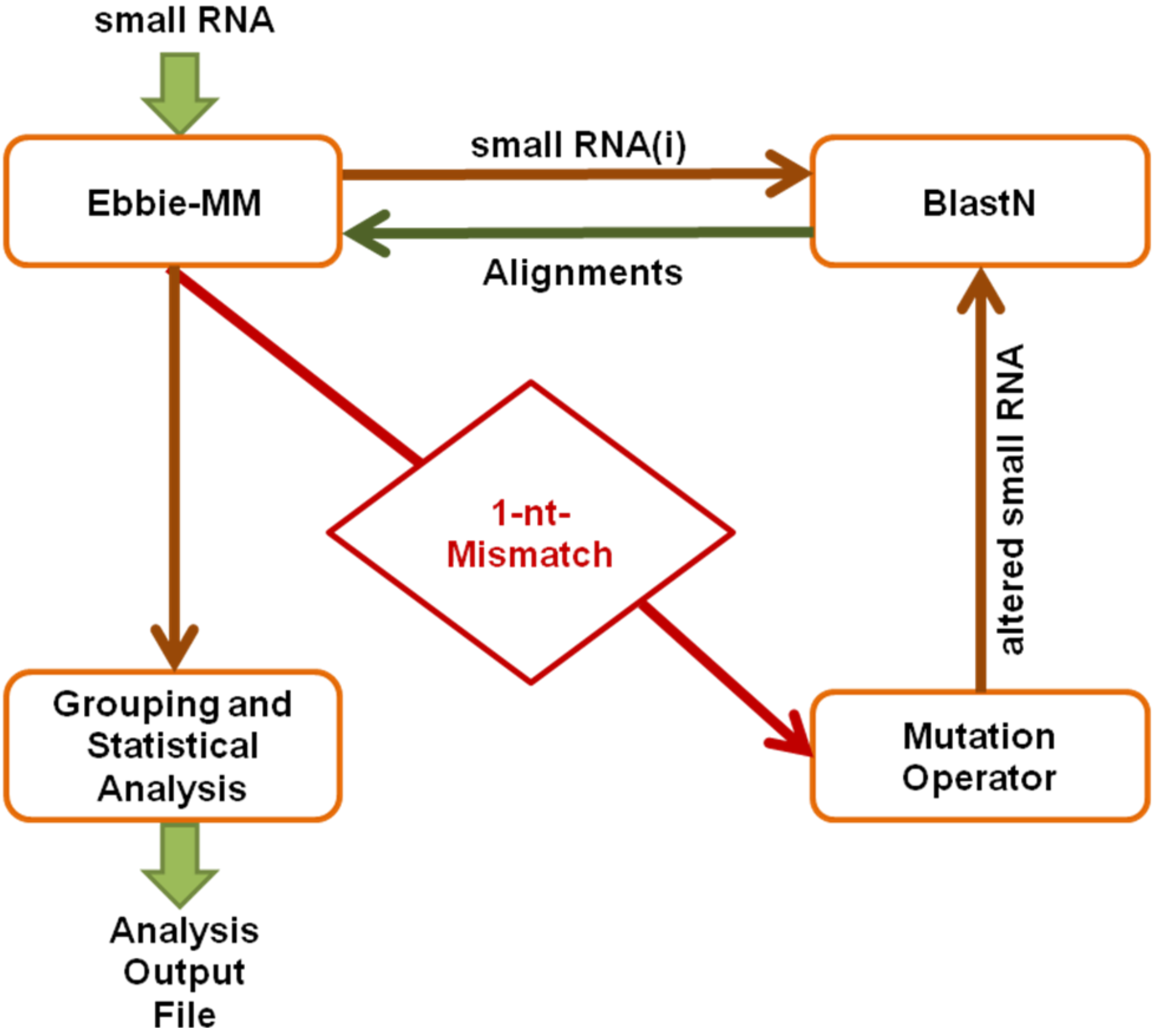
Supplementary Information:

Ebbie-(mis)match Algorithm Description

Design and function of the *Ebbie-(mis)match (Ebbie-MM)* algorithm is shown as flow chart in **Supplementary Figure 1**. *Ebbie-MM* is a software application written in Java 1.6.0, with the pseudo-code describing the algorithm listed in **Supplementary Table 1**. A list of small RNA sequences in FASTA format is prepared as the input to the algorithm. Sequences containing non-ACGUT symbol will also be filtered out and duplicate small RNA sequences will be removed before proceeding. Following the initial screen, BlastN aligns each sequence to its original genome. If a perfect match is found, the event is being counted (fullMatchEvent) and the next sequence will be loaded. In the event where more than one mismatch is found, the algorithm will record the event (othersEvent) and proceed to analyze the next sequence. However, if a single mismatch is found after performing BlastN, a mutation operator will attempt to mutate the small RNA sequence: the mutation operator takes the form of changing the nucleotide where the single mismatch occurred to the other three nucleotides. After a perfect match is successfully found in the same genomic loci as the initial 1-nt-mismatch, the algorithm will record the event (oneMismatchEvent) that lead to a perfect match and continue to the next sequence. *Ebbie-MM* is configured to output a variety of statistics per run. Information such as number of perfect match found, number of one mismatch, and various mutation statistics are reported. While parsing BlastN output, *Ebbie-MM* also removes alignments containing non-ACGUT symbols (including blanks and gap symbol). To speed up *Ebbie-MM*, the BlastN program is continuously kept in memory, reducing redundancy in calling BlastN each time a new sequence is analyzed. We also implemented the -q option as input parameter for BlastN. When aligning the Arabidopsis dataset to its respective micro RNAs, *Ebbie-MM*v4.1 aligned 536 1-nt-mismatched sequences with a default value of -q -3, while *Ebbie-MM*v4.2 using -q -1 aligned 793 1-nt-mismatched sequences in the same experiment. Also, for alignment of the Arabidopsis dataset to all predicted *Arabidopsis thaliana* tRNAs, *Ebbie-MM*v4.2 (-q -1) outperforms *Ebbie-MM*v4.1 (-q -3) with 1,000 and 798 1-nt-mismatch sequences aligned, respectively.

