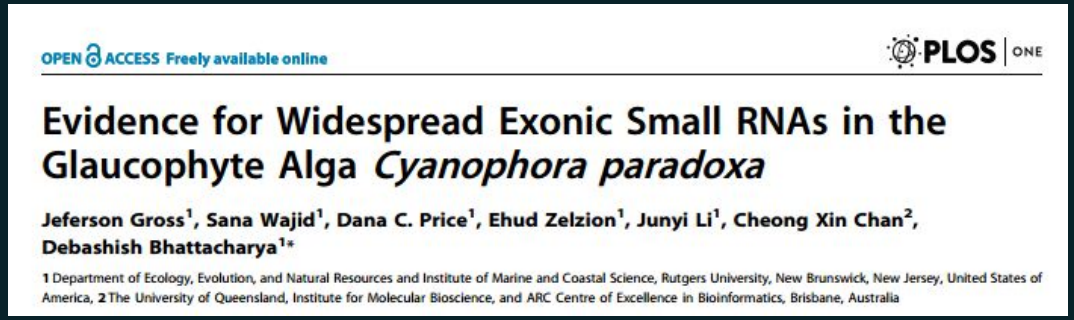
Bioinformatics identification of small RNA targets in the Glaucophyte alga Cyanophora paradoxa: Analysis of ancient evolutionary mechanisms that arose in primordial algae

Sana Wajid
Rutgers University
Nestler Lab Bioinformatician Interview
Icahn School of Medicine at Mount Sinai
August 13, 2014

Overview

- ➔ Background and Introduction
- ➔ Hypothesis and experimental rationale
- ➔ Methods: Bioinformatics pipelines
- ➔ Results
- ➔ Future Objectives
- ➔ References
- ➔ Questions



OPEN ACCESS Freely available online

PLOS ONE

Evidence for Widespread Exonic Small RNAs in the Glaucophyte Alga *Cyanophora paradoxa*

Jeferson Gross[1], Sana Wajid[1], Dana C. Price[1], Ehud Zelzion[1], Junyi Li[1], Cheong Xin Chan[2], Debashish Bhattacharya[1]*

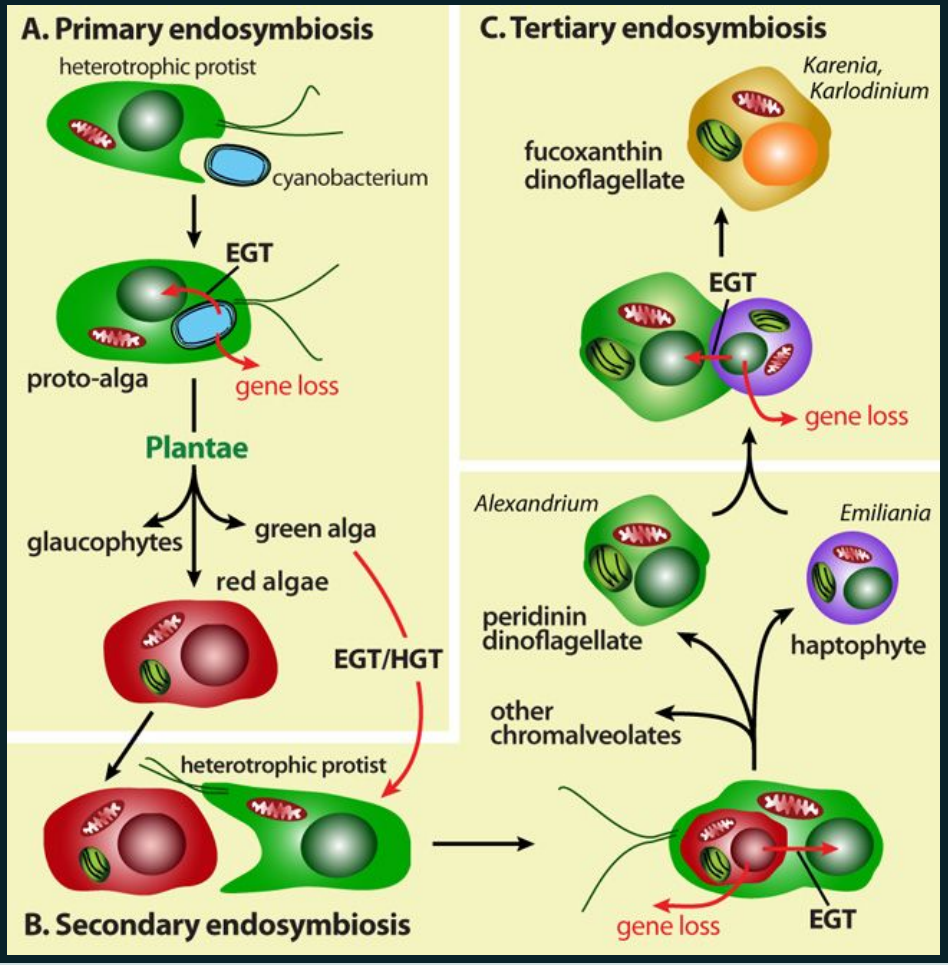
1 Department of Ecology, Evolution, and Natural Resources and Institute of Marine and Coastal Science, Rutgers University, New Brunswick, New Jersey, United States of America, 2 The University of Queensland, Institute for Molecular Bioscience, and ARC Centre of Excellence in Bioinformatics, Brisbane, Australia

Background and Introduction

Cyanophora paradoxa elucidates primary endosymbiosis

- The plastid originated at least once > 1.5 billion years ago, where a cyanobacterial ancestor that was captured by a single-celled protist as food-stuff but remained as an endosymbiont.

- Following many photoautotrophic cyanobacterial capture events, a *novel metabolic-toolkit*, which involved photosynthesis, emerged among microbiological life.



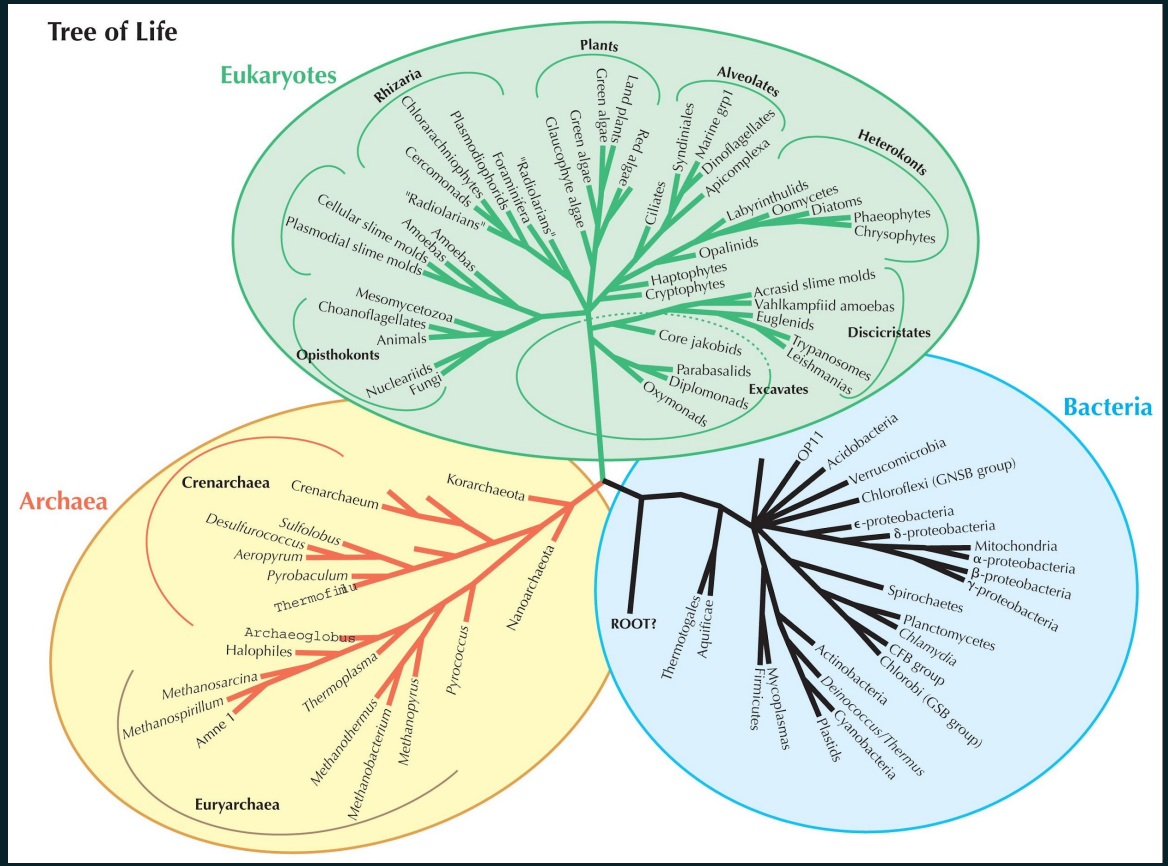
- *Cyanophora paradoxa* is a model organism for endosymbiosis
- Contains two blue-greenish vesicles called cyanelles in its protoplasm which were are thought to be acquired endosymbiotically from cyanobacteria

http://www.nature.com/scitable/topicpage/the-origin-of-plastids-14125758

- Current ***survivors in the tree of life*** possess the "*core*" metabolic processes which allowed for photosynthesis and other features found in plastids which were then transferred to the nuclear genome (8,10,11)

