|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Supplemental Table S1.** List of small RNA sequencing datasets used in the current study. | | |  | |  | | |  |  |
| ***tissues or cell lines*** | ***species*** | ***genome version*** | | ***References*** | | ***GEO #*** | ***libraries*** |
| Ovary, testes, brain, newborn, embryonic cell | Mus musculus | mm9 (UCSC) | | Chiang, H.R. et al.. Genes Dev, 992-1009, 2010 | | GSE20384 | 37 |
| Small RNA libraries from late embryo, larval, pupal and adult head | Drosophila melanogaster | dm6 (UCSC) | | modENCODE project | | GSE12840 | 11 |
| 7 day wild-type P. patens in protonemata; 14 day old wild-type in protonemata and young Gametophores; ~60 day old wild-type mature in gametophores and sporophytes | Physcomitrella patens | Ppatens\_251\_v3 (JGI) | | Axtell MJ, Snyder JA, Bartel DP.  Plant Cell 2007 Jun;19(6):1750-69. PMID: 17601824 | | GSE5103 | 3 |
| NC stays at 25℃ for 0 hour | Chlamydomonas reinhardtii | Creinhardtii\_236 (JGI) | | current study | | SRP091654 | 1 |
| Small RNAs from Chlamydomonas reinhardtii control, phosphate starvation, and sulphate starvation | Chlamydomonas reinhardtii | Creinhardtii\_236 (JGI) | | Loizeau K, Qu Y, Depp S, Fiechter V et al.  Nucleic Acids Res 2014 Mar;42(5):3286-97. | | GSE32457 | 3 |
| Seedlings of Arabidopsis Col-0 and NRG1 transgenic plants that were grow at 22°C or 30°C for 4 weeks. | Arabidopsis thaliana | TAIR 10 | | Zhong SH, Liu JZ, Jin H, Lin L et al.  Proc Natl Acad Sci U S A 2013 May 28;110(22):9171-6. PMID: 23686579 | | GSM632205, GSM632206 , GSM632207, GSM632208, GSM632209 | 5 |
| Seedlings of root, shoot and panicle. | Oryza sativa | MSU7.0 | | Jeong DH, Park S, et al. Plant Cell, 2011, 4185-207, PMID: 22158467 | | GSE32973 | 5 |

**Supplemental Table S2**. A comparison of features in various miRNA identification software packages.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **name** | **open source/ availability** | **language** | **major RNA abundance** | **stem-loop length** | **secondary structure1** | **duplex2** | **bulge3** | **visualization** |
| miRvial | Yes | Java/Bash | Yes | Yes | Yes | Yes | Yes | Yes |
| Shortstack | Yes | PERL | Yes | Yes | No | Yes | Yes | Yes |
| miRA | Yes | C | Yes | Yes | Yes | No | Yes | Yes |
| miRTRAP | Yes | PERL | Yes | Yes | No | No | No | Yes |
| miRdeep2 | Yes | PERL | Yes | No | No | No | No | No |
| miRanalyzer | Yes\* | JAVA | Yes | Yes | Yes | No | No | No |
| miRDeep-P | Yes | PERL/Bash | Yes | No | No | No | No | No |
| miRPlant | Yes | JAVA | Yes | No | No | No | No | No |
| MIReNA | Yes | PERL/C/Bash | Yes | Yes | Yes | No | No | No |
| MIRCAT | Yes\* | JAVA | Yes | Yes | No | No | No | Yes |
| MIRCAT2 | Yes\* | JAVA | Yes | Yes | No | No | No | Yes |
| MIREAP | Yes | PERL | No | No | No | No | Yes | Yes |

1. Analysis of suboptimal RNA secondary structures

2. Requires presence of a duplex with ~2nt 3' overhang

3. Analyze bulge or unpaired nucleotides within the duplex

\* software but not source codes available

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Supplemental Table S3**. Sequences of primers used in validation of new miRNAs in C. reinhardtii | | | |  | |
| **miRNA** | **genomic loci Creinhardtii\_236 (JGI)** | **5P-F** | **3P-F** | |
| Novel-1 | chromosome\_17:6144067-6144266 [-] | TCGGTCGCCTTGTCTGTTTATG | TAAACAGACAAGGCGACCGACA | |
| Novel-2 | chromosome\_8:347741-347940 [+] | GCGAGGACGAAGAGCTTCTTGATT | CGAGGAGCTCTTCGTCCTCG | |
| Novel-3 | chromosome\_8:121799-121998 [+] | TGGCTTTCGTCGGTCCTAGG | GCTAGGACCGACGAAAGCCACT | |
| Novel-4 | chromosome\_6:6776057-6776256 [+] | GCCAAGACCTTCGAGCTCTACC | GCTAGAGCTCGAAGAACTTGGGA | |
| Novel-5 | chromosome\_5:910409-910608 [+] | CTGGAGGTTCCGTCGGCTACA | TAGCCGATGGAACCCCCAG | |
| Std1 |  | GCTATATGCAAGTCCGGCCATAC |  | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Supplemental Table S4**. The performance of miRvial after relaxing the criterion of presence of a duplex. **#. A** lists the number of true positives predicted by miRvial. **#. B** lists the number of predicted positives. Sensitivity (SE) is true positives divided by the number of known miRNA. Precision (PR) is true positives divided by the number of predicted positive miRNAs. F1 score (F1) is 2 \* precision \* sensitivity divided by (precision + sensitivity), which is the harmonic mean of precision and sensitivity. | | | | | | |
|
|
|  |  |  |  |  |  |  |
|  | known | miRvial | | | | |
| Species | #. A | #. B | SE | PR | F1 score |
| M. musculus | 506 | 503 | 3873 | 99.4% | 13.0% | 23.0% |
| D. melanogaste | 150 | 142 | 3712 | 94.7% | 3.8% | 7.4% |
| P. patens | 98 | 97 | 478 | 99.0% | 20.3% | 33.7% |
| N. crassa | 25 | 15 | 37 | 60.0% | 40.5% | 48.4% |
| C. Reinhardtii | 82 | 79 | 1723 | 96.3% | 4.6% | 8.8% |