All analysis runs were performed on a 158 node Linux cluster or a numerical server hosted as research support by the Academic Information and Communication Technologies at the University of Alberta. For benchmarking, an SGI Altix XE250 (3 GHz quad core Xeon processor with 64 GB RAM) was used. As benchmark dataset, we used the Arabidopsis dataset as input and all predicted *Arabidopsis thaliana* tRNAs or all recorded micro RNAs in the MirBase v.12 as reference database (**Supplementary Table 2**).

Supplementary Figure 1: Flowchart of *Ebbie-(mis)match* (*Ebbie-MM*). RNA sequences in FASTA format are the input to *Ebbie-MM*. Using BlastN, each sequence is aligned to the genome of origin. For all perfect matches or mismatches greater than 1 nt, *Ebbie-MM* will count these events and go to the next sequence. If a 1-nt-mismatch is found, a mutational operator is activated which alters the mismatched nucleotide position until a perfect match to the same genome location is obtained. *Ebbie-MM* counts all substitutions and reports them at the end of the analysis.



Supplementary Table 1: Structure of the *Ebbie-(mis)match* algorithm in small RNA sequence analysis.

```
sRNA = a collection of small RNA sequences
For (i=1 to NumberOfSmallRNA) do
   alignmentsList = BlastN (sRNA(i))
   If ( FullMatchFound( alignmentList ) )
       Record fullMatchEvent
   Elseif OneMismatchFound( alignmentList) ) then
       Do
            mutateRNA = Mutate (sRNA(i))
            alignmentList = BlastN(mutateRNA)
       While ( ! FullMatchFound(alignmentList) & mutationAvailable )
        If ( FullMatchFound(alignmentList) )
               Record mutationEvent & oneMismatchEvent
        End If
    Else
       Record othersEvent
    End If
End For
Output application statistics
```

Supplementary Table 2: Benchmark analysis of *Ebbie-MM*.

```
Input:  193024 sequences
database: predicted ath-tRNAs
resources_used.cput=12:35:18
resources_used.mem=2221380kb
resources_used.vmem=4196428kb
```

```
Input:  193024 sequences
database: ath mature micro RNAs
resources_used.cput=11:26:55
resources_used.mem=724200kb
resources_used.vmem=4195660kb
```

Supplementary Table 3: 3,852 non-match from *Oryza sativa* and 193,024 from *Arabidopsis thaliana* were aligned to their respectively genomes, analyzed, and their substitution matrix as determined by *Ebbie-MM* is shown. The fourth column indicates substitutions observed due to reverse transcriptase Superscript II in a *lacZ* forward mutation frequency assay.

Substitution Matrix (database: genome)	***Oryza sativa* small RNAs**	***A. thaliana* small RNAs**	***Superscript II errors***
A-to-C	11	803	
A-to-G	56	5622	
A-to-U	28	1730	
C-to-A	19	1133	prevalent
C-to-G	9	697	
C-to-U	59	5237	
G-to-A	64	5997	rare
G-to-C	9	774	
G-to-U	26	1348	
T-to-A	20	1673	
T-to-C	52	5611	
T-to-G	11	666	prevalent

.

Supplementary Table 4: The rice and Arabidopsis datasets were aligned to their respectively predicted tRNAs, analyzed, and their substitution matrix as determined by *Ebbie-MM* is shown.

Substitution Matrix (database: tRNA)	***Oryza sativa***	***A. thaliana***
A-to-C	1	8
A-to-G	6	135
A-to-U	12	210
C-to-A	0	11
C-to-G	0	5
C-to-U	3	44
G-to-A	15	222
G-to-C	3	74
G-to-U	6	142
T-to-A	1	17
T-to-C	5	100
T-to-G	0	32

Supplementary Table 5: The rice and Arabidopsis datasets were aligned to their respective confirmed micro RNAs (MirBase v.12), analyzed, and their substitution matrix as determined by *Ebbie-MM* is shown.

Substitution Matrix (database: miRNA)	***Oryza sativa***	***A. thaliana***
A-to-C	0	22
A-to-G	1	130
A-to-U	0	56
C-to-A	0	34
C-to-G	0	24
C-to-U	3	77
G-to-A	4	192
G-to-C	0	47
G-to-U	1	57
T-to-A	0	37
T-to-C	2	100
T-to-G	0	17

Supplementary Table 6: Small RNAs from the Arabidopsis dataset were compared to published micro RNAs. Alignments of cloned small RNA (upper sequence) and mature micro RNA (lower sequence) are given in column 1. The first letter of the small RNA sequence name indicates the tissue of origin (F: flower; R: root; S: seed; Q: silique). The second column lists the frequency that the modified RNA identified with respect to the unmodified parental micro RNA. Only sequences present at >0.5% of the amount of the parental micro RNA are listed. Occurrences of cloning of the respective parental micro RNA in the tissue from which the small RNA was cloned is listed with the micro RNA name in the third column. The fourth column is the sum of all occurrences that the parental micro RNA was sequenced from all tissues.

