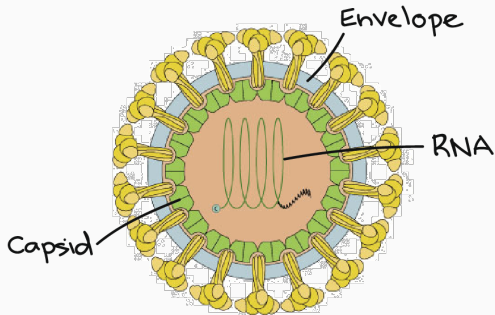


A puzzle with a million pieces: assembling viral genomes from sequencing data

Jasmijn Baaijens

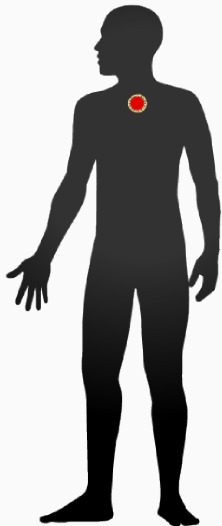
March 31st, 2017

Virus basics



- Diameter: 20-300 nanometres
- Genetic material + protein coat (capsid)
- DNA (A,C,T,G) or *RNA* (A,C,U,G)
- Ebola, Zika, HIV, Hepatitis, ...

