

APPENDIX A

Compiled script to remove redundancy

```
#To remove redundancy in fasta file

use strict;

use warnings;

#read the file into a hash

my %seq;

my $title;

my $infile=shift or die "give me a infile\n";

open (IN,"$infile");

while (<IN>){

    $_=~s/\n//;

    $_=~s/\r//;

    if ($_~/>){

        $title=$_;

        $title=~s/>//;

    }

    else{

        $seq{$_}=$title;

    }

}

close IN;

#remove the abundant sequences

my @seq=keys (%seq);

my @uniqueseq;

my $find=0;

foreach (@seq){

    $find=0;

    my $seq=uc($_);
```

```

foreach (@uniqueseq){
    #if ($seq=~/$_/){
    #    $_=$seq;#replace with longer seq
    #    $find=1;
    #}
    if ($_ =~/$seq/){
        $find=1;
    }
}
if ($find==0){
    push @uniqueseq,$seq;
}
}
#outout the final result
open (OUT,">output.fasta");
foreach (@uniqueseq){
    print OUT ">$seq{$_}\n$_\n";
}
close OUT;

```

**Compiled perl script to convert sequences and to execute targetfinder.pl program in
CleaveLand**

```
if ($#ARGV !=2)
{
print "Wrong Syntax!!!Usage as below:\nperl convert.pl FASTAFILENAME.fa FASTA-
FORMATED-SEQ-File.fasta OutputScriptName.sh\n";
exit;
}
open (FILE,$ARGV[0]);
open (OUTPUTFILE, ">>$ARGV[2]");
$oddcount=0;
while (<FILE>) {
chomp;
$oddcount=$oddcount+1;
if($oddcount%2!=0)
{
($id,$no)=split("\t");
$seq=~s/>/;/;
}
else
{
$seq=$_;
print "Before Convert:\t";
print "ID: $id\t";
print "Seq: $seq\t";
print "No: $no\n";
$seq=~s/T/U/g;
print "After Convert:\t";
print "ID: $id\t";
print "Seq: $seq\t";
```

```
print "No: $no\n";  
#Generate script to execute list of targetfinder.pl program for converted Seq  
print OUTPUTFILE "perl targetfinder.pl -s $seq -d $ARGV[1] -q $id>$id.txt\n";  
}  
}  
close(OUTPUTFILE);  
close(FILE);  
system "chmod +x $ARGV[2]";  
exit;
```

APPENDIX B

Linux Commands

Linux Commands for CleaveLand

CleaveLand3_map2dd.pl

a. usage: perl CleaveLand3_map2dd.pl -d[map_data_file] -t[txome_FASTA_file] -s [length_of_degradome_reads] -f[format] >[deg_density_file] 2>[run_info_file]

i. map_data_file used bowtie format

ii. txome_FASTA_file: The transcriptome that was mapped to FASTA format.

iii. length_of_degradome_reads: 21

iv. format: bowtie

CleaveLand3_analysis.pl

a. Usage: perl CleaveLand3_analysis.pl -d[degradome_density_file] -t[alignment_directory] -p[p-value-cutoff] >[table_output] 2>[log]

i. degradome_density_file: Generated in step above

ii. alignment_directory: The path to a directory containing all targetfinder.pl derived result files that need to be examine.

iii. p-value-cutoff: 0.05

Linux command for SeqTar

Command to run SeqTar if using BLAST

```
$java -cp ./JSmallRNA.jar jsmallrna.seqtar.SeqTar -t all.cDNA -i miRNA.fa -b all-norm-vs-cdna-sorted.blast -l 20 -p 100 -o miRNA-cdna.seqtar -s 60 -m 0.1 > miRNA-vs-cdna.log
```

Command to run SeqTar if using SOAP2 for alignment

```
$java -cp ./JSmallRNA.jar jsmallrna.seqtar.SeqTarSOAP -t all.cDNA -i miRNA.fa -b all-norm-vs-cdna-sorted.soap -l 20 -p 100 -o miRNA-cdna.seqtar -s 60 -m 0.1 > miRNA-vs-cdna.log
```