<table>
<tr><th>Alignment of small RNAs to mature micro RNAs</th><th>Frequency of Cloning</th><th>MIR-name and tissue specific occurrence of cloning</th><th>MIR cloning occurrence in all tissues.</th></tr>
<tr><td>>F6638|10
UGUCAAAGGAGAUUUGCCCUG (1 -- 21)
||&||||||||||||||||||
UGCCAAAGGAGAUUUGCCCUG (1 -- 21)</td><td>**31.250 %**</td><td>**MIR399a**
32</td><td>**79**</td></tr>
<tr><td>>F303688|32
UGACAGAAGAGAGUGGGCAC (1 -- 20)
|||||||||||||||&||||
UGACAGAAGAGAGUGAGCAC (1 -- 20)</td><td>**14.545 %**</td><td>**MIR156**
220</td><td>**6652**</td></tr>
<tr><td>>F264708|31
UGACAGAAGAGAAUGAGCAC (1 -- 20)
||||||||||||&|||||||
UGACAGAAGAGAGUGAGCAC (1 -- 20)</td><td>14.091 %</td><td>MIR156
220</td><td>6652</td></tr>
<tr><td>>R127189|24
UGACAGAGGAGAGUGAGCAC (1 -- 20)
|||||||&||||||||||||
UGACAGAAGAGAGUGAGCAC (1 -- 20)</td><td>10.909 %</td><td>MIR156
220</td><td>6652</td></tr>
<tr><td>>F314124|23
UGACAGAAAAGAGUGAGCAC (1 -- 20)
||||||||&|||||||||||
UGACAGAAGAGAGUGAGCAC (1 -- 20)</td><td>10.455 %</td><td>MIR156
220</td><td>6652</td></tr>
<tr><td>>F32312|23
UUUGGAUUGAAAGGAGCUCUA (1 -- 21)
|||||||||||&|||||||||
UUUGGAUUGAAGGGAGCUCUA (1 -- 21)</td><td>**10.407 %**</td><td>**MIR159a**
221</td><td>**6621**</td></tr>
<tr><td>>F59737|19
UUUGGAUUGAAGGGAACUCUA (1 -- 21)
|||||||||||||||&|||||
UUUGGAUUGAAGGGAGCUCUA (1 -- 21)</td><td>8.597 %</td><td>MIR159a
221</td><td>6621</td></tr>
<tr><td>>F130626|17
UUUGGAUUGAAGAGAGCUCUA (1 -- 21)
||||||||||||&||||||||
UUUGGAUUGAAGGGAGCUCUA (1 -- 21)</td><td>7.692 %</td><td>MIR159a
221</td><td>6621</td></tr>
</table>

<table>
<tr><td>>F79943|6
UGCGGGAAGCAUUUACACAUG (1 -- 21)
||||||||||||||&||||||
UGCGGGAAGCAUUUGCACAUG (1 -- 21)</td><td>7.407 %</td><td>MIR822
81</td><td>1063</td></tr>
<tr><td>>F120551|16
UGACAGAAGAGAGCGAGCAC (1 -- 20)
|||||||||||||&||||||
UGACAGAAGAGAGUGAGCAC (1 -- 20)</td><td>7.273 %</td><td>MIR156
220</td><td>6652</td></tr>
<tr><td>>F62062|7
UAGCCAAGGAUGACUUACCUG (1 -- 21)
||||||||||||||||&||||
UAGCCAAGGAUGACUUGCCUG (1 -- 21)</td><td>6.931 %</td><td>MIR169h
101</td><td>648</td></tr>
<tr><td>>Q191982|5
UCAUUGAGUGCAGCGCUGAUG (1 -- 21)
|||||||||||||||&|||||
UCAUUGAGUGCAGCGUUGAUG (1 -- 21)</td><td>6.757 %</td><td>MIR397a
74</td><td>110</td></tr>
<tr><td>>F58199|13
UUUGAAUUGAAGGGAGCUCUA (1 -- 21)
||||&||||||||||||||||
UUUGGAUUGAAGGGAGCUCUA (1 -- 21)</td><td>5.882 %</td><td>MIR159a
221</td><td>6621</td></tr>
<tr><td>>F82838|119
UGAAGCUACCAGCAUGAUCUA (1 -- 21)
|||||||&|||||||||||||
UGAAGCUGCCAGCAUGAUCUA (1 -- 21)</td><td>4.798 %</td><td>MIR167a
2480</td><td>35681</td></tr>
<tr><td>>Q215750|5
UCAUUGAGUGCGUCGUUGAUG (1 -- 21)
|||||||||||&|||||||||
UCAUUGAGUGCAUCGUUGAUG (1 -- 21)</td><td>4.545 %</td><td>MIR397b
110</td><td>251</td></tr>
<tr><td>>R13207|10
UGACGGAAGAGAGUGAGCAC (1 -- 20)
||||&|||||||||||||||
UGACAGAAGAGAGUGAGCAC (1 -- 20)</td><td>4.545 %</td><td>MIR156
220</td><td>6652</td></tr>
<tr><td>>F74265|10
UUUGGAUUAAAGGGAGCUCUA (1 -- 21)
||||||||&||||||||||||
UUUGGAUUGAAGGGAGCUCUA (1 -- 21)</td><td>4.525 %</td><td>MIR159a
221</td><td>6621</td></tr>
<tr><td>>F5985|108
UGAAGCUGCCAACAUGAUCUA (1 -- 21)
|||||||||||&|||||||||
UGAAGCUGCCAGCAUGAUCUA (1 -- 21)</td><td>4.355%</td><td>MIR167a
2480</td><td>35681</td></tr>
<tr><td>>F6129|102
UGAAACUGCCAGCAUGAUCUA (1 -- 21)
||||&||||||||||||||||
UGAAGCUGCCAGCAUGAUCUA (1 -- 21)</td><td>4.113 %</td><td>MIR167a
2480</td><td>35681</td></tr>
<tr><td>>F12117|53
UGGAGAAGCAGGACACGUGCA (1 -- 21)
||||||||||||&||||||||
UGGAGAAGCAGGGCACGUGCA (1 -- 21)</td><td>3.967 %</td><td>MIR164a
1336</td><td>6857</td></tr>
<tr><td>>F25190|33
UGAAAGUGGCUACAUCGGGGU (1 -- 21)
||||||||&||||||||||||
UGAAAGUGACUACAUCGGGGU (1 -- 21)</td><td>3.952 %</td><td>MIR161
835</td><td>10573</td></tr>
<tr><td>>F84238|32
UGAAAAUGACUACAUCGGGGU (1 -- 21)
|||||&|||||||||||||||
UGAAAGUGACUACAUCGGGGU (1 -- 21)</td><td>3.832 %</td><td>MIR161
835</td><td>10573</td></tr>
</table>