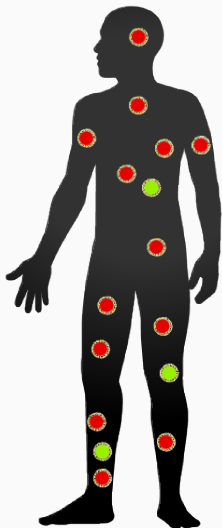
Virus infection



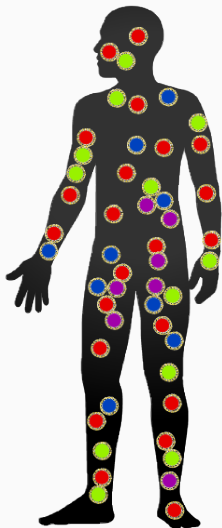
Virus infection



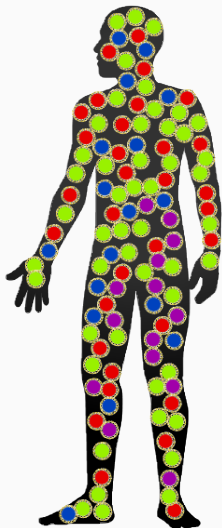
Virus infection



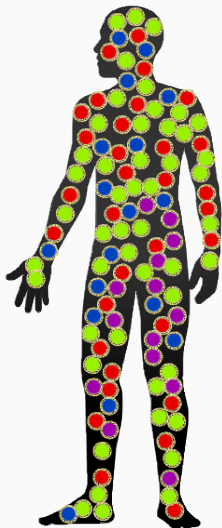
Virus infection



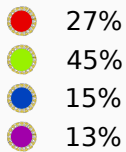
Virus infection



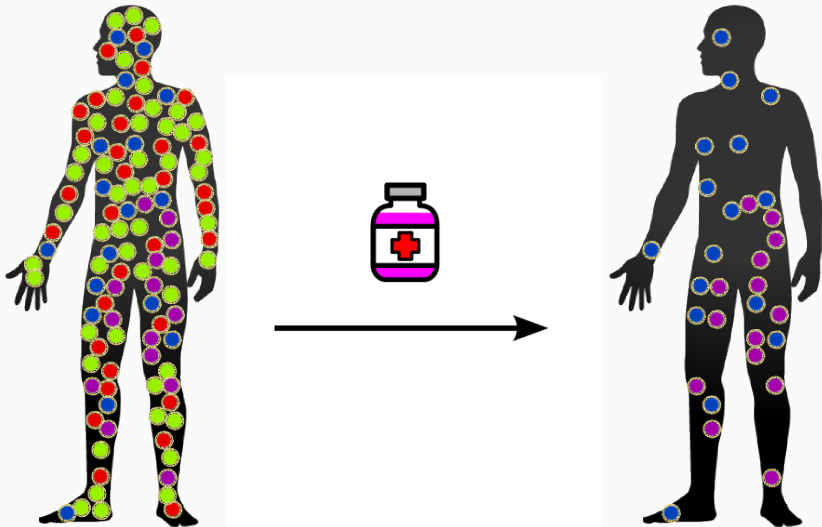
Virus infection



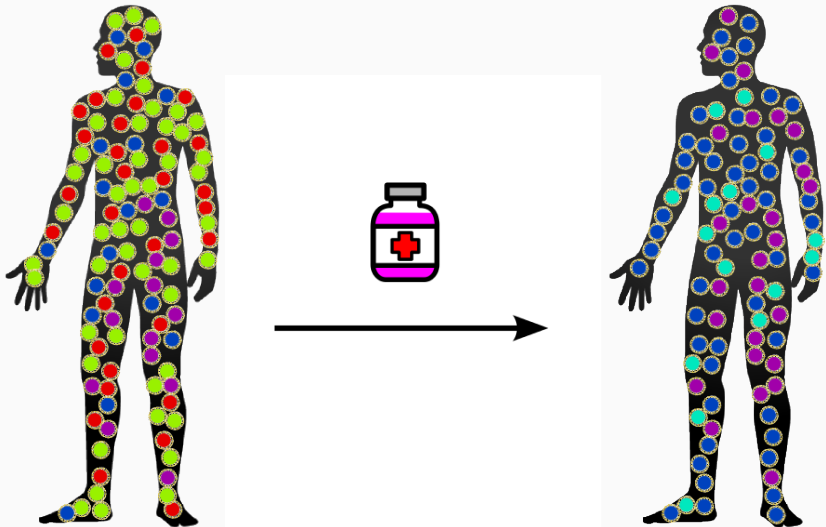
Varying frequencies:



Quasispecies behaviour



Quasispecies behaviour

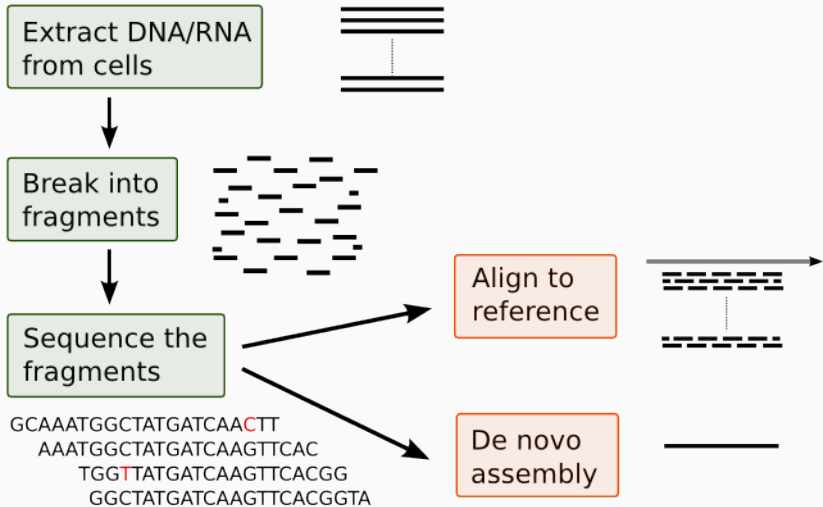


But how do we find out which strains
are present during an infection?