Command line interface,

for BLASTN users:

```
java -classpath JSmallRNA.jar jsmallrna.seqtar.SeqTar
```

for SOAP2 users:

```
java -classpath JSmallRNA.jar jsmallrna.seqtar.SeqTarSOAP
```

APPENDIX C

List of Similar prediction of miRNA targets by CleaveLand, SeqTar and PARESnip

No	Mature known		Precursor known		Precursor Novel	
	miRNA	Target	miRNA	Target	miRNA	Target
1	t0474181_3_miR172	CL1Contig4708	t0001605_164_miR166	CL1Contig6721	IF-m0131_5p	CL12Contig3
2	t0985773_2_miR172	CL1Contig4708	t0002618_118_miR394	scaffold8018	IF-m0131_5p	CL1416Contig1
3	t1280040_2_miR172	CL1Contig4708	t0005201_74_miR166	CL1Contig6721		
4			t0005310_73_miR171	CL6736Contig1		
5			t0005310_73_miR171	scaffold7045		
6			t0012946_40_miR171	CL6736Contig1		
7			t0012946_40_miR171	scaffold7045		
8			t0026047_25_miR156	Contig1220		
9			t0031289_22_miR535	scaffold8787		
10			t0036915_20_miR166	CL1Contig6721		
11			t0040001_19_miR172	CL1Contig4708		
12			t0040001_19_miR172	scaffold9904		
13			t0040348_19_miR166	CL1Contig6721		
14			t0044311_18_miR168	CL1Contig3831		
15			t0045092_17_miR159	CL6066Contig1		
16			t0054871_15_miR156	CL1Contig2758		
17			t0054871_15_miR156	Contig1220		
18			t0071542_13_miR396	C88454		
19			t0071542_13_miR396	CL1Contig193		
20			t0071542_13_miR396	CL1Contig7158		
21			t0071542_13_miR396	scaffold2199		
22			t0076566_12_miR156	CL1Contig2758		
23			t0076566_12_miR156	Contig1220		
24			t0090912_11_miR166	CL1Contig6721		
25			t0092642_11_miR166	CL1Contig6721		
26			t0098708_10_miR156	CL1Contig2758		
27			t0098708_10_miR156	Contig1220		
28			t0104231_10_miR156	CL1Contig2758		
29			t0104231_10_miR156	Contig1220		
30			t0106813_10_miR166	CL1Contig6721		
31			t0107574_10_miR156	CL1Contig2758		
32			t0107574_10_miR156	Contig1220		
33			t0108037_10_miR156	CL1Contig2758		
34			t0108037_10_miR156	Contig1220		
35			t0129372_9_miR156	CL1Contig2758		
36			t0129372_9_miR156	Contig1220		