<table>
<tr><td>>F21516|95
UGAAGCCGCCAGCAUGAUCUA (1 -- 21)
||||||&||||||||||||||
UGAAGCUGCCAGCAUGAUCUA (1 -- 21)</td><td>3.831 %</td><td>MIR167a
2480</td><td>35681</td></tr>
<tr><td>>F66516|80
UGAAGCUGCCAGCAUGGUCUA (1 -- 21)
||||||||||||||||&||||
UGAAGCUGCCAGCAUGAUCUA (1 -- 21)</td><td>3.226 %</td><td>MIR167a
2480</td><td>35681</td></tr>
<tr><td>>R305334|7
UGACAGAAGAGGGUGAGCAC (1 -- 20)
|||||||||||&||||||||
UGACAGAAGAGAGUGAGCAC (1 -- 20)</td><td>3.182 %</td><td>MIR156
220</td><td>6652</td></tr>
<tr><td>>F40228|76
UGAAGCUGCCAGCAUAAUCUA (1 -- 21)
|||||||||||||||&|||||
UGAAGCUGCCAGCAUGAUCUA (1 -- 21)</td><td>3.065 %</td><td>MIR167a
2480</td><td>35681</td></tr>
<tr><td>>F23089|53
UGAUUGAGCCGCACCAAUAUC (1 -- 21)
||||||||||||&||||||||
UGAUUGAGCCGCGCCAAUAUC (1 -- 21)</td><td>2.923 %</td><td>MIR171a
1813</td><td>7571</td></tr>
<tr><td>>F13604|65
UGAAGCUGCCAGCACGAUCUA (1 -- 21)
||||||||||||||&||||||
UGAAGCUGCCAGCAUGAUCUA (1 -- 21)</td><td>2.621 %</td><td>MIR167a
2480</td><td>35681</td></tr>
<tr><td>>F131735|19
UGAAAGUAACUACAUCGGGGU (1 -- 21)
|||||||&|||||||||||||
UGAAAGUGACUACAUCGGGGU (1 -- 21)</td><td>2.275 %</td><td>MIR161
835</td><td>10573</td></tr>
<tr><td>>F235549|19
UGAAAGCGACUACAUCGGGGU (1 -- 21)
||||||&||||||||||||||
UGAAAGUGACUACAUCGGGGU (1 -- 21)</td><td>2.275 %</td><td>MIR161
835</td><td>10573</td></tr>
<tr><td>>R67763|6
UGAUCGAGCCGUGUCAAUAUC (1 -- 21)
||||&||||||||||||||||
UGAUUGAGCCGUGUCAAUAUC (1 -- 21)</td><td>2.151 %</td><td>MIR170
279</td><td>1378</td></tr>
<tr><td>>F6836|24
UGCCUGGCUCCCUAUAUGCCA (1 -- 21)
|||||||||||||&|||||||
UGCCUGGCUCCCUGUAUGCCA (1 -- 21)</td><td>2.020 %</td><td>MIR160
1188</td><td>5891</td></tr>
<tr><td>>S128677|5
AGAAUCUUGAUGAUGUUGCAU (1 -- 21)
|||||||||||||||&|||||
AGAAUCUUGAUGAUGCUGCAU (1 -- 21)</td><td>1.880 %</td><td>MIR172
266</td><td>4189</td></tr>
<tr><td>>R120994|12
UUUGGAUUGAAGAGAGCUCUU (1 -- 21)
||||||||||||&||||||||
UUUGGAUUGAAGGGAGCUCUU (1 -- 21)</td><td>1.875 %</td><td>MIR159b
640</td><td>982</td></tr>
<tr><td>>F8018|21
UCGGACCAAGCUUCAUUCCCC (1 -- 21)
||||||||&||||||||||||
UCGGACCAGGCUUCAUUCCCC (1 -- 21)</td><td>1.727 %</td><td>MIR166
1216</td><td>3300</td></tr>
<tr><td>>F19104|22
UGGAGAAGCAGGGCACAUGCA (1 -- 21)
||||||||||||||||&||||
UGGAGAAGCAGGGCACGUGCA (1 -- 21)</td><td>1.647 %</td><td>MIR164a
1336</td><td>6857</td></tr>
</table>

