

# Table of contents

---

- [Table of contents](#)
- [User guide](#)
  - [PhasiHunter workflow](#)
  - [Dependencies](#)
  - [Installation](#)
    - [Manual Installation](#)
    - [Docker image](#)
    - [Conda/mamba configure file](#)
  - [Demo data](#)
  - [Executing PhasiHunter with step-by-step submodules.](#)
  - [Executing PhasiHunter with one-command module](#)
    - [Default config.yaml file](#)
    - [The main output file](#)
- [Copyright](#)

## User guide

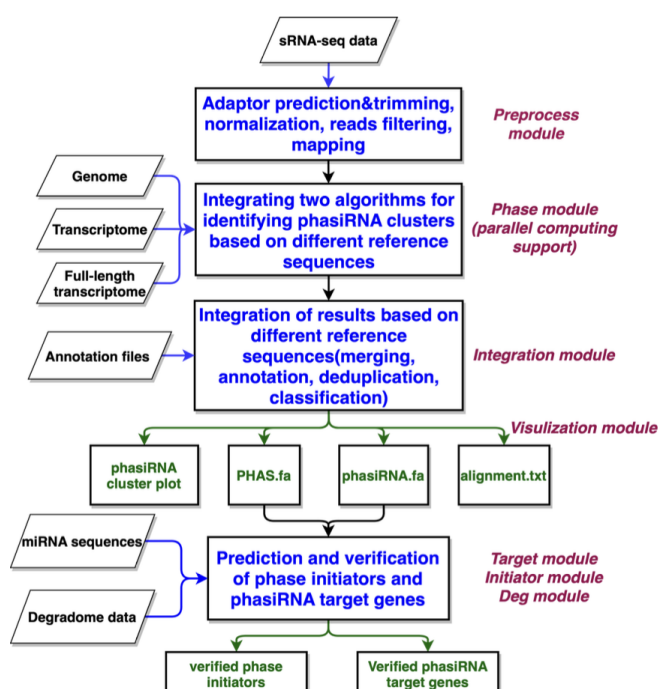
---

Welcome to phasihunter 😊

A multithreaded program for mining phasiRNA regulation pathways based on multiple reference sequences.

## PhasiHunter workflow

---



# Dependencies

---

phasihunter is a CLI program running on linux platform. The correct running of phasihunter depends on some existing softwares.

- Bowtie (Langmead, et al., 2009. Genome Biol)
- Biopython (Cock, et al., 2009. Bioinformatics)
- Bedtools (Quinlan and Hall, 2010. Bioinformatics)
- Dnapi (Tsuji and Weng, 2016. PloS One)
- Trim\_galore (<https://github.com/FelixKrueger/TrimGalore>)
- Seqkit (Shen, et al., 2016. PloS One)
- Perl5 (<https://www.perl.org>)
- Fasta36 (Pearson and Lipman, 1988. Proc Natl Acad Sci U S A)
- TarHunter (Ma, et al., 2018. Bioinformatics)

## Installation

---

### Manual Installation

---

1. Install all dependencies
2. Clone phasihunter

```
git clone https://github.com/HuangLab-CBI/PhasiHunter.git .
```

3. Setting environment variable in ~/.bashrc

```
echo "export PATH=\$PATH:<phasihunter PATH> >> ~/.bashrc"
```

example:

```
echo "export PATH=\$PATH:/home/user/volumes/PhasiHunter >> ~/.bashrc"
```

4. type `phasihunter -h` to check phasihunter whether installation correct. If phasiHunter is installed correctly you will see the following content.

```
Usage:
  phasiHunter subcommand options

subcommand:
  run:          one command executing mode, config.yaml file required
  preprocess:   generating map file for phasiRNA cluster prediction
  phase:        predicting phasiRNA cluster based on multiple reference sequences
  integration:  integrating phase module output
  visualization: for phasiRNA cluster visualization
  target:       for sRNA target gene prediction
  initiator:    for phasiRNA initiator prediction
  deg:          for phasiRNA initiator or phasiRNA target gene verification based on degradome verification

type phasiHunter subcommand -h for more subcommand detail
```

### Docker image

---

For convenience, we also provide a Docker image at

```
https://hub.docker.com/repository/docker/zacksfeng/phasihunter
```

The Docker image has been configured with all the dependencies required for running phasiHunter.