No	Mature known		Precursor known		Precursor Novel	
	miRNA	Target	miRNA	Target	miRNA	Target
37			t0131056_9_miR159	CL6066Contig1		
38			t0136478_8_miR172	CL1Contig4708		
39			t0136478_8_miR172	scaffold9904		
40			t0136535_8_miR156	CL1Contig2758		
41			t0136535_8_miR156	Contig1220		
42			t0139379_8_miR156	CL1Contig2758		
43			t0139379_8_miR156	Contig1220		
44			t0155859_8_miR156	CL1Contig2758		
45			t0155859_8_miR156	Contig1220		
46			t0171757_7_miR168	CL1Contig3831		
47			t0172653_7_miR156	CL1Contig2758		
48			t0172653_7_miR156	Contig1220		
49			t0177819_7_miR172	CL1Contig4708		
50			t0177819_7_miR172	scaffold9904		
51			t0184979_7_miR3706	CL1Contig2539		
52			t0190106_7_miR159	CL6066Contig1		
53			t0194607_7_miR3706	CL1Contig2539		
54			t0196088_7_miR168	CL1Contig3831		
55			t0198720_6_miR156	CL1Contig2758		
56			t0198720_6_miR156	Contig1220		
57			t0205805_6_miR156	CL1Contig2758		
58			t0205805_6_miR156	Contig1220		
59			t0210998_6_miR166	CL1Contig6721		
60			t0216987_6_miR156	CL1Contig2758		
61			t0216987_6_miR156	Contig1220		
62			t0217413_6_miR166	CL1Contig6721		
63			t0224046_6_miR156	CL1Contig2758		
64			t0224046_6_miR156	Contig1220		
65			t0232713_6_miR166	CL1Contig6721		
66			t0248389_6_miR156	CL1Contig2758		
67			t0248389_6_miR156	Contig1220		
68			t0252171_5_miR171	CL6736Contig1		
69			t0252171_5_miR171	scaffold7045		
70			t0264996_5_miR159	CL6066Contig1		
71			t0271747_5_miR172	CL1Contig4708		
72			t0271747_5_miR172	scaffold9904		
73			t0280850_5_miR167	CL1Contig3147		
74			t0280850_5_miR167	CL1Contig6273		
75			t0280850_5_miR167	scaffold3361		
76			t0281198_5_miR166	CL1Contig6721		
77			t0285020_5_miR166	CL1Contig6721		

No	Mature known		Immature known		Immature Novel	
	miRNA	Target	miRNA	Target	miRNA	Target
78			t0285209_5_miR159	CL1Contig6829		
79			t0285209_5_miR159	CL2056Contig1		
80			t0289689_5_miR156	CL1Contig2758		
81			t0289689_5_miR156	Contig1220		
82			t0290868_5_miR396	C88454		
83			t0290868_5_miR396	CL1Contig193		
84			t0290882_5_miR156	CL1Contig2758		
85			t0290882_5_miR156	Contig1220		
86			t0292918_5_miR156	CL1Contig2758		
87			t0292918_5_miR156	Contig1220		
88			t0297628_5_miR396	C88454		
89			t0297628_5_miR396	CL1Contig193		
90			t0297628_5_miR396	CL1Contig7158		
91			t0297628_5_miR396	scaffold2199		
92			t0307917_5_miR166	CL1Contig6721		
93			t0316769_5_miR156	CL1Contig2758		
94			t0316769_5_miR156	Contig1220		
95			t0321952_5_miR535	scaffold8787		
96			t0322151_5_miR156	CL1Contig2758		
97			t0322151_5_miR156	Contig1220		
98			t0330820_5_miR159	CL6066Contig1		
99			t0336947_5_miR396	C88454		
100			t0336947_5_miR396	CL1Contig193		
101			t0336947_5_miR396	CL1Contig7158		
102			t0336947_5_miR396	scaffold2199		
103			t0338505_5_miR396	C88454		
104			t0338505_5_miR396	CL1Contig193		
105			t0338505_5_miR396	CL1Contig7158		
106			t0339162_5_miR172	CL1Contig4708		
107			t0339162_5_miR172	scaffold9904		
108			t0348904_4_miR172	CL1Contig4708		
109			t0348904_4_miR172	scaffold9904		
110			t0353389_4_miR156	CL1Contig2758		
111			t0353389_4_miR156	Contig1220		
112			t0357256_4_miR166	CL1Contig6721		
113			t0361561_4_miR172	CL1Contig4708		
114			t0361561_4_miR172	scaffold9904		
115			t0364959_4_miR172	CL1Contig4708		
116			t0364959_4_miR172	scaffold9904		
117			t0365694_4_miR156	CL1Contig2758		
118			t0365694_4_miR156	Contig1220		