<table>
<tr><td>>F13950|29
UGAUUAAGCCGCGCCAAUAUC (1 -- 21)
|||||&|||||||||||||||
UGAUUGAGCCGCGCCAAUAUC (1 -- 21)</td><td>1.600 %</td><td>MIR171a
1813</td><td>7571</td></tr>
<tr><td>>F20312|28
UGAUUGAACCGCGCCAAUAUC (1 -- 21)
|||||||&|||||||||||||
UGAUUGAGCCGCGCCAAUAUC (1 -- 21)</td><td>1.544 %</td><td>MIR171a
1813</td><td>7571</td></tr>
<tr><td>>F2423|20
UGGAGAAACAGGGCACGUGCA (1 -- 21)
|||||||&|||||||||||||
UGGAGAAGCAGGGCACGUGCA (1 -- 21)</td><td>1.497 %</td><td>MIR164a
1336</td><td>6857</td></tr>
<tr><td>>F201853|35
UGAAGCUGCUAGCAUGAUCUA (1 -- 21)
|||||||||&|||||||||||
UGAAGCUGCCAGCAUGAUCUA (1 -- 21)</td><td>1.411 %</td><td>MIR167a
2480</td><td>35681</td></tr>
<tr><td>>F102678|17
UGGAGAAGCAGGGCGCGUGCA (1 -- 21)
||||||||||||||&||||||
UGGAGAAGCAGGGCACGUGCA (1 -- 21)</td><td>1.272 %</td><td>MIR164a
1336</td><td>6857</td></tr>
<tr><td>>F296568|17
UGGAGAAGCGGGGCACGUGCA (1 -- 21)
|||||||||&|||||||||||
UGGAGAAGCAGGGCACGUGCA (1 -- 21)</td><td>1.272 %</td><td>MIR164a
1336</td><td>6857</td></tr>
<tr><td>>F31816|15
UGCCCGGCUCCCUGUAUGCCA (1 -- 21)
||||&||||||||||||||||
UGCCUGGCUCCCUGUAUGCCA (1 -- 21)</td><td>1.263 %</td><td>MIR160
1188</td><td></td></tr>
<tr><td>>F11872|15
UCGGACCAGACUUCAUUCCCC (1 -- 21)
|||||||||&|||||||||||
UCGGACCAGGCUUCAUUCCCC (1 -- 21)</td><td>1.234 %</td><td>MIR166
1216</td><td>3300</td></tr>
<tr><td>>F199187|29
UGAAGCUGUCAGCAUGAUCUA (1 -- 21)
||||||||&||||||||||||
UGAAGCUGCCAGCAUGAUCUA (1 -- 21)</td><td>1.169 %</td><td>MIR167a
2480</td><td>35681</td></tr>
<tr><td>>F60317|25
UGAAGUUGCCAGCAUGAUCUA (1 -- 21)
|||||&|||||||||||||||
UGAAGCUGCCAGCAUGAUCUA (1 -- 21)</td><td>1.008 %</td><td>MIR167a
2480</td><td>35681</td></tr>
<tr><td>>F36704|18
UGAUUGGGCCGCGCCAAUAUC (1 -- 21)
||||||&||||||||||||||
UGAUUGAGCCGCGCCAAUAUC (1 -- 21)</td><td>0.993 %</td><td>MIR171a
1813</td><td>7571</td></tr>
<tr><td>>F35494|11
UCGGACCAGGCUUCACUCCCC (1 -- 21)
|||||||||||||||&|||||
UCGGACCAGGCUUCAUUCCCC (1 -- 21)</td><td>0.905 %</td><td>MIR166
1216</td><td>3300</td></tr>
<tr><td>>F77837|12
UGGAGAAGCAGAGCACGUGCA (1 -- 21)
|||||||||||&|||||||||
UGGAGAAGCAGGGCACGUGCA (1 -- 21)</td><td>0.898 %</td><td>MIR164a
1336</td><td>6857</td></tr>
<tr><td>>F239860|10
UGCCUGGCCCCCUGUAUGCCA (1 -- 21)
||||||||&||||||||||||
UGCCUGGCUCCCUGUAUGCCA (1 -- 21)</td><td>0.842 %</td><td>MIR160
1188</td><td>5891</td></tr>
</table>