- Molecular phylogenetics and comparative genomics toolkits

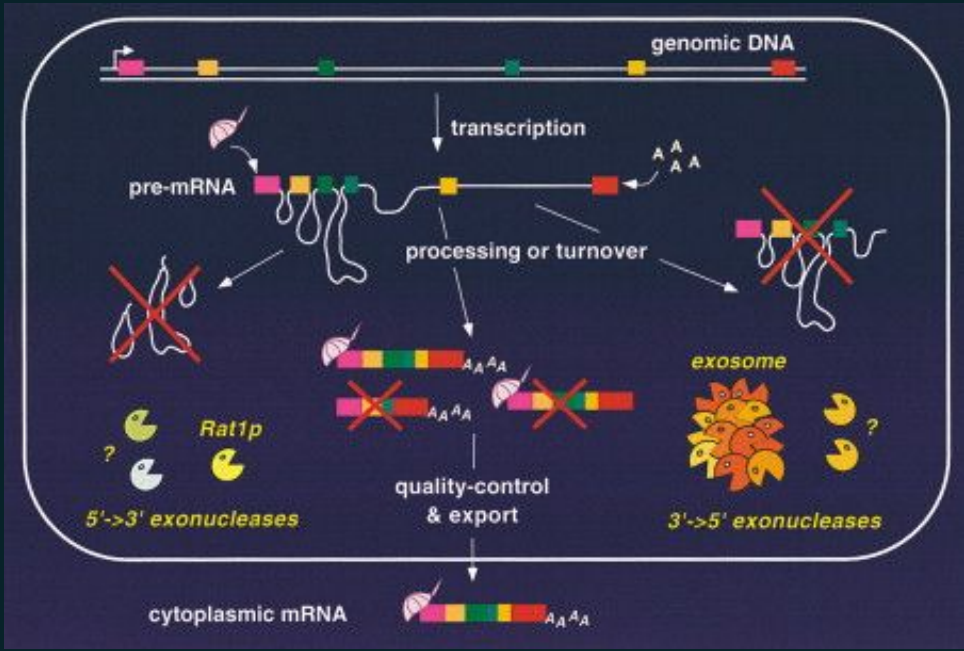
- Inferences on evolutionary past has been bolstered by NGS platforms (31)



http://phylogenomics.blogspot.com/2008/10/twisted-tree-of-life-award-2-science.html

Post-transcriptional silencing is a form of negative regulation

- Silencing can occur through:
  - translational repression then RNA decay or
  - endonucleolytic cleavage through argonaute slicing mechanisms

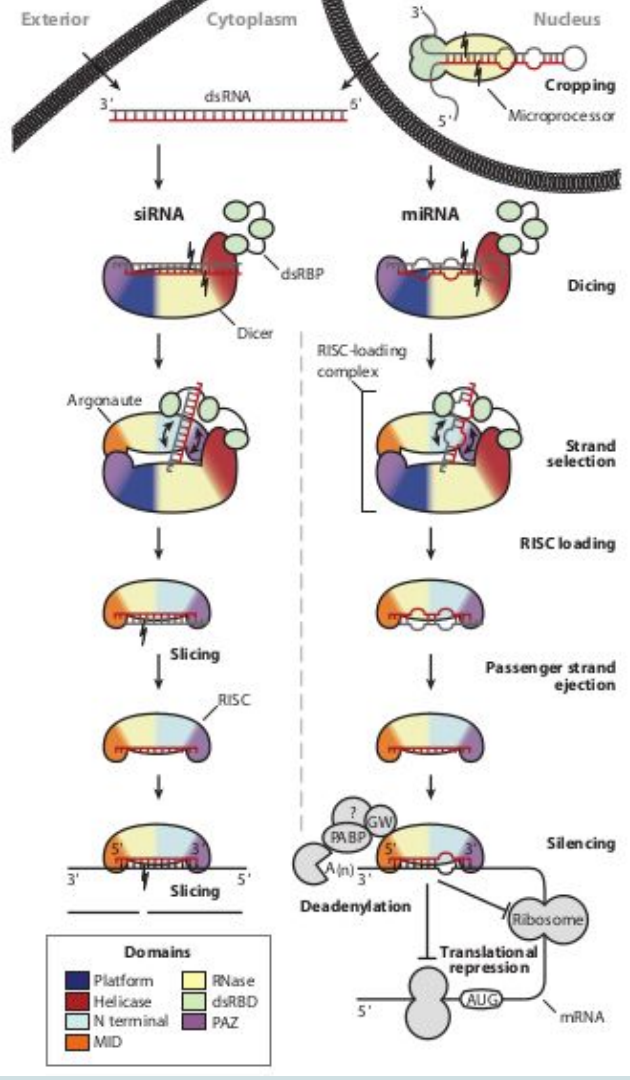


Moore, Melissa J. "Nuclear RNA turnover." Cell 108.4 (2002): 431-434.

RNA Turnover mechanisms utilize RNA interference

- ➔ RNAi/siRNA is primarily used for gene silencing (i.e. inhibition of gene expression)
- ➔ Double-stranded RNA (dsRNA) intermediates are tagged then undergo degradation

(24, 29)



RNAi is specific to eukaryotes

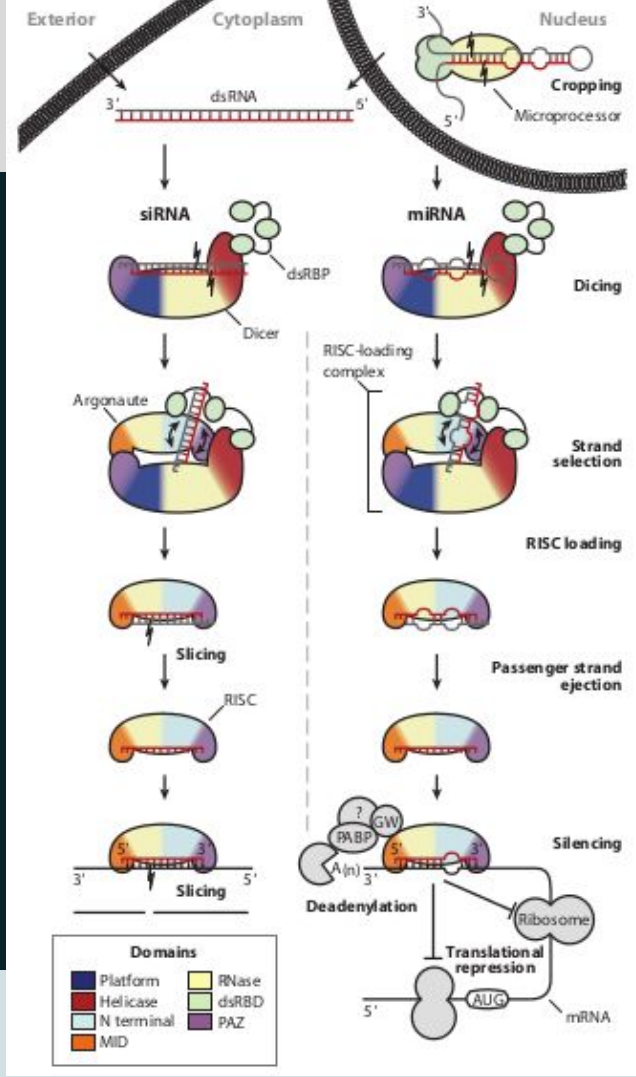
- ➔ RNAi is an ancient process that likely existed before the split of prokaryotes and eukaryotes
- ➔ RNAi may have served primarily as an anti-viral and transposon defense mechanism
- ➔ For example, in early evolution, genomic defense against self and non-self RNA became an important tool to abate pervasive and parasitic self-replicating autonomous genomes

(24, 25)

Outcomes of the RNAi gene-protein network

- ➔ Canonical RNAi mechanism is triggered by the formation of dsRNA duplex precursor which is cleaved by Ribonuclease III enzyme Dicer at both termini and then loaded onto the Argonaute complex
- ➔ Products of Dicer cleavage are small interfering RNAs (siRNAs) of variable sizes (21-25 nt) with siRNA 3'-2 nt overhang pairing and 5'-monophosphates

(24, 29)



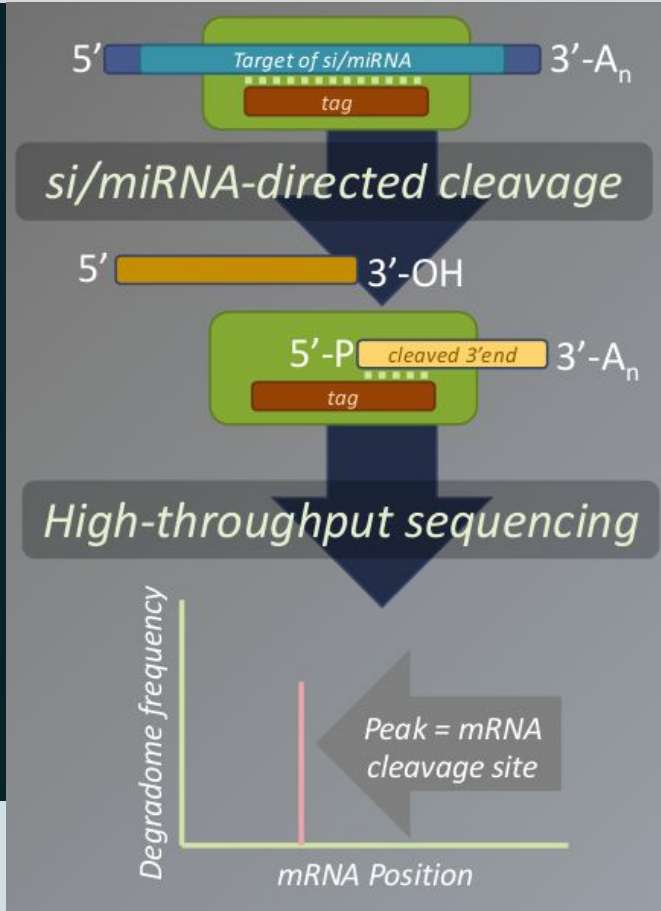
Evolutionary origins of sRNA pathways in eukaryotes

- ➔ Comparative genomic studies based on existence of the three main components of RNAi by Cerutti and Mollano (2006) and Shabalina and Koonin (2008) using also parsimonious evaluation reveal the following.
  - ◆ Several unicellular eukaryotes which have lost RNAi functionalities, perhaps independently and multiple times: Opisthokonta (*Saccharomyces cerevisiae*), Excavata, Archaeplastida, Chromalveolata (20, 22).
- ➔ It is likely that the RNAi mechanism is not necessary for unicellular eukaryotic life whereas it remains an integral part of multicellular eukaryotic developmental gene regulation (e.g. complex body pattern formation) (20).