No	Mature known		Immature known		Immature Novel	
	miRNA	Target	miRNA	Target	miRNA	Target
119			t0379139_4_miR156	CL1Contig2758		
120			t0379139_4_miR156	Contig1220		
121			t0416602_4_miR166	CL1Contig6721		
122			t0423214_4_miR166	CL1Contig6721		
123			t0438404_4_miR168	CL1Contig3831		
124			t0443445_4_miR172	CL1Contig4708		
125			t0443445_4_miR172	scaffold9904		
126			t0447750_4_miR156	CL1Contig2758		
127			t0447750_4_miR156	Contig1220		
128			t0460615_4_miR156	CL1Contig2758		
129			t0460615_4_miR156	Contig1220		
130			t0490707_4_miR156	CL1Contig2758		
131			t0490707_4_miR156	Contig1220		
132			t0491401_4_miR172	CL1Contig4708		
133			t0491401_4_miR172	scaffold9904		
134			t0512962_3_miR156	CL1Contig2758		
135			t0512962_3_miR156	Contig1220		
136			t0532972_3_miR156	CL1Contig2758		
137			t0532972_3_miR156	Contig1220		
138			t0560329_3_miR156	CL1Contig2758		
139			t0560329_3_miR156	Contig1220		
140			t0576056_3_miR159	CL6066Contig		
141			t0580730_3_miR168	CL1Contig3831		
142			t0587741_3_miR172	CL1Contig4708		
143			t0587741_3_miR172	scaffold9904		
144			t0596282_3_miR172	CL1Contig4708		
145			t0596282_3_miR172	scaffold9904		
146			t0608090_3_miR156	CL1Contig2758		
147			t0608090_3_miR156	Contig1220		
148			t0624702_3_miR396	CL1Contig193		
149			t0660439_3_miR172	CL1Contig4708		
150			t0660439_3_miR172	scaffold9904		
151			t0663130_3_miR172	CL1Contig4708		
152			t0663130_3_miR172	scaffold9904		
153			t0676852_3_miR172	CL1Contig4708		
154			t0676852_3_miR172	scaffold9904		

APPENDIX D

List of Category I miRNA target pairs predicted by SeqTar

No	Mature Known miRNA		Precursor Known miRNA		Precursor Novel miRNA	
1	t0410587_3_miR1507	C93886	t0258665_5_miR1132	CL1405Contig1	IF-m0071_5p	CL12Contig3
2	t0474181_3_miR172	scaffold9904	t0258665_5_miR1132	CL6423Contig1	IF-m0071_5p	CL1416Contig1
3	t0474181_3_miR172	CL1Contig4708	t0258665_5_miR1132	CL469Contig1	IF-m0058_3p	scaffold10585
4	t0474181_3_miR172	scaffold3468	t0511283_3_miR1507	C93886	IF-m0001_5p	CL12Contig3
5	t0668799_3_miR396	CL1Contig193	t0791213_3_miR1507	C93886	IF-m0001_5p	CL1416Contig1
6	t0668799_3_miR396	C88454	t3124657_1_miR5024	CL1Contig8908	IF-m0093_5p	CL12Contig3
7	t0668799_3_miR396	CL1Contig7187	t3352443_1_miR3445	CL1Contig4377	IF-m0093_5p	CL1416Contig1
8	t0668799_3_miR396	CL2840Contig1	t3777187_1_miR1088	C93996	IF-m0022_5p	CL12Contig3
9	t0668799_3_miR396	CL427Contig1	t3777187_1_miR1088	CL1Contig1233	IF-m0022_5p	CL1416Contig1
10	t0668799_3_miR396	CL1Contig7158	t3777187_1_miR1088	CL576Contig1	IF-m0090_5p	CL12Contig3
11	t0692989_2_miR842	Contig432	t3777187_1_miR1088	CL320Contig1	IF-m0090_5p	CL1416Contig1
12	t0985773_2_miR172	scaffold9904	t3852900_1_miR3468	CL1Contig2539	IF-m0037_3p	scaffold10585
13	t0985773_2_miR172	CL1Contig4708	t3868342_1_miR5029	CL1Contig8875	IF-m0049_5p	CL12Contig3
14	t1139835_2_miR157	CL1Contig2758	t4142425_1_miR3468	Contig516	IF-m0049_5p	CL1416Contig1
15	t1139835_2_miR157	CL6899Contig1	t4142425_1_miR3468	Contig1070	IF-m0097_3p	scaffold10585
16	t1139835_2_miR157	scaffold8787	t4294464_1_miR5051	CL1Contig1735	IF-m0131_5p	CL12Contig3
17	t1139835_2_miR157	Contig1220	t4446170_1_miR1173	C137724	IF-m0131_5p	CL1416Contig1
18	t1139835_2_miR157	scaffold7197	t5374781_1_miR831	CL576Contig1	IF-m0072_5p	CL12Contig3
19	t1280040_2_miR172	scaffold9904	t6013175_1_miR1507	C93886	IF-m0072_5p	CL1416Contig1
20	t1280040_2_miR172	CL1Contig4708	t7244964_1_miR157	CL1Contig2758	IF-m0058_3p	CL1Contig5119
21	t1520340_2_miR414	CL7Contig1	t7244964_1_miR157	CL6899Contig1	IF-m0086_5p	C137468

