



NRC-CMRC

Plant
Biotechnology
Institute

Brassica napus small RNAs profiling and predictions

Daiqing Huang, Kevin Koh, Larissa Ramsay,
Allan Feurtado, Mathieu Cloutier, Edwin Wang,
Jacek Nowak, Andrew Shape, Adrian Cutler.



National Research
Council Canada

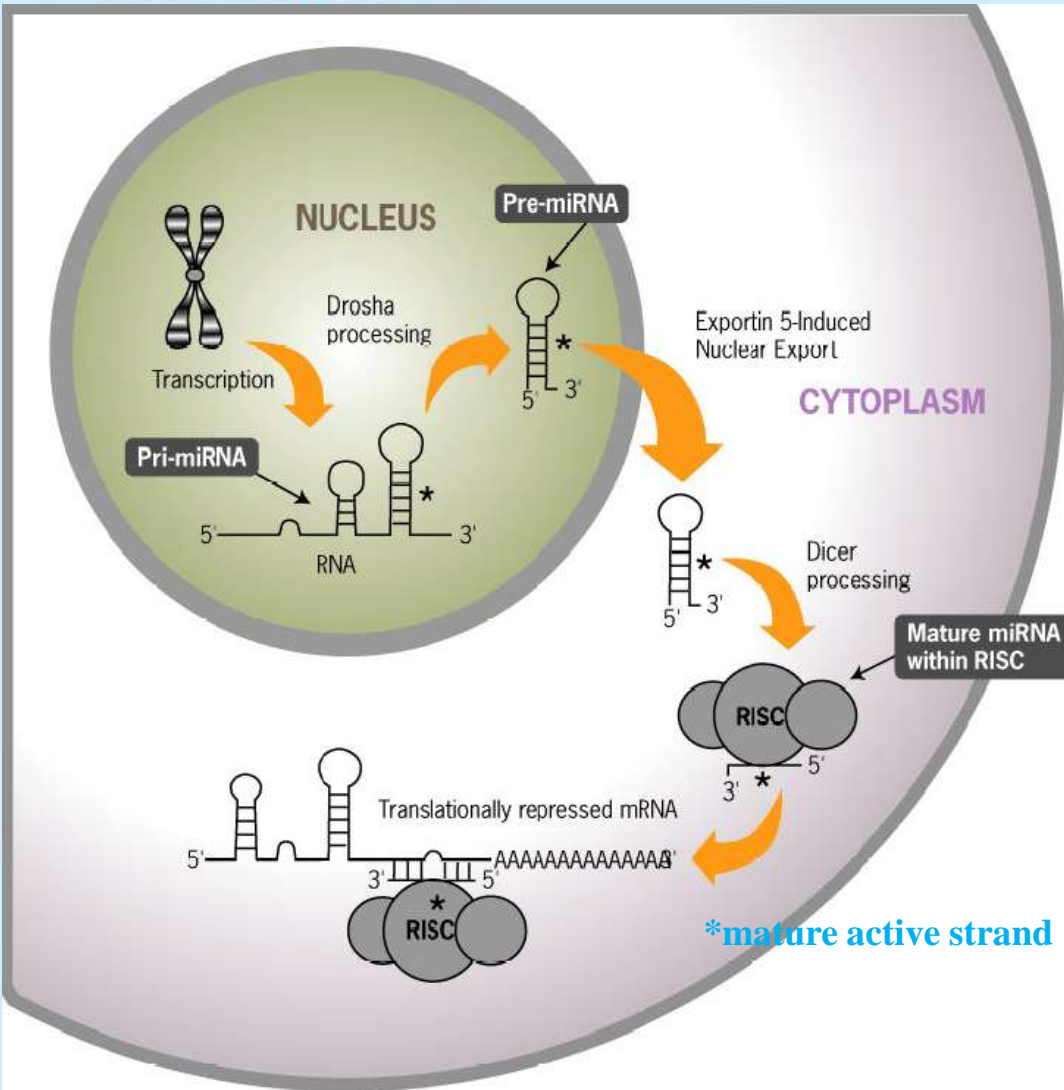
Conseil national
de recherches Canada

Canada

Different classes of small RNAs

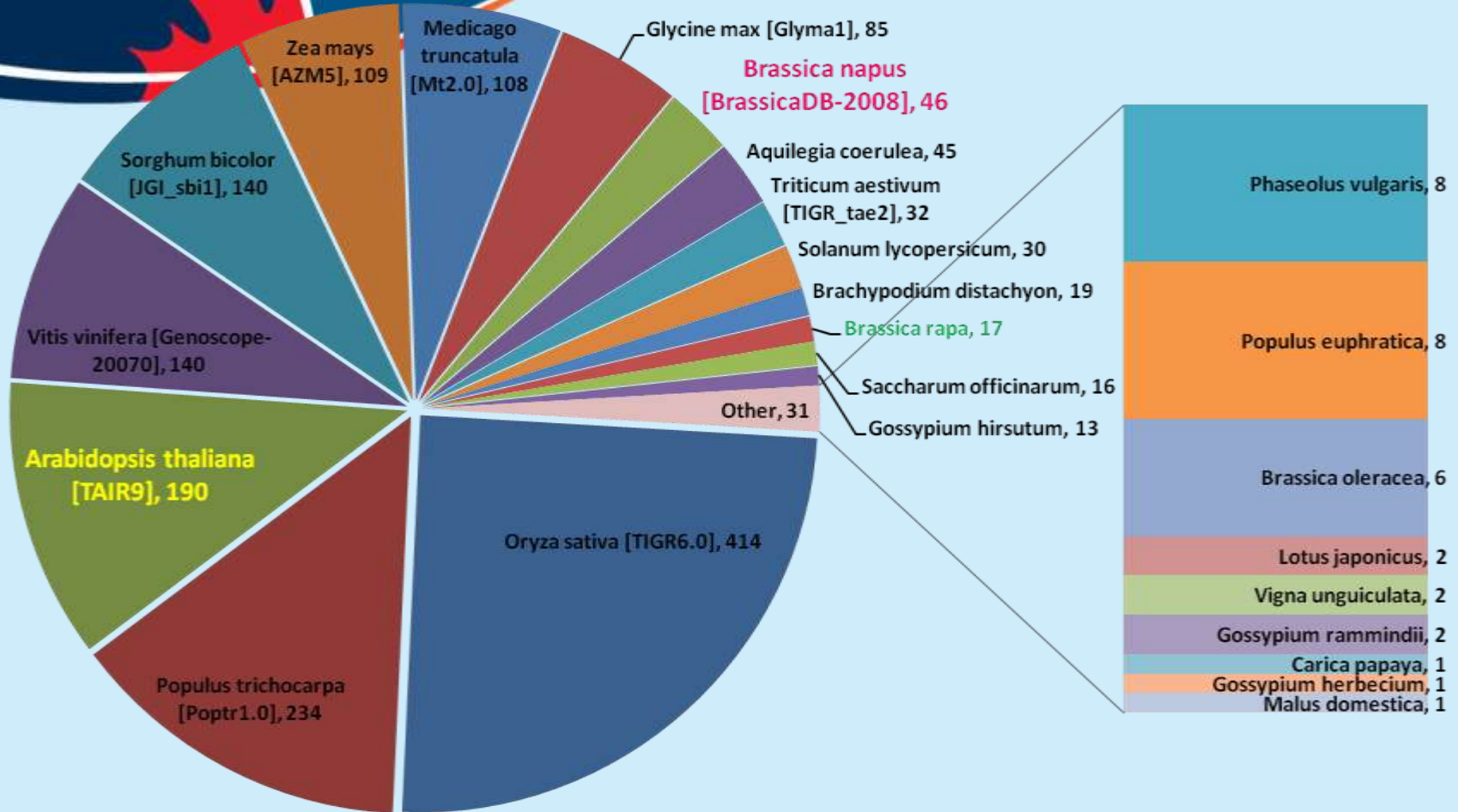
Class of small RNA	Size of mature form (nt)	Structure of precursor	Biogenesis	Mechanism of action
miRNA	20–23	Imperfect hairpin	Successive cleavage by Drosha and Dicer resulting in a mature form with defined sequence	Translational repression, mRNA cleavage
asiRNA	23–28	Long dsRNA	Processing of long dsRNA by Dicer resulting in multiple short RNAs	Regulation of chromatin structure, transcriptional silencing shown in fission yeast and plants
Endogenous siRNA (ta-siRNA, nat-siRNA etc)	20–23	Long dsRNA	Processing of long dsRNA by Dicer. Biogenesis requires RdRP activity	mRNA cleavage
