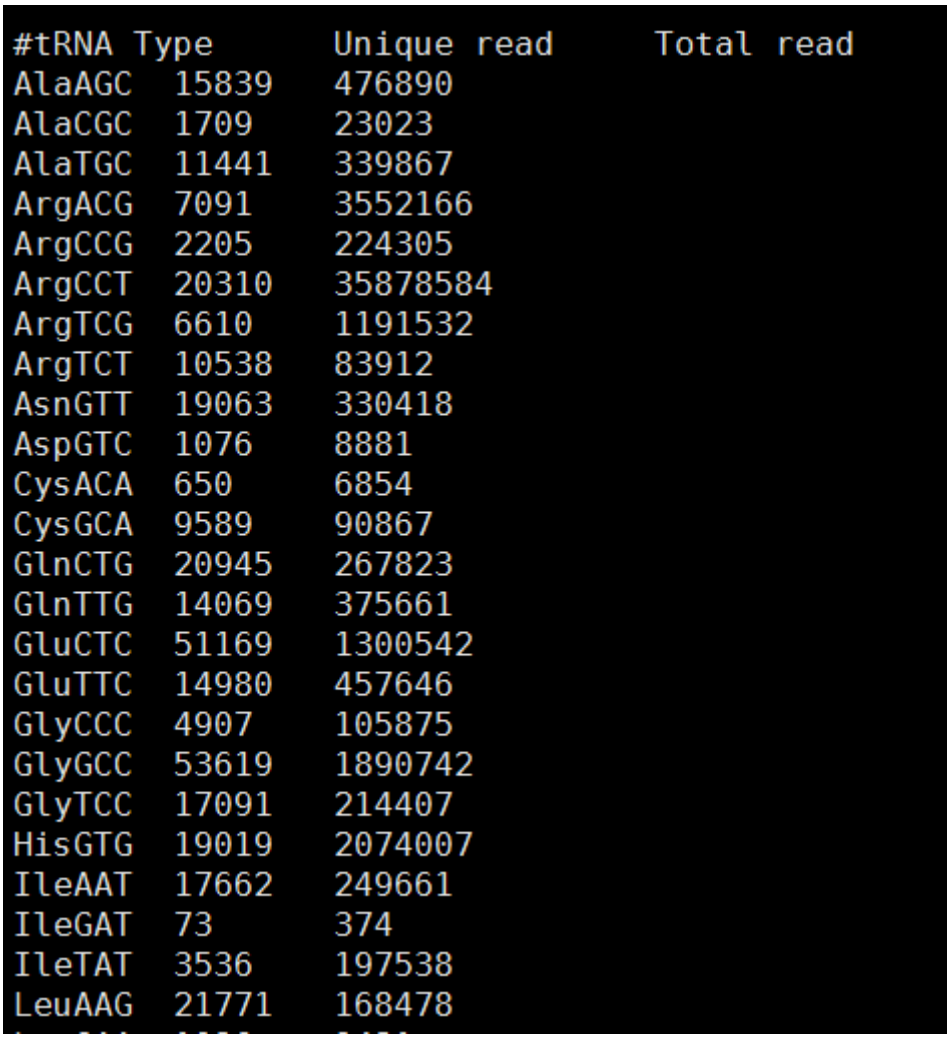
#tRNA loci	Unique read	Total read
5'end	72854	43374223
Inner	330153	6294070
3'end	153742	4269830
3'trailer	31	41

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	CGTGGAGACCTGGGCTC	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	GGCCGTGGAGACCTGGGCT	0.00	0.791698546289212
tRF-2	scaffold_8.trna6-HisGTG	20	36-55	3402	1380.60169	GGCCGTGGAGACCTGGGCTC	0.00	0.75751503006012
tRF-5	scaffold_1.trna66-LeuCAA	23	1-23	87	35.30639	GCCTTGGTGGTGAAATGGTAGAC	0.00	0.813084112149533
tRF-5	scaffold_15.trna5-IleGAT	19	1-19	80	32.46565	CCGACCTTAGCTCAGTGGC	0.00	0.571428571428571
tRF-3	scaffold_15.trna5-IleGAT	17	59-75	186	75.48263	TCCGGCAGGTCGGACCA	0.00	0.978947368421053
tRF-3	scaffold_14.trna12-ArgTCT	17	60-76	376	152.58855	TCCCACCGGGCGTGCCA	0.00	0.397883597883598
tRF-2	scaffold_14.trna12-ArgTCT	21	36-56	133	53.97414	TAATCAAGCGATTGTGGGTTC	0.00	0.685567010309278
tRF-2	scaffold_14.trna12-ArgTCT	17	40-56	112	45.45191	CAGGCGACTGTGGGTTC	0.00	0.7
tRF-5	scaffold_4.trna32-GlnTTG	23	1-23	1092	443.15610	GGTTCTATGGTGTAGTGGTTAGC	0.00	0.420161600615621
tRF-2	scaffold_16.trna16-MetCAT	19	38-56	50	20.29103	TTGGCGAATTCGTAGGTTC	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	GGTTCCATGGTCTAGCGG	0.00	0.306825297432686
tRF-2	scaffold_12.trna2-AlaCGC	21	36-56	1932	784.04541	CTTGCGAGAGGCACGGGGTTC	0.00	0.808030112923463
tRF-2	scaffold_12.trna2-AlaCGC	21	36-56	418	169.63301	CGTGCGAGAGGCACGGGGTTC	0.00	0.910675381263617
tRF-2	scaffold_12.trna2-AlaCGC	20	37-56	130	52.75668	TTGCGAGAGGCACGGGGTTC	0.00	0.474452554744526
tRF-2	scaffold_12.trna2-AlaCGC	20	36-55	123	49.91593	CTTGCGAGAGGCACGGGGTT	0.00	0.75
tRF-2	scaffold_12.trna2-AlaCGC	20	37-56	104	42.20534	CTGCGAGAGGCACGGGGTTC	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: /sncrnabench/db/mirna/ptc_pri.fa
  miRNA sequence file: /sncrnabench/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

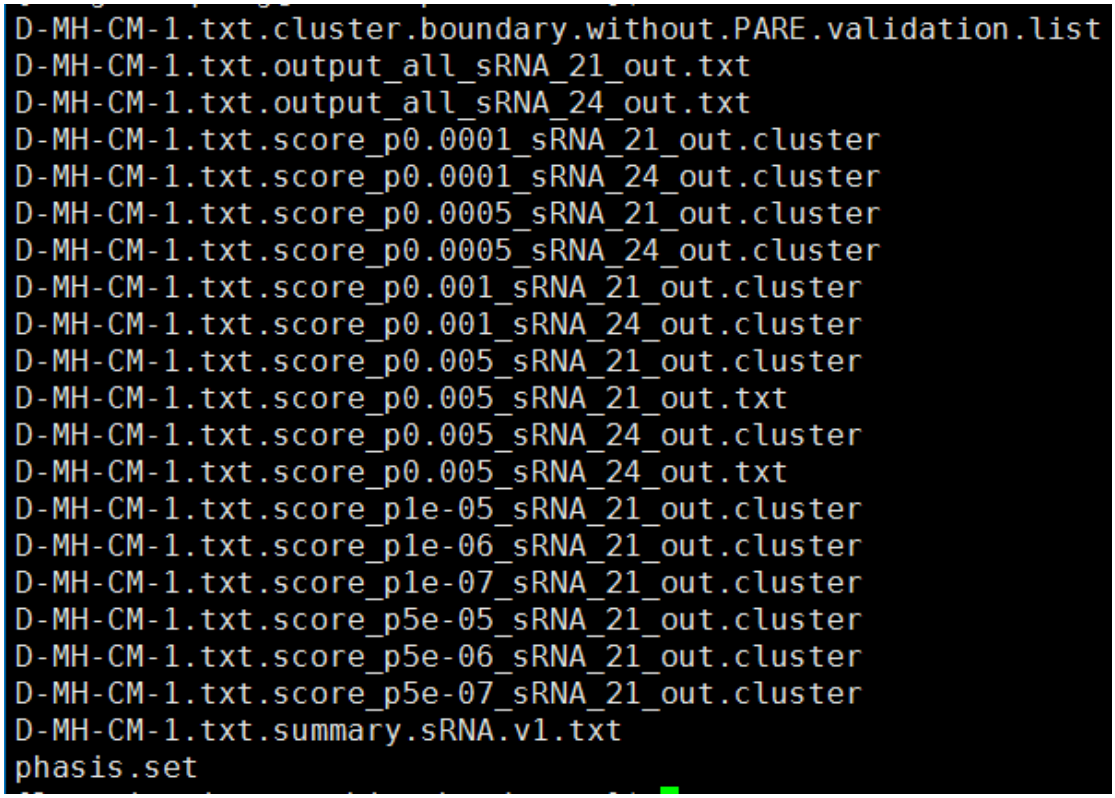
This step we take PHASIS suite for phased clusters discovery, comparison, annotation and to identify miRNA triggers. PHASIS consists of the following three components:

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.

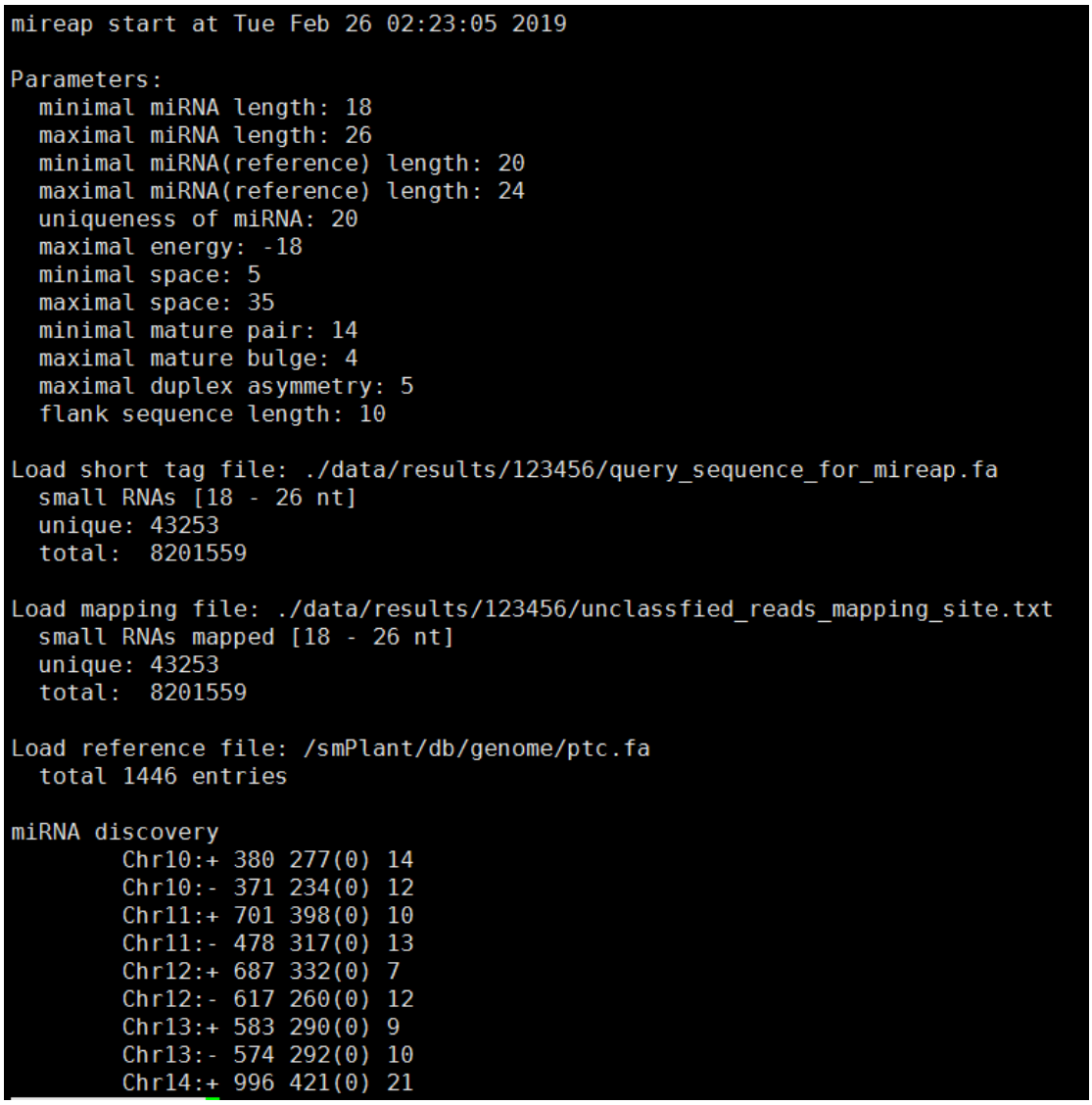
The analysis will create three directories which named phasdetect, phasmerge and phastrigs respectively, the results files as shown in the figure below:



Phasing analysis result files

3.1.4.5 Novel miRNAs Results

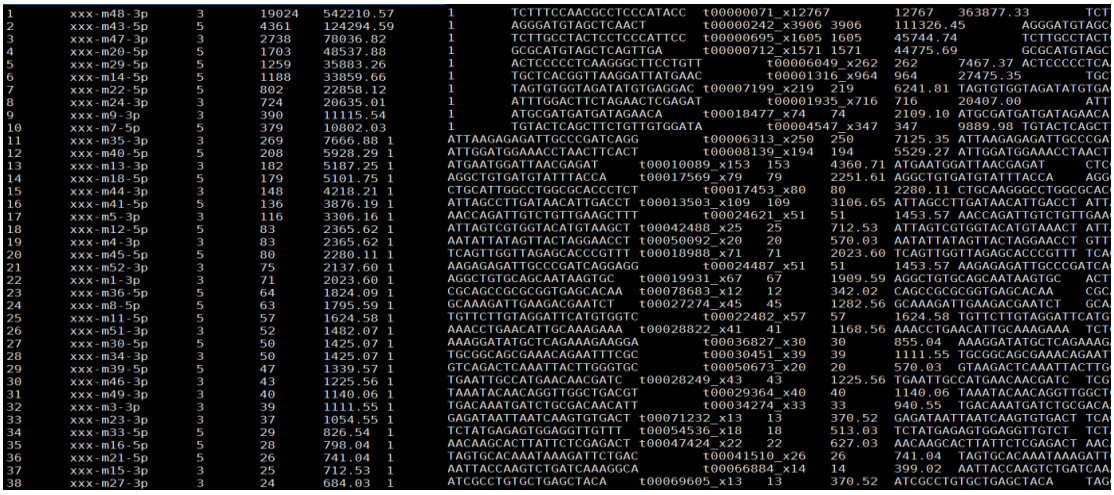
Mireap novel miRNA detect and expression analysis were performed in this step, and there have some intermediate files such as novel_hairpin.fa, novel_mature.fa, mireap-xxx.aln, mireap-xxx.gff, and mireap-xxx.log, the log file which contains the information of alignments result, the information of annotation 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:  8201559

Load mapping file: ./data/results/123456/unclassfied_reads_mapping_site.txt
  small RNAs mapped [18 - 26 nt]
  unique: 43253
  total:  8201559

