Mississippi: a galaxy toolsuite for small RNA analysis

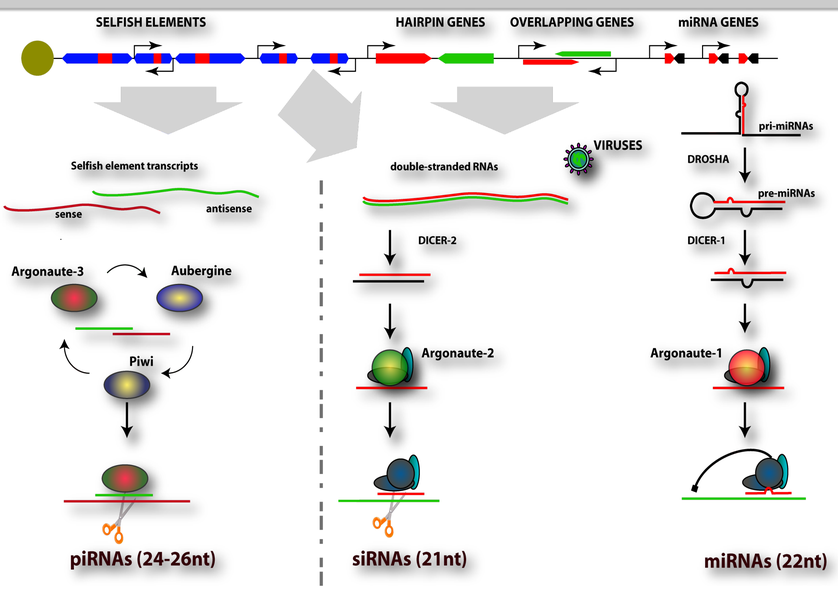
http://mississippi.fr
Marius van den Beek and Christophe Antoniewski
Université Pierre et Marie Curie - Bât. C, 5ème étage – Drosophila Genetic and Epigenetic Lab. 9, Quai Saint Bernard - 75252 PARIS CEDEX 05
email: marius.van-den-beek@snv.jussieu.fr, christophe.antoniewski@snv.jussieu.fr

Background

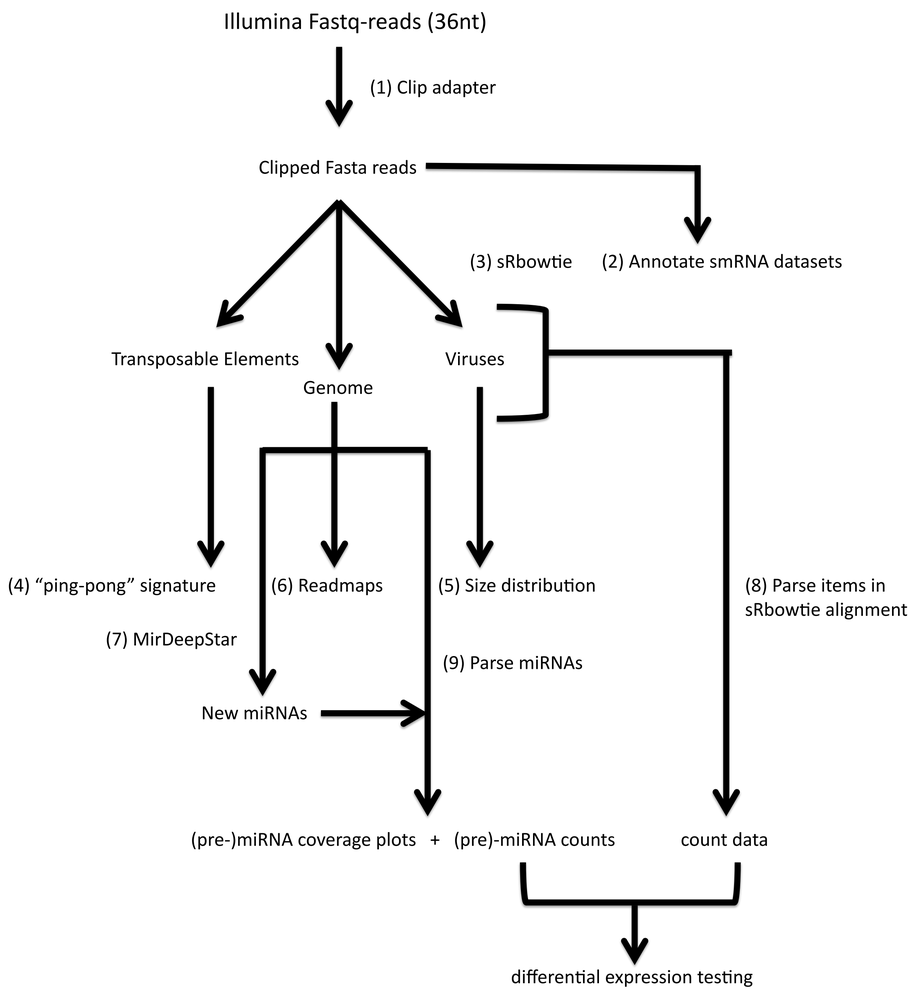
Non-coding small RNAs (miRNA, siRNA, piRNA, …) are involved in the regulation of genes and transposable elements as well as in the defense against viral infections. Their discovery and their functional characterization rely heavily on high throughput RNA sequencing.