Next-generation sequencing of non-coding RNAs

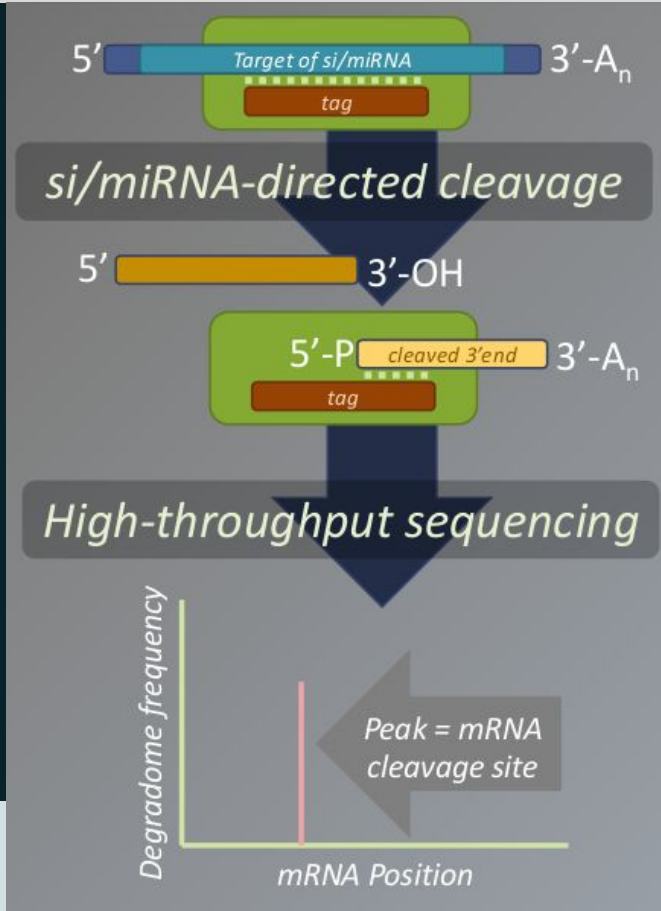
- Genome wide analysis of small RNAs in *C. paradoxa* for a clearer picture of the composition of epigenomes (i.e. RNA expression)

- Other reasons include, Look for essential sRNA hotspots regulating stress responses, growth and development



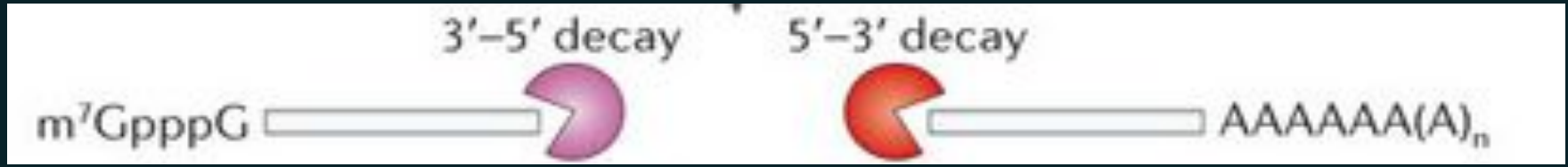
The 5’ monophosphate is an important biochemical marker

- The 5’ monophosphate is fundamental to bioinformatic studies because it allows us to trace the *degradome*
- Decapping prevents mRNA from translation process



Sequencing the degradome

- ➔ A snapshot of the transcriptome is achieved using methods that exist for degradome sequencing and identification of si/miRNA cleavage sites by locating the 5’ monophosphate and study 5’->3 decay.
- ➔ Genome-wide mapping of uncapped transcripts (GMUCT) is a protocol which samples the 5’-ends of uncapped mRNAs but not 3’-5’ decay (27).



Schoenberg, Daniel R., and Lynne E. Maquat. "Regulation of cytoplasmic mRNA decay." Nature Reviews Genetics 13.4 (2012): 246-259.

Hypothesis and experimental rationale

Experimental Rationale

- ➔ For this study, a means of visually analyzing RNA-seq degradome data from *Cyanophora paradoxa* is proposed.
- ➔ Our lab has recently sequenced genome of *Cyanophora paradoxa* and this was used as a template for preliminary evolutionary bioinformatic checks to indicate presence of the RNAi molecular toolkit.
- ➔ According to Gross et. al (2013), *Cyanophora paradoxa* nuclear genome contains gene model matches with high BLASTp e-values to three RNAi components: Dicer, Ago and RDRp.

(1,7)

Methods

Sequencing of sRNA & degradome

Read profiling, trimming raw seq data and expression analysis

- ➔ Three template sequence libraries of *C. paradoxa* exist: genomic contigs, EST contigs (nt >= 200) and and CDS.
- ➔ Mapping filtered sRNAs against these three sequence libraries with 100% identity produced the small RNA and degradome library data used for this analysis
- ➔ Using CLC Genomics, reads were mapped to *C. paradoxa* contigs, ESTS and CDS
- ➔ Expression analysis of reads mapped to CDS of different conditions

Results & Discussion

```
> CDS
GAlignments with 1762675 alignments and 1 metadata column:
                                                seqnames strand       cigar
                                                   <Rle>  <Rle> <character>
       [1] Glaucophyta-Cyanophora_paradoxa_dxContig10000x1      -         18M
       [2] Glaucophyta-Cyanophora_paradoxa_dxContig10002x1      +         16M
       [3] Glaucophyta-Cyanophora_paradoxa_dxContig10002x1      -         22M
       [4] Glaucophyta-Cyanophora_paradoxa_dxContig10002x1      +         16M
       [5] Glaucophyta-Cyanophora_paradoxa_dxContig10002x1      +         23M
       ...                                             ...    ...        ...
 [1762671]  Glaucophyta-Cyanophora_paradoxa_dxContig9997x3      +         16M
 [1762672]  Glaucophyta-Cyanophora_paradoxa_dxContig9997x3      +         24M
 [1762673]  Glaucophyta-Cyanophora_paradoxa_dxContig9998x3      +         16M
 [1762674]  Glaucophyta-Cyanophora_paradoxa_dxContig9998x3      +         19M
 [1762675]  Glaucophyta-Cyanophora_paradoxa_dxContig9998x3      +         21M
            qwidth     start       end     width     njunc   |
         <integer> <integer> <integer> <integer> <integer>   |
       [1]        18       332       349        18         0   |
       [2]        16        56        71        16         0   |
       [3]        22       160       181        22         0   |
       [4]        16       178       193        16         0   |
       [5]        23       261       283        23         0   |
       ...       ...       ...       ...       ...       ... ...
 [1762671]        16        59        74        16         0   |
 [1762672]        24       208       231        24         0   |
 [1762673]        16       194       209        16         0   |
 [1762674]        19       210       228        19         0   |
 [1762675]        21       213       233        21         0   |
                          seq
                <DNAStringSet>
       [1]       AGGCGGCCGAGGCGACGG
       [2]         CAAGAACAACGGGACC
       [3]   GCAGATTCGGAAGCATCATCCG
       [4]         TCCGTGCCGTTGGACG
       [5]  CTCCGGAGGATTTGGGGGCTTGA
       ...                      ...
 [1762671]         ATGCGGACGAGGGAAC
 [1762672] GCGATTGGCTCAGTCCGCAGCTAC
 [1762673]         GCCACGGGGGGATGAA
 [1762674]      GTCGTCCCGCAACCTACTC
 [1762675]    GTCCCGCAACCTACTCCAGAT
 ---
```

➔ SAM file <-> `GappedAlignment` object stores the following information:

- Contig Name
- Strand as + or -
- Start
- End
- Width
- Sequence
- etc.

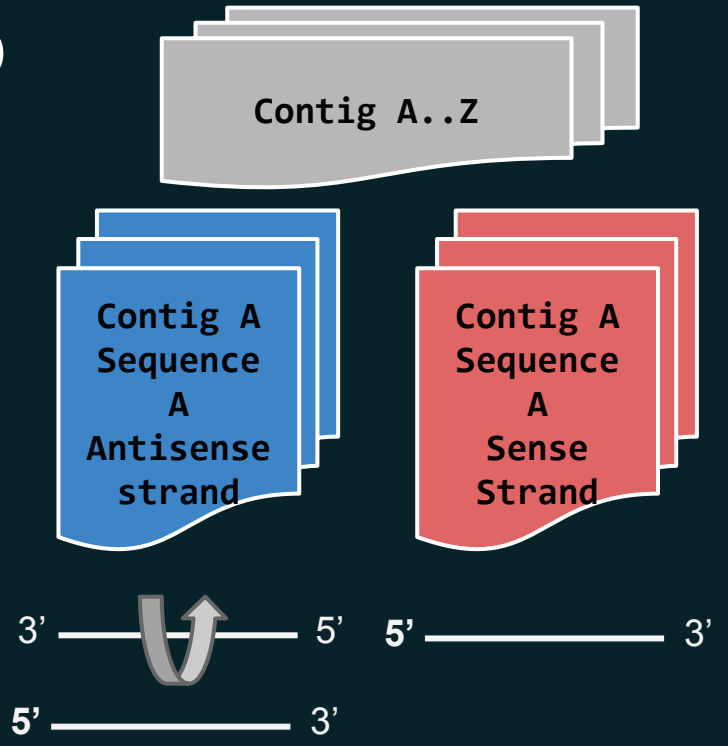
samtools.github.io/hts-specs/SAMv1.pdf

5’ position CDS and DEG tag density

processSeqs <- function (takes in mapping)

For each contig (subset):

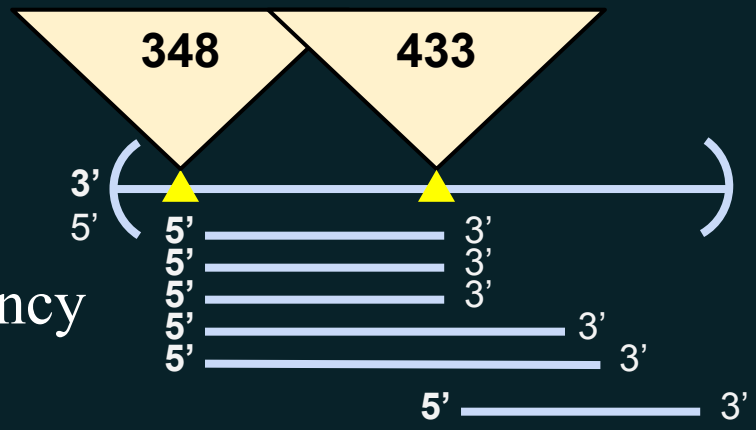
1. **Separate sequences based on + or -**
   - sense
   - antisense
2. **Modify start position for antisense as plot will consist of 5’ ends’ first position only**
   - start = end
   - end = end + width



