Supplementary File 1. Comparison of runtime, sensitivity and accuracy of miRDP2 and other five tools.

To compare runtime and performance of miRDP2 and other five tools, miRDeep-P (Yang and Li, 2011), miRPlant (An, et al., 2014), miR-PREFeR (Lei and Sun, 2014), miRA (Evers, et al., 2015), miReNA (Mathelier and Carbone, 2010), we installed all six tools in cluster server with Cent OS release 6.5 system. These programs are run with same input sequencing files and genomes with 2x Intel Xeon Processor E5-2670 v2 10C 2.5GHz. All programs are run using 1 thread, and 40Gb memory in computing node of our server. Specially, miRPlant is controlled from GUI written in Java and is not able to run on the server. We instead test miRPlant on a PC with Windows 10 system, Intel Core i7-4720HQ 2.6GHz and 16Gb memory. We have also tested miRDP2 and miRDeep-P on this PC. There are no significant difference on time consumption between programs running on PC and on server.

For small genomes as *Arabidopsis thaliana*, *Oryza sativa*, and *Solanum lycopersium*, all the programs could run properly. However, for large genomes of *Zea mays* and *Triticum aestivum* (including *Solanum lycopersium* for miRA), some of the programs depleted all computing resource and break down halfway. miReNA, miRA, and miR-PREFeR have failed to generate results for some or all of the input sequencing files, probably due to memory deficiency while dealing with .sam files or intermediate files. miRPlant has consumed too much space in C:/, possible temporary files, that are not able to run on our PC.

The commands (including preprocessing steps) and parameters for miRA, MIReNA, miR-PREFeR, and miRPlant are listed as the following.

miReNA:
Formatting of reads file using custom perl script.
Running MIReNA.sh with `-D' option.

miR-PREFeR:
Formatting of reads file using custom perl script.
Running miR_PREFeR.py using `-L pipeline' option.

Parameters:
PRECURSOR_LEN = 300
READS_DEPTH_CUTOFF = 20
NUM_OF_CORE = 1
MAX_GAP = 100
MIN_MATURE_LEN = 18
MAX_MATURE_LEN = 24
ALLOW_NO_STAR_EXPRESSION = Y

ALLOW_3NT_OVERHANG = N
CHECKPOINT_SIZE = 3000

miRPlant:
Formatting of reads file using custom perl script.

Parameters:
Adapter =
precursor length = 200
min loop length = 20
flank length = 10
max inconRead ratio = 0.1
miR Lnegth = 18 to 23
min phred = 20
max multimap = 101
min reads = 5
min score = -10

miRA:
Formatting of reads file using custom perl script.
Mapping reads using bowtie, with option -a -v 0 -S.
Running miRA pipeline with `full' option.

Parameters:
log_level = 2
openmp_thread_count = 1
cluster_gap_size = 10
cluster_min_reads = 10
cluster_flank_size = 200
cluster_max_length = 2000
min_precursor_length= 50
max_precursor_length= 0
max_mfe_per_nt = -0.2
max_hairpin_count = 4
min_double_strand_length = 18
permutation_count = 100
max_pvalue = 0.01
min_coverage = 0.01
min_paired_fraction = 0.55
min_duplex_length = 18
max_duplex_length = 30
allow_three_mismatches = 1
allow_two_terminal_mismatches= 1

create_coverage_plots = 1
create_structure_plots = 1
create_structure_coverage_plots = 1
cleanup_auxiliary_files = 1

References

An, J.*, et al.* miRPlant: an integrated tool for identification of plant miRNA from RNA sequencing data. *BMC Bioinformatics* 2014;15:275.

Evers, M.*, et al.* miRA: adaptable novel miRNA identification in plants using small RNA sequencing data. *BMC Bioinformatics* 2015;16:370.

Lei, J. and Sun, Y. miR-PREFeR: an accurate, fast and easy-to-use plant miRNA prediction tool using small RNA-Seq data. *Bioinformatics* 2014;30(19):2837-2839.

Mathelier, A. and Carbone, A. MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data. *Bioinformatics* 2010;26(18):2226-2234.

Yang, X. and Li, L. miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics* 2011;27(18):2614-2615.