tncRNA	19–23	Unknown	Processing of tiny non-coding RNA depends on Dicer but no hairpin precursor can be identified.	Unknown

miRNA biogenesis



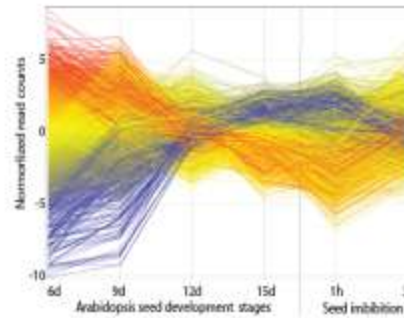
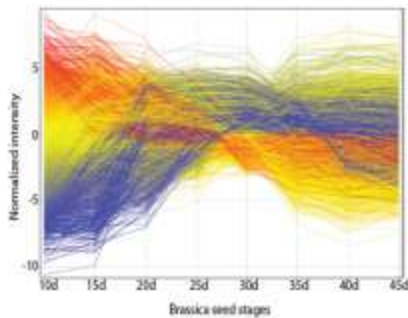
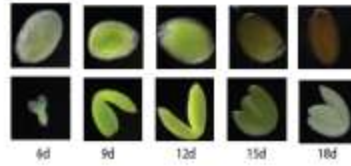
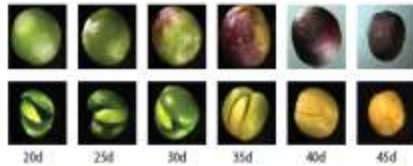
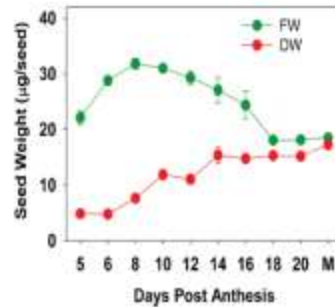
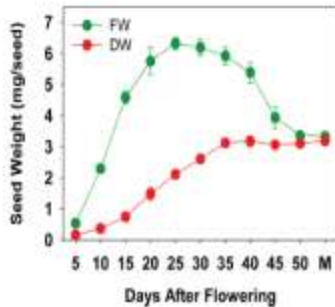
1. Transcription
2. Hairpin release in the nucleus
3. Export to cytoplasm
4. Dicer processing
5. Strand selection by RISC
(RNA-induced silencing complex)
6. Translational repression or mRNA cleavage

Known Plant miRNAs



Release 14 of the database contains 10883 entries representing hairpin precursor miRNAs, expressing 10581 mature miRNA products, in 115 species. 190 in Arabidopsis and 46 in Brassica napus.

Datasets



B. napus and *Arabidopsis* seed maturation stages. Use point of maximum fresh weight (8 days in *Arabidopsis* and 25 days in *Brassica*) as reference point for comparisons.