5' position CDS and DEG tag density

3. **For each start position**
   But separately per - and +
   a. `nPosition` = normalized position*
   b. `pFrequency` = actual frequency at position

   c. **`nFrequency`** = normalized frequency



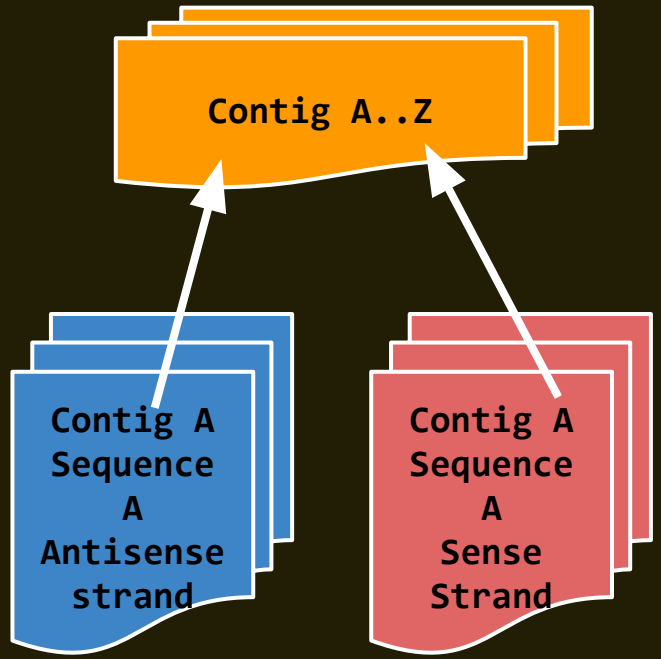
$$\texttt{nFrequency} = \text{Frequency}_i \,/\, \sum_{1,i} \text{Frequency}_i$$

where **i** = position of 5' end of sRNA or deg tag

5’ position CDS and DEG tag density

4. Bind separated subsets

```
> sub.antisense.freq
  nPosition pFrequency nFrequency   cName
1        38          1          1 55596x1
2       346          1          1 55596x1
3       348          5          5 55596x1
4       433          1          1 55596x1
5       639          1          1 55596x1
6       716          1          1 55596x1
```