## Conda/mamba configure file

---

We also provide a conda/mamba environment configuration file. User can install all the required dependencies with command `conda/mamba install -f /foo/PhasiHunter/bin/env.yaml`

## Demo data

---

### Download link

[https://cbi.njau.edu.cn/PhasiHunter\\_demo\\_data/test\\_osa.tar.gz](https://cbi.njau.edu.cn/PhasiHunter_demo_data/test_osa.tar.gz)

## Executing PhasiHunter with step-by-step submodules.

---

Parameter in < > means necessary; parameter in [ ] means optional

### 1. Data pre-process

```
1 phasiHunter preprocess -m r -i /home/user/test_osa/SRR5049781.fastq.gz -mi
  19 -ma 25 -e 1 -n 1000000 -o
  /home/user/test_osa/SRR5049781_processed_cdna.map -in
  /home/user/test_osa/index/oryza_sativa_cdna_index
2
3 phasiHunter preprocess -m m -i
  /home/user/test_osa/SRR5049781_trimmed_format_filter.fa -mi 19 -ma 25 -e 1 -
  n 1000000 -o /home/user/test_osa/SRR5049781_processed_gdna.map -in
  /home/user/test_osa/index/oryza_sativa_gdna_index
```

- preprocess module usage

```
1 Help message:
2 options:
3 # necessary options:
4 -m: string -- mode: r | c | m;
5 raw(mode): trim adaptor --> normalization --> length and
abundance filter --> mapping
6 clean(mode): normalization --> length and abundance
filter --> mapping
7 mapping(mode): mapping
8 -i: file -- for r mode: fastq file or fastq.gz file
9 for c mode: fasta file or fasta.gz file
10 for m mode: length and abundance filter fasta file
11 -r: file -- reference sequence fasta file
12 -in: string -- index prefix, -r option will be ignored when -in enable
13 -o: outfile -- outfile name
14
15 # options with default value
16 -j: int -- adaptor trim parallel cores; <8 is recommend, only need
in r mode, default=1
```

```

17     -bj: int    -- bowtie parallel cores; defalut=1
18     -mh: int    -- max hits when mapping to ref sequence, default=10
19     -mi: int    -- minimal sRNA reads length cutoff, default=19
20     -ma: int    -- maxmial sRNA reads length cutoff, default=25
21     -e: float   -- sRNA reads cpm cutoff, default=1
22     -n: int     -- normalization base, default=1000000
23
24
25     # other
26     -v:         -- print version information
27     -h:         -- print help information

```

## 2. PhasiRNA and PHAS loci prediction

```

1  phasiHunter phase -cm /home/user/test_osa/SRR5049781_processed_cdna.map -c
   /home/user/test_osa/oryza_sativa_cdna.fa -gm
   /home/user/test_osa/SRR5049781_processed_gdna.map -g
   /home/user/test_osa/oryza_sativa_gdna.fa -fm None -f None -fa
   /home/user/test_osa/SRR5049781_trimmed_format_filter.fa -a
   /home/user/test_osa/phase_a.txt -o /home/user/test_osa/phase_o.txt -me b -il
   5 -pl 21 -pn 4 -mh 10 -j 20 -pv 0.001 -ps 15 -pr 0.4 -cl y

```

- phase module usage

```

1  phase usage:
2  option:
3     -cm: file  -- map file based on reference transcriptome sequence
4     -c:  file  -- reference transcritome sequence, fasta file
5     -gm: file  -- map file based on reference genome sequence
6     -g:  file  -- reference genome sequence, fasta file
7     -fm: file  -- map file based on full length transcriptome sequence
8     -f:  file  -- full length transcriptome sequence, fasta file
9     -fa: file  -- sRNA file
10     -a: out   -- allsiRNA cluster output file, default name is phase_a.txt
11     -o: out   -- phasiRNA cluster output file, default name is phase_o.txt
12     -me: str  -- phasiRNA prediction method, h(hypergeometric test) |
   p(phase score) | b (both), default=b
13     -il: int  -- phasiRNA cluster island, default=5
14     -pl: int  -- phase length, 21 | 24, default=21
15     -pn: int  -- phase number, default=4
16     -mh: int  -- max hits when mapping to ref sequence, default=10
17     -j:  int  -- parallel number, default=1
18     -pv: float -- pvalue cutoff, default=0.001, only function with h/b
   method applied
19     -ps: float -- phase score cutoff, default=15, only function with p/b
   method applied
20     -pr: float -- phase ratio cutoff, default=0.4, only function with p/b
   method applied
21     -cl: str  -- delete .phasiHuter_bowtieIndex, y|n, default=y
22     -v:         -- print version information
23     -h:         -- print help information