<table>
<tr><td>>F226618|7
UGAAAGUGACUAUAUCGGGGU (1 -- 21)
||||||||||||&||||||||
UGAAAGUGACUACAUCGGGGU (1 -- 21)</td><td>0.838 %</td><td>MIR161
835</td><td>10573</td></tr>
<tr><td>>R105543|6
UUUGGACUGAAGGGAGCUCUU (1 -- 21)
||||||&||||||||||||||
UUUGGAUUGAAGGGAGCUCUU (1 -- 21)</td><td>0.838 %</td><td>MIR159b
640</td><td>982</td></tr>
<tr><td>>F13301|11
UGGAAAAGCAGGGCACGUGCA (1 -- 21)
||||&||||||||||||||||
UGGAGAAGCAGGGCACGUGCA (1 -- 21)</td><td>0.823 %</td><td>MIR164a
1336</td><td>6857</td></tr>
<tr><td>>F42449|10
UGGAGAAGUAGGGCACGUGCA (1 -- 21)
||||||||&||||||||||||
UGGAGAAGCAGGGCACGUGCA (1 -- 21)</td><td>0.749 %</td><td>MIR164a
1336</td><td>6857</td></tr>
<tr><td>>F106104|9
UCGGACCAGGCUUCGUUCCCC (1 -- 21)
||||||||||||||&||||||
UCGGACCAGGCUUCAUUCCCC (1 -- 21)</td><td>0.740 %</td><td>MIR166
1216</td><td>3300</td></tr>
<tr><td>>F23216|13
UGAUUGAGCCGCGCCAGUAUC (1 -- 21)
||||||||||||||||&||||
UGAUUGAGCCGCGCCAAUAUC (1 -- 21)</td><td>0.717 %</td><td>MIR171a
1813</td><td>7571</td></tr>
<tr><td>>F269395|6
AGAAUCUUGGUGAUGCUGCAU (1 -- 21)
|||||||||&|||||||||||
AGAAUCUUGAUGAUGCUGCAU (1 -- 21)</td><td>0.713 %</td><td>MIR172
841</td><td>4189</td></tr>
<tr><td>>R8810|32
UGAAAGUGACUACGUCGGGGU (1 -- 21)
|||||||||||||&|||||||
UGAAAGUGACUACAUCGGGGU (1 -- 21)</td><td>0.697 %</td><td>MIR161
4593</td><td>10573</td></tr>
<tr><td>>F61240|8
UGCCUGGCUUCCUGUAUGCCA (1 -- 21)
|||||||||&|||||||||||
UGCCUGGCUCCCUGUAUGCCA (1 -- 21)</td><td>0.673 %</td><td>MIR160
1188</td><td>5891</td></tr>
<tr><td>>F30889|12
UGAUUGAGCCGCGCCGAUAUC (1 -- 21)
|||||||||||||||&|||||
UGAUUGAGCCGCGCCAAUAUC (1 -- 21)</td><td>0.662 %</td><td>MIR171a
1813</td><td>7571</td></tr>
<tr><td>>R911|22
UCGCUUGGCGCAGGUCGGGAA (1 -- 21)
||||||||&||||||||||||
UCGCUUGGUGCAGGUCGGGAA (1 -- 21)</td><td>0.654 %</td><td>MIR168
3364</td><td>5039</td></tr>
<tr><td>>R166606|8
UGGAGAGGCAGGGCACGUGCA (1 -- 21)
||||||&||||||||||||||
UGGAGAAGCAGGGCACGUGCA (1 -- 21)</td><td>0.599 %</td><td>MIR164a
1336</td><td>6857</td></tr>
<tr><td>>R96325|5
AGAAUCUUGAUGAUGCCGCAU (1 -- 21)
||||||||||||||||&||||
AGAAUCUUGAUGAUGCUGCAU (1 -- 21)</td><td>0.599 %</td><td>MIR172a
835</td><td>4189</td></tr>
<tr><td>>F110613|5
AGAAUCUUGAUAAUGCUGCAU (1 -- 21)
|||||||||||&|||||||||
AGAAUCUUGAUGAUGCUGCAU (1 -- 21)</td><td>0.595 %</td><td>MIR172
841</td><td>4189</td></tr>
</table>

<table>
<tr><td>>R84800|12
UGUGUCCUCAGGUCACCCCUG (1 -- 21)
|||||&|||||||||||||||
UGUGUUCUCAGGUCACCCCUG (1 -- 21)</td><td>0.529 %</td><td>MIR398b
2267</td><td>3932</td></tr>
<tr><td>>R9804|12
UGUGUUCUCAGGCCACCCCUG (1 -- 21)
||||||||||||&||||||||
UGUGUUCUCAGGUCACCCCUG (1 -- 21)</td><td>0.529 %</td><td>MIR398b
2267</td><td>3932</td></tr>
</table>