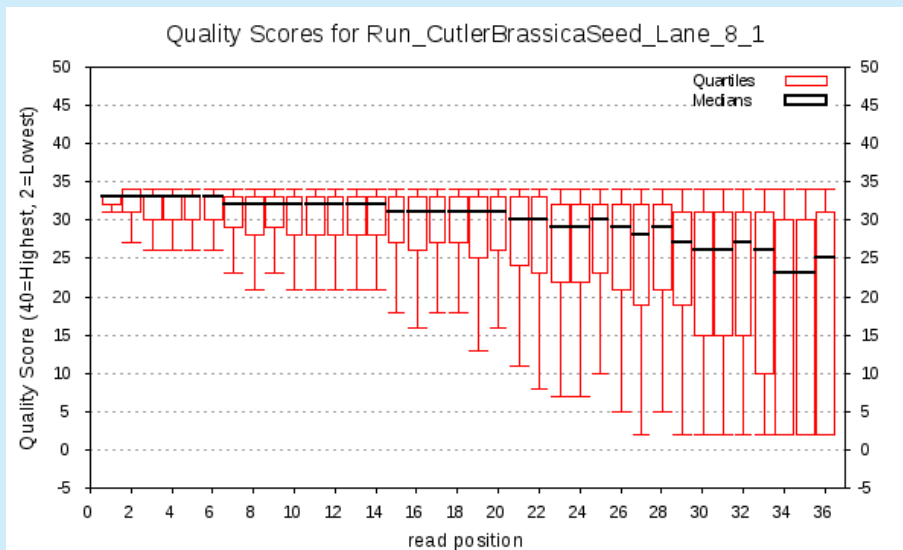
17 Illumina libraries, ~300M 36nts

10 pooled Solid libraries, ~100M 35nts

10 pooled Roche libraries, ~1M reads 50-120nts.

Data Processing

- Filtering out low quality reads.

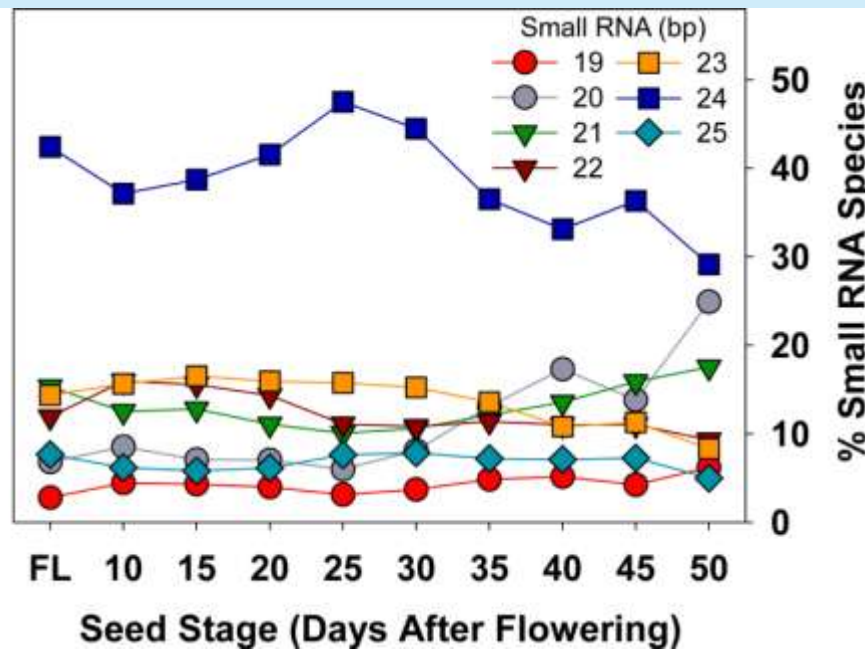
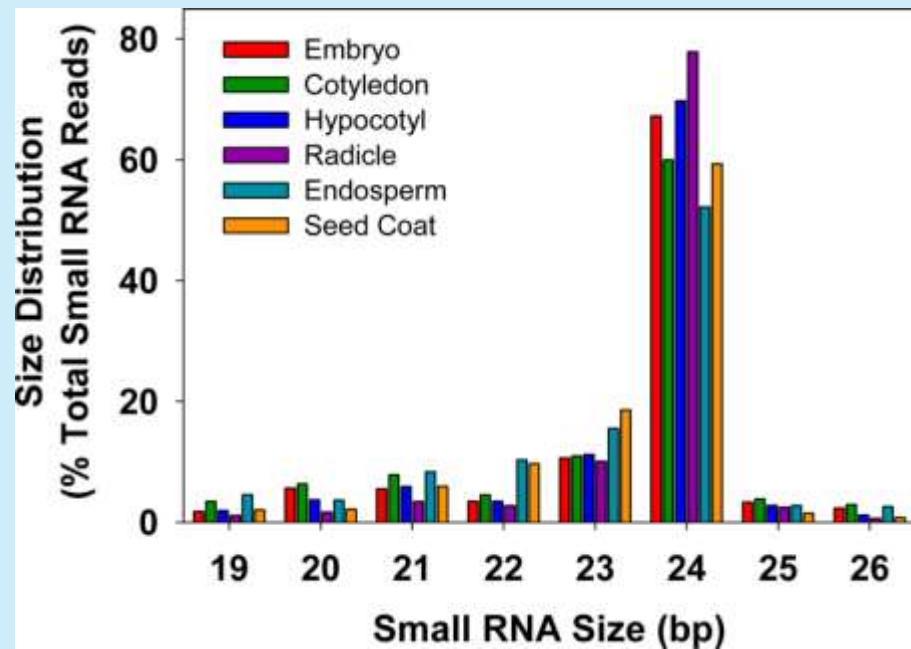