```

### 3. PhasiRNA and PHAS loci result integration

```
1 phasiHunter integration -io /home/user/test_osa/phase_o.txt -ia
/home/user/test_osa/phase_a.txt -an
/home/user/test_osa/oryza_sativa_gdna.gff3 -g y -o
/home/user/test_osa/integration_o.txt -a
/home/user/test_osa/integration_a.txt -s
/home/user/test_osa/integration_s.txt -po
/home/user/test_osa/integration_p.txt -ao
/home/user/test_osa/as_apa_related_result.txt -j 1 -pn 4 -pl 21 -pv 0.001 -
il 5
```

- integration module usage

```
1 integration usage:
2 option:
3 # necessary options:
4 -io: file -- phase module -o output file
5 -ia: file -- phase module -a output file
6 -an: file -- reference genome gff3 file
7 -g: str -- y | n, whether exist gdna based PHAS Loci
8
9 # options with default value
10 -o: out -- integration phasiRNA cluster, default name is
integration_o.txt
11 -a: out -- integration all siRNA cluster, default name is
integration_a.txt
12 -s: out -- integration summary, default name is integration_s.txt
13 -po: out -- PHAS Loci information, default name is integration_p.txt
14 -ao: out -- alternative splicing/alternative polyadenylation related
PHAS gene, optional
15 -j: int -- parallel number, default=1
16 -pn: int -- phase number, default=4
17 -pl: int -- phase length, 21 | 24, default=21
18 -pv: float -- pvalue cutoff, default=0.001
19 -il: int -- phasiRNA cluster island, default=5
20 -dp: str -- y | n, discard only P method result, default=y
21
22 # optional options
23 -fn: file -- full length transcript annotation file
24
25 # other
26 -v: -- print version information
27 -h: -- print help information
```

### 4. Print phasiRNA\_cluster plot, phasiRNA.fa, PHAS.fa

```
1 phasiHunter visulization -io /home/user/test_osa/integration_o.txt -ia
/home/user/test_osa/integration_a.txt -ip
/home/user/test_osa/integration_p.txt -a /home/user/test_osa/alignment.txt -
o /home/user/test_osa/phasiRNA.fa -p /home/user/test_osa/PHAS.fa -pl 21 -m
```

```
10 -c /home/user/test_osa/oryza_sativa_cdna.fa -g
/home/user/test_osa/oryza_sativa_gdna.fa -f None -pc y -pg y -pf n
```

- visulization module usage

```
1 visulization usage:
2 option:
3 # necessary options:
4 -io: file -- integration -io outputfile
5 -ia: file -- integration -ia outputfile
6 -ip: file -- integration -po outputfile
7 -a: out -- alignment file, default name is alignment.txt
8 -o: out -- phasiRNA fasta file, default name is phasiRNA.fa
9 -p: out -- PHAS Gene fasta file; Format:
>geneid/chr\tphasiRNA_cluster_region(start end)\tseq_region(start end),
default name is PHAS.fa
10
11 # options with default value
12 -pl: int -- phase length, 21 | 24, default=21
13 -m: float -- the number for reducing the size of Y-axis. default=10
14
15 # optional options
16 -c: file -- reference transcriptome sequence, fasta file, enable cdna
based phasiRNA.fa, PHAS.fa, Alignmen, Plot output
17 -g: file -- reference genome sequence, fasta file, enable gdna based
phasiRNA.fa, PHAS.fa, Alignmen, Plot output
18 -f: file -- full length transcriptome sequence, fasta file, enable
flnc based phasiRNA.fa, PHAS.fa, Alignmen, Plot output
19 -pc: str -- plot cdna based phasiRNA cluster, y | n, default=y
20 -pg: str -- plot gdna based phasiRNA cluster, y | n, default=y
21 -pf: str -- plot flnc based phasiRNA cluster, y | n, default=y
22
23 # other
24 -v: -- print version information
25 -h: -- print help information
```