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

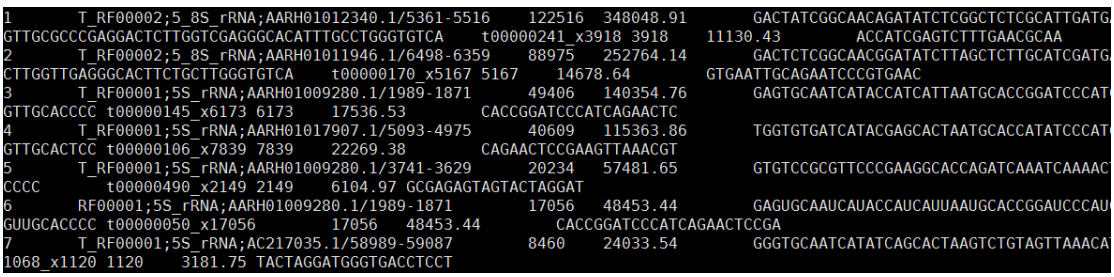
The final result file which named novel_mirna_expression.list. All unclassified reads were considered for detecting candidate novel miRNA genes. Sequence of predicted putative miRNA and miRNA star along with the corresponding tag number, tag count and hairpin structure are provided. The file contains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



novel_mirna_expression.list

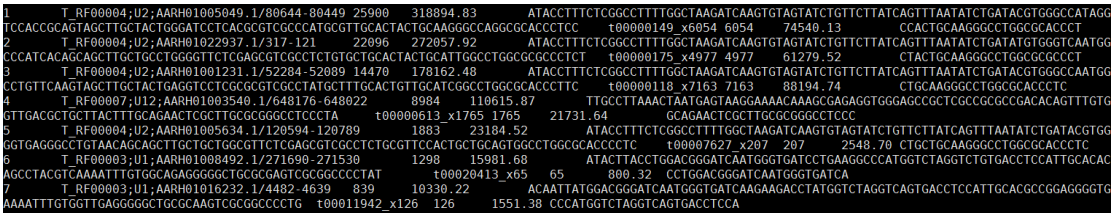
3.1.4.6 Expression Results

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 sunch 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.



rRNA_stat_table.result

snoRNA_stat_table.result



snRNA_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	TAGAACCCACAAAGGGTGTTGGTC	1746	99.49	NUM
t00009025_x174	AAGTTGGGGGCTCGAAGACGATCA	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	AGCTTGAACGTCTCTCTTGGCAAC	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

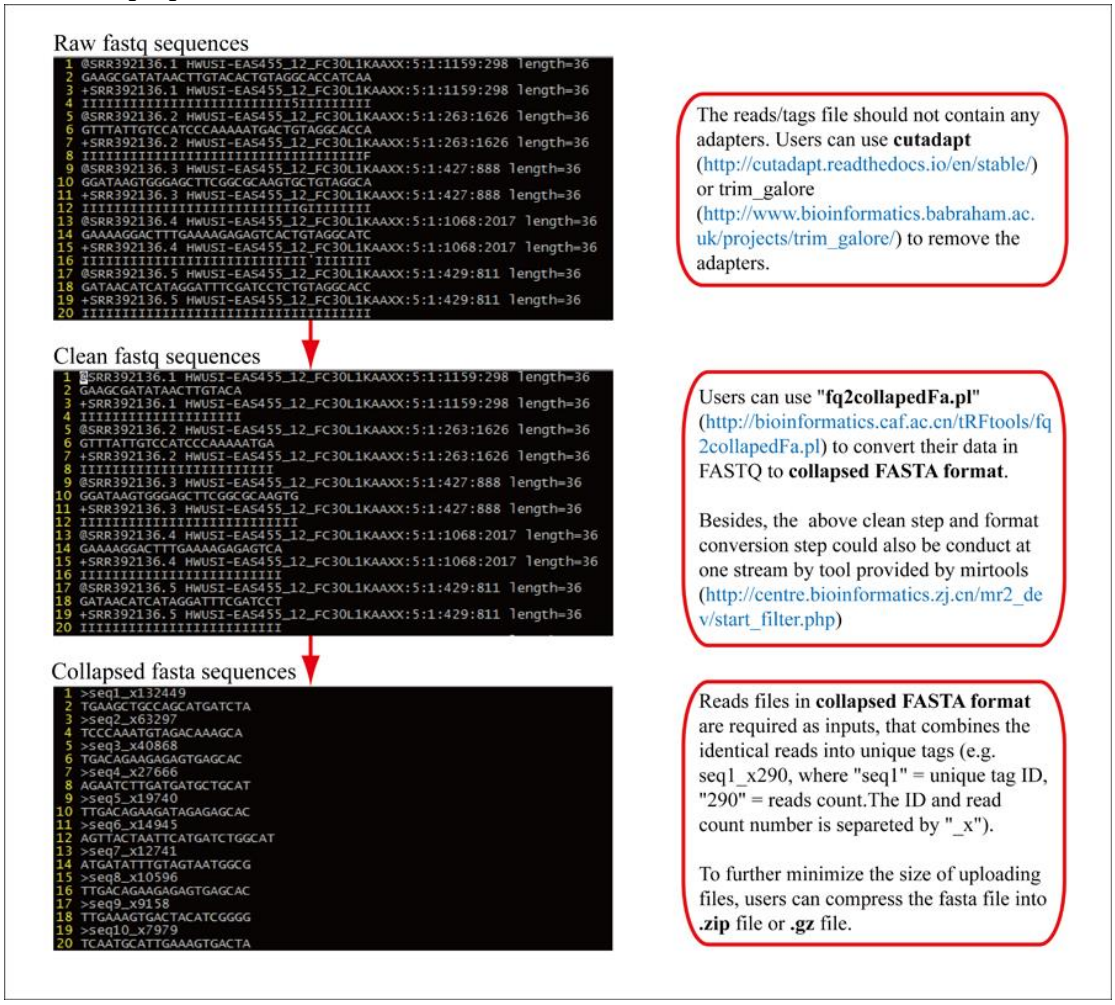
```
748625  439398
```

24nt.unmapped.stat.txt

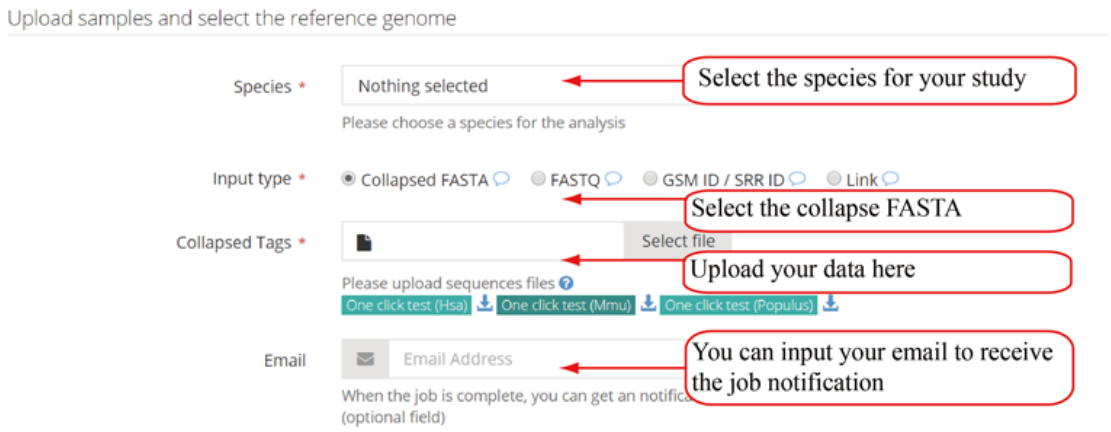
3.2 Launch Analysis Use Website

3.2.1 Prepare the inputs

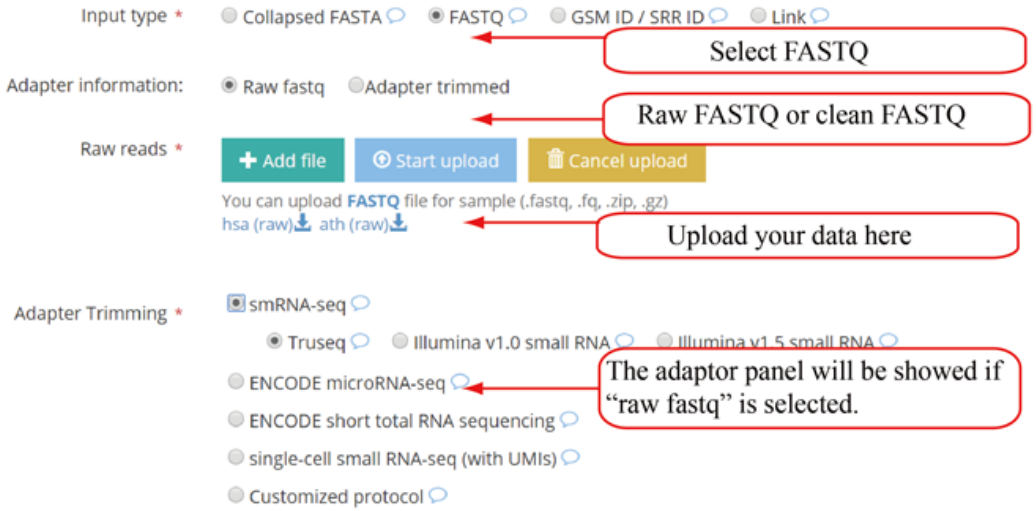
Data Preprocessing: We can do the following steps which below the figure in the red border to preprocess the data.



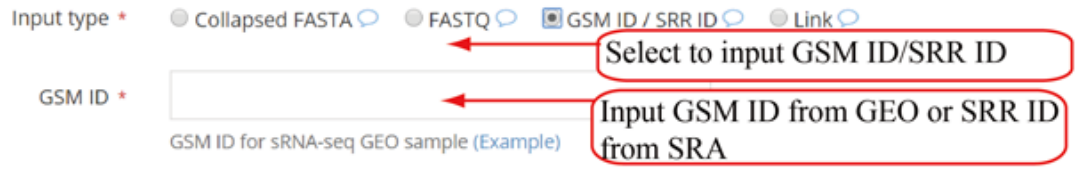
Inputs data of collapsed FASTA



Inputs data of FASTQ

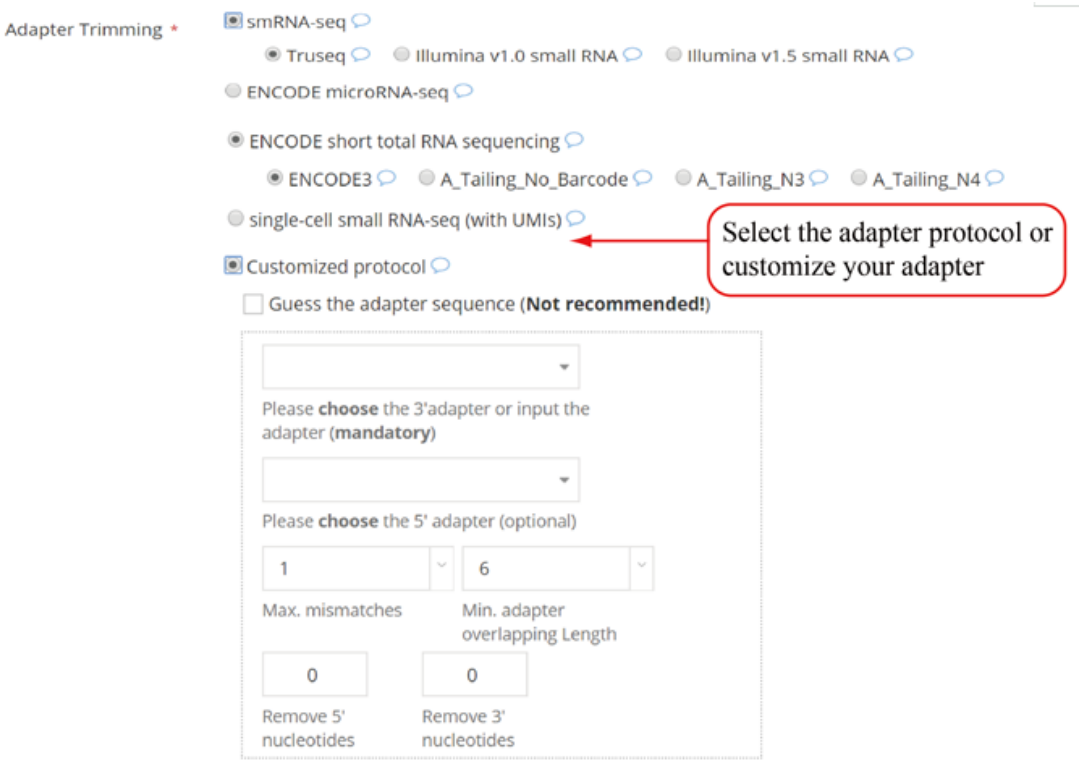


Inputs of GSM ID/SRR ID



Inputs of data link

Adapter parameters



3.2.2 Analysis how-to

Parameters for single case study:

Common parameters

Length of interval (bp) *
18
Shortest tag Length
45
Largest tag Length

Set the length range of tags used to perfom the analysis

Allowed mismatch in tags mapping *
1
Mismatch 0-3

Maximum mismatch in tags mapping

Maximum of tag mapping times *
20
Default mapping time threshold for multiple mapping

Maximum times of tag multiple mapping

miRNA analysis parameters

MEGABLAST Parameters
0.01 E value
5 B value
5 V value

Parameters of megablast used to align read to miRNA

Novel miRNA prediction tool
miRDeep miReap

miReap L value *
100
Flank sequence length of the query. Range 0 - 200

Choose from miReap and miRDeep to predict novel miRNA

Maximum of tag mapping times *
10
Maximum of mapping times for for miRNA

Maximum times of tag multiple mapping for miRNA

tRFs analysis parameters

Tag abundance cut-off *
20
Min. tag abundance (Read Rer Million)

Set tag minimum abundance to call high confident tRFs

P-value cut-off *
0.01
P-value inferred based on Binomial statistics method to distinguish random fragments

Set the p-value when call tRFs

Maximum of tag mapping times *
30
m: Maximum of mapping times for tRNA&tRF

Maximum times of tag multiple mapping for tRNA and tRF

piRNA/24nt siRNA analysis parameters

Minimum depth *
10
Minimum read depth of 24-nt siRNAs to call clusters.

