PHASER

version 1.0

- Documentation -

1. Scope

PHASER is a tool to analyze the 3’ - 5’ distances of mapped sequence reads. It has been recently described that secondary piRNA biogenesis (piRNA ping-pong) can induce Zucchini-dependent primary processing of targeted transcripts resulting in the production of so-called phased piRNAs (Han et al. 2015, Mohn et al. 2015). In this process, the target molecule is sliced consecutively starting from a ping-pong target site, and each downstream cleavage position determines the 3’ and 5’ end of adjacent (trail-) piRNAs, respectively. The amount of phased piRNAs can be determined when analyzing 3’ - 5’ distances of mapped sequence reads where a distance of 1 indicates a pair of phased piRNAs.

2. Getting started

Running PHASER on your local machine requires the installation of a Perl interpreter. Perl is pre-installed on common Linux and Mac systems. For Windows you can download and install either StrawberryPerl (www.strawberryperl.com) or ActivePerl (www.activestate.com/activeperl/downloads). Before you can use PHASER you must map your sequence reads to a genome or reference sequence. For this, you can use the sRNAmapper tool provided at http://www.smallrnagroup.uni-mainz.de/software.html. Alternatively you can use SeqMap (Yiang and Wong 2008) with the option /output_all_matches. The output file produced by sRNAmapper or SeqMap is the input file for PHASER. You can optionally apply the reallocate tool to apportion read counts according to estimated local transcription rates before using PHASER (Rosenkranz 2015).

```
perl phaser.pl -input input.map [-option value]
```

For example:

```
perl phaser.pl -input input.map -output results.txt -range 150
```

If no output file is specified, PHASER will print the results to STDOUT. You can specify the range of interest [bp] with the option `-range` (`-r`, default=100) which is the maximum distance of mapped sequence reads to be reported in the results.

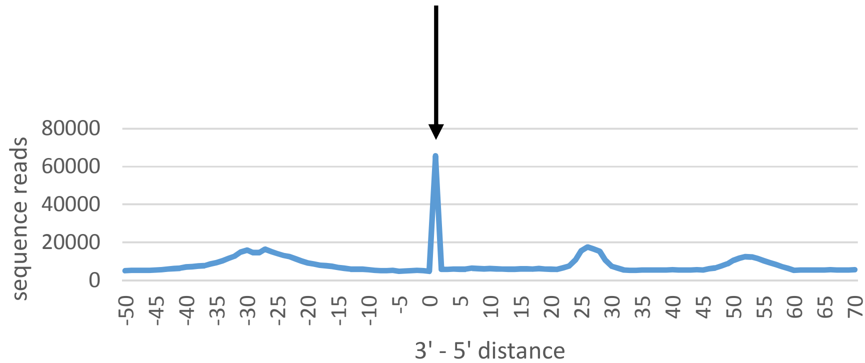
3. Results

PHASER will output a results table that look like this:

```
-10     5640.27288229279
-9      5303.09978049404
-8      5025.98318741945
-7      5146.21762421253
-6      5236.14537412041
-5      4689.74330053654
-4      4949.41064507736
-3      5110.58491286223
-2      5231.13344875739
-1      4997.60828685566
0       4786.87440515233
1       65712.58874425118
2       5659.07840432595
3       5811.22862024895
4       5923.03704194346
5       5907.22518277875
6       5930.50797198025
7       6394.73326087992
8       6260.82587838821
9       6034.44219112269
10      6245.04392236349
```

In the presence of phased piRNAs you will observe a clear peak at 1 bp distance followed by some broadening peaks every ~28 bp (length of a typical piRNA in your dataset). There should be also a rather broad peak from ~-32 to ~-24 which reflects the fact that many piRNAs share identical 5’ ends while showing variation in sequence length.

```
                TACGCTATTAGCGCGCTATATCG
                TACGCTATTAGCGCGCTATATCGCT
                TACGCTATTAGCGCGCTATATCGCTA
                TACGCTATTAGCGCGCTATATCGCTATGCTAGCGATACCGGATAGCGGTTCGATATCGTCGATACGTATAGCGC
                TACGCTATTAGCGCGCTATATCGCTATGCTAGCGATACCGGATAGCGGTTCGATATCGTCGATACGTATAGCGCAGC
5’-GCGTACGGCTATACGCTATTAGCGCGCTATATCGCTATGCTAGCGATACCGGATAGCGGTTCGATATCGTCGATACGTATAGCGCAGCTTG-3’
```



4. Contact

If you have any questions or comments or found any bugs in the software please do not hesitate to contact:

David Rosenkranz
Institute of Anthropology, small RNA group
Johannes Gutenberg University Mainz, Germany
Email: rosenkranz@uni-mainz.de
Web: http://www.smallRNAgroup-mainz.de