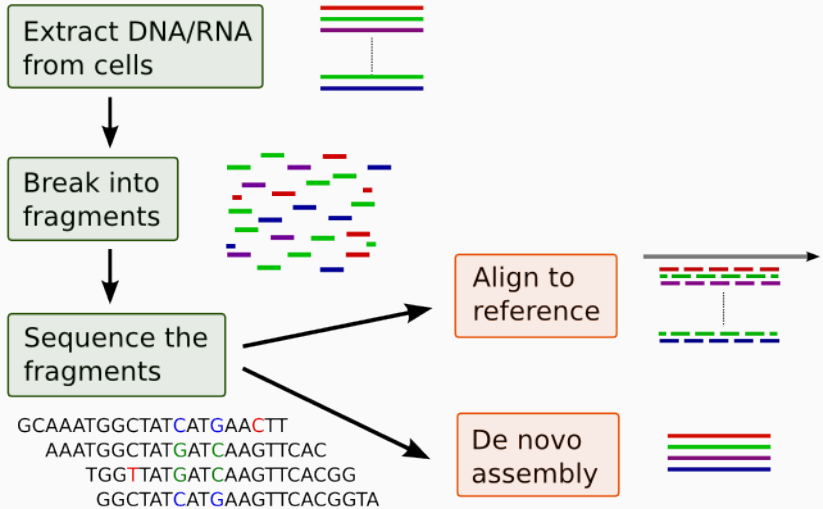
Genome sequencing



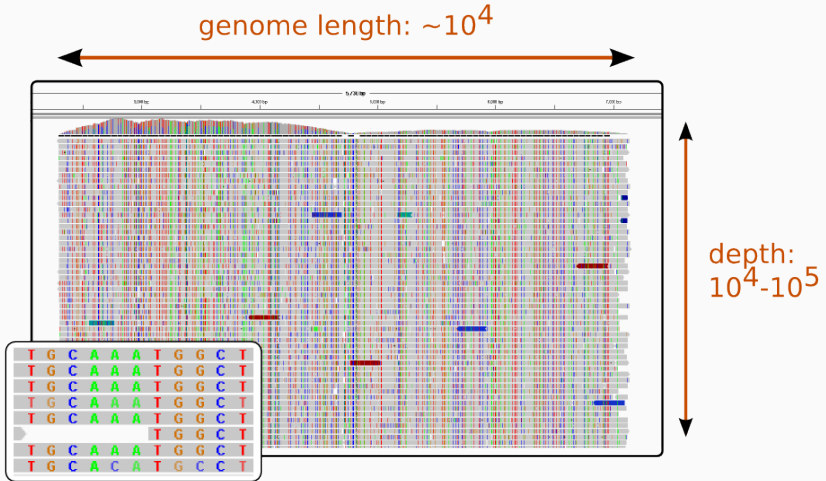
Genome sequencing



Genome sequencing



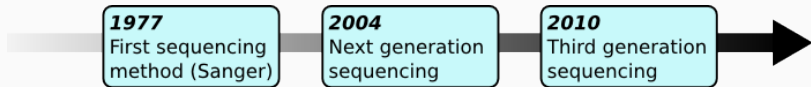
Data: sequencing reads



Goal:

Reconstruct each of the individual
virus strains (haplotypes)
without using a reference genome

De novo genome assembly



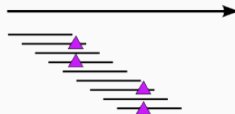
Overlap graphs

nodes: sequencing reads
edges: suffix-prefix overlaps



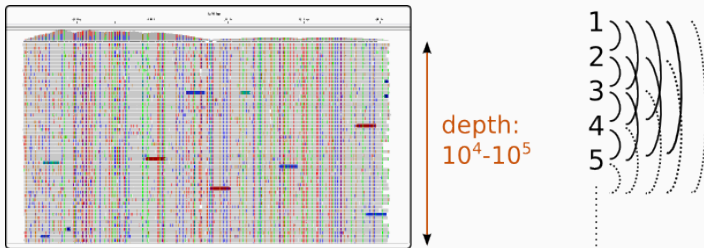
De Bruijn graphs

length k substrings
length $k-1$ overlaps

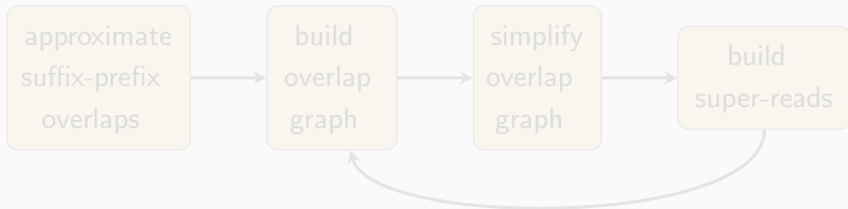


Overlap graphs help to distinguish sequencing errors from true mutations!