Cluster padding distance *
150
The neighbor clusters with distance smaller than this value will be merged. With step: 50

Parameter to call piRNA/24nt siRNAs producing loci

Maximum of tag mapping times *
50
Maximum of mapping times for piRNA/24nt siRNA tags

Maximum times of tag multiple mapping for piRNA/24nt-siRNA

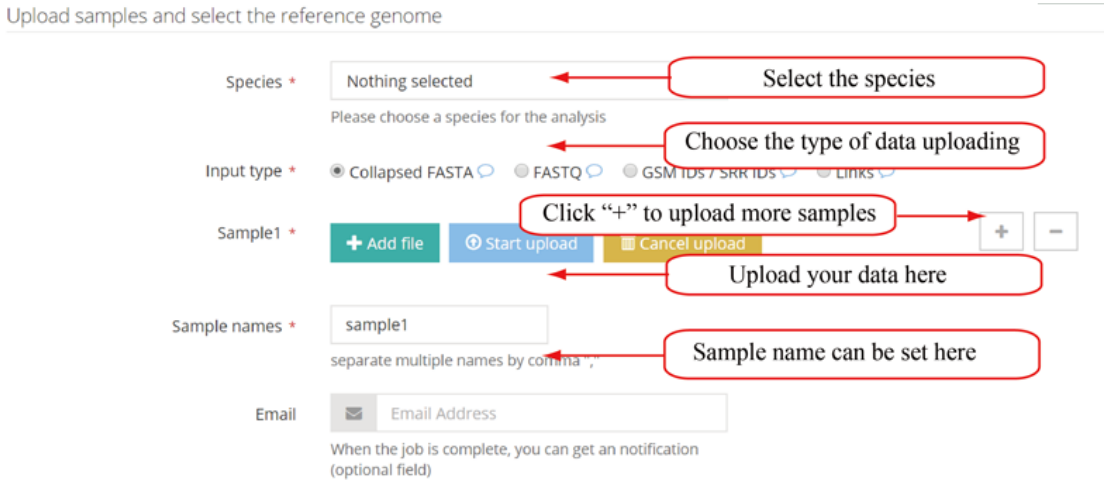
PhasiRNA analysis parameters

Phased length *
21,24
Periodicitic length of phased siRNAs (bp)

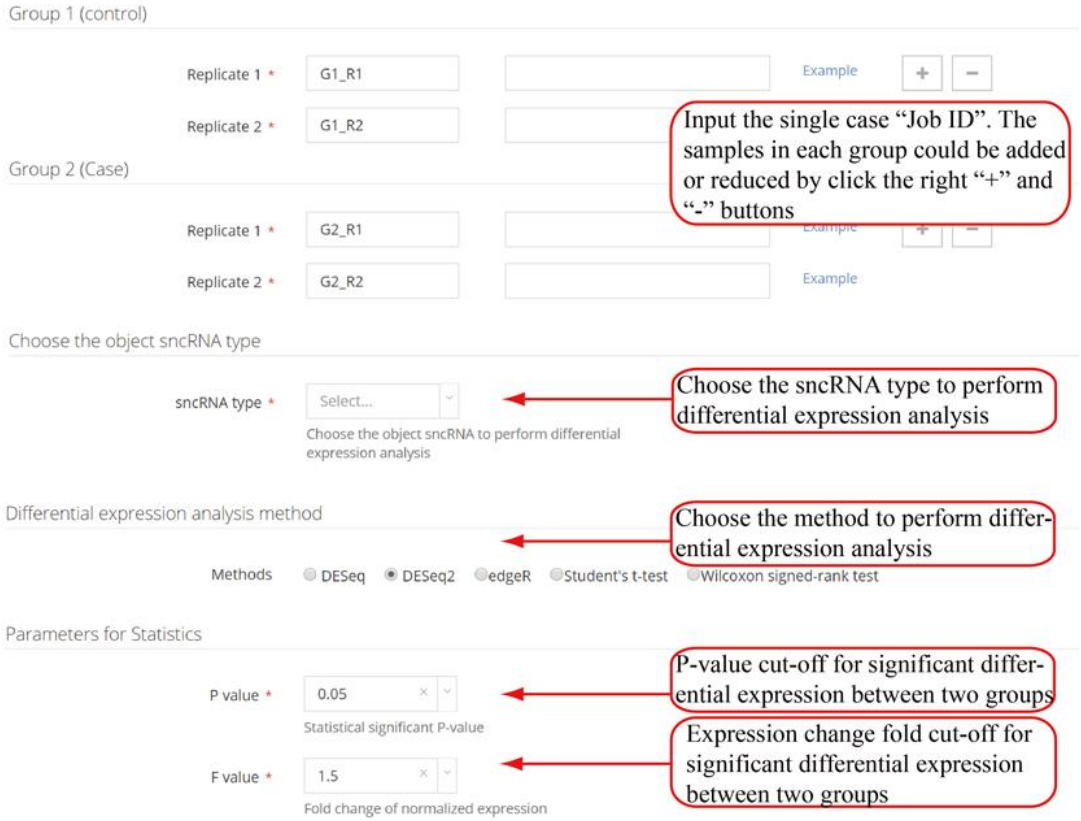
Minimum depth *
3
Minimum read depth for p-value computations.

PhasiRNA and loci calling parameters

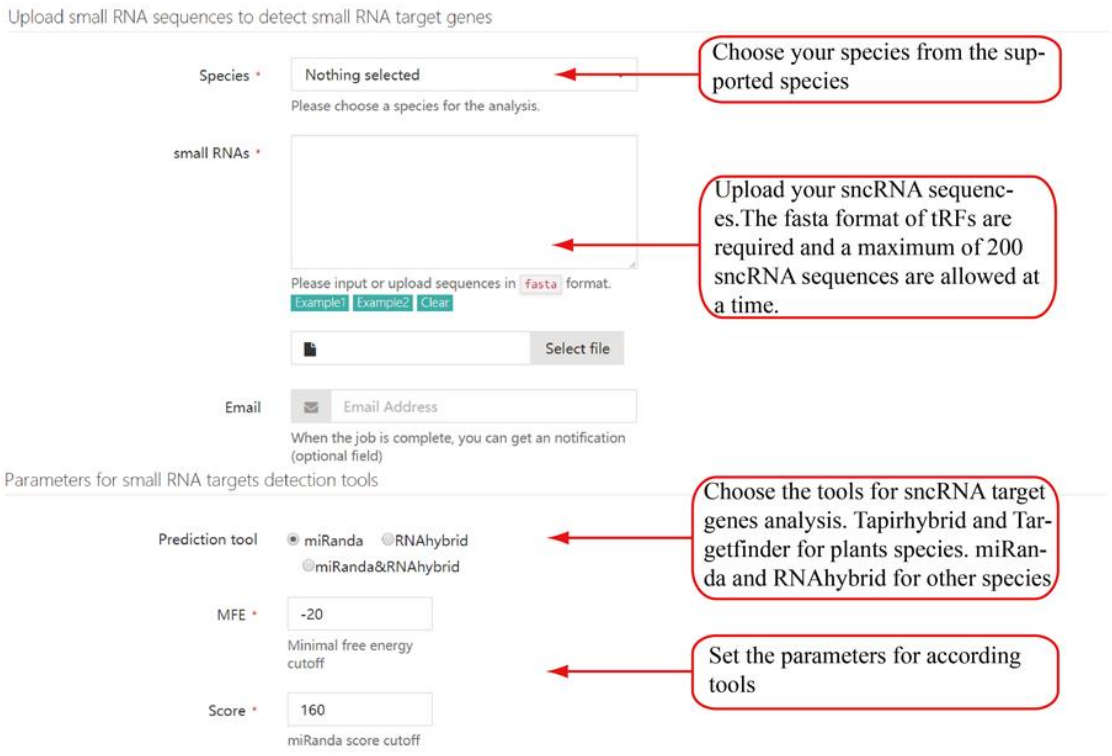
Parameters for batch case study



Parameters for group case study



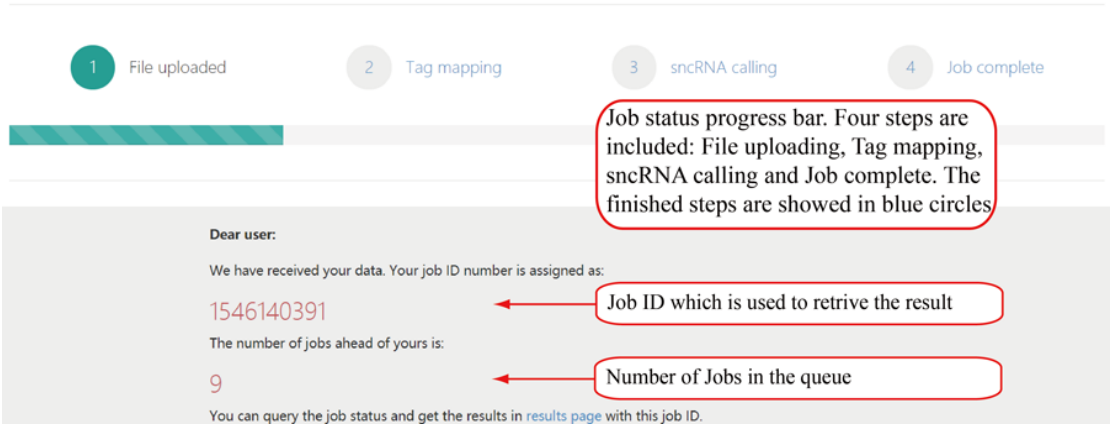
Parameters for sncRNA targets gene analysis



Parameters for sncRNA modification analysis

3.2.3 Retrieve the results

Query the job status and retrieve the results

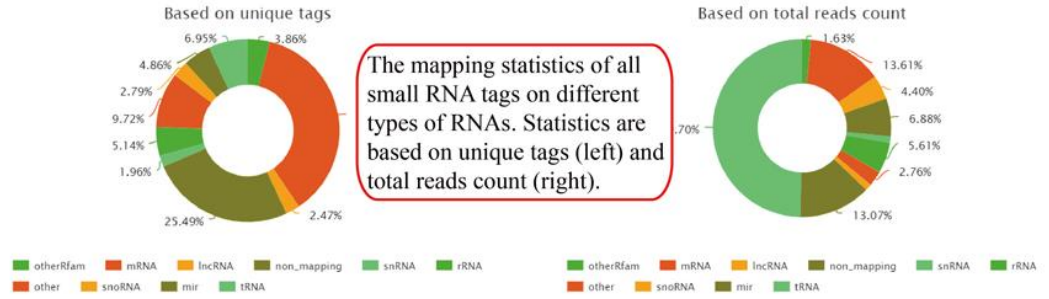


Query the job status

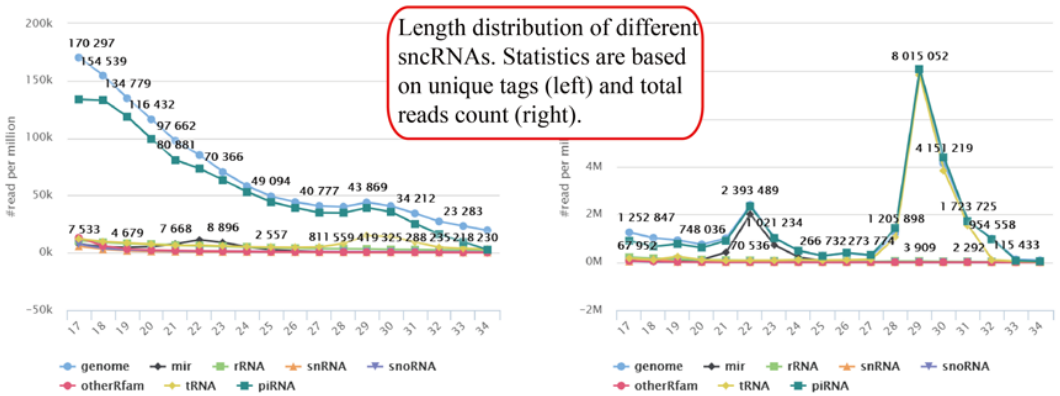
3.2.4 Explain the results

Sequence contamination evaluation by miRTrace

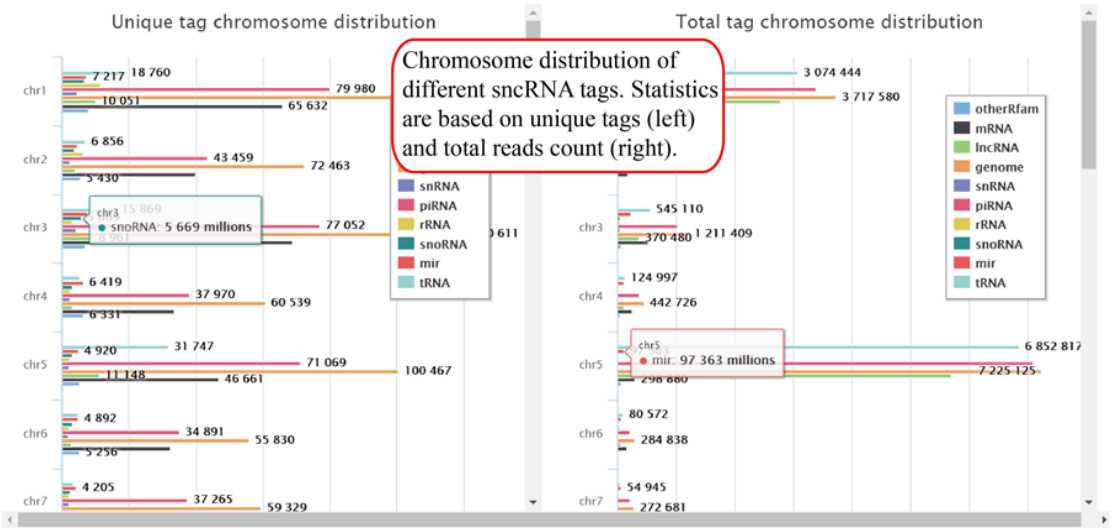
Tags mapping statistics on different types of RNAs, including miRNA, tRNA, rRNA, snRNA, snoRNA, other RFAM ncRNA, mRNA, lncRNA, other mapping



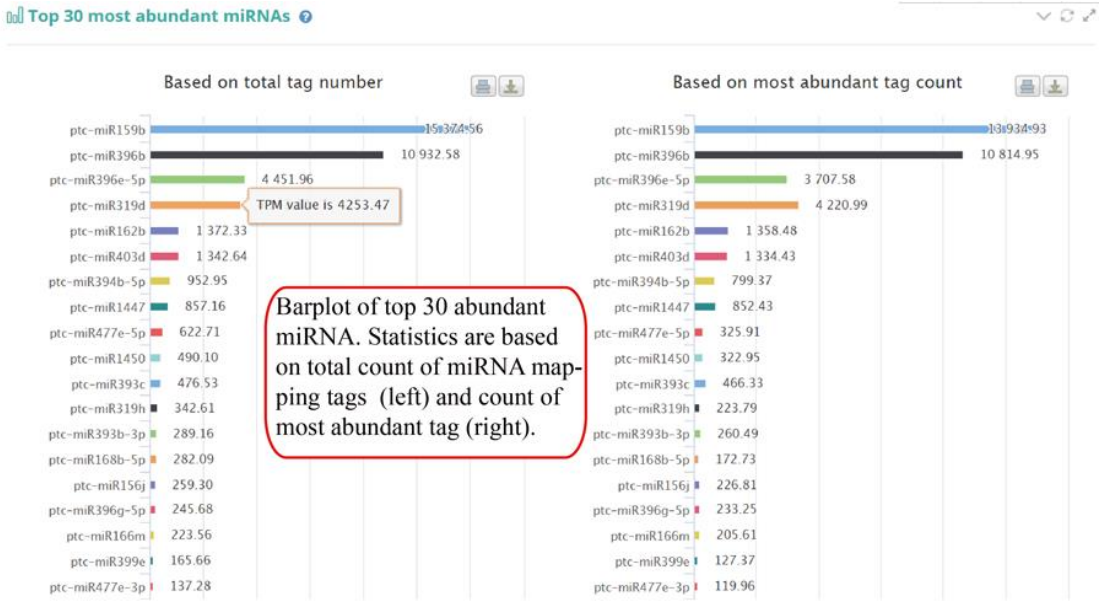
Tag length distribution of different sncRNAs. Statistics are based on unique tags and total tag counts



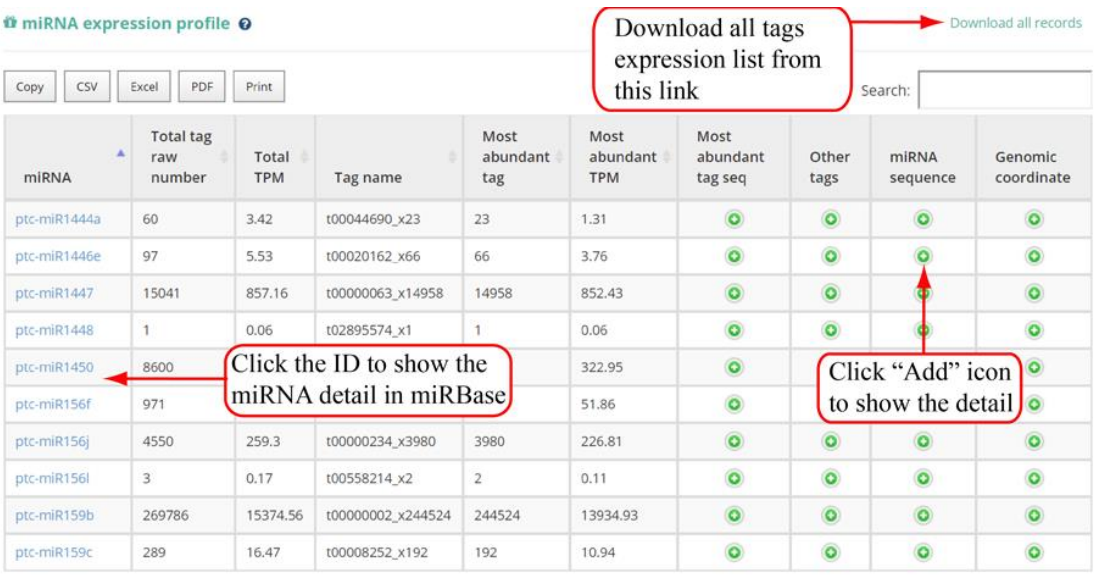
Tag length distribution of different sncRNAs. Statistics are based on unique tags and total tag counts