## 5. Initiator prediction and verification

```
1 phasiHunter target -q /home/user/test_osa/osa.miRbase.fa -b
/home/user/test_osa/PHAS.fa -o /home/user/test_osa/miR_target.txt -T 10
2
3 phasiHunter initiator -i /home/user/test_osa/integration_o.txt -j
/home/user/test_osa/miR_target.txt -ip /home/user/test_osa/integration_p.txt
-pd 5 -pl 21 -ps 1 -o /home/user/test_osa/initiator.txt
4
5 phasiHunter deg -i /home/user/test_osa/deg/GSM1040649_format_filter.map -q
/home/user/test_osa/osa.miRbase.fa -j /home/user/test_osa/initiator.txt -t
/home/user/test_osa/oryza_sativa_cdna.fa -o GSM1040649_MTI_deg.txt -s 1 -m 0
-p y -in y -pl 1 -pf MTI_deg --lib GSM1040649 -less
```

- target module usage

```

1  Usage:
2      perl /home/user/volumes/PhasiHunter/bin/TarHunterL_Modified.pl -q
   <mir_file> -b <targ_file> -o <out_file> [Options]
3
4  Required arguments:
5      -q (--qmir):      query miRNA file
6      -b (--targ):      target file
7      -o (--output):    output file
8
9  Options:
10
11      -M (--total_misp): max. total mispairs          [Default: off]
12      -m (--seed_misp):  max. seed mispairs           [Default: off]
13      -f (--score):      score cutoff                  [Default: 4 ]
14
15      -I (--mimics):      eTM search                   [Default: off]
16      -i (--mimics_str):  eTM stringency
17                        (0: strict, 1: relaxed)         [Default: 0 ]
18
19      -T (--threads):     FASTA threads                [Default: 1 ]
20      -t (--tab):         tabular format output        [Default: off]
21      -h (--help):       help information
22
23  Dependencies:
24      fasta36

```

- initiator module usage

```

1  initiator option:
2      -i [str]integration -o output
3      -j [str]the target predicted by psRNA target server or target module
4      -ip [str]integration -po output
5      -pd [int]the microRNA distance away to phase border, default=105(21) or
   120 (24), optional
6      -pl [int]21 or 24, the phase length of 21 or 24, default=21
7      -ps [int]0 or 1, the position of cleavage at 10(0) or 9-11 (1), default=1
8      -o [str]outputfilename.
9      -h print the version and details of the usage

```

- deg module usage

```

1  // function: verified the sRNA - Target interaction with degradome data
2
3  options:
4      -i: <inputfilename>      -- mapping file for degradome data mapping
   transcripts, by bowtie
5      -q: <sRNA fasta>         -- small RNA sequences used for target
   prediction, fasta
6      -j: <inputfilename>      -- from psRNATarget batch download file or
   initiator output
7      -t: <inputfilename>      -- transcripts file, fasta

```

```

8      -o: <outputfilename>      --      matched map file with only matched records
9      -s: <shift_number>        --      if shifts=0 then cleaved exactly at pos.10,
default=1
10     -m: <minum deg_num>        --      minum number of degradome reads, int,
default=0
11     -p: <T-plot function>      --      enable the plot function, y | n, default='n'
12     -in: <bool>                --      y | n, use initiator output information
13     -pl [int]                  --      1,plot only category 1; 2, plot categories 1
and 2, default=1
14     -pf [str]                  --      output folder name, for exporting t-plot
images and outputfile
15     --lib [str]                --      library name
16     -less                      --      only output cat_1 and cat_2 information
17
18     *****
19     //About the categories:
20     Cat #1, degradome read at the cleavage site is most abundant.
21     Cat #2, the read is less than the most abundant one, but higher than the
median.
22     Cat #3, the read is less than the median, but high than 1
23     Cat #4, the read is identical or less than 1 (if degradome data is
normalized)

```

## 6. PhasiRNA target prediction and verification

```

1  phasiHunter target -q /home/user/test_osa/phasiRNA.fa -b
/home/user/test_osa/oryza_sativa_cdna.fa -o
/home/user/test_osa/phasiRNA_target.txt -T 10
2
3  phasiHunter deg -i /home/user/test_osa/deg/GSM1040649_format_filter.map -q
/home/user/test_osa/phasiRNA.fa -j /home/user/test_osa/phasiRNA_target.txt -
t /home/user/test_osa/oryza_sativa_cdna.fa -o GSM1040649_PTI_deg.txt -s 1 -m
0 -p y -in n -pl 1 -pf PTI_deg --lib GSM1040649 -less