No	Mature Known miRNA		Precursor Known miRNA		Precursor Novel miRNA	
22	t1520340_2_miR414	scaffold4460	t7244964_1_miR157	scaffold8787	IF-m0037_3p	CL1Contig5119
23	t2222779_1_miR774	CL1845Contig1	t7244964_1_miR157	scaffold5419	IF-m0097_3p	CL1Contig5119
24	t2222779_1_miR774	CL1Contig49	t7942865_1_miR3449	CL1Contig7125	IF-m0027_5p	C137468
25	t2731798_1_miR4378	CL148Contig2	t8428879_1_miR159	CL6066Contig1	IF-m0123_5p	C137468
26	t3446235_1_miR5080	Contig517	t8931141_1_miR4348	CL7601Contig1	IF-m0062_5p	C137468
27	t3998496_1_miR474	scaffold11841	t9189336_1_miR529	CL1Contig2758	IF-m0044_5p	C137468
28	t4945335_1_miR1535	CL1Contig8389			IF-m0115_5p	C67286
29	t5586009_1_miR854	CL1Contig25			IF-m0035_5p	C67286
30	t5804383_1_miR157	CL1Contig7938			IF-m0056_5p	C67286
31	t5804383_1_miR157	C65942			IF-m0133_5p	C137468
32	t6007800_1_miR1085	CL2842Contig1				
33	t6007800_1_miR1085	CL1Contig6893				
34	t6007800_1_miR1085	CL1Contig6478				
35	t6007800_1_miR1085	scaffold8817				
36	t6546439_1_miR1868	CL1Contig5306				
37	t7434667_1_miR1024	C187123				
38	t7507061_1_miR1507	C93886				
39	t8991801_1_miR528	CL1Contig1205				

APPENDIX E

Supplementary Tables S1, S2 and S3

APPENDIX F

Summary of mapped degradomes

Mature Female Flower Degradome

Seeded quality full-index search: 00:16:43

reads processed: 2444724

reads with at least one reported alignment: 1650517 (67.51%)

reads that failed to align: 794207 (32.49%)

Reported 1650517 alignments to 1 output stream(s)

Time searching: 00:16:44

Overall time: 00:16:44

Immature Female Flower Degradome

Seeded quality full-index search: 00:22:09

reads processed: 3237523

reads with at least one reported alignment: 2246016 (69.37%)

reads that failed to align: 991507 (30.63%)

Reported 2246016 alignments to 1 output stream(s)

Time searching: 00:22:11

Overall time: 00:22:11

APPENDIX G

Number of sequences in miRNAs and number of predicted targets according to CleaveLand, SeqTar and PARESnip

miRNAs	Number of Sequences			Number of Predicted Targets by tools		
	Total	Redundant	Non-redundant	CleaveLand	SeqTar	PARESnip
Mature Known	31205	5896	25309	261	170932	5775
Mature Novel	163	114	49	0	39579	13
Precursor Known	31535	5392	26143	1041	329145	6723
Precursor Novel	145	102	43	2	123662	21