The ~20:30nt length of small RNAs raises specific challenges for meaningful read mapping and analysis, so that standard RNAseq analysis methods have to be adapted.
Based on our own experience and through interactions with collaborating teams, we provide an integrated set of galaxy tools that should streamline the most frequent small RNA analysis workflows, while being usable by novice users.
The toolsuite includes a modified bowtie-wrapper and workflows that allow users to quickly and reproducibly interrogate various aspects of small RNA biology.
We provide tools for the discovery and differential expression analysis of miRNAs and a way for genome-wide visualization of miRNA precursors that complements Trackster.
Furthermore we provide tools to detect the “ping-pong” biogenesis signature of piRNAs, to detect piRNA-producing loci in the genome and to study and visualize the impact of piRNAs and siRNAs on transposable elements.

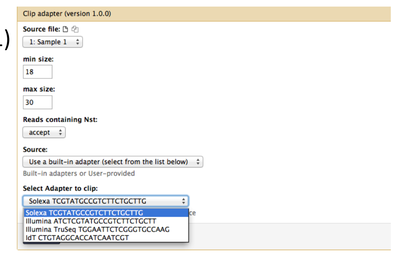
A demo server is available at http://mississippi.fr, and the mississippi toolsuite can be installed from the testtoolshed.



A typical mississippi analysis workflow

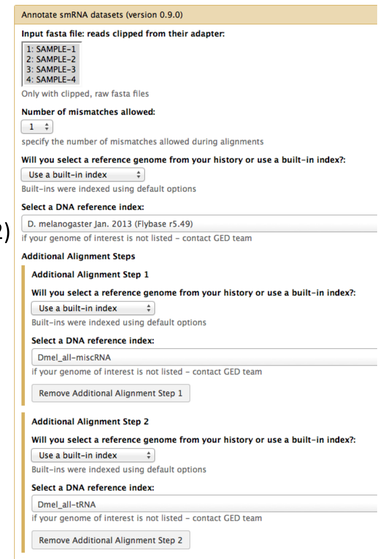


(1)



A simple and easy to use read clipper. Read clipping is very important, since unclipped reads would result in misinterpretation of small RNA size, and hence the biological role of the small RNA.

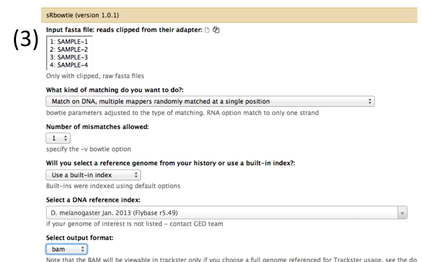
(2)



alignment reference	sample_1.fasta	sample_2.fasta	sample_3.fasta	sample_4.fasta
D. melanogaster Jan. 2013 (Flybase r5.49)	2,728,219	3,222,565	3,991,609	4,643,928
dme_miR_r20	295,494	307,743	505,760	469,219
Dmel_all-miscRNA	732,351	1,053,787	1,035,978	1,239,209
Ensembl_Dmel_extended_tRNA	17,587	21,687	23,327	35,660
Dmel_all-ncRNA	14,418	15,979	15,783	29,145
Dmel_all-transposon	1,195,910	1,316,928	1,753,348	2,110,280
Dmel_all-intron	121,776	139,139	179,298	236,484
Ensembl_Dmel_extended_tRNA	121,776	139,139	179,298	236,484
Remaining Unmatched	350,683	367,302	478,115	523,931

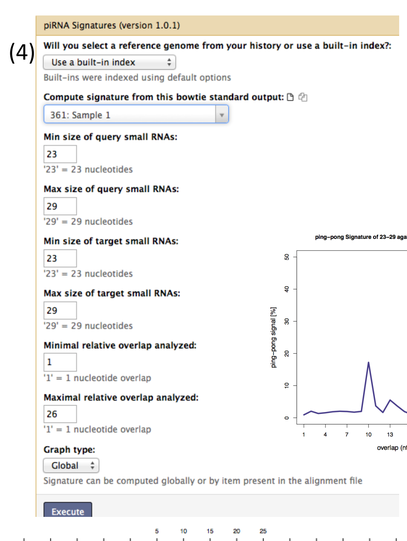
A tool for rapid iterative mapping and counting of small RNA reads against various references. This allows for a rapid check of library quality and can be used to obtain library normalization factors.