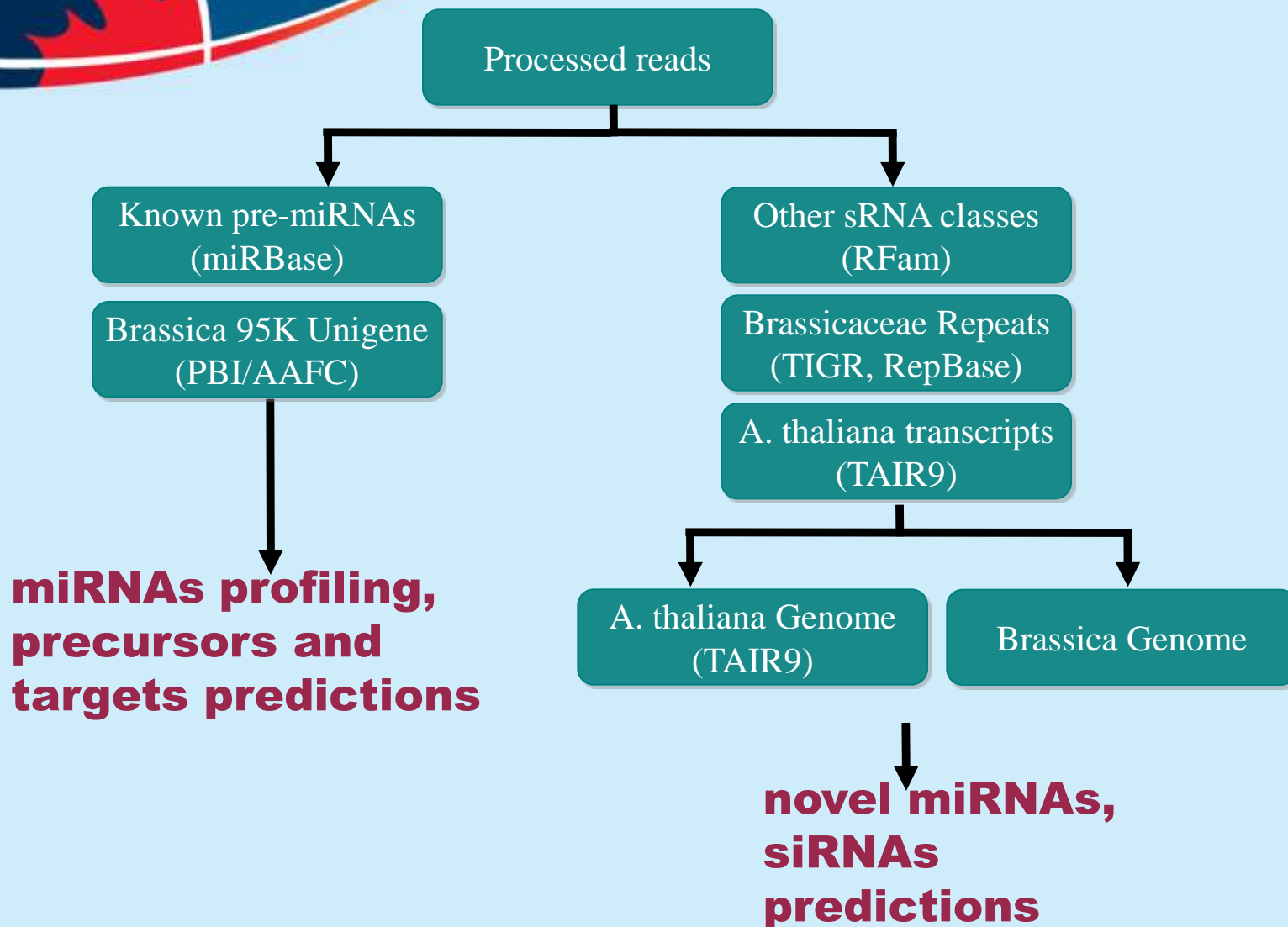
- $\text{Quality_Score}(Q) = -10\log(\text{Err})$, $Q > 15$.
- $Q = 20 \Rightarrow 1\%$ probability of sequencing error.
- $Q = 2 \Rightarrow$ read end indicator (Illumina).

Data Processing

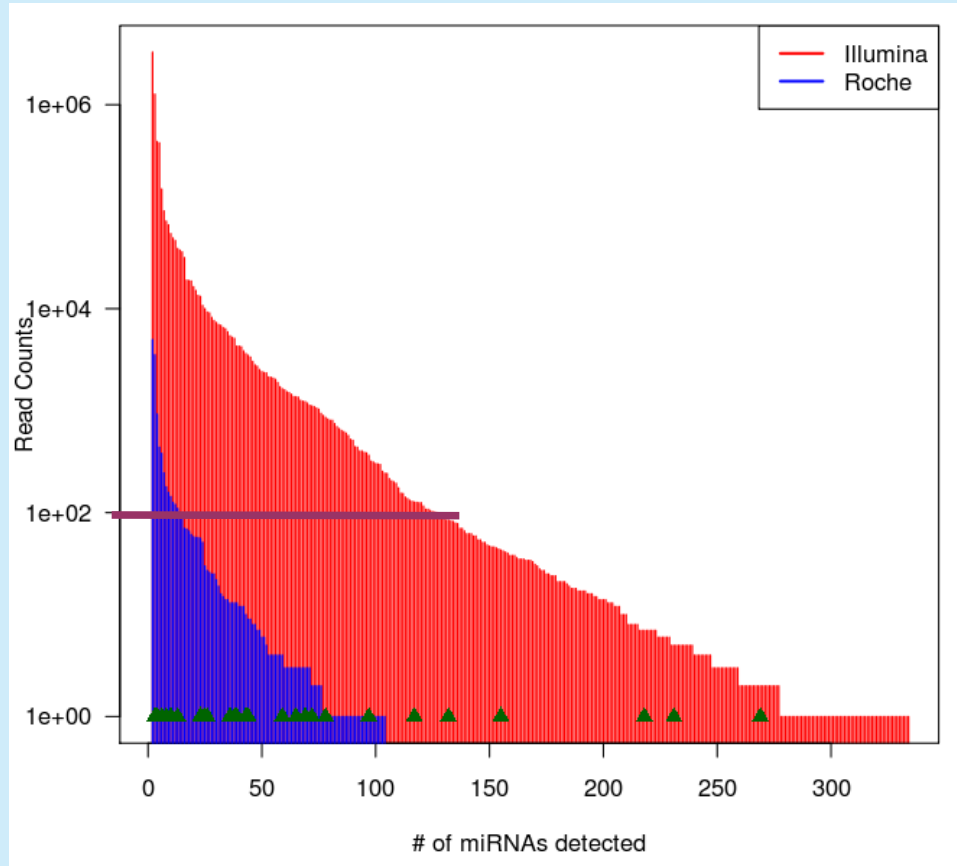
- Trimming 3' adaptors, adaptor-only reads and removing single nucleotide repeats.
- Tools: FASTX ToolKit, Bowtie.
- Computational inexpensive. Balance between the length of 3'-adaptor region, base qualities and number of mismatches.
- Normalized to total number of reads per library per million reads.