```

## Executing PhasiHunter with one-command module

### One-command module usage

```

1  One command executing mode
2
3  Usage:
4      phasiHunter run [-i] [config file]
5      phasiHunter run -d
6
7  option:
8      -i: yaml format config file
9      -d: using the default config, defalut config file is
/fooo/PhasiHunter/bin/config.yaml
10     -h: print help information
11
12  WARNIG: make sure choose the correct config file before run this command

```



## Default config.yaml file

---

Some INPUT and OUTPUT still need modified when using.

```
1  # Please provide the full path to the input file
2
3  # Configure the modules that need to be run
4  # y means enable, n means disable
5  Runing_module:
6    preprocess: y
7    phase: y
8    integration: y
9    visulization: y
10   initiator_prediction_and_verification:
11     target: y
12     initiator: y
13     deg: y
14   phasiRNA_target_prediction_and_verification:
15     phasiRNA_target: y
16     phasiRNA_deg: y
17
18  # Configure the preprocess module
19  preprocess:
20    # raw(mode): trim adaptor --> normalization --> length and abundance
21    # clean(mode): normalization --> length and abundance filter --> mapping
22    # mapping(mode): mapping
23    mode: r # [r | c | m]
24
25    # for r mode: fastq file or fastq.gz file
26    # for c mode: fasta file or fasta.gz file
27    # for m mode: length and abundance filter fasta file
28    # ** INPUT **
29    inputfile: /home/user/test_osa/SRR5049781.fastq.gz
30
31    # reference sequence fasta file
32    # ** INPUT **
33    reference_fasta: # disable when index parameter enable, multiple sequence
34    # - /home/user/test_osa/oryza_sativa_cdna.fa
35    # - /home/user/test_osa/oryza_sativa_gdna.fa
36
37    # index prefix, reference_fasta option will be ignored when index enable,
38    # multiple index can provided here
39    # ** INPUT **
40    index:
41      - /home/user/test_osa/index/oryza_sativa_cdna_index
42      - /home/user/test_osa/index/oryza_sativa_gdna_index
43
44    # outfile name, relative path is work for outputfile, but absolute path
45    # is still recommended. The number must be the same as the number of
46    # reference_fasta or indexes
```

```
44 # ** OUTPUT **
45 outfile_name:
46   - /home/user/test_osa/SRR5049781_processed_cdna.map
47   - /home/user/test_osa/SRR5049781_processed_gdna.map
48
49 # adaptor trim parallel cores; <8 is recommend, only need in r mode
50 trim_adaptor_cores: 1
51
52 # bowtie parallel cores
53 bowtie_mapping_cores: 1
54
55 # max hits when mapping to ref sequence
56 bowtie_max_hits_cutoff: 10
57
58 # minimal sRNA reads length cutof
59 minimal_sRNA_length_cutoff: 19
60
61 # maxmial sRNA reads length cutoff
62 maxmial_sRNA_length_cutoff: 25
63
64 # sRNA reads cpm cutoff
65 sRNA_expression_cutoff: 1
66
67 # normalization base
68 library_normalization_base: 1000000
69
70
71 # Configure the phase module
72 # predicting with only one reference sequence or multiple reference
  sequences
73 phase:
74   # map file based on reference transcriptome sequence
75   # ** INPUT **
76   mapped_cdna_file: /home/user/test_osa/SRR5049781_processed_cdna.map
77
78   # map file based on reference genome sequence
79   # ** INPUT **
80   mapped_gdna_file: /home/user/test_osa/SRR5049781_processed_gdna.map
81
82   # map file based on full length transcriptome sequence
83   # ** INPUT **
84   mapped_flnc_file:
85
86   # reference transcritome sequence, fasta file
87   # ** INPUT **
88   cdna_fasta: /home/user/test_osa/oryza_sativa_cdna.fa
89
90   # reference genome sequence, fasta file
91   # ** INPUT **
92   gdna_fasta: /home/user/test_osa/oryza_sativa_gdna.fa
93
94   # full length transcriptome sequence, fasta file
95   # ** INPUT **
```

```
96 flnc_fasta:
97
98 # sRNA file
99 # ** INPUT **
100 sRNA_fa: /home/user/test_osa/SRR5049781_trimmed_format_filter.fa
101
102 # allsiRNA cluster output
103 # ** OUTPUT **
104 allsiRNA_cluster_output: /home/user/test_osa/phase_a.txt
105
106 # phasiRNA cluster output file
107 # ** OUTPUT **
108 phasiRNA_cluster_output: /home/user/test_osa/phase_o.txt