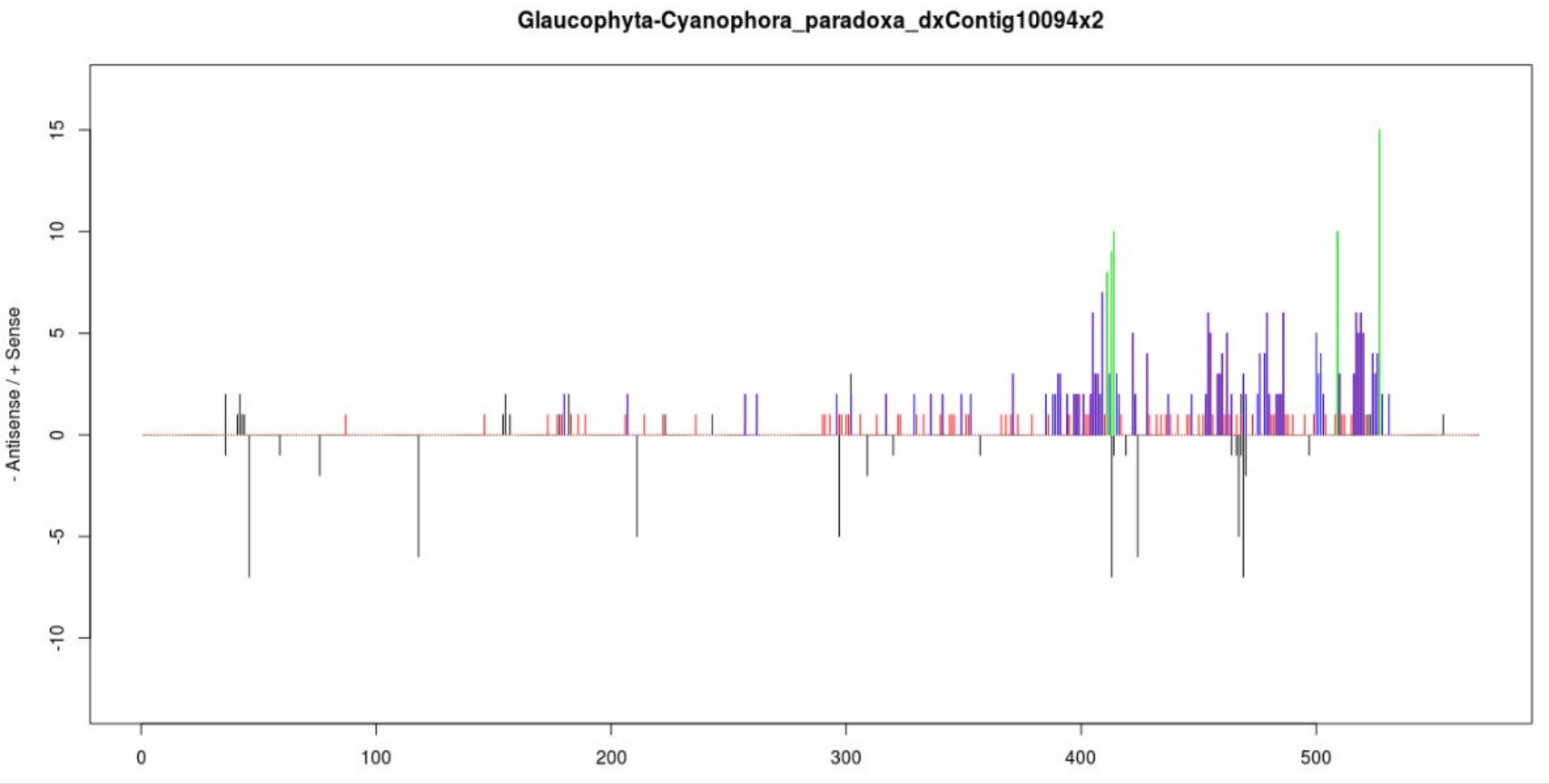
5. Plot histogram

contigSize - position

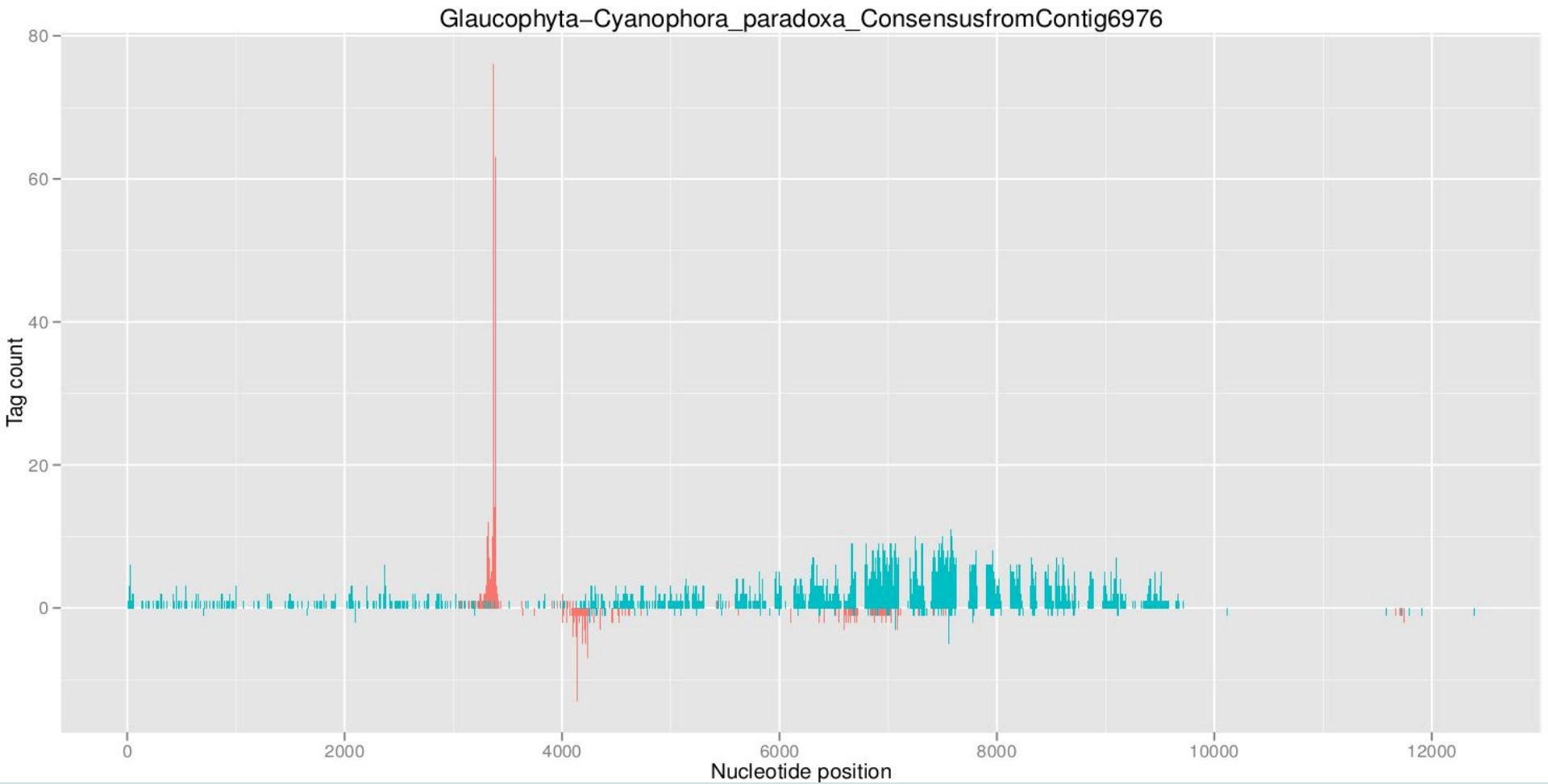
5’ position CDS and DEG tag density

5. **Plot histogram**

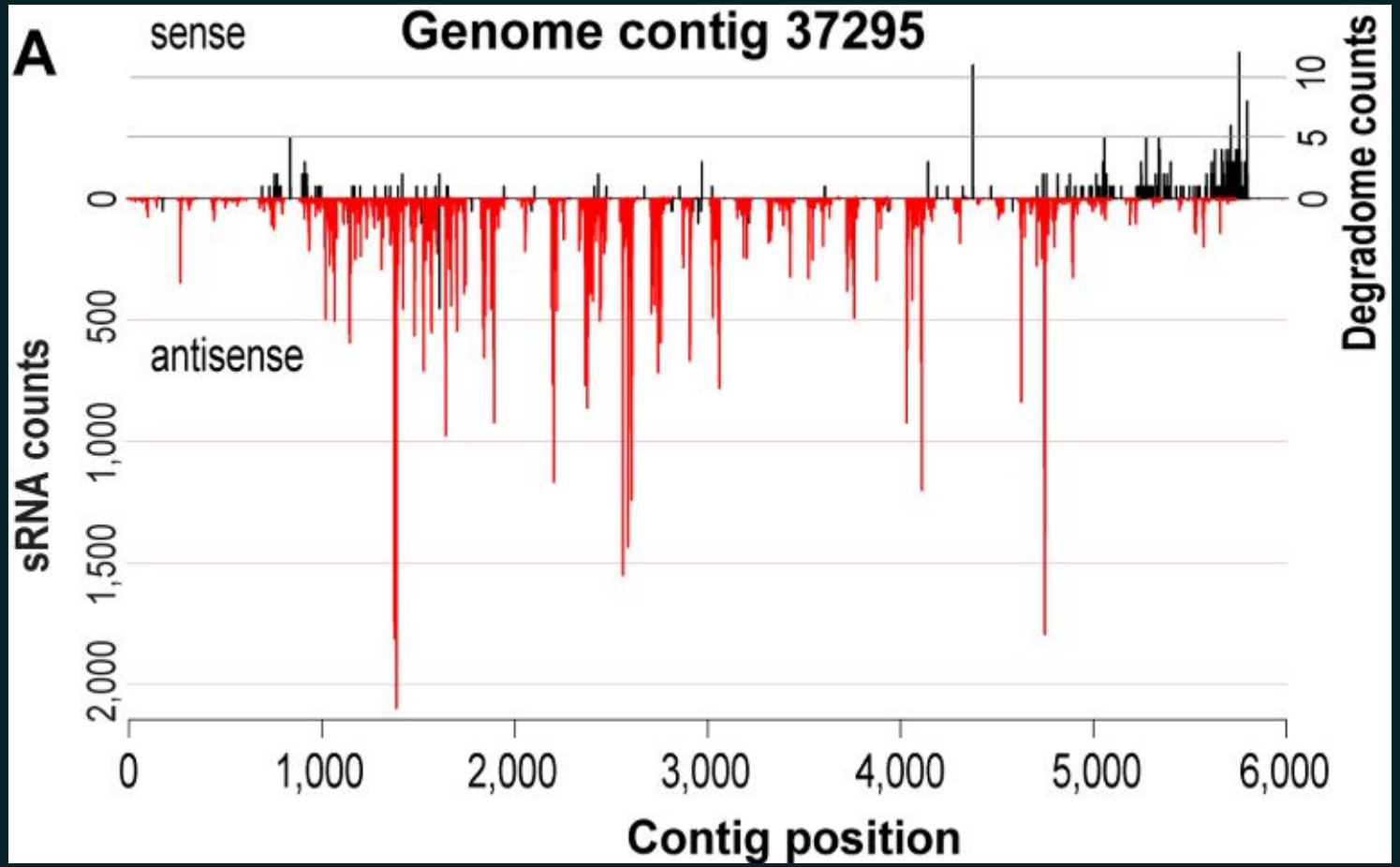
```
plotContig <- function(cds.sense.freq,cds.antisense.freq,deg.sense.freq,cds.antisense.freq){
#SIMPLE DEG + CDS PLOT
      plot(cds.sense.freq$nPosition, cds.sense.freq$pFrequency, type = "h", ylim =
c(((max(deg.antisense.freq$pFrequency)/2)*-1), max(deg.sense.freq$pFrequency)/2)) #, xlim =
c(1350,1450))
      points(cds.antisense.freq$nPosition, cds.antisense.freq$pFrequency*-1, type = "h")
      points(deg.antisense.freq$nPosition, deg.antisense.freq$pFrequency * -1, type = "h", col =
"red")
      points(deg.sense.freq$nPosition, deg.sense.freq$pFrequency, type = "h", col = "red")
      title(main=as.data.frame(seqnames(CDS))[1,])
}
cutoff <- mean(deg.sense.freq$nFrequency)
```



Top: sense; Bottom: antisense



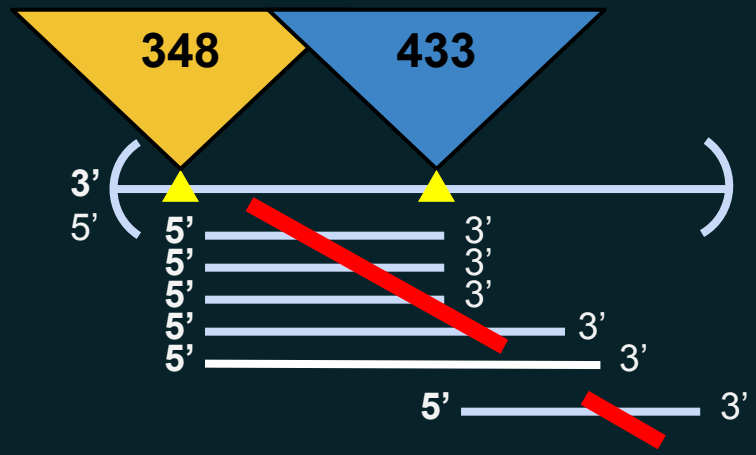
DEG = Red, CDS = Blue



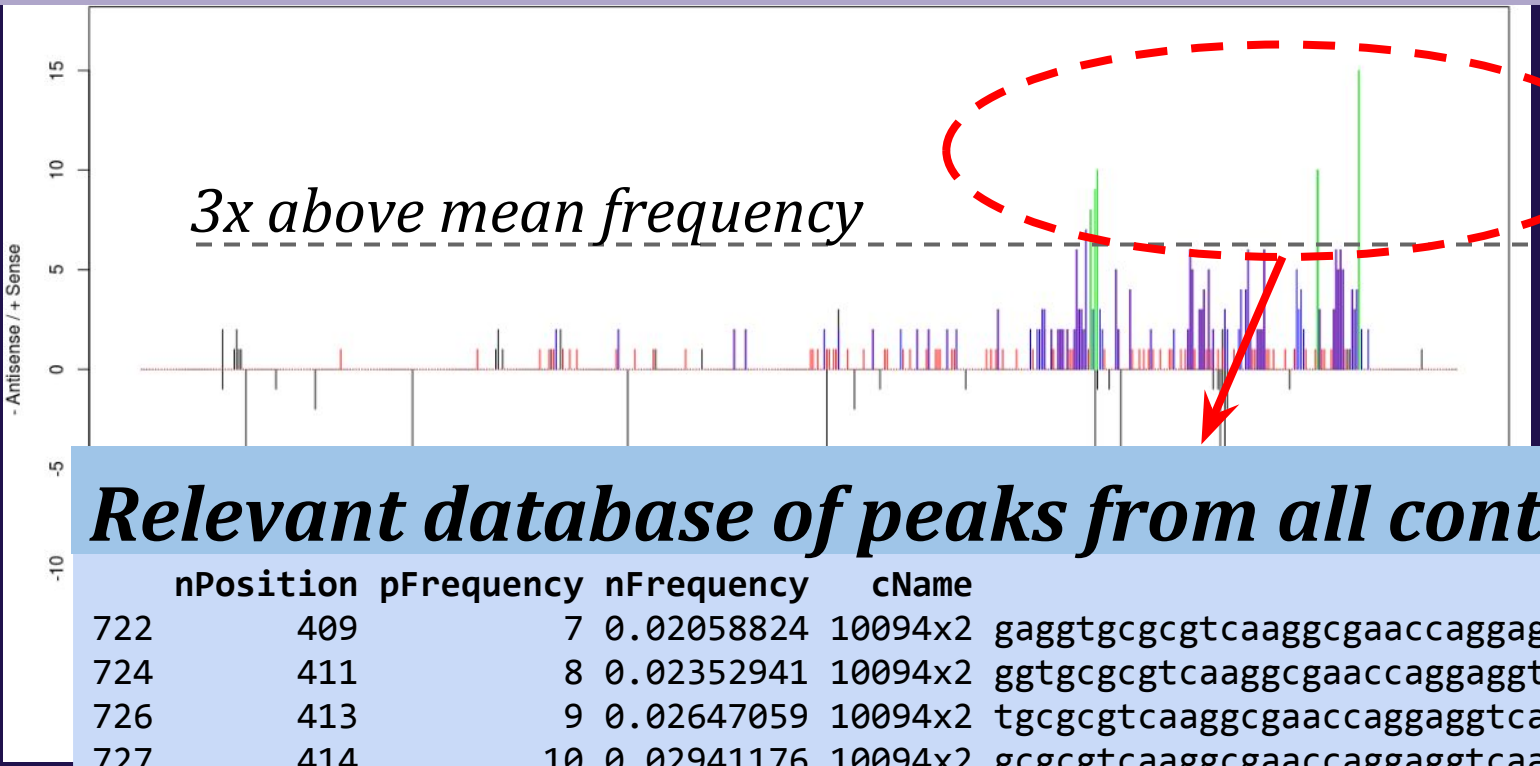
DEG = Red, CDS = Blue

Cleavage tag abundance and noise reduction of dataset

1. **For each contig**
   a. Take the mean
      i. Mean = mean of contig
      ii. Filter peaks that are X above mean
2. **For each position**
   a. Take the mean across all contigs