Data Processing





Detection and Profiling *B. napus* miRNAs

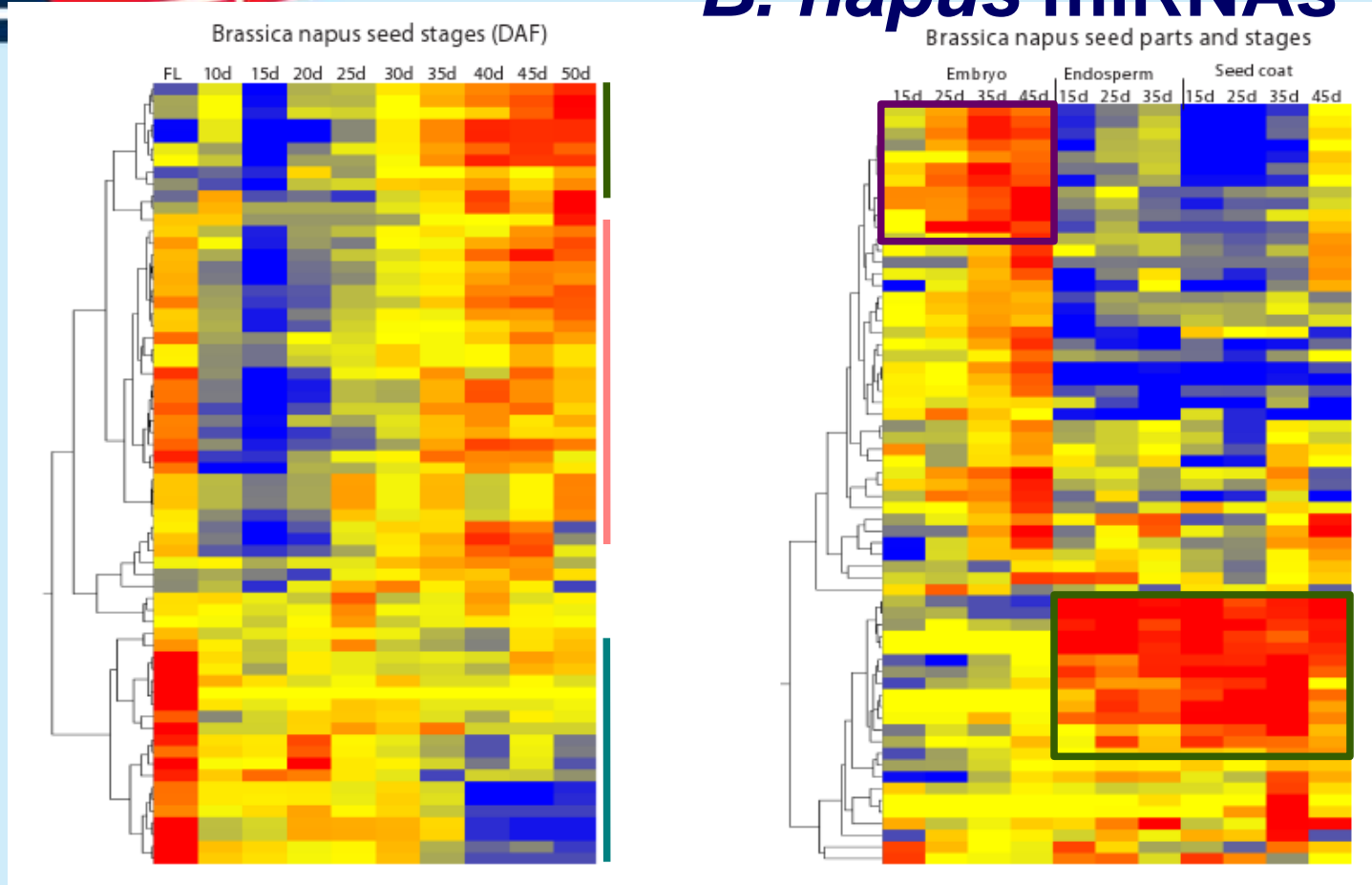


- 300M Illumina reads (approx. 1 flow-cell)
- 1M Roche reads (one plate)

Detection and Profiling *B. napus* miRNAs

- Sequence similarity search against miRBase (v16).
- Illumina libraries (~300M)
 - detected 333 conserved miRNAs, includes all 46 known *B. napus* miRNAs.
 - 101 miRNAs profiles with read counts ≥ 300 .
- Roche libraries (~1M)
 - detected 100 conserved miRNAs.
 - 12 with read counts ≥ 100 .
- No evidence showing if one platform performs better than the other.
- Sequencing depth is crucial for expression profiling.