Top 30 most abundant miRNAs

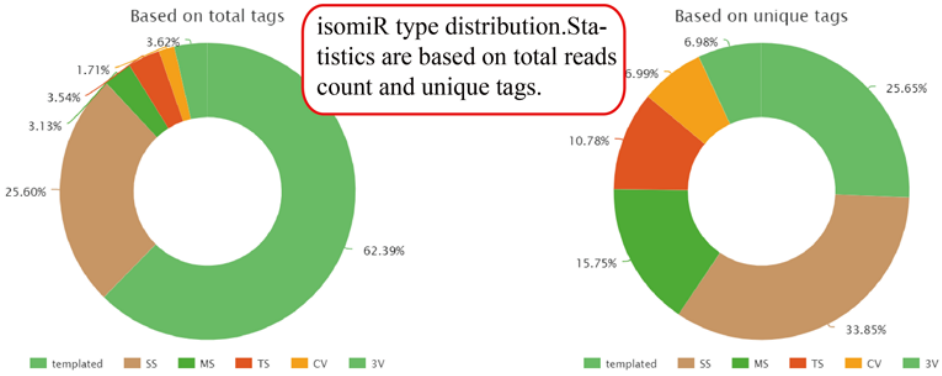


miRNA expression list

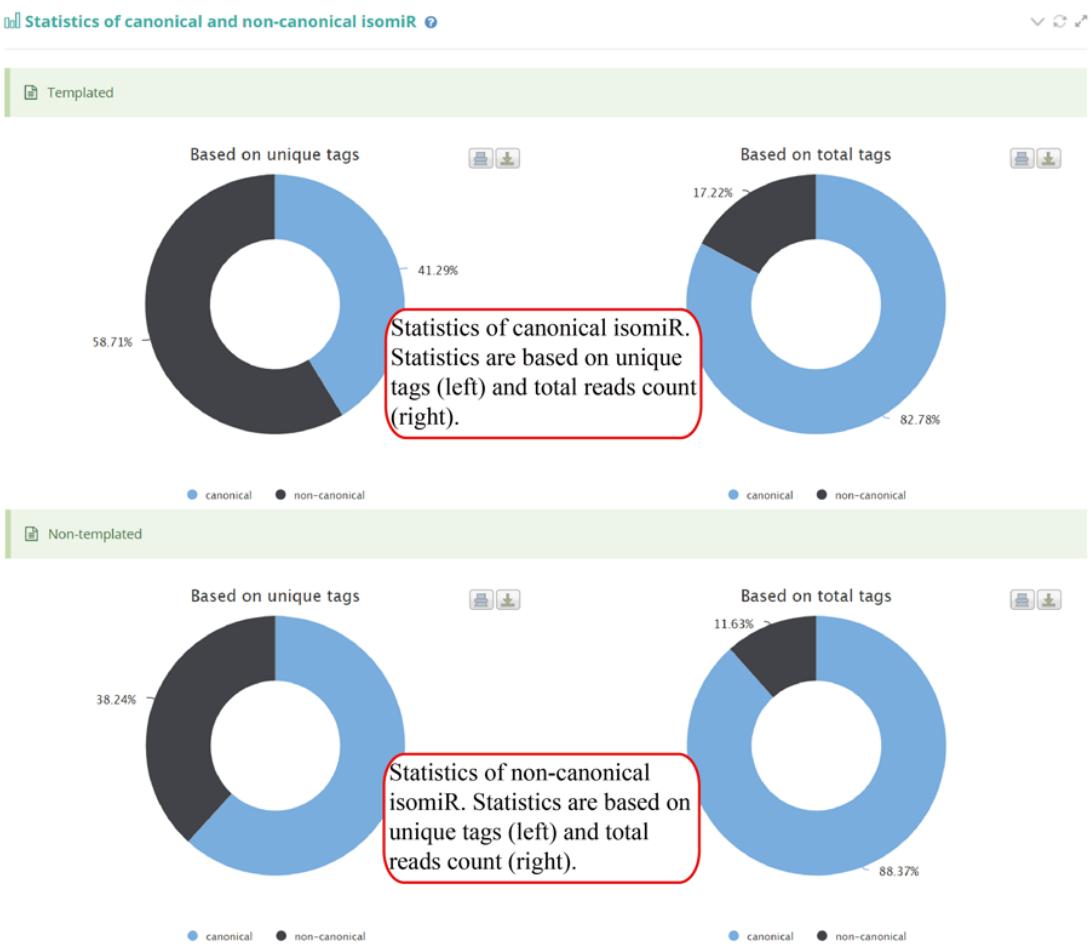


miRNA	Total tag raw number	Total TPM	Tag name	Most abundant tag	Most abundant TPM	Most abundant tag seq	Other tags	miRNA sequence	Genomic coordinate
ptc-miR1444a	60	3.42	t00044690_x23	23	1.31				
ptc-miR1446e	97	5.53	t00020162_x66	66	3.76				
ptc-miR1447	15041	857.16	t00000063_x14958	14958	852.43				
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				

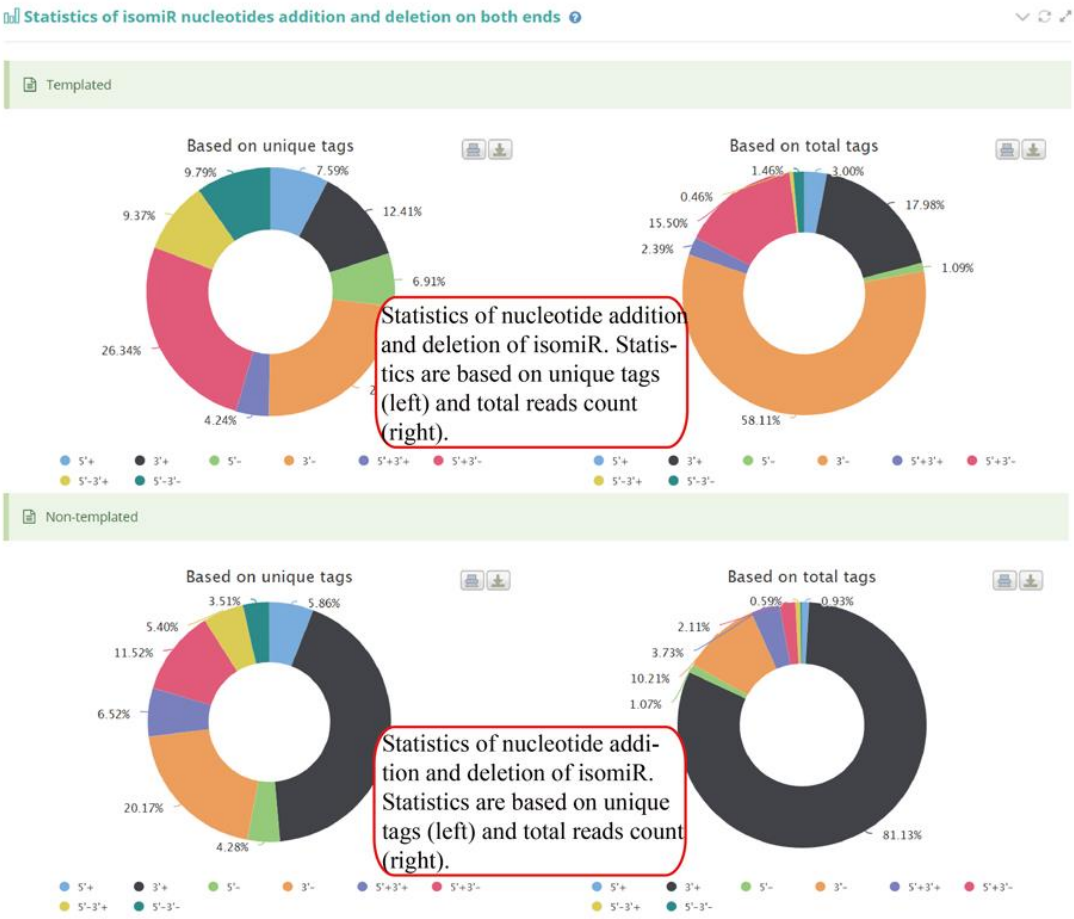
Tag distribution of isomiR, including templated isomiR and non-templated isomiR (SS, MS, TS, CV and 3V)



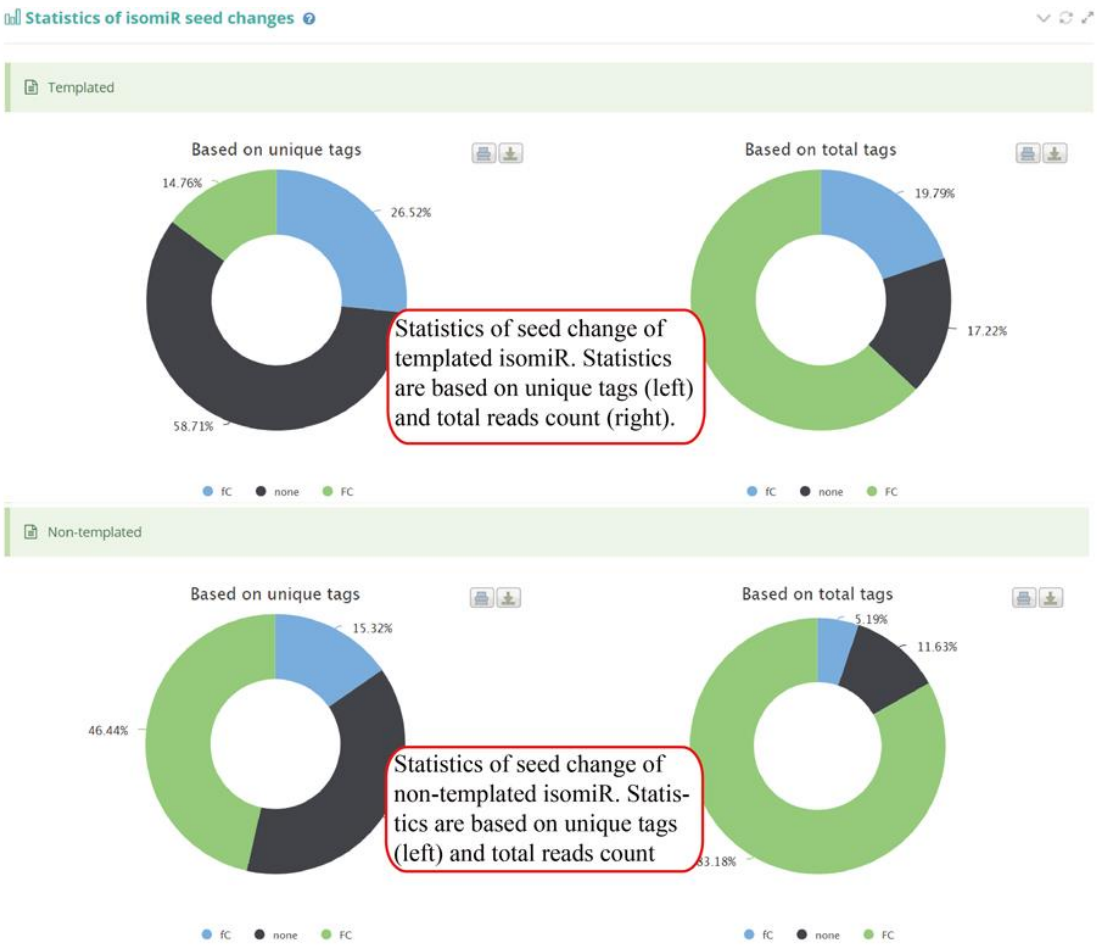
Statistics of canonical and non-canonical isomiR



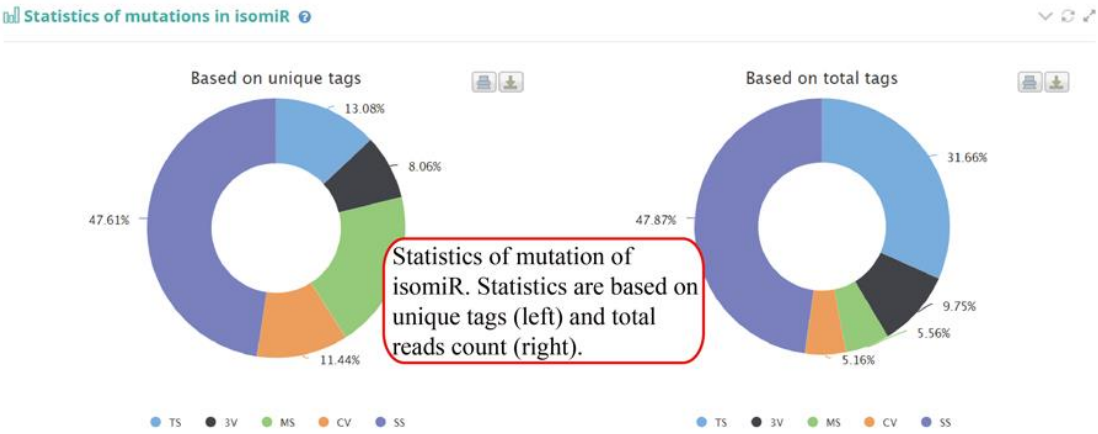
Statistics of isomiR nucleotides addition and deletion on both ends