Building and filtering a "peak" database



Relevant database of peaks from all contigs

	nPosition	pFrequency	nFrequency	cName	seq
722	409	7	0.02058824	10094x2	gaggtgcgcgtcaaggcgaaccaggaggtca
724	411	8	0.02352941	10094x2	ggtgcgcgtcaaggcgaaccaggaggtcaag
726	413	9	0.02647059	10094x2	tgcgcgtcaaggcgaaccaggaggtcaaggc
727	414	10	0.02941176	10094x2	gcgcgtcaaggcgaaccaggaggtcaaggcc
784	509	10	0.02941176	10094x2	gcaaggtcaccaagctcccggtcgtcacggg
798	527	15	0.04411765	10094x2	cggtcgtcacgggcgactcgctcggctcggg

Building and filtering a "peak" database

***Normalized relative frequency** of an X-multiple small RNA across this contig in that position % away from the stop codon (1-this.contig_length)*

Contig Name

***Genomic (DNA) sequence -/+15 bases around first base position of sense strand of sRNA:** used for generating fasta file when/if needed*

Relevant database of peaks from all contigs

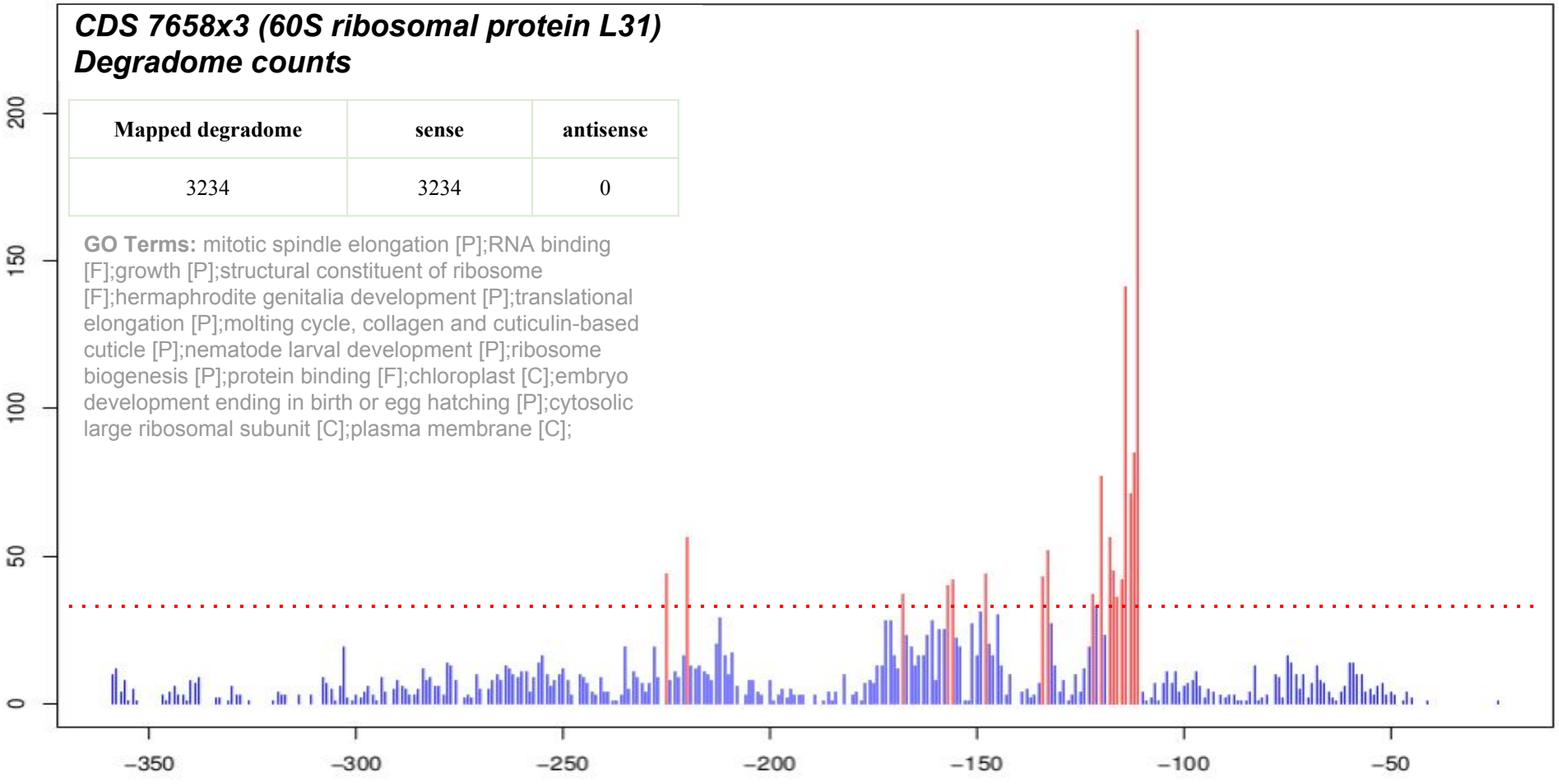
	pPosition	pFrequency	nFrequency	cName	seq
722	409	7	0.02058824	10094x2	gaggtgcgcgtcaaggcgaaccaggaggtca
724	411	8	0.02352941	10094x2	ggtgcgcgtcaaggcgaaccaggaggtcaag
726			0.02647059	10094x2	tgcgcgtcaaggcgaaccaggaggtcaaggc
727			0.02941176	10094x2	gcgcgtcaaggcgaaccaggaggtcaaggcc
784			0.02941176	10094x2	gcaaggtcaccaagctcccggtcgtcacggg
798			0.04411765	10094x2	cggtcgtcacgggcgactcgctcggctcggg

Number of multiples of this sRNA (regardless of length) that start at this position

CDS 7658x3 (60S ribosomal protein L31)
Degradome counts

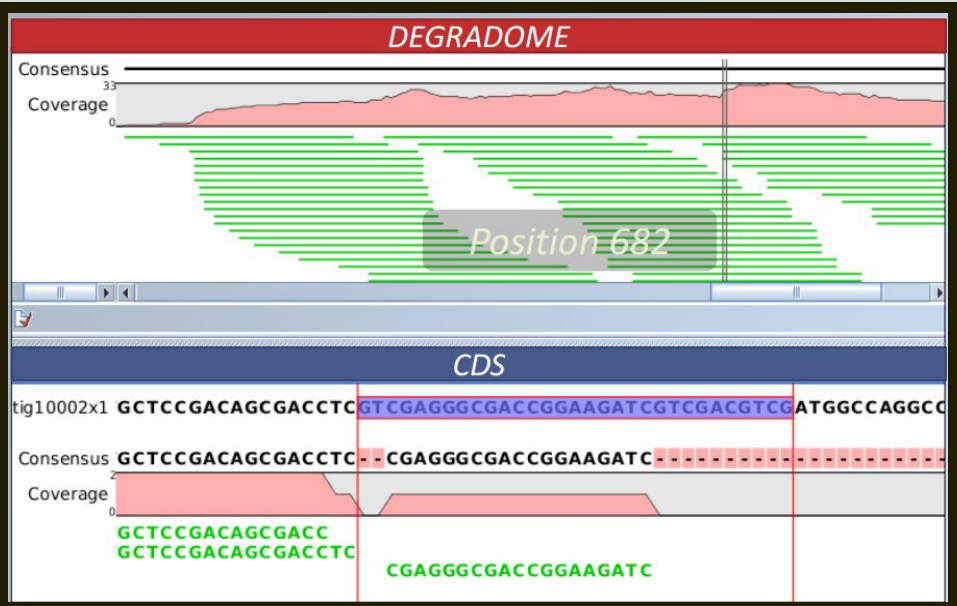
Mapped degradome	sense	antisense
3234	3234	0

GO Terms: mitotic spindle elongation [P];RNA binding [F];growth [P];structural constituent of ribosome [F];hermaphrodite genitalia development [P];translational elongation [P];molting cycle, collagen and cuticulin-based cuticle [P];nematode larval development [P];ribosome biogenesis [P];protein binding [F];chloroplast [C];embryo development ending in birth or egg hatching [P];cytosolic large ribosomal subunit [C];plasma membrane [C];



Red = 3x above mean

Peak "Mining" with BLAST



1. Generate fasta from non-redundant peak database top sequences: CDS and Degradome
2. Align with BLAST
3. Same with randomized 30 nt (+/- 15 nt from first position → randomize `nPosition`)
4. Align randomized peak dataset with BLAST
5. Look for significance (actual vs. random peaks)

```
10002x1 , gtcgagggcgaccggaagatcgtcgacgtcg , 682
10027x1 , cgagctggtggaccatccgtttgagacacga , 477
10038x2 , acggaatcctcttcgtttacaatcccgagga , 275
10051x1 , ggtaaacaacagcgtgttgatgagcttctgg , 985
10051x1 , aaacaacagcgtgttgatgagcttctggttc , 988
10056x1 , actgcgttccctattcctattgcaagcgctc , 760
10056x1 , ctcgaatatggtggcactcgaagatctgaaa , 846
10056x1 , ccagaccccgaaaacgcttcagtgaagggtc , 964
10056x1 , tgcaagcaagtcccttgctggtcgcaagttt , 1041
10056x1 , aagcaagtcccttgctggtcgcaagtttgca , 1044
10064x1 , cacacctcgctcgtgaagatcgagggcgtgc , 82
10064x1 , cgcaacgaggtcgacttctacctcggcaagc , 118
10064x1 , caacgaggtcgacttctacctcggcaagcgc , 120
10064x1 , cggcaacaacggcatcgtccgcgccaagttc , 279
10064x1 , catcgtccgcgccaagttccggaagaacctg , 291
10064x1 , atcgtccgcgccaagttccggaagaacctgc , 292
10064x1 , tcgtccgcgccaagttccggaagaacctgcc , 293
10064x1 , cgtccgcgccaagttccggaagaacctgccg , 294
10094x2 , gaggtgcgcgtcaaggcgaaccaggaggtca , 409
10094x2 , ggtgcgcgtcaaggcgaaccaggaggtcaag , 411>
```