109
110 # phasiRNA prediction method, h(hypergeometric test) | p(phase score) | b
    (both)
111 phasiRNA_prediction_method: b
112
113 # phasiRNA cluster island
114 phasiRNA_cluster_island: 5
115
116 # phase length
117 phase_length: 21
118
119 # phase number
120 phase_number_cutoff: 4
121
122 # max hits when mapping to ref sequence
123 bowtie_max_hits_cutoff: 10
124
125 # parallel number
126 parallel_cores: 20
127
128 # pvalue cutoff, only function with h/b method applied
129 pvalue_cutoff: 0.001
130
131 # phase score cutoff, only function with p/b method applied
132 phase_score_cutoff: 15
133
134 # phase ratio cutoff, only function with p/b method applied
135 phase_ratio_cutoff: 0.4
136
137 # delete .phasiHuter_bowtieIndex, y|n
138 delete_index: y
139
140
141 # Configure the integration module
142 integration:
143     # phase module phasiRNA_cluster_output
144     # ** INPUT **
145     o_inputfile: /home/user/test_osa/phase_o.txt
146
147     # phase module allsiRNA_cluster_output
```

```
148 # ** INPUT **
149 a_inputfile: /home/user/test_osa/phase_a.txt
150
151 # reference genome gff3 file
152 # ** INPUT **
153 gff3: /home/user/test_osa/oryza_sativa_gdna.gff3
154
155 # y | n, whether exist gdna based PHAS Loci
156 gdna_based_PHAS_Loci: y
157
158 # integration phasiRNA cluster
159 # ** OUTPUT **
160 integration_phasiRNA_cluster: /home/user/test_osa/integration_o.txt
161
162 # integration all siRNA cluste
163 # ** OUTPUT **
164 integration_allsiRNA_cluster: /home/user/test_osa/integration_a.txt
165
166 # integration summary
167 # ** OUTPUT **
168 integration_summary: /home/user/test_osa/integration_s.txt
169
170 # PHAS Loci information
171 # ** OUTPUT **
172 integration_PHAS_Loci_info: /home/user/test_osa/integration_p.txt
173
174 # alternative splicing/alternative polyadenylation related PHAS gene,
optional
175 as_apa_out: /home/user/test_osa/as_apa_related_result.txt
176
177 # parallel number
178 parallel_cores: 1
179
180 # phase number
181 phase_number_cutoff: 4
182
183 # phase length
184 phase_length: 21
185
186 # pvalue cutoff
187 pvalue_cutoff: 0.001
188
189 # phasiRNA cluster island
190 phasiRNA_cluster_island: 5
191
192 # y | n, discard only P method result
193 discard_only_P_method_result: y
194
195 # full length transcript annotation file
196 flnc_annotation_file:
197
198
199 # Configure the visulization module
```

```

200 visualization:
201     # integration module integration_phasiRNA_cluster
202     # ** INPUT **
203     o_inputfile: /home/user/test_osa/integration_o.txt
204
205     # integration module integration_allsiRNA_cluster
206     # ** INPUT **
207     a_inputfile: /home/user/test_osa/integration_a.txt
208
209     # integration integration_PHAS_Loci_info
210     # ** INPUT **
211     p_inputfile: /home/user/test_osa/integration_p.txt
212
213     # alignment file
214     # ** OUTPUT **
215     output_alignment_file: /home/user/test_osa/alignment.txt
216
217     # phasiRNA fasta file
218     # ** OUTPUT **
219     output_phasiRNA_fa: /home/user/test_osa/phasiRNA.fa
220
221     # PHAS Gene fasta file, Format:
222     >geneid/chr\tphasiRNA_cluster_region(start end)\tseq_region(start end)
223     # ** OUTPUT **
224     output_PHAS_fa: /home/user/test_osa/PHAS.fa
225
226     # phase length
227     phase_length: 21
228
229     # the number for reducing the size of Y-axis
230     Y_axis: 10
231
232     # reference transcriptome sequence, fasta file, enable cdna based
233     phasiRNA.fa, PHAS.fa, Alignmen, Plot output
234     # ** INPUT **
235     cdna_fasta: /home/user/test_osa/oryza_sativa_cdna.fa
236
237     # reference genome sequence, fasta file, enable gdna based phasiRNA.fa,
238     PHAS.fa, Alignmen, Plot output
239     # ** INPUT **
240     gdna_fasta: /home/user/test_osa/oryza_sativa_gdna.fa
241
242     # full length transcriptome sequence, fasta file, enable flnc based
243     phasiRNA.fa, PHAS.fa, Alignmen, Plot output
244     # ** INPUT **
245     flnc_fasta:
246
247     # plot cdna based phasiRNA cluster, y | n
248     plot_cdna_based_phasiRNA_cluster: y