Supplementary Table 7: Output of *Ebbie-MM-ago*. The perl script is searching for small RNAs of a specific sequence in the Mi-dataset which contains cloning frequencies of small RNAs from their associated AGO complexes. Then *Ebbie-MM-ago* removes the 5' nucleotide and adds UU to the 3' terminus and searches again the Mi-dataset. As input, all *Arabidopsis thaliana* micro RNAs from MirBase v12 were used.

```
>ath-miR822 MIMAT0004239
TGCGGGAAGCATTTGCACATG

TGCGGGAAGCATTTGCACATG* is found in AGO1 1065 times
TGCGGGAAGCATTTGCACATG* is found in AGO2 5 times
TGCGGGAAGCATTTGCACATG* is found in AGO4 8 times
TGCGGGAAGCATTTGCACATG* is found in AGO5 19 times
minus 1 plus UU
GCGGGAAGCATTTGCACATGTT* is found in AGO1 124 times
GCGGGAAGCATTTGCACATGTT* is found in AGO2 0 times
GCGGGAAGCATTTGCACATGTT* is found in AGO4 120 times
GCGGGAAGCATTTGCACATGTT* is found in AGO5 1 times
9*****************************************************

>ath-miR156g MIMAT0001012
CGACAGAAGAGAGTGAGCAC

CGACAGAAGAGAGTGAGCAC* is found in AGO1 5 times
CGACAGAAGAGAGTGAGCAC* is found in AGO2 0 times
CGACAGAAGAGAGTGAGCAC* is found in AGO4 1 times
CGACAGAAGAGAGTGAGCAC* is found in AGO5 0 times
minus 1 plus UU
GACAGAAGAGAGTGAGCACTT* is found in AGO1 53 times
GACAGAAGAGAGTGAGCACTT* is found in AGO2 0 times
GACAGAAGAGAGTGAGCACTT* is found in AGO4 0 times
GACAGAAGAGAGTGAGCACTT* is found in AGO5 0 times
39*****************************************************

>ath-miR158b MIMAT0001014
CCCCAAATGTAGACAAAGCA

CCCCAAATGTAGACAAAGCA* is found in AGO1 91 times
CCCCAAATGTAGACAAAGCA* is found in AGO2 1 times
CCCCAAATGTAGACAAAGCA* is found in AGO4 1 times
CCCCAAATGTAGACAAAGCA* is found in AGO5 4 times
minus 1 plus UU
CCCAAATGTAGACAAAGCATT* is found in AGO1 52 times
CCCAAATGTAGACAAAGCATT* is found in AGO2 0 times
```

```
CCCAAATGTAGACAAAGCATT* is found in AGO4 0 times
CCCAAATGTAGACAAAGCATT* is found in AGO5 0 times
71*****************************************************

>ath-miR399f MIMAT0000956
TGCCAAAGGAGATTTGCCCGG

TGCCAAAGGAGATTTGCCCGG* is found in AGO1 99 times
TGCCAAAGGAGATTTGCCCGG* is found in AGO2 0 times
TGCCAAAGGAGATTTGCCCGG* is found in AGO4 2 times
TGCCAAAGGAGATTTGCCCGG* is found in AGO5 1 times
minus 1 plus UU
GCCAAAGGAGATTTGCCCGGTT* is found in AGO1 6 times
GCCAAAGGAGATTTGCCCGGTT* is found in AGO2 0 times
GCCAAAGGAGATTTGCCCGGTT* is found in AGO4 0 times
GCCAAAGGAGATTTGCCCGGTT* is found in AGO5 0 times
95*****************************************************

>ath-miR172b MIMAT0000205
AGAATCTTGATGATGCTGCAT

AGAATCTTGATGATGCTGCAT* is found in AGO1 9796 times
AGAATCTTGATGATGCTGCAT* is found in AGO2 470 times
AGAATCTTGATGATGCTGCAT* is found in AGO4 1793 times
AGAATCTTGATGATGCTGCAT* is found in AGO5 717 times
minus 1 plus UU
GAATCTTGATGATGCTGCATTT* is found in AGO1 2 times
GAATCTTGATGATGCTGCATTT* is found in AGO2 5 times
GAATCTTGATGATGCTGCATTT* is found in AGO4 14 times
GAATCTTGATGATGCTGCATTT* is found in AGO5 9 times
131*****************************************************

>ath-miR780.1 MIMAT0004218
TCTAGCAGCTGTTGAGCAGGT

TCTAGCAGCTGTTGAGCAGGT* is found in AGO1 17 times
TCTAGCAGCTGTTGAGCAGGT* is found in AGO2 0 times
TCTAGCAGCTGTTGAGCAGGT* is found in AGO4 1 times
TCTAGCAGCTGTTGAGCAGGT* is found in AGO5 33 times
minus 1 plus UU
CTAGCAGCTGTTGAGCAGGTTT* is found in AGO1 1 times
CTAGCAGCTGTTGAGCAGGTTT* is found in AGO2 0 times
CTAGCAGCTGTTGAGCAGGTTT* is found in AGO4 0 times
CTAGCAGCTGTTGAGCAGGTTT* is found in AGO5 4 times
141*****************************************************

>ath-miR157b MIMAT0000173
```

```
TTGACAGAAGATAGAGAGCAC

TTGACAGAAGATAGAGAGCAC* is found in AGO1 5104 times
TTGACAGAAGATAGAGAGCAC* is found in AGO2 2 times
TTGACAGAAGATAGAGAGCAC* is found in AGO4 1059 times
TTGACAGAAGATAGAGAGCAC* is found in AGO5 86 times
minus 1 plus UU
TGACAGAAGATAGAGAGCACTT* is found in AGO1 104 times
TGACAGAAGATAGAGAGCACTT* is found in AGO2 1 times
TGACAGAAGATAGAGAGCACTT* is found in AGO4 5 times
TGACAGAAGATAGAGAGCACTT* is found in AGO5 0 times
143*****************************************************

>ath-miR157d MIMAT0000175
TGACAGAAGATAGAGAGCAC