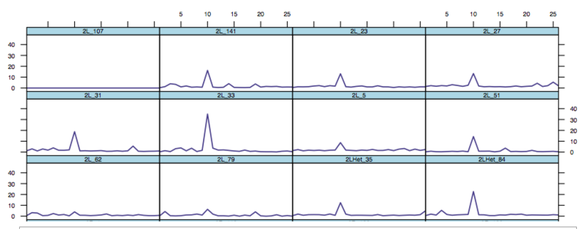
(3)



An alternative wrapper for Bowtie 1, focused on small RNA mapping. Allows fasta reads as input, hides commandline options and provides the possibilty to (additionally) obtain mapped and unmapped reads as fasta files.

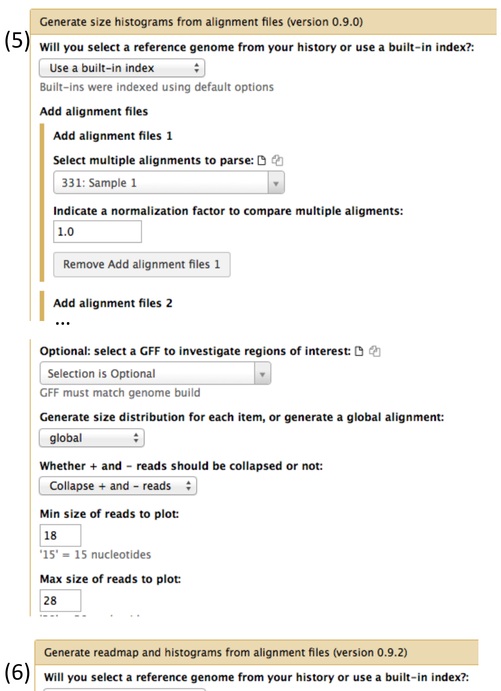
(4)



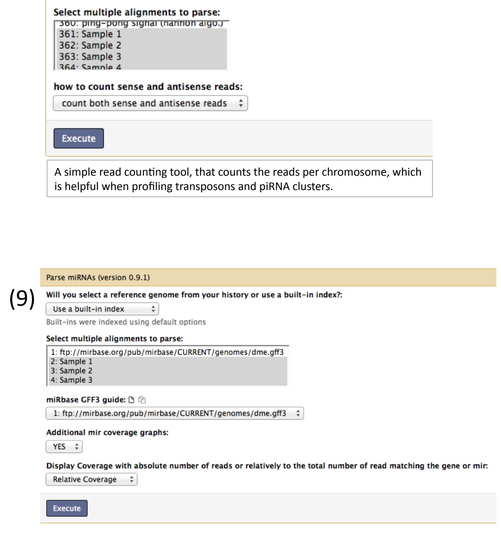


The piRNA signatures tool calculates the overlap frequency between mapped reads at a specific position and reads that are in the immediate vicinity. Certain genomic clusters give rise to reads preferentially overlapping by 10 nucleotides, a consequence of the biogenesis. The user can choose to display the tendency for all mapped reads or split by chromosome header (controlled by the “Graph type” option).

(5)

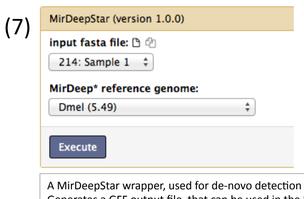


(6)



The size distribution and readmap tools allow the rapid analysis of the size distribution and location of mapping for hundreds of loci for multiple sampleswith a single click. The tools can optionally be guided by a GFF file describing the coordinates of interesting features, such as transposable elements, piRNA clusters, tRNA loci, a set of up- or downregulated genes

(7)