```

```
249 # plot flnc based phasiRNA cluster, y | n
250 plot_flnc_based_phasiRNA_cluster: n
251
252
253 # Configure the target module
254 target:
255     # query miRNA file, fasta format
256     # ** INPUT **
257     query_fa: /home/user/test_osa/osa.miRbase.fa
258
259     # PHAS.fa/transcript.fa, fasta file
260     # ** INPUT **
261     subject_fa: /home/user/test_osa/PHAS.fa
262
263     # output file
264     # ** OUTPUT **
265     output: /home/user/test_osa/miR_target.txt
266
267     # max. total mispairs
268     total_misp: off
269
270     # max. seed mispairs
271     seed_misp: off
272
273     # score cutoff
274     score: 4
275
276     # eTM search
277     mimics: off
278
279     # eTM stringency, (0: strict, 1: relaxed)
280     mimics_str: 0
281
282     # fasta36 threads
283     threads: 10
284
285
286 # Configure the initiator module
287 initiator:
288     # integration module integration_phasiRNA_cluster
289     # ** INPUT **
290     i_input_file: /home/user/test_osa/integration_o.txt
291
292     # the target predicted by psRNAtarget server or target module
293     # ** INPUT **
294     j_input_file: /home/user/test_osa/miR_target.txt
295
296     # integration module integration_PHAS_Loci_info
297     # ** INPUT **
298     p_input_file: /home/user/test_osa/integration_p.txt
299
300     # the microRNA distance away to phase border, default=105(21) or 120 (24)
301     sRNA_distance: 5
```

```

302
303     # 21 or 24, the phase length of 21 or 24,
304     phase_length: 21
305
306     # 0 or 1, the position of cleavage at 10(0) or 9-11 (1)
307     cleavage_shift: 1
308
309     # outputfilename
310     # ** OUTPUT **
311     outputfile: /home/user/test_osa/initiator.txt
312
313
314     # Configure the deg module
315     deg:
316         # mapping file for degradome data mapping transcripts, by bowtie
317         # ** INPUT **
318         inputfile:
319             - /home/user/test_osa/deg/GSM1040649_format_filter.map
320             - /home/user/test_osa/deg/GSM1040650_format_filter.map
321
322         # miRNA sequences used for target prediction, fasta
323         # ** INPUT **
324         query_fa: /home/user/test_osa/osa.miRbase.fa
325
326         # initiator module outputfile
327         # ** INPUT **
328         STI_result: /home/user/test_osa/initiator.txt
329
330         # transcripts file, fasta
331         # ** INPUT **
332         transcript_fa: /home/user/test_osa/oryza_sativa_cdna.fa
333
334         # matched map file with only matched records
335         # filename only, do not input directory
336         # ** OUTPUT **
337         output:
338             - GSM1040649_MTI_deg.txt
339             - GSM1040650_MTI_deg.txt
340
341         # if shifts=0 then cleaved exactly at pos.10
342         shift: 1
343
344         # minum number of degradome reads, int
345         minum_deg_abun: 0
346
347         # enable the plot function, y | n
348         T_plot: y
349
350         # y | n, use initiator output information
351         initiator: y
352
353         # 1,plot only category 1; 2, plot categories 1 and 2
354         plot_categories: 1