TGACAGAAGATAGAGAGCAC* is found in AGO1 325 times
TGACAGAAGATAGAGAGCAC* is found in AGO2 1 times
TGACAGAAGATAGAGAGCAC* is found in AGO4 9 times
TGACAGAAGATAGAGAGCAC* is found in AGO5 4 times
minus 1 plus UU
GACAGAAGATAGAGAGCACTT* is found in AGO1 263 times
GACAGAAGATAGAGAGCACTT* is found in AGO2 0 times
GACAGAAGATAGAGAGCACTT* is found in AGO4 1 times
GACAGAAGATAGAGAGCACTT* is found in AGO5 0 times
167*****************************************************

>ath-miR157a MIMAT0000172
TTGACAGAAGATAGAGAGCAC

TTGACAGAAGATAGAGAGCAC* is found in AGO1 5104 times
TTGACAGAAGATAGAGAGCAC* is found in AGO2 2 times
TTGACAGAAGATAGAGAGCAC* is found in AGO4 1059 times
TTGACAGAAGATAGAGAGCAC* is found in AGO5 86 times
minus 1 plus UU
TGACAGAAGATAGAGAGCACTT* is found in AGO1 104 times
TGACAGAAGATAGAGAGCACTT* is found in AGO2 1 times
TGACAGAAGATAGAGAGCACTT* is found in AGO4 5 times
TGACAGAAGATAGAGAGCACTT* is found in AGO5 0 times
173*****************************************************

>ath-miR156a MIMAT0000166
TGACAGAAGAGAGTGAGCAC

TGACAGAAGAGAGTGAGCAC* is found in AGO1 3007 times
TGACAGAAGAGAGTGAGCAC* is found in AGO2 1 times
TGACAGAAGAGAGTGAGCAC* is found in AGO4 62 times
```

```
TGACAGAAGAGAGTGAGCAC* is found in AGO5 8 times
minus 1 plus UU
GACAGAAGAGAGTGAGCACTT* is found in AGO1 53 times
GACAGAAGAGAGTGAGCACTT* is found in AGO2 0 times
GACAGAAGAGAGTGAGCACTT* is found in AGO4 0 times
GACAGAAGAGAGTGAGCACTT* is found in AGO5 0 times
195*****************************************************

>ath-miR391 MIMAT0000933
TTCGCAGGAGAGATAGCGCCA

TTCGCAGGAGAGATAGCGCCA* is found in AGO1 560 times
TTCGCAGGAGAGATAGCGCCA* is found in AGO2 2 times
TTCGCAGGAGAGATAGCGCCA* is found in AGO4 94 times
TTCGCAGGAGAGATAGCGCCA* is found in AGO5 0 times
minus 1 plus UU
TCGCAGGAGAGATAGCGCCATT* is found in AGO1 0 times
TCGCAGGAGAGATAGCGCCATT* is found in AGO2 0 times
TCGCAGGAGAGATAGCGCCATT* is found in AGO4 27 times
TCGCAGGAGAGATAGCGCCATT* is found in AGO5 0 times
271*****************************************************

>ath-miR172e MIMAT0001019
GGAATCTTGATGATGCTGCAT

GGAATCTTGATGATGCTGCAT* is found in AGO1 543 times
GGAATCTTGATGATGCTGCAT* is found in AGO2 2 times
GGAATCTTGATGATGCTGCAT* is found in AGO4 56 times
GGAATCTTGATGATGCTGCAT* is found in AGO5 16 times
minus 1 plus UU
GAATCTTGATGATGCTGCATTT* is found in AGO1 2 times
GAATCTTGATGATGCTGCATTT* is found in AGO2 5 times
GAATCTTGATGATGCTGCATTT* is found in AGO4 14 times
GAATCTTGATGATGCTGCATTT* is found in AGO5 9 times
303*****************************************************

>ath-miR156f MIMAT0000171
TGACAGAAGAGAGTGAGCAC

TGACAGAAGAGAGTGAGCAC* is found in AGO1 3007 times
TGACAGAAGAGAGTGAGCAC* is found in AGO2 1 times
TGACAGAAGAGAGTGAGCAC* is found in AGO4 62 times
TGACAGAAGAGAGTGAGCAC* is found in AGO5 8 times
minus 1 plus UU
GACAGAAGAGAGTGAGCACTT* is found in AGO1 53 times
GACAGAAGAGAGTGAGCACTT* is found in AGO2 0 times
GACAGAAGAGAGTGAGCACTT* is found in AGO4 0 times
```

```
GACAGAAGAGAGTGAGCACTT* is found in AGO5 0 times
409*****************************************************

>ath-MIR408     25
ATGCACTGCCTCTTCCCTGGC

ATGCACTGCCTCTTCCCTGGC* is found in AGO1 6 times
ATGCACTGCCTCTTCCCTGGC* is found in AGO2 549 times
ATGCACTGCCTCTTCCCTGGC* is found in AGO4 0 times
ATGCACTGCCTCTTCCCTGGC* is found in AGO5 12 times
minus 1 plus UU
TGCACTGCCTCTTCCCTGGCTT* is found in AGO1 21 times
TGCACTGCCTCTTCCCTGGCTT* is found in AGO2 1 times
TGCACTGCCTCTTCCCTGGCTT* is found in AGO4 0 times
TGCACTGCCTCTTCCCTGGCTT* is found in AGO5 1 times
18*******************************************************
>F42411 91
TGCACTGCCTCTTCCCTGGCTT

TGCACTGCCTCTTCCCTGGCTT* is found in AGO1 21 times
TGCACTGCCTCTTCCCTGGCTT* is found in AGO2 1 times
TGCACTGCCTCTTCCCTGGCTT* is found in AGO4 0 times
TGCACTGCCTCTTCCCTGGCTT* is found in AGO5 1 times
minus 1 plus UU
GCACTGCCTCTTCCCTGGCTTTT* is found in AGO1 0 times
GCACTGCCTCTTCCCTGGCTTTT* is found in AGO2 0 times
GCACTGCCTCTTCCCTGGCTTTT* is found in AGO4 0 times
GCACTGCCTCTTCCCTGGCTTTT* is found in AGO5 0 times
20*******************************************************
```