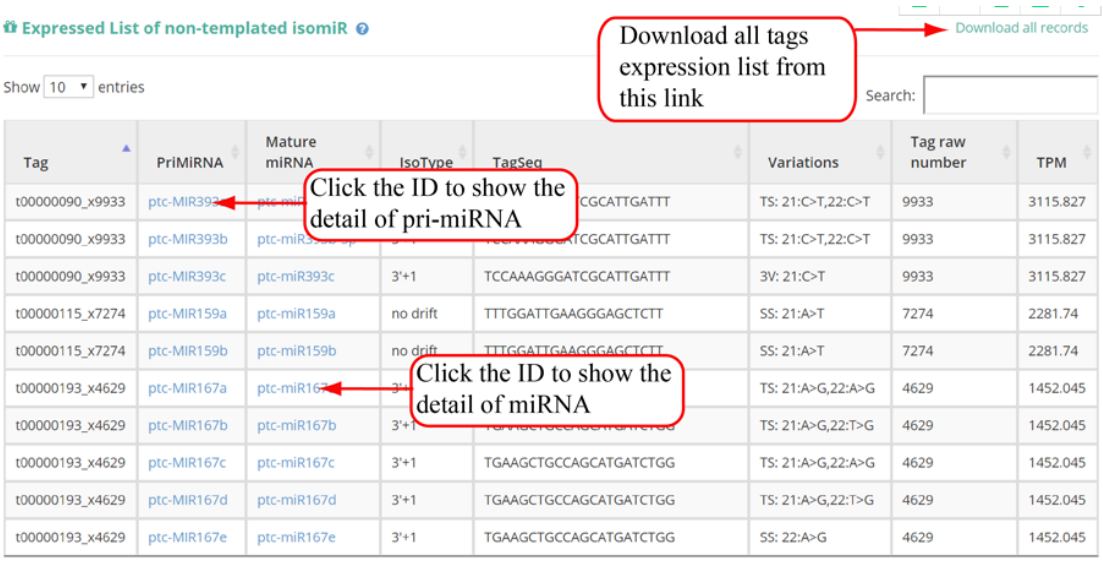
Statistics of isomiR seed changes



Statistics of mutations in isomiR

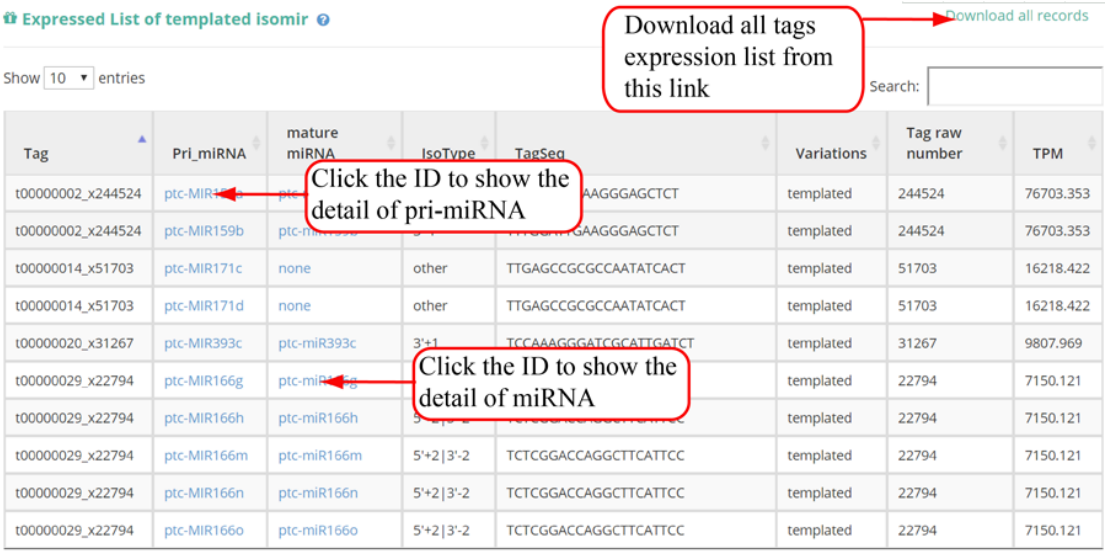


Non-templated isomiR list



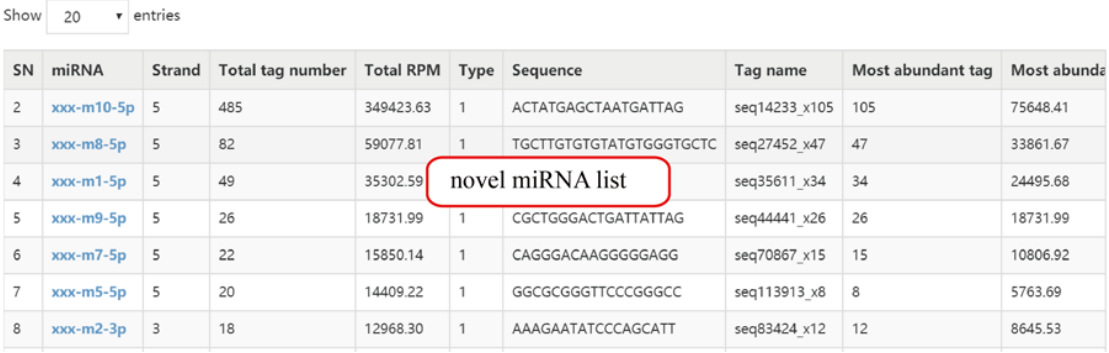
Tag	PriMiRNA	Mature miRNA	IsoType	TagSeq	Variations	Tag raw number	TPM
t00000090_x9933	ptc-MIR393	ptc-miR		CGCATTGATTT	TS: 21:C>T,22:C>T	9933	3115.827
t00000090_x9933	ptc-MIR393b	ptc-miR393b-5p		TCGCATTGATTT	TS: 21:C>T,22:C>T	9933	3115.827
t00000090_x9933	ptc-MIR393c	ptc-miR393c	3'+1	TCCAAAGGGATCGCATTGATTT	3V: 21:C>T	9933	3115.827
t00000115_x7274	ptc-MIR159a	ptc-miR159a	no drift	TTTGGATTGAAGGGAGCTCTT	SS: 21:A>T	7274	2281.74
t00000115_x7274	ptc-MIR159b	ptc-miR159b	no drift	TTTGGATTGAAGGGAGCTCTT	SS: 21:A>T	7274	2281.74
t00000193_x4629	ptc-MIR167a	ptc-miR167	3'+1		TS: 21:A>G,22:A>G	4629	1452.045
t00000193_x4629	ptc-MIR167b	ptc-miR167b	3'+1		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

Templated isomiR list



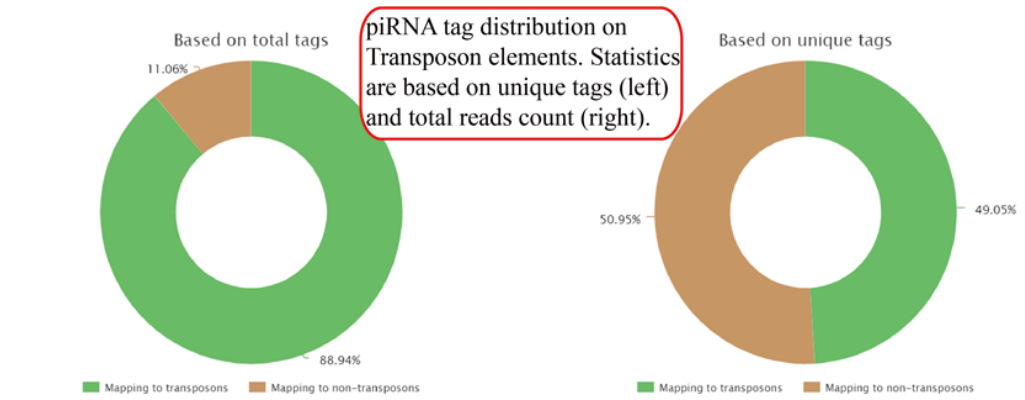
Tag	Pri_miRNA	mature miRNA	IsoType	TagSeq	Variations	Tag raw number	TPM
t00000002_x244524	ptc-MIR159a	ptc-miR		AAGGGAGCTCT	templated	244524	76703.353
t00000002_x244524	ptc-MIR159b	ptc-miR159b-5p		TGAAGGGAGCTCT	templated	244524	76703.353
t00000014_x51703	ptc-MIR171c	none	other	TTGAGCCGCGCCAATATCACT	templated	51703	16218.422
t00000014_x51703	ptc-MIR171d	none	other	TTGAGCCGCGCCAATATCACT	templated	51703	16218.422
t00000020_x31267	ptc-MIR393c	ptc-miR393c	3'+1	TCCAAAGGGATCGCATTGATCT	templated	31267	9807.969
t00000029_x22794	ptc-MIR166g	ptc-miR166g			templated	22794	7150.121
t00000029_x22794	ptc-MIR166h	ptc-miR166h			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

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:

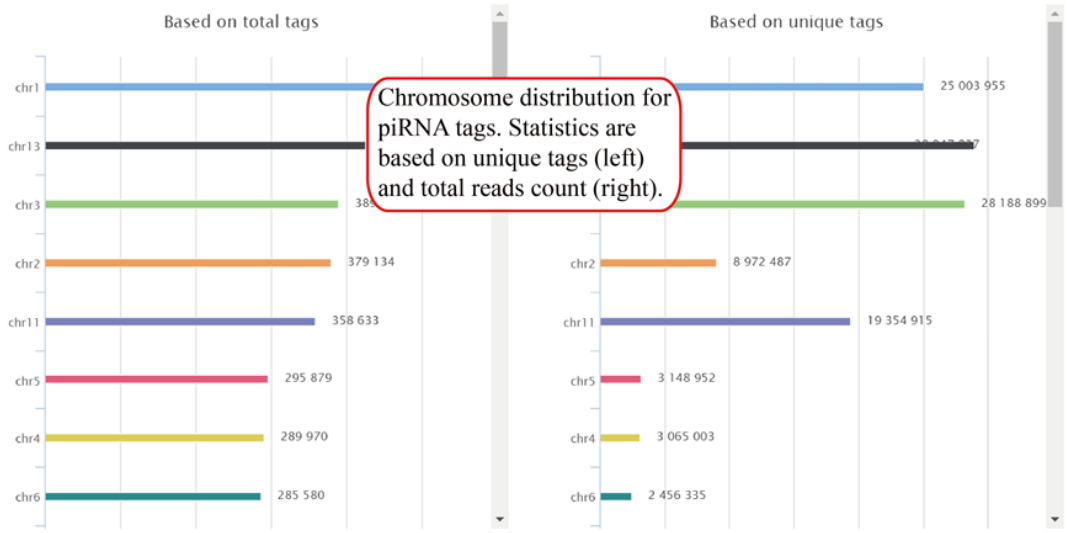


SN	miRNA	Strand	Total tag number	Total RPM	Type	Sequence	Tag name	Most abundant tag	Most abunda
2	xxx-m10-5p	5	485	349423.63	1	ACTATGAGCTAATGATTAG	seq14233_x105	105	75648.41
3	xxx-m8-5p	5	82	59077.81	1	TGCTTGTGTGTATGTGGGTGCTC	seq27452_x47	47	33861.67
4	xxx-m1-5p	5	49	35302.59			seq35611_x34	34	24495.68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

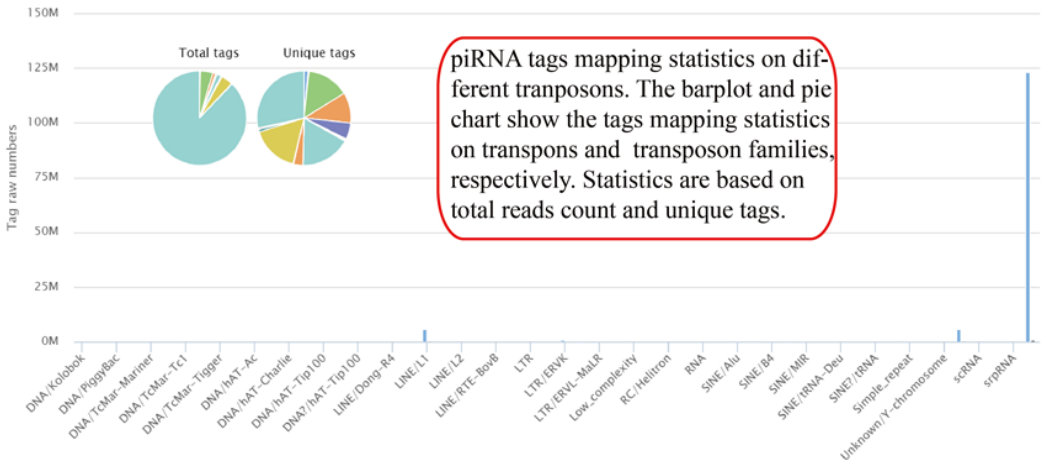
Tag distribution of piRNA on Transposon elements



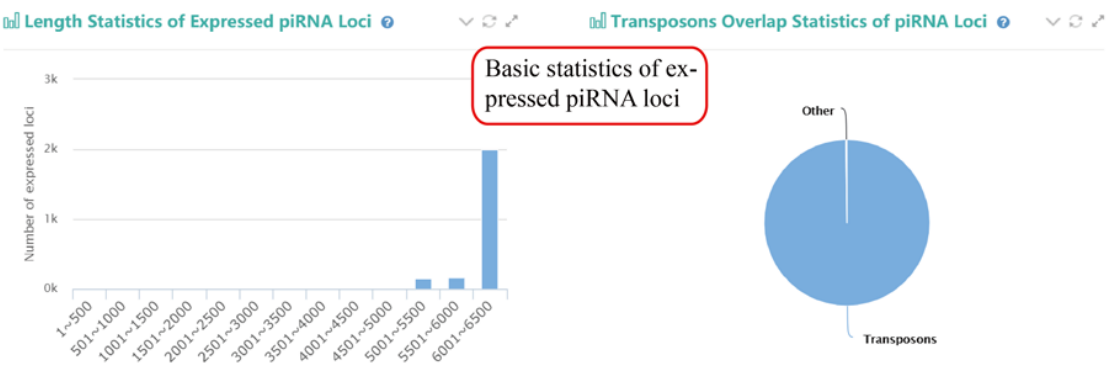
Tag distribution on different chromosomes for piRNA



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

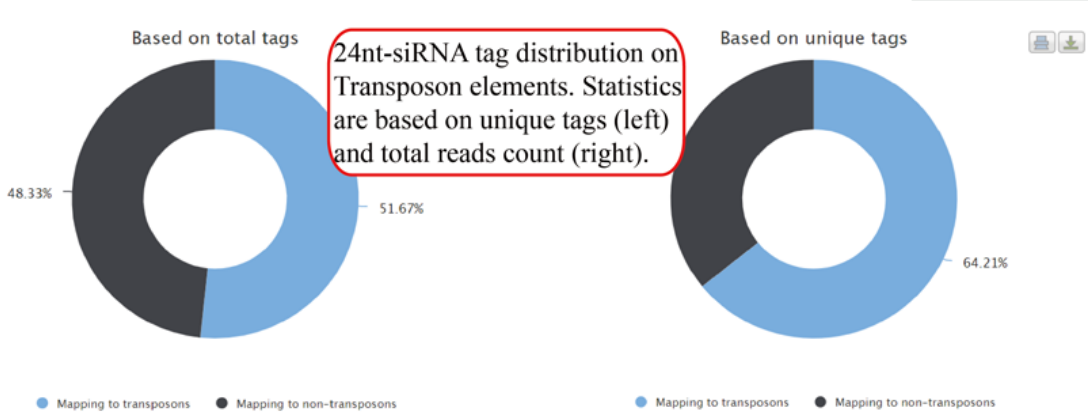
Expressed piRNA list (top 200)

Show 20 entries

Tag name	Sequence	Tag raw number	RPM	Mapping locus
piR-mmu-64542836	GTTTCCGTAGTGTAGTGGTTATCACGTTCG	209861	7136.59	NA
piR-mmu-2004832	GCATGGGTGGTTCAGTGGTAGAATTCTC	209174	7113.22	NA
piR-mmu-58055787	TCCCTGTGGTCTAGTGGTTAGG	3	5399.10	NA
piR-mmu-55164778	AAGAAAGATTGCAAGA	71561	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	GCCCGGCTAGCTCAGTCGGTAGAGCATGGGA	44288	1506.07	NA