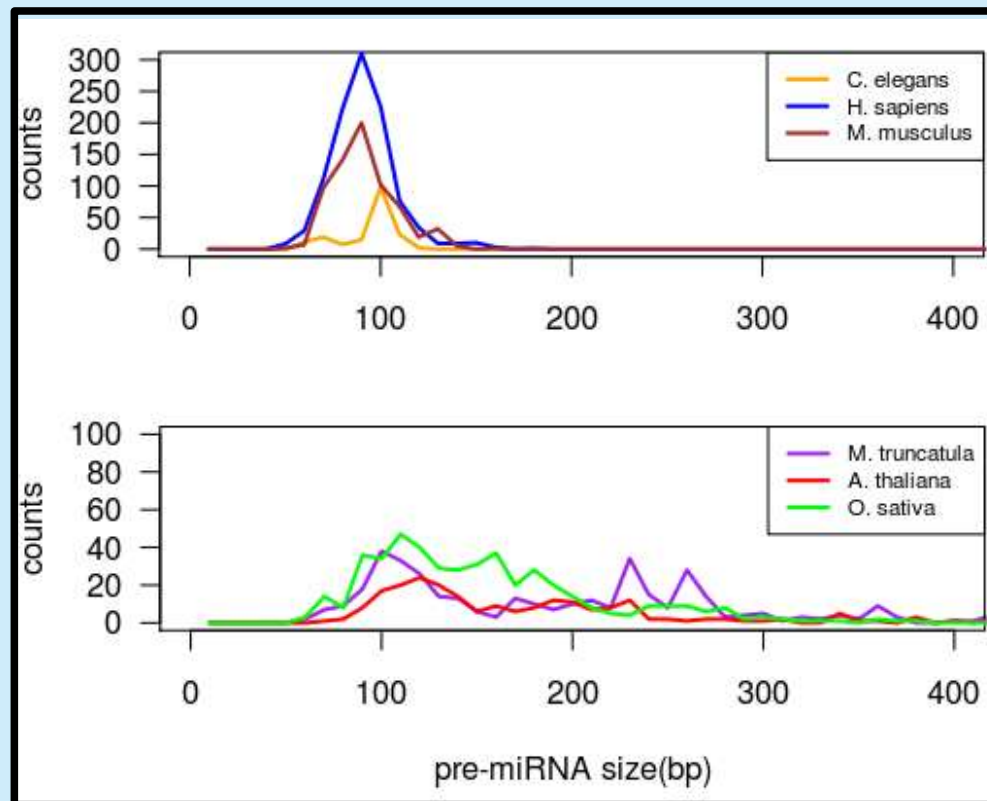
Detection and Profiling *B. napus* miRNAs



- Clustering analyses (GeneSpring) of the most abundant miRNAs.
- Log₂ expression values are plotted and visualized on the color scale: red, high; yellow, moderate; blue, low expression.

Novel miRNAs Predictions

- Plants pre-miRNAs are more heterogenous in sizes compared to mammals. For example, *A.thaliana* (169nt \pm 83nt) versus *H. sapiens* (83nt \pm 16nt).



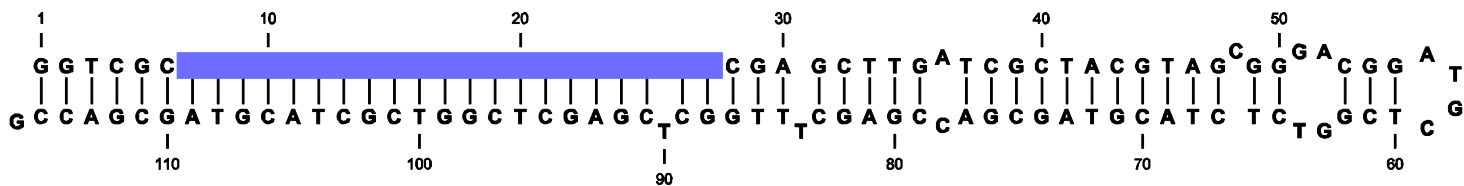
Novel miRNAs Predictions

- Scanning genomic region in a sliding window, fold each sub-region into 2nd structure and examine the thermodynamic stability (i.e. free energy, ΔG), and a stem-loop structure exists (i.e. mature-star sequence pair).
- 2nd pre-miRNA structure (imperfect hairpin) is relatively stable (less bulge and high degree base-pairing) compared to other RNAs.
- Tools:
 - novoMir (designed for plants)
 - HMMMir, TRIPLET-SVM (designed for plants)
 - miRDeep (designed and trained using known *C. elegans* miRNAs)

Visualization (CLCBio Genomic Workbench)