```

```
355
356 # output folder name, for exporting t-plot images and outputfile
357 plot_folder: MTI_deg
358
359 # library name
360 library:
361   - GSM1040649
362   - GSM1040650
363
364 # only output cat_1 and cat_2 information
365 less: y
366
367
368 # Configure the phasiRNA_target module
369 phasiRNA_target:
370   # query phasiRNA file, fasta format
371   # ** INPUT **
372   query_fa: /home/user/test_osa/phasiRNA.fa
373
374   # target file, fasta file
375   # ** INPUT **
376   subject_fa: /home/user/test_osa/oryza_sativa_cdna.fa
377
378   # output file
379   # ** OUTPUT **
380   output: /home/user/test_osa/phasiRNA_target.txt
381
382   # max. total mispairs
383   total_misp: off
384
385   # max. seed mispairs
386   seed_misp: off
387
388   # score cutoff
389   score: 4
390
391   # eTM search
392   mimics: off
393
394   # eTM stringency, (0: strict, 1: relaxed)
395   mimics_str: 0
396
397   # fasta36 threads
398   threads: 10
399
400
401 # Configure the phasiRNA_deg module
402 phasiRNA_deg:
403   # mapping file for degradome data mapping transcripts, by bowtie
404   # ** INPUT **
405   inputfile:
406     - /home/user/test_osa/deg/GSM1040649_format_filter.map
407     - /home/user/test_osa/deg/GSM1040650_format_filter.map
```



```

408
409 # phasiRNA sequences used for target prediction, fasta
410 # ** INPUT **
411 query_fa: /home/user/test_osa/phasiRNA.fa
412
413 # psRNATarget/target outputfile
414 # ** INPUT **
415 STI_resultt: /home/user/test_osa/phasiRNA_target.txt
416
417 # transcripts file, fasta
418 # ** INPUT **
419 transcript_fa: /home/user/test_osa/oryza_sativa_cdna.fa
420
421 # matched map file with only matched records
422 # filename only, do not input directory
423 # ** OUTPUT **
424 output:
425     - GSM1040649_PTI_deg.txt
426     - GSM1040650_PTI_deg.txt
427
428 # if shifts=0 then cleaved exactly at pos.10
429 shift: 1
430
431 # minum number of degradome reads, int
432 minum_deg_abun: 0
433
434 # enable the plot function, y | n
435 T_plot: y
436
437 # y | n, use initiator output information, for phasiRNA_deg, it must be n
438 initiator: n
439
440 # 1,plot only category 1; 2, plot categories 1 and 2
441 plot_categories: 1
442
443 # output folder name, for exporting t-plot images and outputfile
444 plot_folder: PTI_deg
445
446 # library name
447 library:
448     - GSM1040649
449     - GSM1040650
450
451 # only output cat_1 and cat_2 information
452 less: y

```

## The main output file

---

- preprocess module
  - preprocessed fasta file
  - alignment file generated by bowtie

- phase module
  - redundant allsiRNA cluster output
    - table header: gene, strand, sRNA\_position, sRNA\_abundance, sRNA\_record, sRNA\_sequence, sRNA\_length, pvalue, phase\_ratio, phase\_number, phase\_abundance, phase\_score, marker
  - redundant phasiRNA cluster output
    - table header: PHAS\_gene, strand, phasiRNA\_position, phasiRNA\_abundance, phasiRNA\_record, phasiRNA\_sequence, phasiRNA\_length, pvalue, phase\_ratio, phase\_number, phase\_abundance, phase\_score, marker
- integration module
  - integrated PHAS loci information
  - integration summary information
  - integrated allsiRNA cluster output
    - table header: gene, strand, sRNA\_position, sRNA\_abundance, sRNA\_record, sRNA\_sequence, sRNA\_length, phase\_ratio, phase\_number, phase\_abundance, phase\_score, pvalue, gene\_annotation, marker
  - integrated phasiRNA cluster output
    - table header: PHAS\_gene, strand, phasiRNA\_position, phasiRNA\_abundance, phasiRNA\_record, phasiRNA\_sequence, phasiRNA\_length, phase\_ratio, phase\_number, phase\_abundance, phase\_score, pvalue, PHAS\_gene\_annotation, marker
  - as\_apa\_related PHAS gene
- visulization module
  - phasiRNA fasta file
    - id description: recorder\_\_PHAS\_gene\_\_position\_\_abundance\_strand\_order
  - PHAS loci fasta file
    - id description: recorder\_\_PHAS\_gene\_\_[start]\_\_[end]\_\_[extend\_start]\_\_[extend\_end]\_\_[marker]
  - phasiRNA alignment result
  - phasiRNA cluster plot
- initiator\_prediction\_and\_verification
  - miRNA-PHAS\_loci interaction output
  - Predicted phase initiator output
  - verified phase initiator output with degradome data
  - degradome verification t-plot
- phasiRNA\_target\_prediction\_and\_verification
  - phasiRNA-target interaction output
  - verified phasiRNA-target interaction with degradome data
  - degradome verification t-plot

Copyright © Crop Bioinformatics Group (CBI), College of Agricultural, Nanjing Agricultural University.

Free for academic use. For commercial use, please contact us ([huangji@njau.edu.cn](mailto:huangji@njau.edu.cn))