

srnaMapper is an exact, exhaustive, and efficient short RNA read mapper.

Matthias Zytnicki, Christine Gaspin

Short RNA seq

- mostly <50 bp long reads
- highly redundant reads
- no dedicated mapping tool

Strategy

- collapse the reads into a tree
- store the genome into a suffix array + BWT

Results

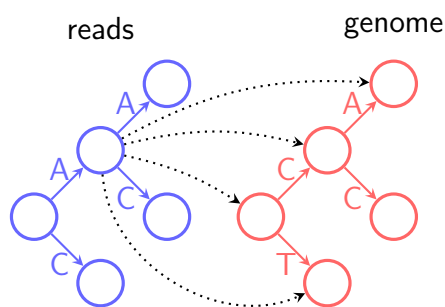
15.5 million reads, mapped to *A. thaliana*.

Results in minutes.

tool	all	collapsed
bwa aln	16	6
bwa aln -N	≈ 140	35
bwa mem	5	1.5
bwa mem -a	6	1.8
srnaMapper	7	—

Algorithm

- tree inclusion with error
- find all the best matches within distance
- compute at each read node the genome nodes at distance 0, at distance 1, etc.



1-differences of prefix A of the reads on the genome.

- the corresponding genome nodes of the reads nodes are computed recursively, using the information computed for the parent node

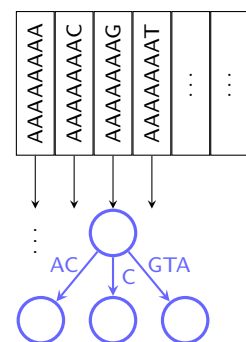
Gory details

Optimizations

- discard small reads (<8 bp)
- discard low complexity reads

Implementation

- Genome structure uses the BWA API.
- The n -differences of the 8-mers are pre-computed.
- Reads are stored in a radix tree.



Structure of the reads tree.

Future work

- multiple threading
- use a penalty scheme dedicated to miRNAs (which may contain edition at their ends)
- other algorithmic and hardware-related optimisations



← Get srnaMapper