Secondary structure: $\Delta G = -92.1\text{kcal/mol}$

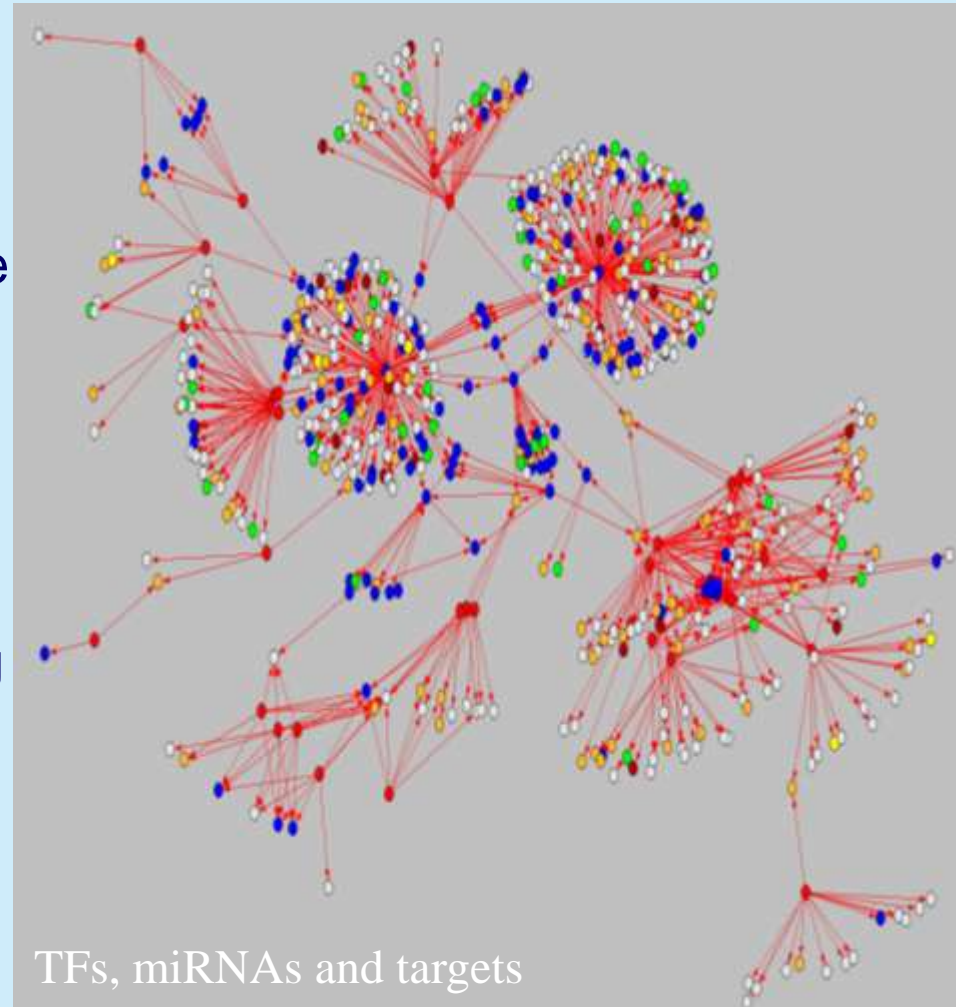


miRNA Target Predictions

- Tool: TargetFinder (v.1.6).
 - high degree of complementarity in 5' and central positions between plants miRNAs their targets.
 - use plant-based position-dependent scoring metric to computationally predict plant miRNA targets.
- Brassica 95K unigene set (PBI/AAFC).
 - >70% of Brassica gene-space.
 - target predictions for miRNAs detected in the NGS experiments.
 - 1,202 targets were predicted from the 101 conserved miRNAs, includes a number transcription factors (SPL2, ARF17, etc.)

Interpreting 'omics' data

- Constructing networks to visualize gene/ protein/ metabolite relationships
 - *De novo*: letting the data define the relationships without preconditions.
 - *Based on prior knowledge*: constructing a network framework by integrating literature information and using genomics data to complement, confirm and enhance it.
- Overlaying different types of 'omics' data

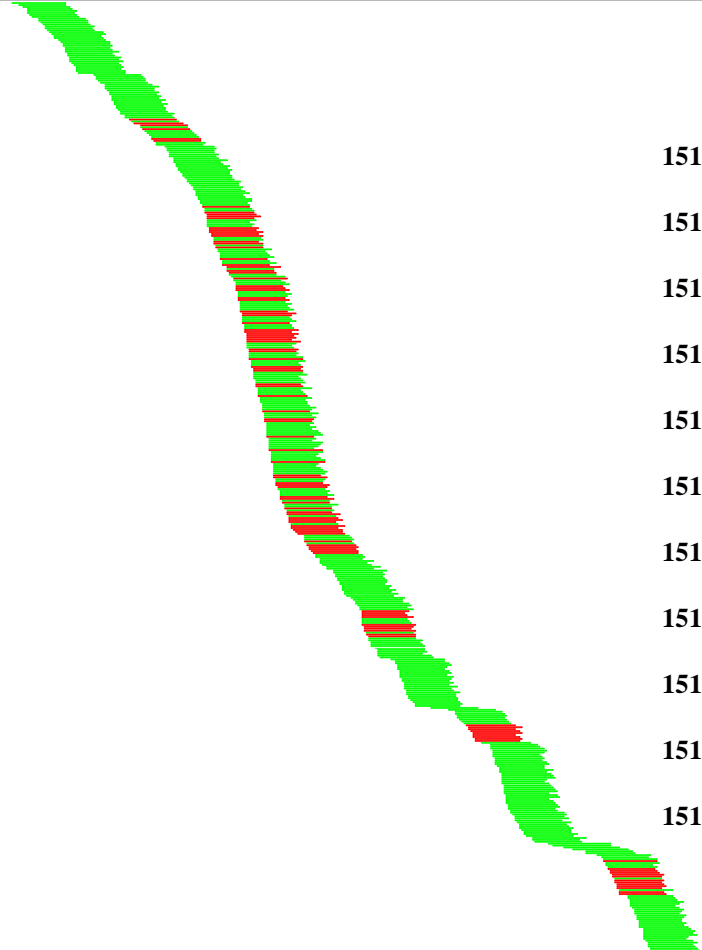
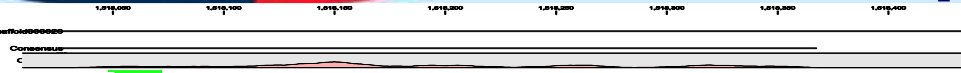


trans-acting siRNAs Predictions

- TAS loci in plant genomes encode ta-siRNAs that regulate expression of a number of genes.
- In Arabidopsis, TAS3 precursor is controlled by two miR390 target sites flanking two ta-siRNA sequences targeting Auxin Response Factor (ARF) transcription factors [lateral root growth related].
- Cleavage of the 3'-miR390-site initiates ta-siRNAs biogenesis.