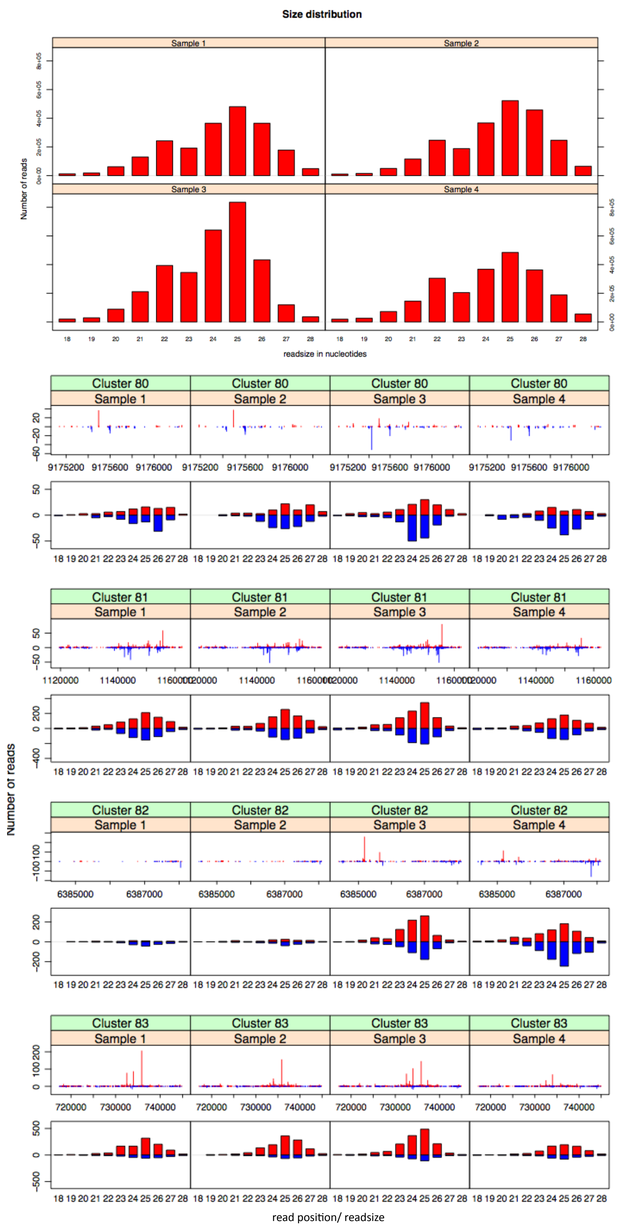
A MirDeepStar wrapper, used for de-novo detection of miRNAs. Generates a GFF output file, that can be used in the Parse miRNA tool to profile miRNA expression.

(8)

A simple read counting tool, that counts the reads per chromosome, which is helpful when profiling transposons and piRNA clusters.

(9)

A tool that allows the visual inspection of pre-miRNA arm usage. Aso outputs count lists that can be used for differential expression testing.



Seqid	Source	Type	Start	End	Score	Strand	Phase	Attributes
##gff-version 3								
##Seqid	Source	Type	Start	End	Score	Strand	Phase	Attributes
##								
2L	MirDeep*	hairPin_loci	64510	64590	-0.86	+	.	ID=novelMiR_0;99_reads;gactcatcaccg
2L	MirDeep*	hairPin_loci	857550	857626	-4.18	+	.	ID=novelMiR_1;10_reads;cgaattacttggg
2L	MirDeep*	hairPin_loci	2115635	2115713	-8.18	+	.	ID=novelMiR_2;11_reads;tctttttaactcctt
2L	MirDeep*	hairPin_loci	2116814	2116890	-7.87	+	.	ID=novelMiR_3;14_reads;gatcaaaaacttc
2L	MirDeep*	hairPin_loci	3208007	3208083	46.64	+	.	ID=novelMiR_4;40_reads;tctgaaagttctt
2L	MirDeep*	hairPin_loci	4606972	4607048	46.64	+	.	ID=novelMiR_5;40_reads;tctgaaagttctt
2L	MirDeep*	hairPin_loci	4665352	4665430	-8.18	+	.	ID=novelMiR_6;11_reads;tctttttaactcctt
2L	MirDeep*	hairPin_loci	4666531	4666607	-7.87	+	.	ID=novelMiR_7;14_reads;gatcaaaaacttc
2L	MirDeep*	hairPin_loci	4920474	4920554	-3.91	+	.	ID=novelMiR_8;10_reads;gtcagcattatgg
2L	MirDeep*	hairPin_loci	5344491	5344567	-3.33	+	.	ID=novelMiR_9;13_reads;cgaaccaggtg
2L	MirDeep*	hairPin_loci	5641073	5641149	4.35	+	.	ID=novelMiR_10;9_reads;ccacatttctgagt
2L	MirDeep*	hairPin_loci	5642117	5642195	-0.26	+	.	ID=novelMiR_11;52_reads;ccaacttgcctt
2L	MirDeep*	hairPin_loci	6799729	6799807	-3.88	+	.	ID=novelMiR_12;76_reads;tagaataaacc
2L	MirDeep*	hairPin_loci	6991625	6991705	-2.22	+	.	ID=novelMiR_13;19_reads;gcagcaggcg