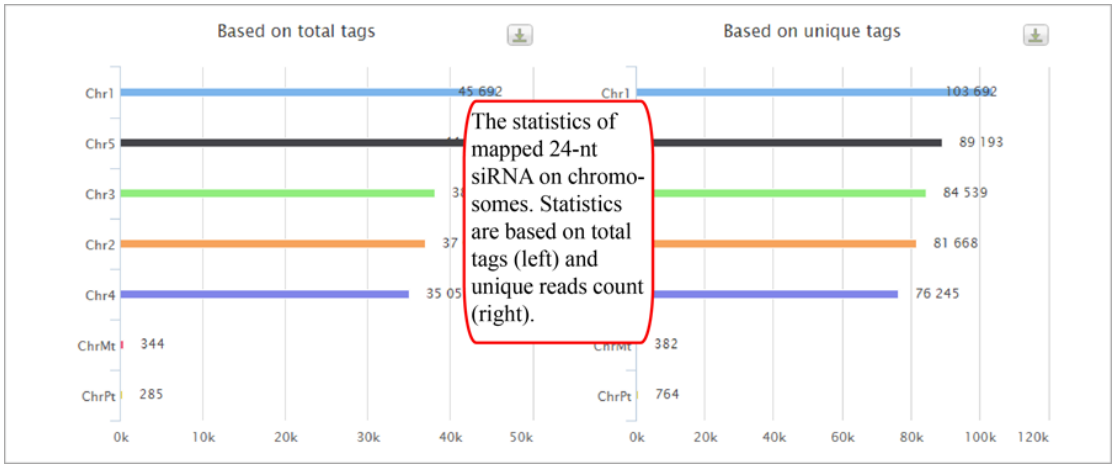
Top 200 list of expressed piRNA

Tag distribution of 24nt-siRNA on Transposon elements

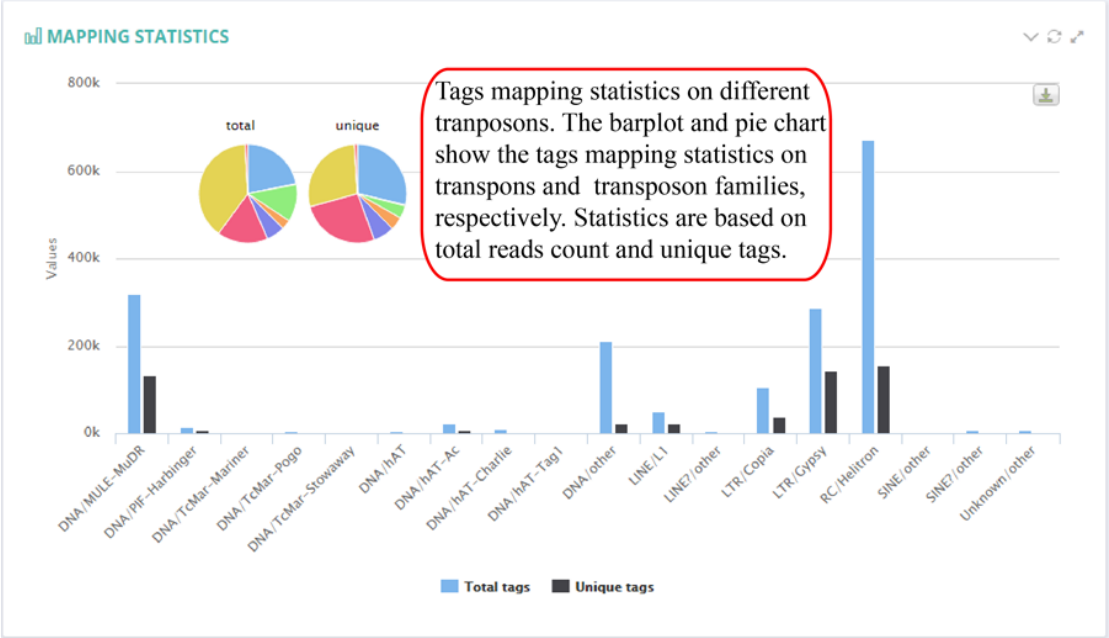


24nt-siRNA tag distribution on Transposon elements. Statistics are based on unique tags (left) and total reads count (right).

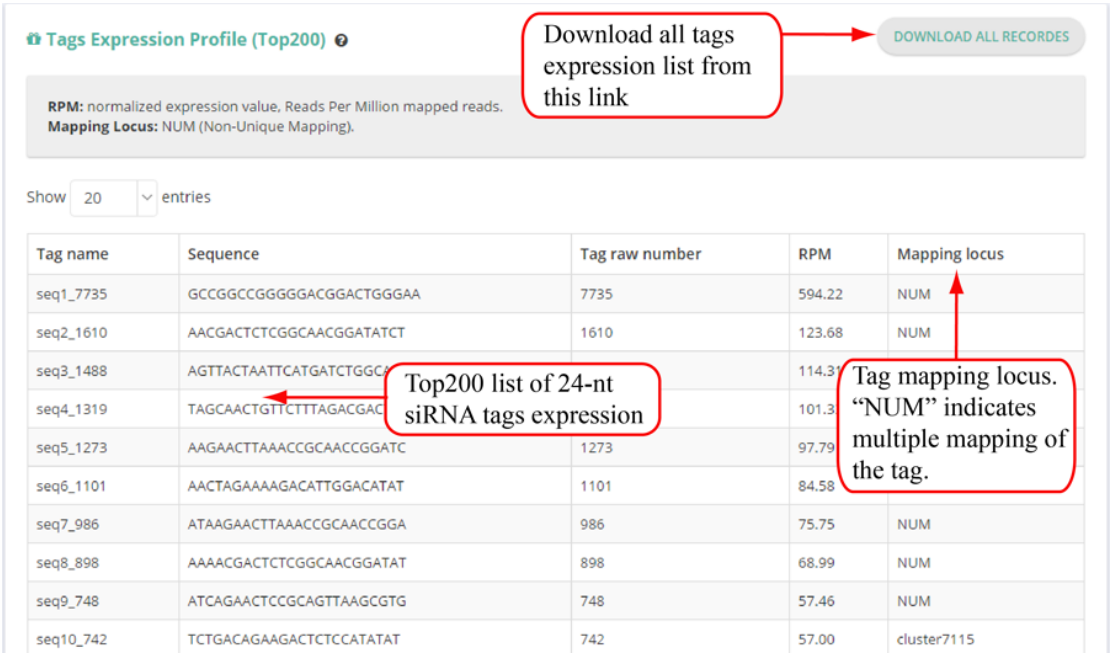
Distribution of 24-nt siRNA on different chromosomes



24nt siRNA tag distribution on different kind of transposon elements

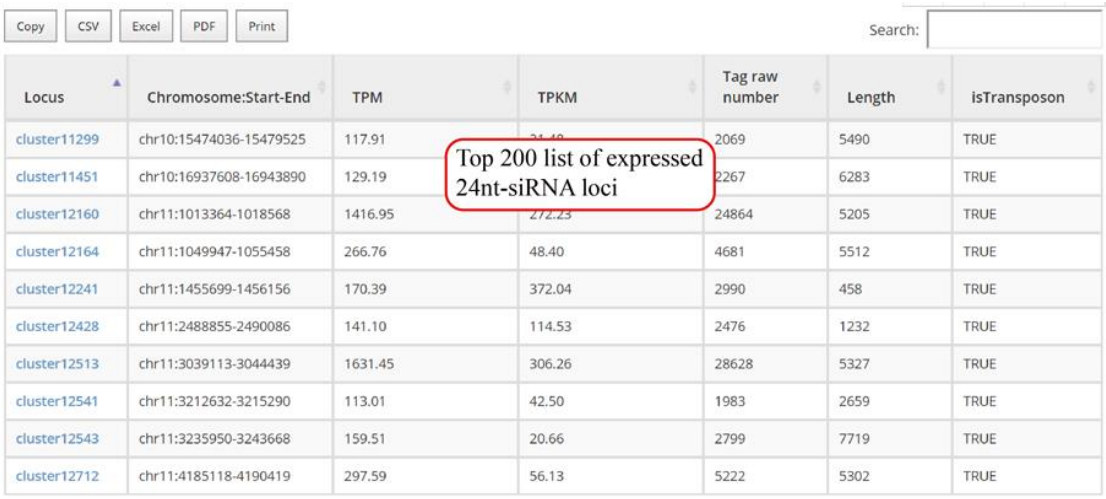


24nt siRNA expression list



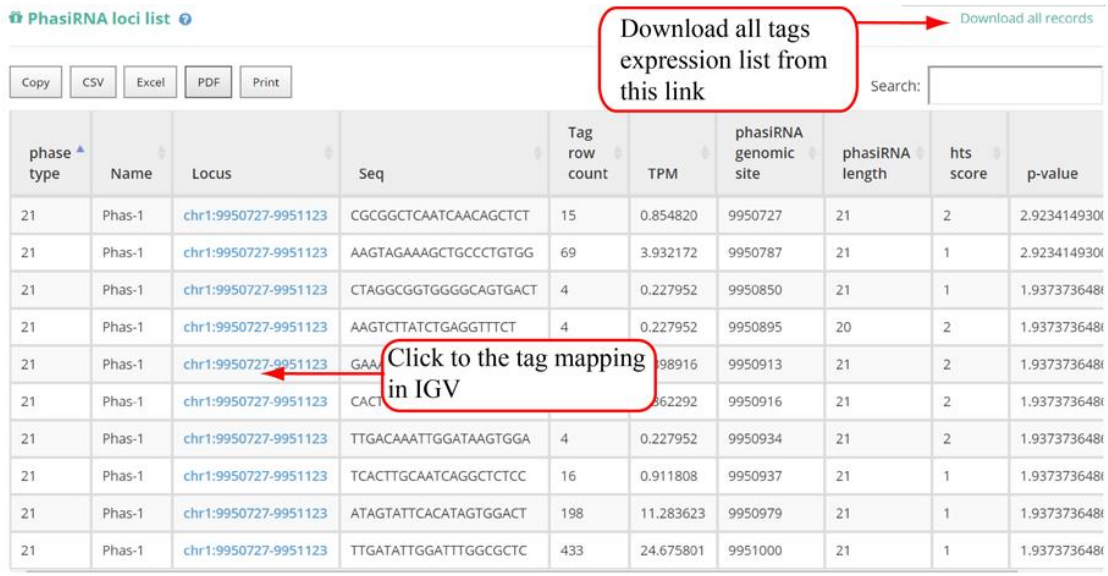
Tag name	Sequence	Tag raw number	RPM	Mapping locus
seq1_7735	GCCGGCCGGGGGACGGACTGGGAA	7735	594.22	NUM
seq2_1610	AACGACTCTCGGCAACGGATATCT	1610	123.68	NUM
seq3_1488	AGTTACTAATTCATGATCTGGCA		114.3	
seq4_1319	TAGCAACTGTTCTTTAGACGAC		101.3	
seq5_1273	AAGAACTTAAACCGCAACCGGATC	1273	97.79	
seq6_1101	AACTAGAAAAGACATTGGACATAT	1101	84.58	
seq7_986	ATAAGAACTTAAACCGCAACCGGA	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

Expressed piRNA producing loci list (top 200)



Locus	Chromosome:Start-End	TPM	TPKM	Tag raw number	Length	isTransposon
cluster11299	chr10:15474036-15479525	117.91		2069	5490	TRUE
cluster11451	chr10:16937608-16943890	129.19		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

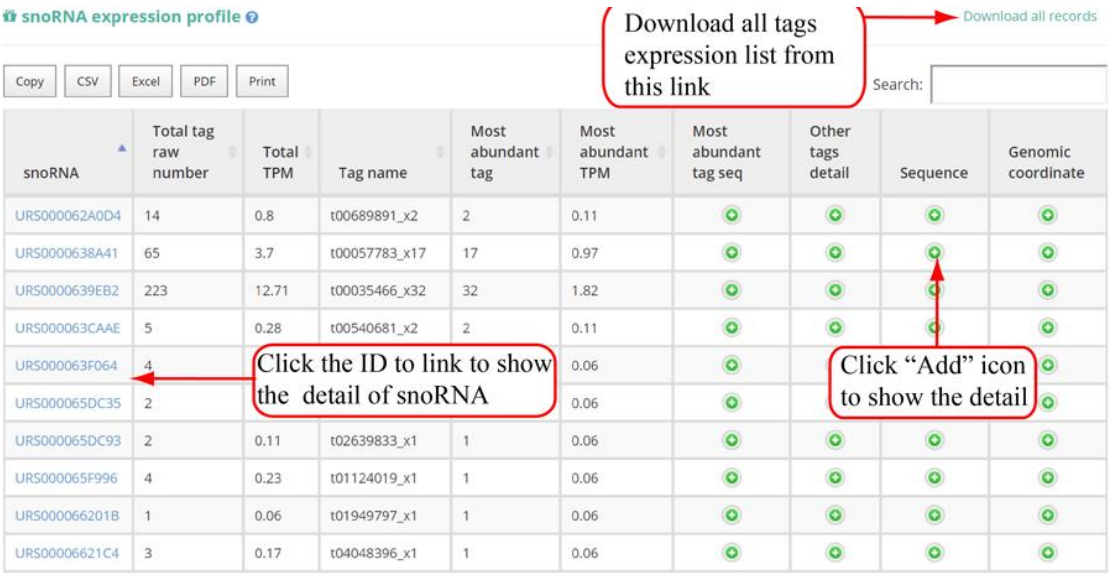
phasiRNA and loci list



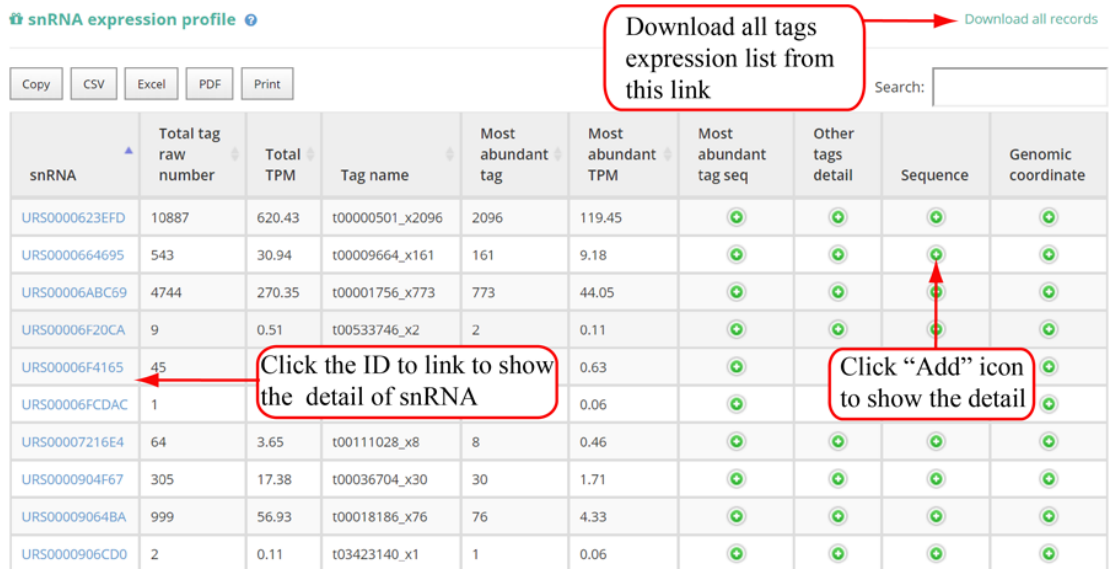
PhasiRNA induced miRNA (trigger) list

rRNA expression list

snoRNA expression list

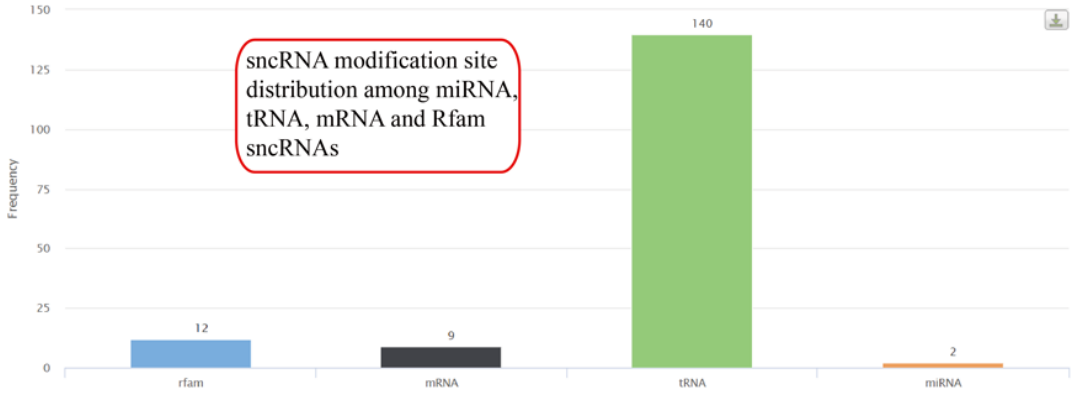


snRNA expression list



Other sncRNA expression list

Distribution of sncRNA with high confident modification



Modification site list of sncRNA

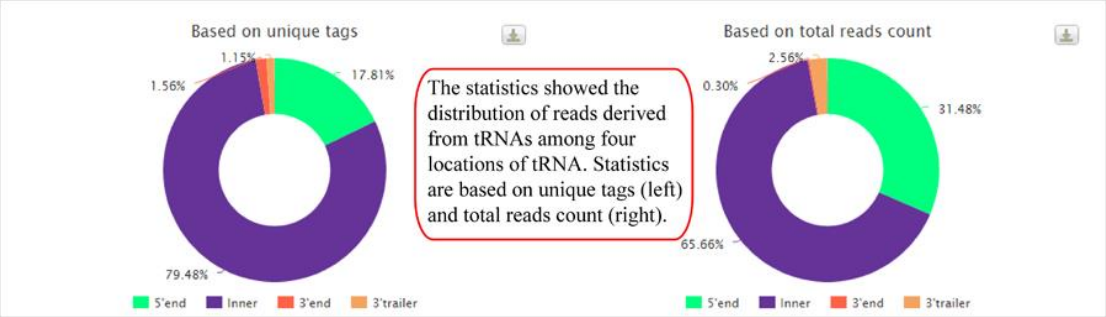
Download Table

Search:

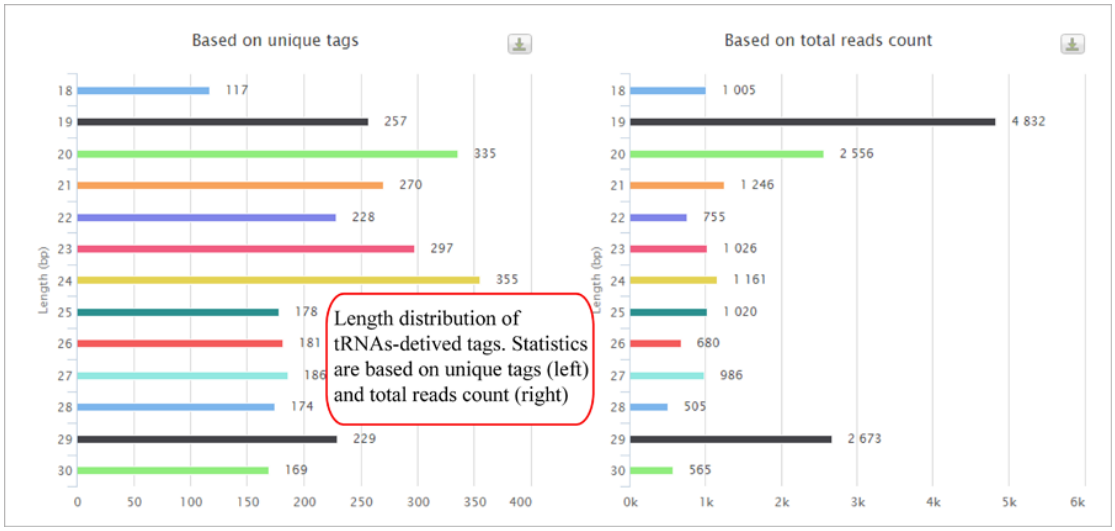
sncRNA	Modification site	coverage_input	count_input	faivalue_input	coverage_treated	count_treate
chr1.trna197-LysTTT|chr1:133033962-133034034:+|76	47	548	34	0.062043795620438	379	171
chr1.trna281-ProAGG|chr1:165641634-165641705:+|75	46			0.0658307210031348	189	92
chr1.trna451-AsnGTT|chr1:171033942-171034015:-|77	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

sncRNA modification site list. P value < 0.05 is used as the threshold to define the modification sites

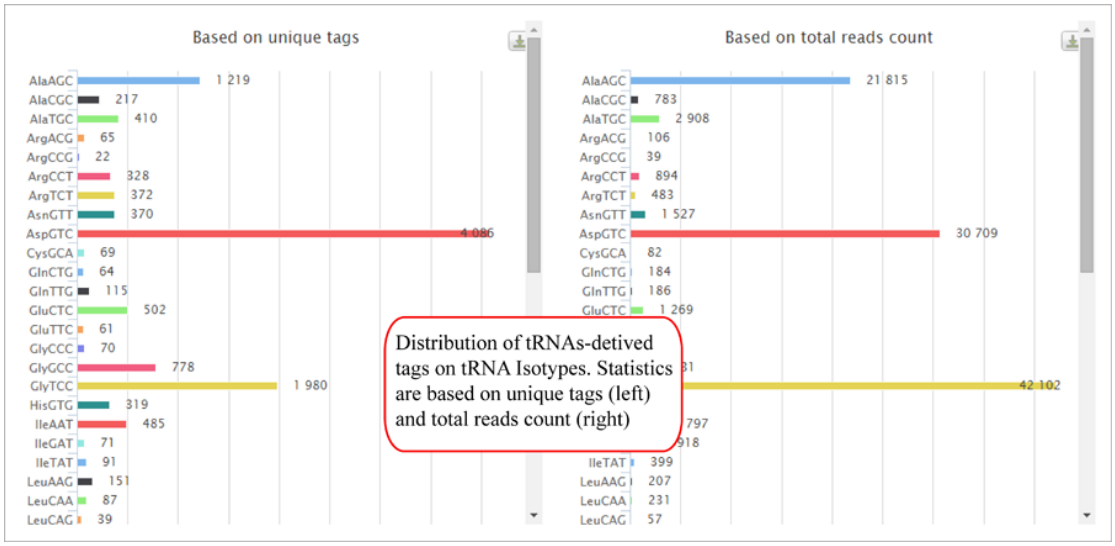
Tags mapping statistics on different regions