trans-acting siRNAs Predictions

Scaffold000020

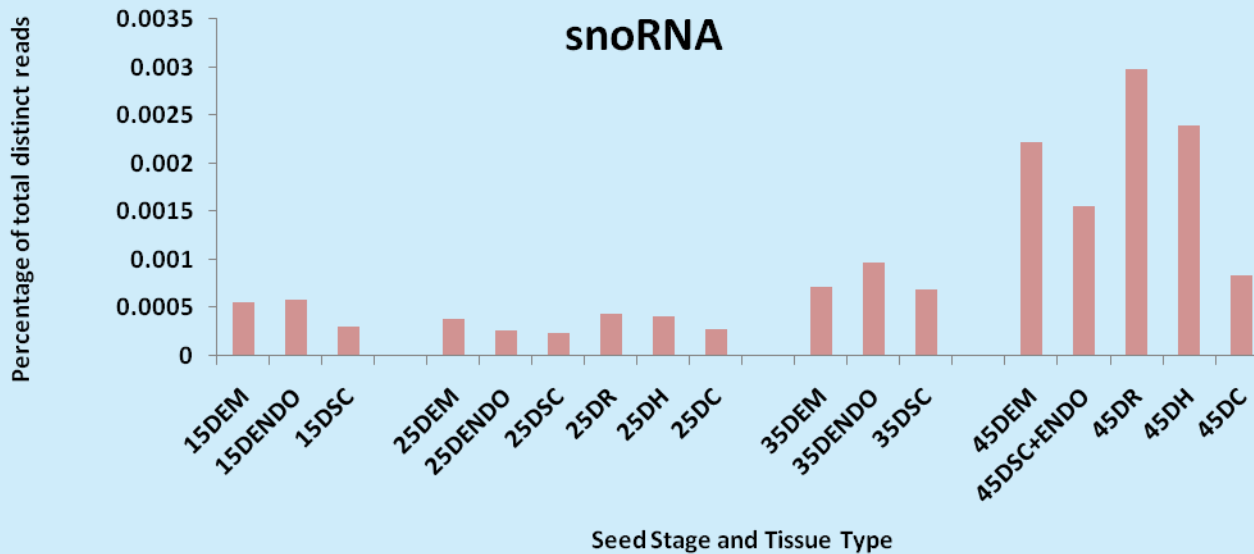
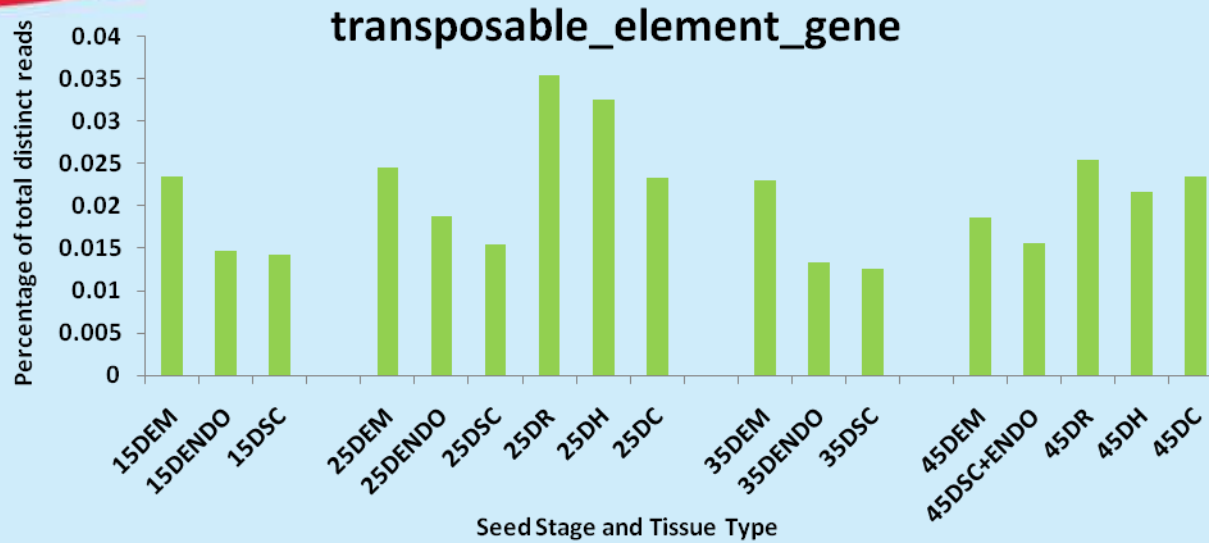


1518140	ngs_44917256_x373	22 +	CAAGATTCTGTAGTGTCAGAG C
1518142	ngs_25848678_x313	22 +	AGATTCTGTAGTGTCAGAGCG G
1518143	ngs_65478652_x634	22 +	GATTCTGTAGTGTCAGAGCGG C
1518144	ngs_42565030_x2394	21 +	ATTCTGTAGTGTCAGAGCGGC
1518145	ngs_97121721_x326	22 +	TTCTGTAGTGTCAGAGCGGCT T
1518148	ngs_94359362_x913	21 +	TGTAGTGTCAGAGCGGCTTCG
1518159	ngs_27403947_x433	21 +	AGCGGCTTCGTCCCTCGTGG A
1518245	ngs_53410554_x1224	22 +	CGGCTTGACATCAATCATGG C
1518249	ngs_97783640_x743	22 +	TTGGACATCAATCATGGCTAC G
1518250	ngs_92886477_x339	22 +	TGGACATCAATCATGGCTACG C
1518344	ngs_96320231_x427	21 -	GGTCTTCGTCGAGTTCTTGAA



**21nt phased siRNAs

Other small RNAs



Tools

- FASTX ToolKit (http://hannonlab.cshl.edu/fastx_toolkit/)
- Bowtie (<http://bowtie-bio.sourceforge.net/index.shtml>)
- TargetFinder (<http://jcclab.science.oregonstate.edu/node/view/56334>)
- BLAST (<http://www.ncbi.nih.gov>)
- NovoMir (<http://www.biophys.uni-duesseldorf.de/novomir/>)
- MirDeep(<http://www.mdc-berlin.de/>)
- CLCBio (<http://www.clcbio.com>)
- IGV (<http://www.broadinstitute.org/software/igv/>)
- MiRBase (<http://www.mirbase.org>)
- Rfam (<http://rfam.sanger.ac.uk/>)
- UEA Plant sRNA tools (<http://srna-tools.cmp.uea.ac.uk/plant/cgi-bin/srna-tools.cgi>)

NRC CNRC

*Plant
Biotechnology
Institute*

Science
— at work for —
Canada



National Research
Council Canada

Conseil national
de recherches Canada

Canada

small RNAs biogenesis and classification