Project standard for naming blast related files:

FASTA:

> 2-letter-orgName.3-letter-identifier | strand | contigName | startPosition | endPosition | coverage or frequency

[Sequence]

Reads Mapped to sRNAs in CDS

	CDS data	
Sense reads (+)	36,7709	20.86%
Antisense reads (-)	1,394,966	79.14%
Total reads	1,762,675	
Contigs	31,895	

Construction of peak database

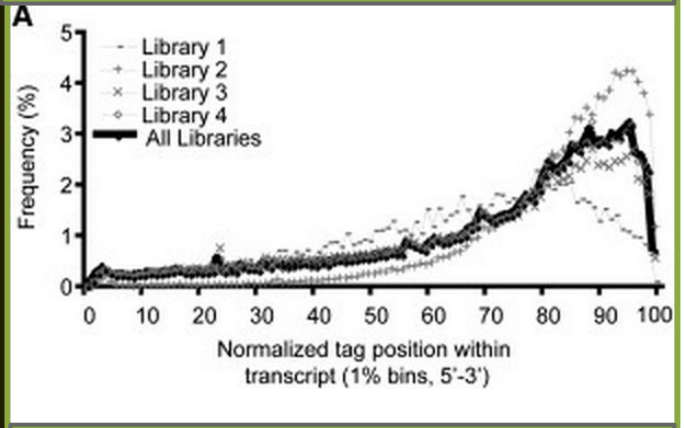
Construction of a non-redundant database from RNA-seq SAM file

- ➔ Compresses database from 1,762,675 to 322,039 rows or down to 18.2699% of the original database
- ➔ Keeps account of frequency or coverage of each unique small RNA

	rname	strand	start	seq	cigar	qwidth	end	width	ngap	freq
1	10000x1	-	332	AGGCGGCCGAGGCGACGG	18M	18	349	18	0	1
2	10000x2	+	178	CTGCTCACGGCCGCCCAG	18M	18	195	18	0	1
3	10002x1	-	160	GCAGATTCGGAAGCATCATCCG	22M	22	181	22	0	1
4	10002x1	+	178	TCCGTGCCGTTGGACG	16M	16	193	16	0	1
5	10002x1	+	261	CTCCGGAGGATTTGGGGGCTTGA	23M	23	283	23	0	1
6	10002x1	+	266	GAGGATTTGGGGGCTTGA	18M	18	283	18	0	1

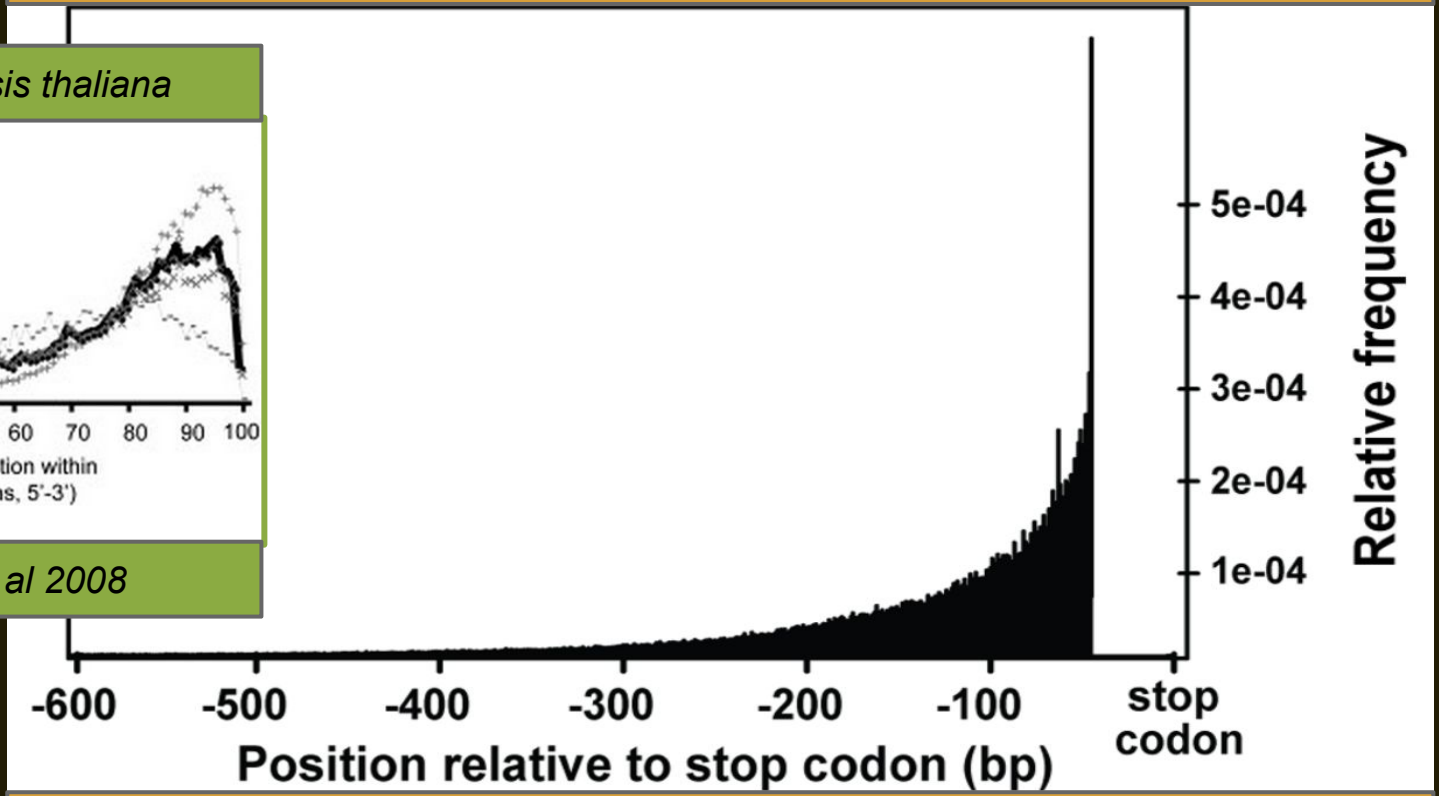
Using R/lattice to generate an aggregate exonucleolytic profile: 5'->3'

3' bias in Arabidopsis thaliana



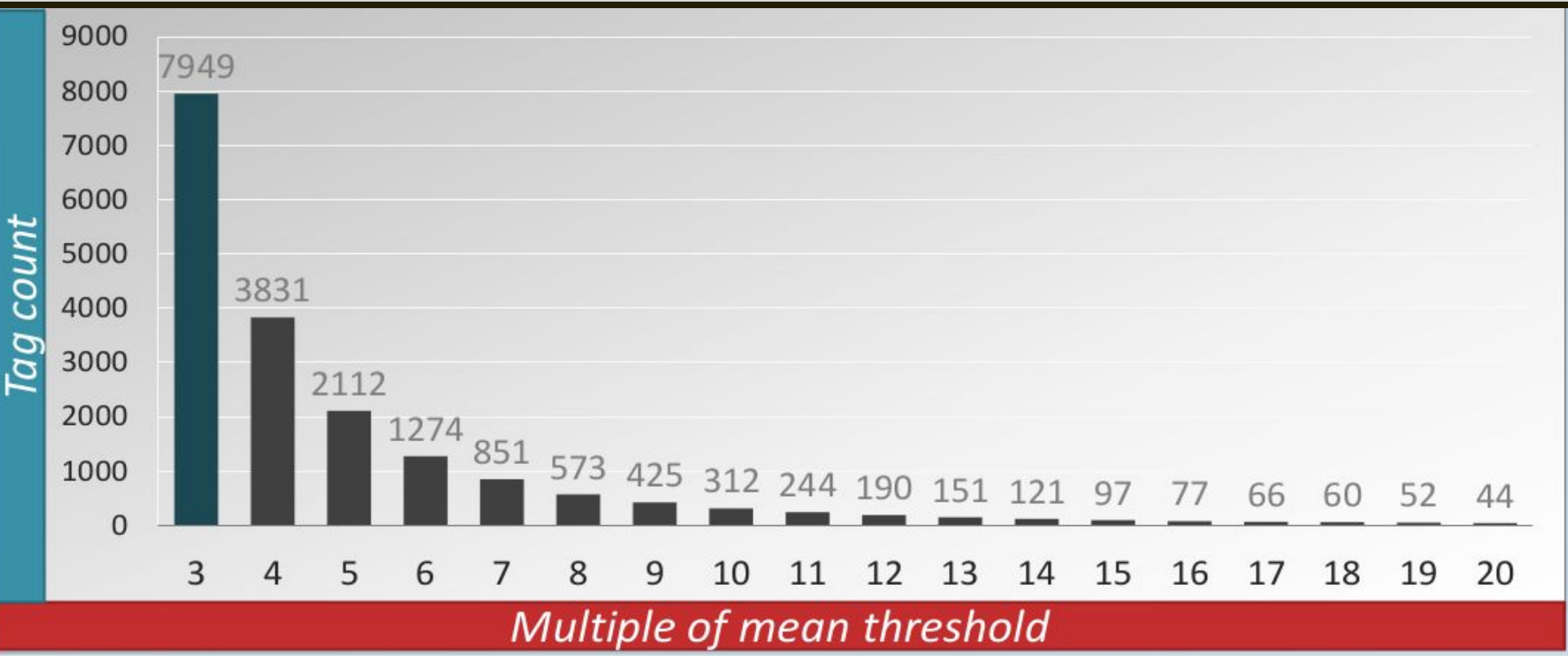
Addo-Quaye et. al 2008

Our generated profile of the 3' bias in Cyanophora paradoxa



Gross et. al 2013

Cleavage tag abundance and noise reduction of dataset



Degradome peaks nX above mean

Discussion

- ➔ Degradome tags are biased to the 3’ end of transcripts
  - ◆ Indicated 3’ bias due to overrepresentation of 5’->3’ decay

- ➔ Bias for antisense small RNAs
  - ◆ In the CDS data: 31,895 contigs had at least one read mapped from a total pool of 1,762,675 reads
  - ◆ 1,394,966 (79.14%) were antisense and 367,709 (20.86%) were sense.

