Endovir

Report NCBI Visiting Bioinf

Jan P. Buchmann

NCBI, NLM, NIH, Bethesda, MD, USA
The University of Sydney, Sydney, Australia

2017-12-08

Exogenous / Endogenous viruses

Exogenous virus

- Integration into host genome not required for replication
- Can occur occasionally, e.g. Human Hepres virus 6
- Clearly separated form host genome

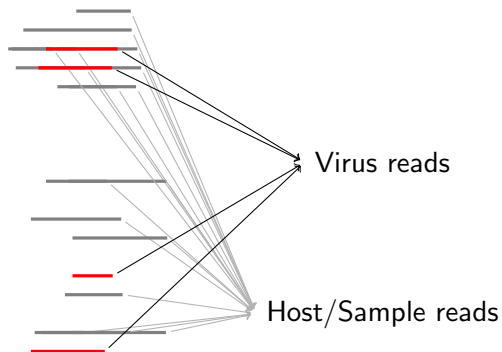3′ — [N] — [P] — [LNYV] — [M] — [G] — [L] — 5′

Endogenous virus

- Integration into host genome mandatory for replication
- Retroviridae, e.g. HIV
- Part of the host genome

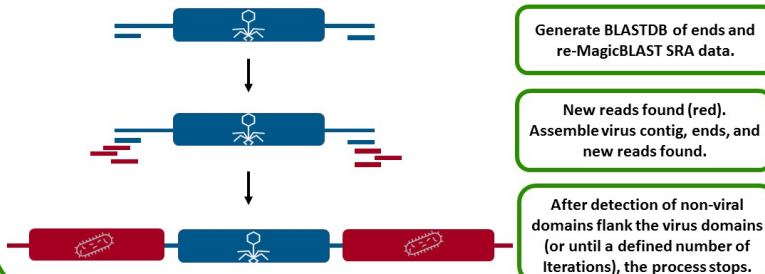5′ ———————— [ORF1]-[ORF2]-[ORF3] ———————— 3′

SRA data

- SRA data contains host and virus sequences
- Endogenous and putative novel viruses

Virus reads

Host/Sample reads

Previous work

- ViruSpy [1]
- Logic mostly implemented in Bash and Perl
  $\rightarrow$ Complex to expand and adjust
- BUD algorithm

Source: https://github.com/NCBI-Hackathons/ViruSpy
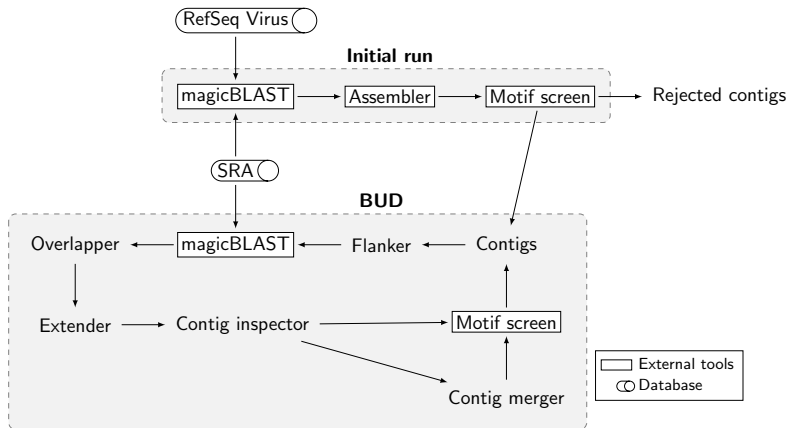
[1] https://github.com/NCBI-Hackathons/ViruSpy
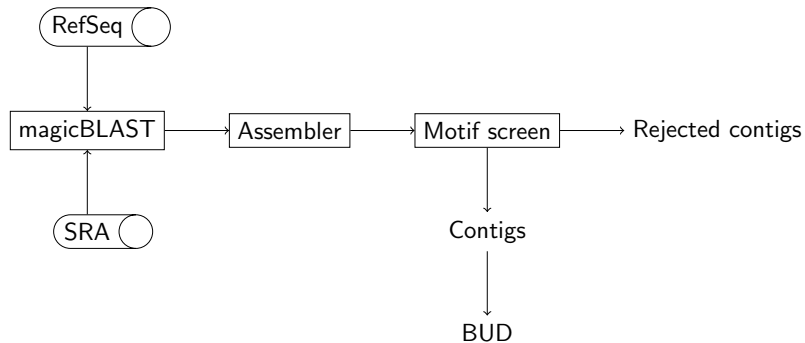
Endovir

Usage

```
usage: endovir.py [-h] [-srr SRR] [--wd WD] [--max_cpu MAX_CPU]

Endovir

optional arguments:
  -h, --help              show this help message and exit
  -srr SRR                SRR number, e.g. SRR5150787
  --wd WD                 Working directory for analysis
  --max_cpu MAX_CPU, -p MAX_CPU
                          Max number of cores to use. NOT YET IMPLEMENTED
```
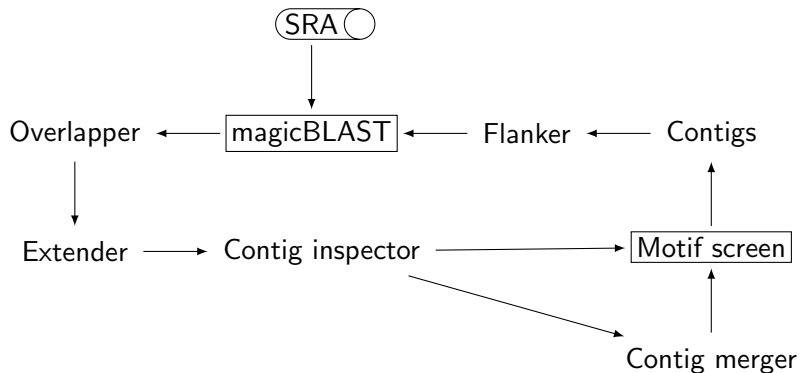
Init

Bud

Flanker