Statistics of tRNA family match results

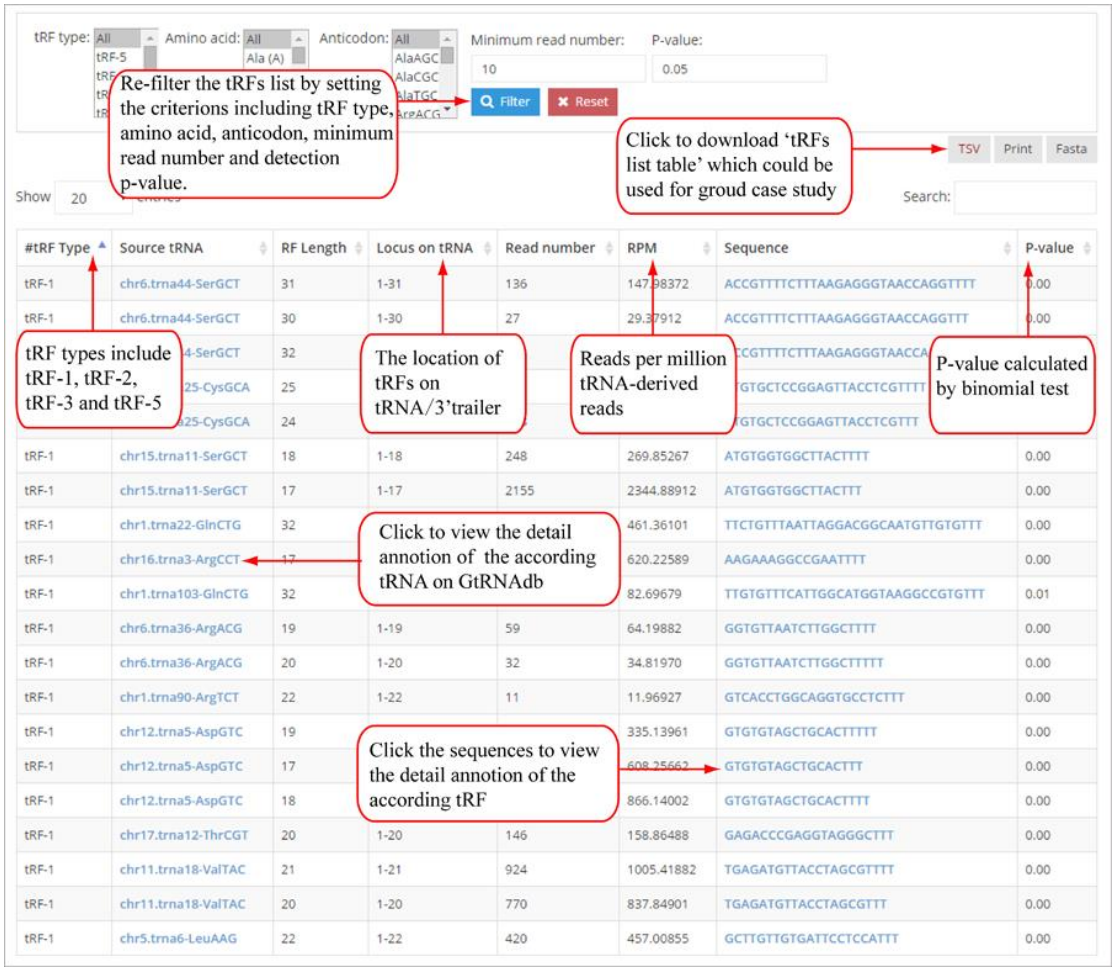


Statistics of tRNA family match results

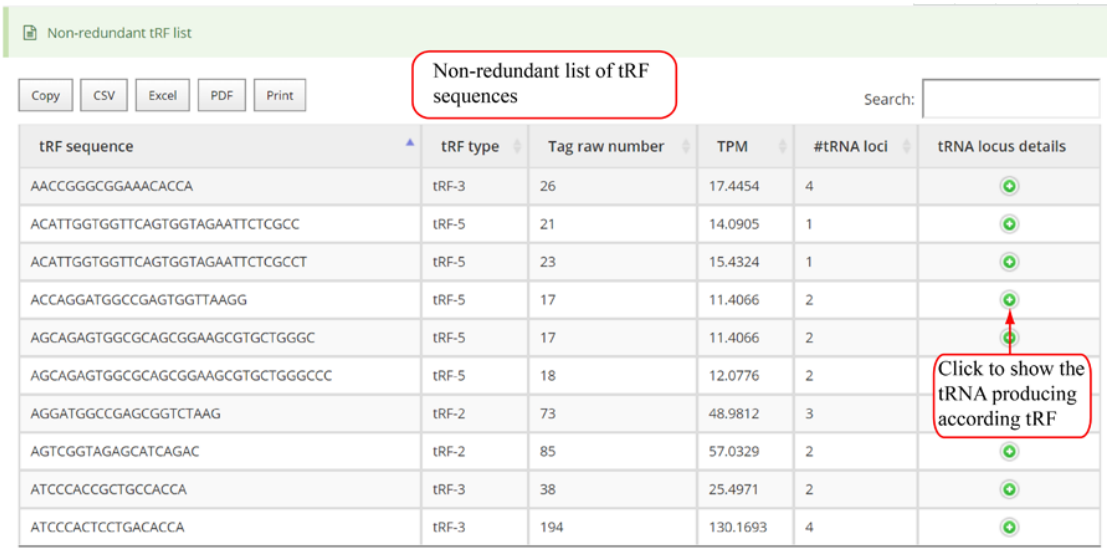


Statistics of tRNA match results

tRFs result list

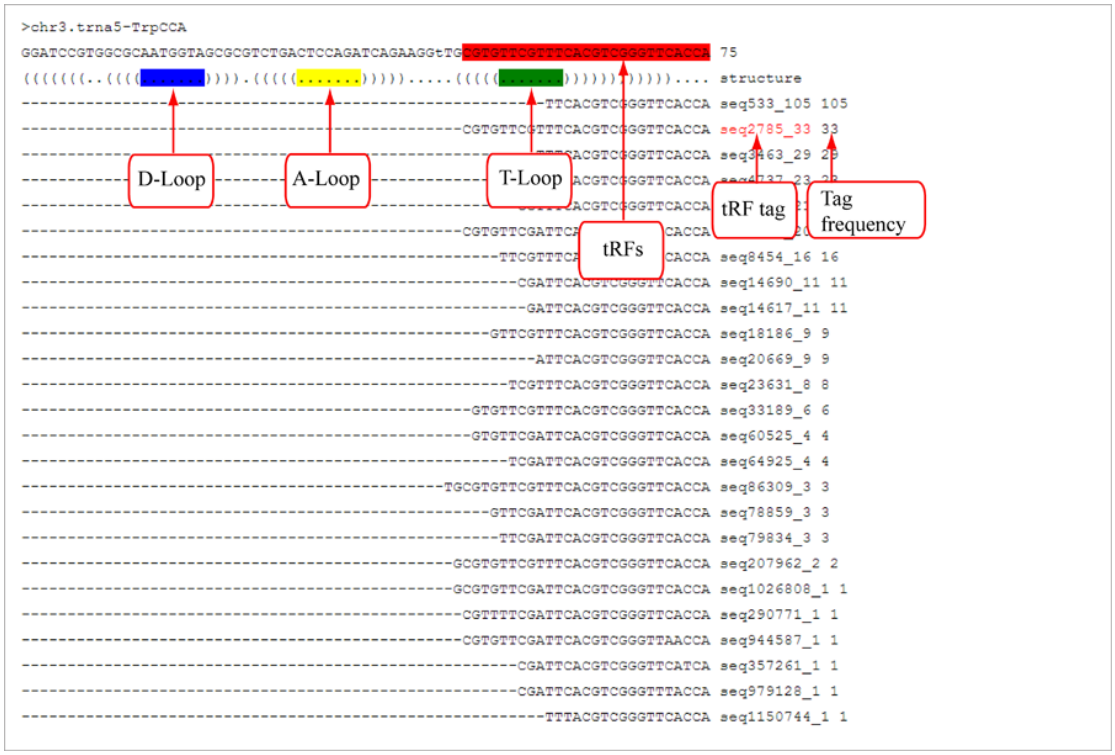


Non-redundant tRFs list

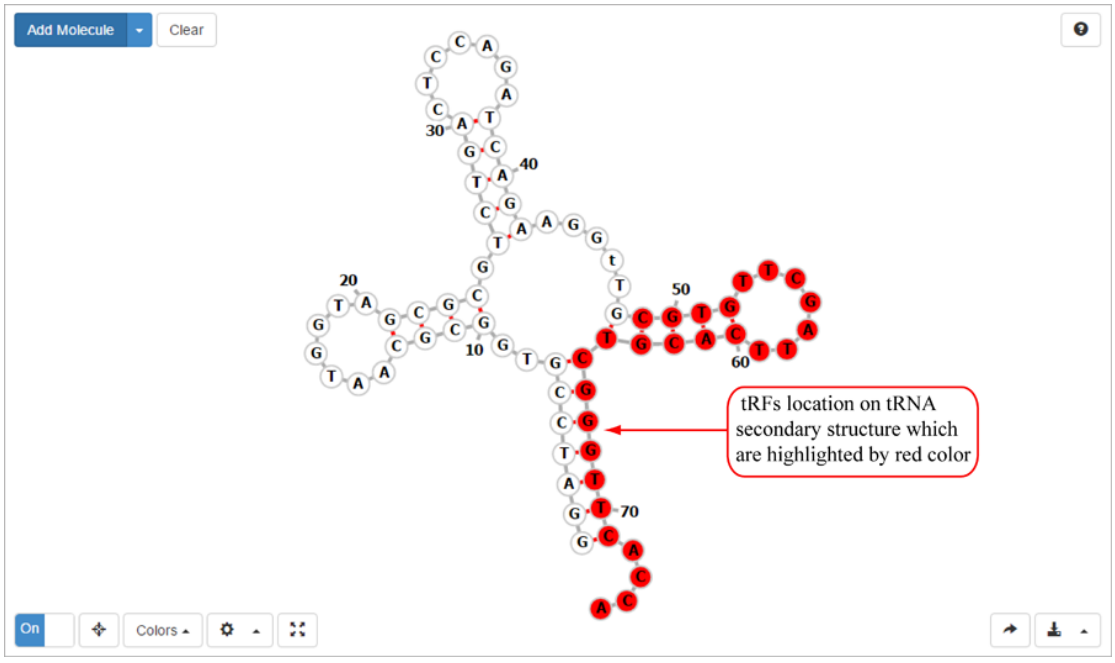


Basic information of tRFs

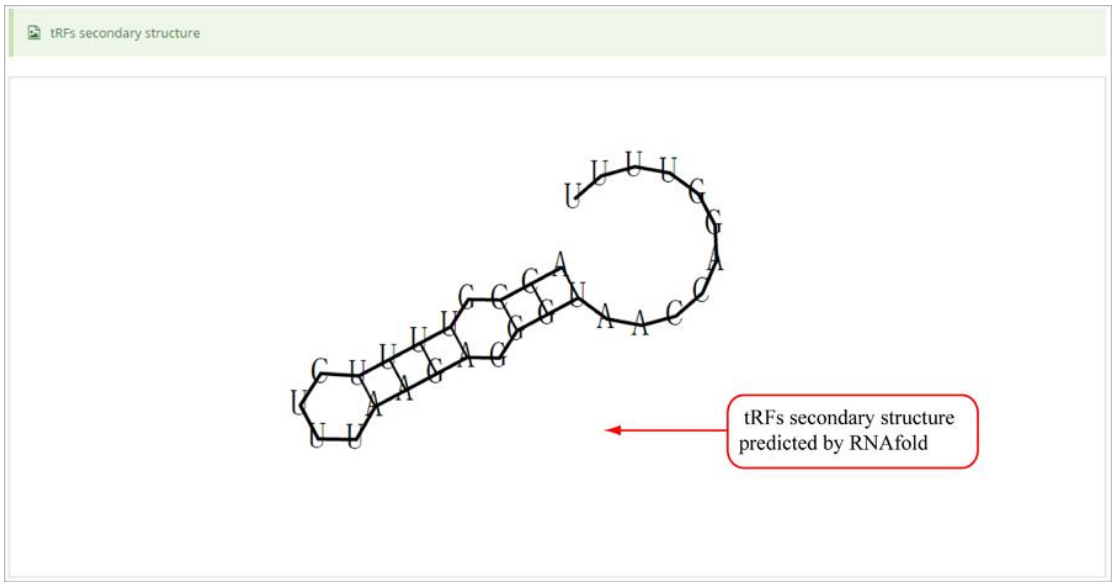
Tags alignments



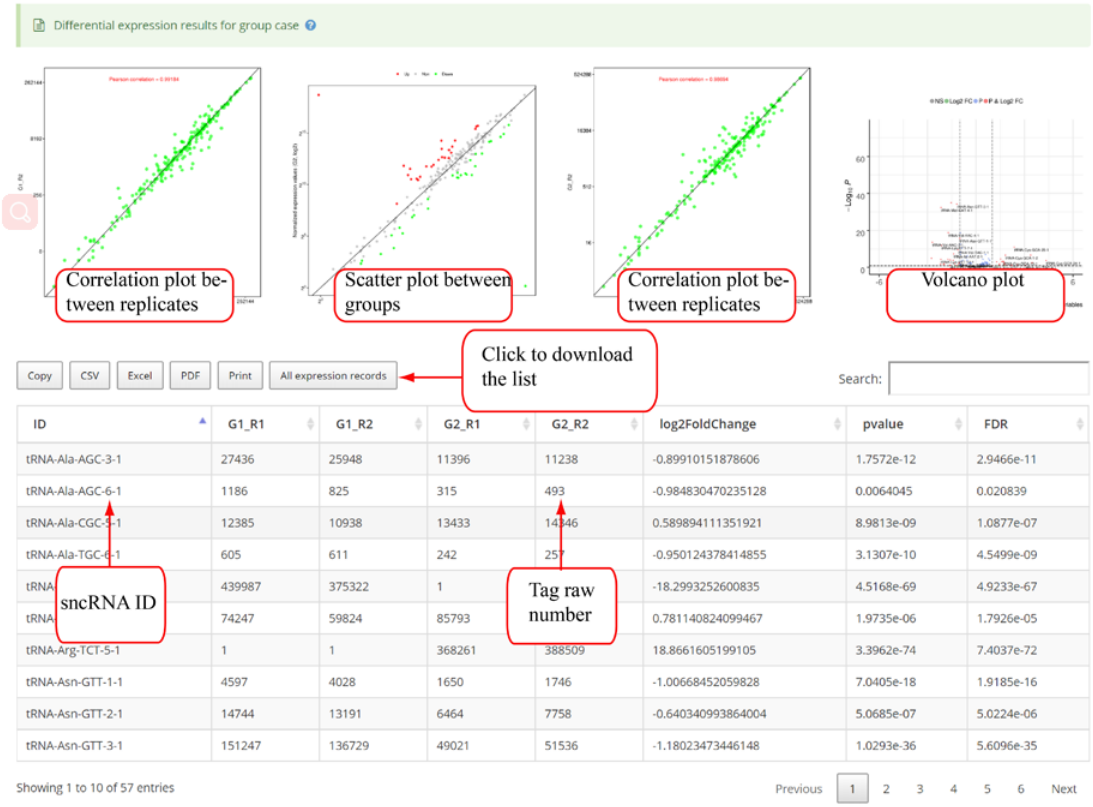
tRFs location on tRNA secondary structure



tRFs secondary structure



Differential expressed sncRNA in group case analysis



ID	G1_R1	G1_R2	G2_R1	G2_R2	log2FoldChange	pvalue	FDR
tRNA-Ala-AGC-3-1	27436	25948	11396	11238	-0.89910151878606	1.7572e-12	2.9466e-11
tRNA-Ala-AGC-6-1	1186	825	315	493	-0.984830470235128	0.0064045	0.020839
tRNA-Ala-CGC-5-1	12385	10938	13433	14346	0.589894111351921	8.9813e-09	1.0877e-07
tRNA-Ala-TGC-6-1	605	611	242	257	-0.950124378414855	3.1307e-10	4.5499e-09
tRNA	439987	375322	1		-18.2993252600835	4.5168e-69	4.9233e-67
tRNA	74247	59824	85793		0.781140824099467	1.9735e-06	1.7926e-05
tRNA-Arg-TCT-5-1	1	1	368261	388509	18.8661605199105	3.3962e-74	7.4037e-72
tRNA-Asn-GTT-1-1	4597	4028	1650	1746	-1.00668452059828	7.0405e-18	1.9185e-16
tRNA-Asn-GTT-2-1	14744	13191	6464	7758	-0.640340993864004	5.0685e-07	5.0224e-06
tRNA-Asn-GTT-3-1	151247	136729	49021	51536	-1.18023473446148	1.0293e-36	5.6096e-35

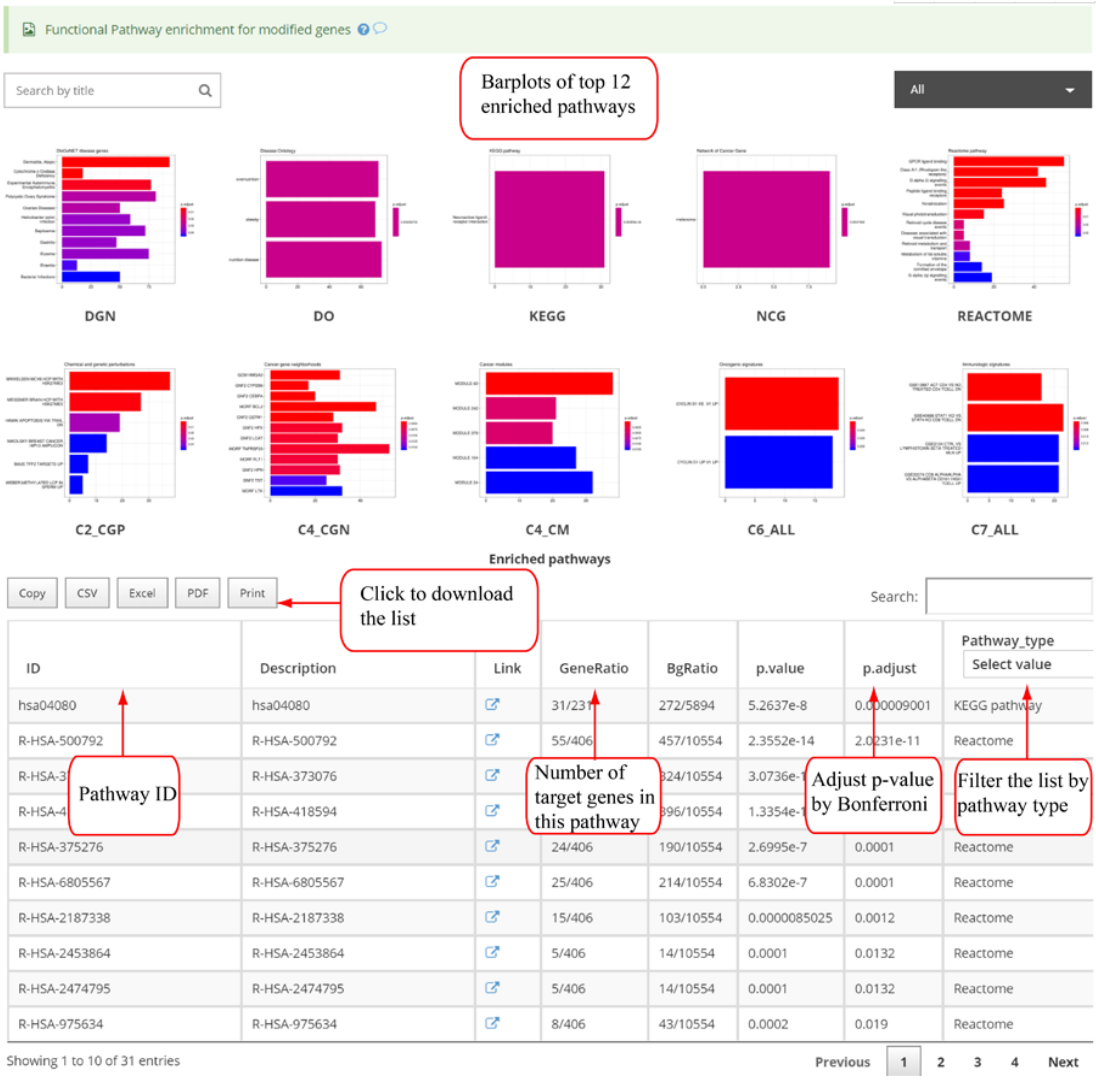
sncRNA target gene lists

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
			T1G01080,AT1G01110,AT1G01220,AT1G01225,AT1G01390,AT1G01440,AT1G01460,AT1G015
			T1G01080,AT1G01110,AT1G01220,AT1G01230,AT1G01460,AT1G01560,AT1G01660,AT1G016
			T1G01070,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

Gene Ontology enrichments of target genes

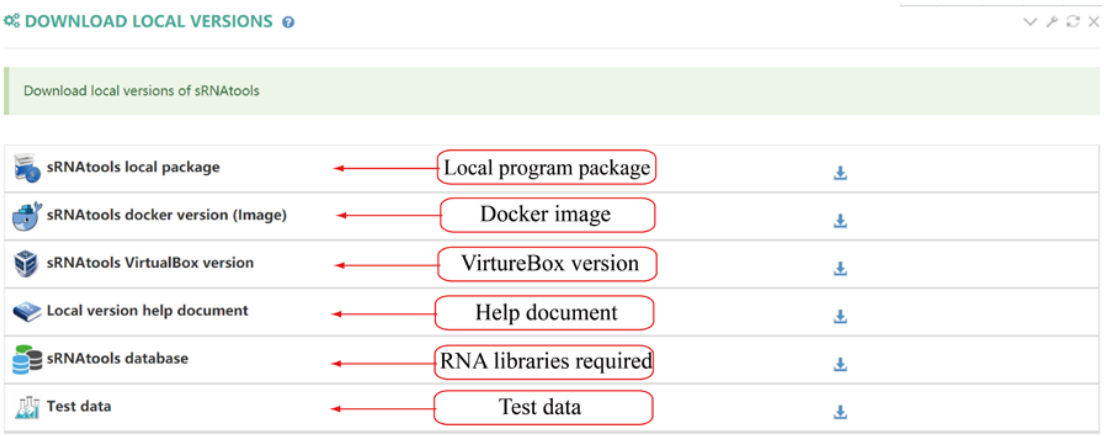
ID	Description	GeneRatio	BgRatio	p.value	p.adjust	GO
GO:0001523	retinoid metabolic process	17/690	86/18493	1.5545e-8	0.0001	BP
GO:0001653	peptide receptor activity	20/691	151/17632	0.0000018754	0.0005	MF
GO:	hormone activity		121/17632	0		MF
GO:	structural constituent of eye lens		20/17632	0.0009		MF
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

Pathway enrichments of target genes



Funcitonal gene network of target genes

Download local pipeline of sncRNAbench

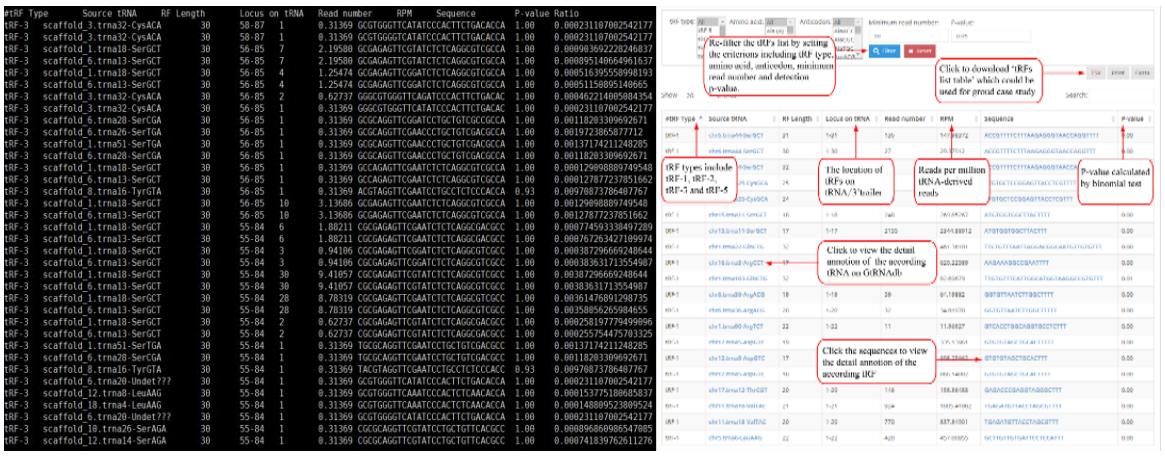


Browser tested

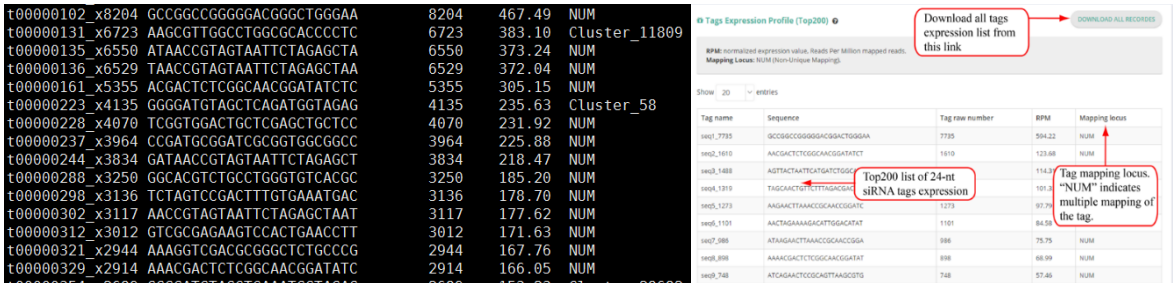
Browser name	Version
Internet Explorer	11
Chrome	70.0.3538.80
Firefox	66.0
Safari	5.1.7

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. Pln24NT: a web resource for plant 24-nt siRNA producing loci, Bioinformatics 2017;33:2065-2067.