Take Aways

- ➔ There are specific exonic hotspots of sRNA production
  - ◆ Reads mapped to the *Cyanophora paradoxa* datasets (CDS, EST, Genomic) are exonic: predominantly associated with mRNA

- ➔ Possible secondary siRNA pathway may exist that is also seen in plants
  - ◆ May account for “blocking pattern” seen in some contigs
  - ◆ Transitivity may be a common ancestral trait in Plantae group

Conclusion & Future Objectives

EpiMapR: Graphical tool for exploring degradome and epigenome data

Input

Dataset processing and analysis

FASTA

SAM/BAM

R & Bioconductor

GenomicRanges

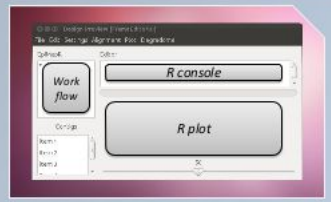
seqinR

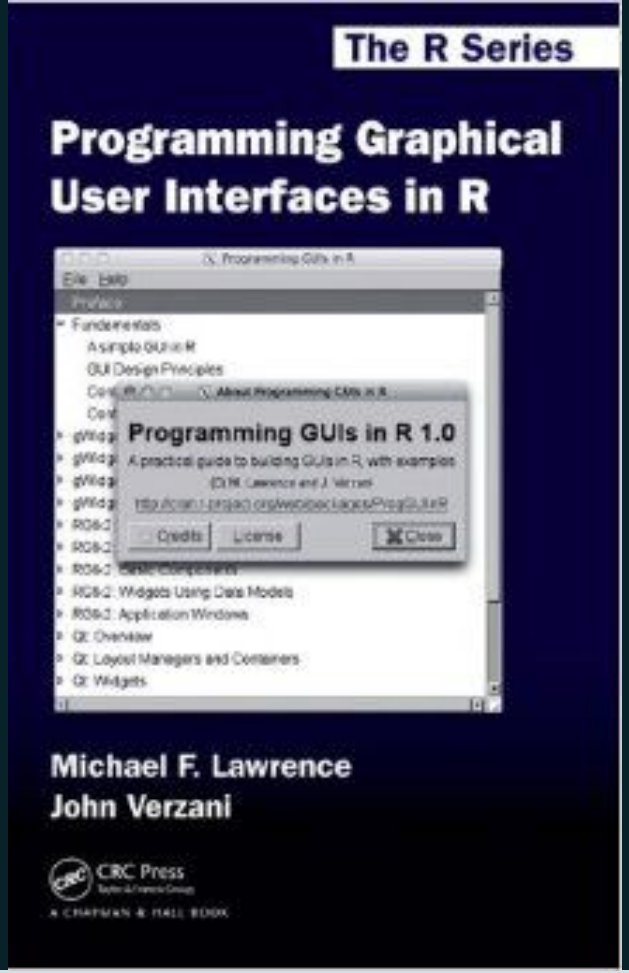
Rsamtools

High-level plotting

lattice

User-friendly Java based UI





User-friendly tool for exploring epigenomic data

References

1. Gross, J., **Wajid, S.**, Price, D. C., Zelzion, E., Li, J., Chan, C. X., & Bhattacharya, D. (2013). Evidence for Widespread Exonic Small RNAs in the Glaucophyte Alga Cyanophora paradoxa. PloS one, 8(7), e67669.
2. Bhattacharya, D., Price, D. C., Chan, C. X., Gross, J., Steiner, J. M., & Löffelhardt, W. (2014). Analysis of the Genome of Cyanophora paradoxa: An Algal Model for Understanding Primary Endosymbiosis. In Endosymbiosis (pp. 135-148). Springer Vienna. Chicago
3. Reyes-Prieto, A., Weber, A. P., & Bhattacharya, D. (2007). The origin and establishment of the plastid in algae and plants. Annu. Rev. Genet., 41, 147-168.
4. Price, D. C., Chan, C. X., Yoon, H. S., Yang, E. C., Qiu, H., Weber, A. P., Schwacke, R. & Bhattacharya, D. (2012). Cyanophora paradoxa genome elucidates origin of photosynthesis in algae and plants. Science, 335(6070), 843-847.
5. Lee, R. E. (2008). Phycology. Cambridge University Press.
6. Seckbach, J. (Ed.). (2002). Symbiosis: mechanisms and model systems (Vol. 4). Springer.
7. Gross, J., & Bhattacharya, D. (2011). Endosymbiont or host: who drove mitochondrial and plastid evolution. Biol Direct, 6, 12.
8. Wise, R. R. (2006). The diversity of plastid form and function. In The structure and function of plastids (pp. 3-26). Springer Netherlands.

References

9. Facchinelli, F., Pribil, M., Oster, U., Ebert, N. J., Bhattacharya, D., Leister, D., & Weber, A. P. (2013). Proteomic analysis of the Cyanophora paradoxa muroplast provides clues on early events in plastid endosymbiosis. Planta, 237(2), 637-651.
10. Falkowski, P. G., Fenchel, T., & Delong, E. F. (2008). The microbial engines that drive Earth's biogeochemical cycles. Science, 320(5879), 1034-1039.
11. Kim, J. D., Senn, S., Harel, A., Jelen, B. I., & Falkowski, P. G. (2013). Discovering the electronic circuit diagram of life: structural relationships among transition metal binding sites in oxidoreductases. Philosophical Transactions of the Royal Society B: Biological Sciences, 368(1622).
12. Mathews, K., van Holde and Ahern. (2000) Biochemistry Third Edition. Benjamin Cummings.
13 Floener, L., Danneberg, G., & Bothe, H. (1982). Metabolic activities in Cyanophora paradoxa and its cyanelles. Planta, 156(1), 70-77.
14. Zvelebil, M. J., & Baum, J. O. (2008). Understanding bioinformatics. Garland Science.
15. Brown, T. A. (2006). Genomes 3. Wiley-Liss, Oxford.
16. Ghildiyal, M., & Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. Nature Reviews Genetics, 10(2), 94-108.
17. Gao, Z., Wang, M., Blair, D., Zheng, Y., & Dou, Y. (2014). Phylogenetic Analysis of the Endoribonuclease Dicer Family. PloS one, 9(4), e95350.

References

18 - Zhang, X., Xia, J., Lii, Y. E., Barrera-Figueroa, B. E., Zhou, X., Gao, S., Lu, L., & Jin, H. (2012). Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. Genome Biol, 13(3), R20.
19. -
20. Shabalina, S. A., & Koonin, E. V. (2008). Origins and evolution of eukaryotic RNA interference. Trends in Ecology & Evolution, 23(10), 578-587.
21. Jackowiak, P., Nowacka, M., Strozycki, P. M., & Figlerowicz, M. (2011). RNA degradome—its biogenesis and functions. Nucleic acids research, 39(17), 7361-7370.
22. Cerutti, H., & Casas-Mollano, J. A. (2006). On the origin and functions of RNA-mediated silencing: from protists to man. Current genetics, 50(2), 81-99.
23. Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X. & Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nature methods, 6(5), 377-382.
24. Nakanishi, K., Weinberg, D. E., Bartel, D. P., & Patel, D. J. (2012). Structure of yeast Argonaute with guide RNA. Nature, 486(7403), 368-374.
25. Joshua-Tor, L., & Hannon, G. J. (2011). Ancestral roles of small RNAs: an Ago-centric perspective. Cold Spring Harbor perspectives in biology, 3(10), a003772.
26. Tuck, A. C., & Tollervey, D. (2011). RNA in pieces. Trends in genetics, 27(10), 422-432

References

27. Poulsen, C., Vaucheret, H., & Brodersen, P. (2013). Lessons on RNA silencing mechanisms in plants from eukaryotic argonaute structures. The Plant Cell Online, 25(1), 22-37.
28. Kuhn, C. D., & Joshua-Tor, L. (2013). Eukaryotic Argonautes come into focus. Trends in biochemical sciences, 38(5), 263-271.
29. -
30. Perrineau, M. M., Zelzion, E., Gross, J., Price, D. C., Boyd, J., & Bhattacharya, D. (2014). Evolution of salt tolerance in a laboratory reared population of Chlamydomonas reinhardtii. Environmental microbiology.
31. Baldauf, S. L. (2003). Phylogeny for the faint of heart: a tutorial. TRENDS in Genetics, 19(6), 345-351.
32. Llave, C., Xie, Z., Kasschau, K. D., & Carrington, J. C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science, 297(5589), 2053-2056.
33. Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y. Y., Sieburth, L., & Voinnet, O. (2008). Widespread translational inhibition by plant miRNAs and siRNAs. Science, 320(5880), 1185-1190.
34. Willmann, M. R., Berkowitz, N. D., & Gregory, B. D. (2013). Improved genome-wide mapping of uncapped and cleaved transcripts in eukaryotes—GMUCT 2.0. Methods.

References

35. Cech, T. R., & Steitz, J. A. (2014). The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. Cell, 157(1), 77-94.
36. Addo-Quaye, C., Miller, W., & Axtell, M. J. (2009). CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. Bioinformatics, 25(1), 130-131.
37. Endres, M. W., Cook, R. T., & Gregory, B. D. (2011). A high-throughput sequencing-based methodology to identify all uncapped and cleaved RNA molecules in eukaryotic genomes. In MicroRNAs in Development (pp. 209-223). Humana Press.
38. Frohman, M. A. (1990). RACE: rapid amplification of cDNA ends. PCR protocols: A guide to methods and applications, 28.
40. Wu, A. R., Neff, N. F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M. E., ... & Quake, S. R. (2014). Quantitative assessment of single-cell RNA-sequencing methods. Nature methods, 11(1), 41-46.
41. SAM Format Specication Working Group. The sam format specification (v1. 4-r985).
42. Perdomo, C., Campbell, J., & Schembri, F. (2014). Detecting Noncoding RNA Expression: From Arrays to Next-Generation Sequencing. In Non-coding RNAs and Cancer (pp. 25-44). Springer New York.
43. Addo-Quaye, C., Eshoo, T. W., Bartel, D. P., & Axtell, M. J. (2008). Endogenous siRNA and miRNA Targets Identified by Sequencing of the Arabidopsis

Thank you for your time!
Questions?