Overlap graph construction & evaluation

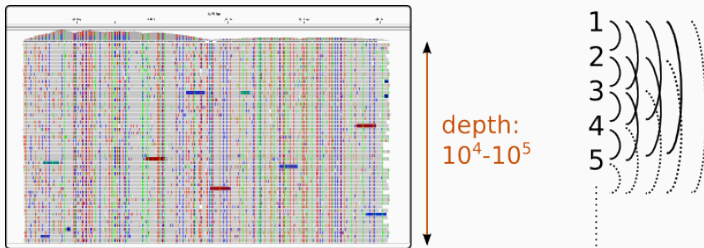


Computational challenge: $\# \text{ overlaps} = O(\text{depth}^2)$

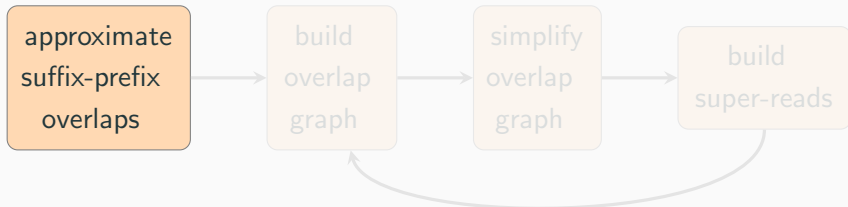


SAVAGE: *Strain Aware VirAI Genome assEmbly*

Overlap graph construction & evaluation

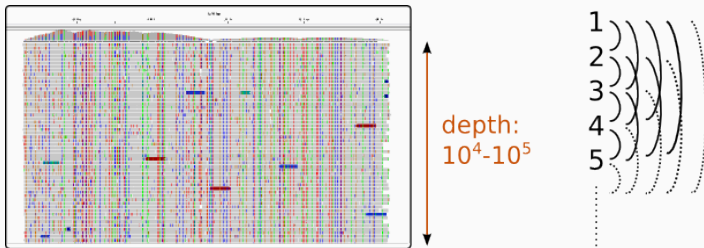


Computational challenge: $\# \text{ overlaps} = O(\text{depth}^2)$

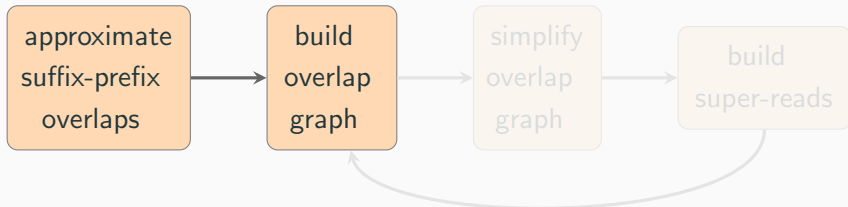


SAVAGE: *Strain Aware VirAI Genome assEmbly*

Overlap graph construction & evaluation

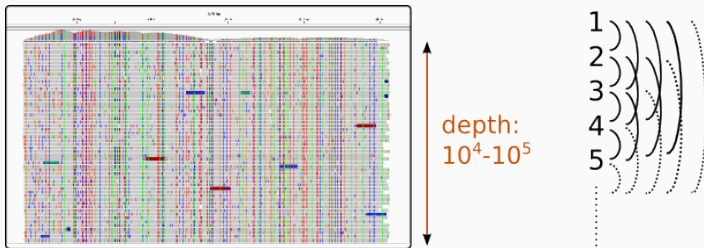


Computational challenge: $\# \text{ overlaps} = O(\text{depth}^2)$

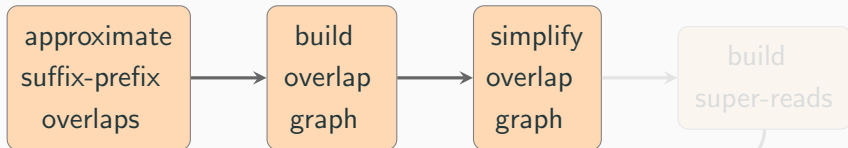


SAVAGE: *Strain Aware VirAI Genome assEmbly*

Overlap graph construction & evaluation

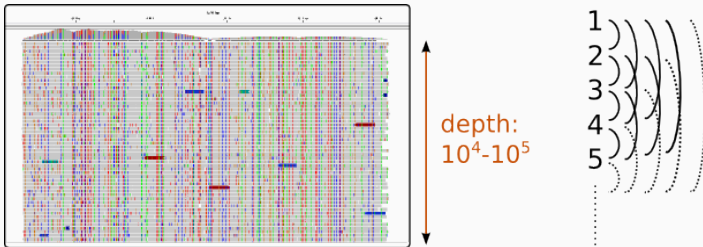


Computational challenge: $\# \text{ overlaps} = O(\text{depth}^2)$

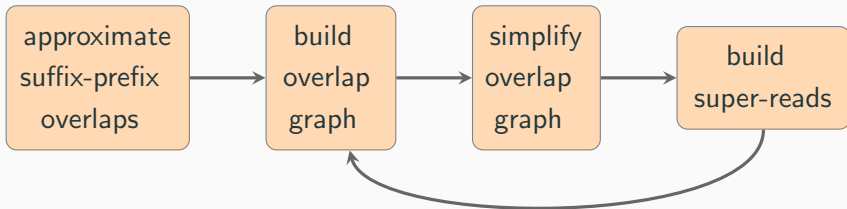


SAVAGE: *Strain Aware VirAI Genome assEmbly*

Overlap graph construction & evaluation

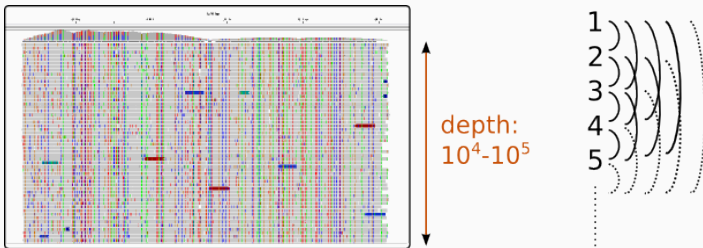


Computational challenge: $\# \text{ overlaps} = O(\text{depth}^2)$

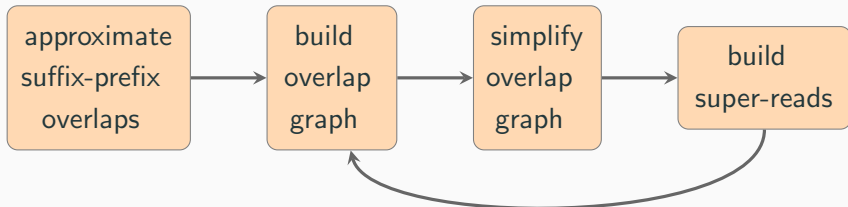


SAVAGE: Strain Aware VirAI Genome assEmbly

Overlap graph construction & evaluation



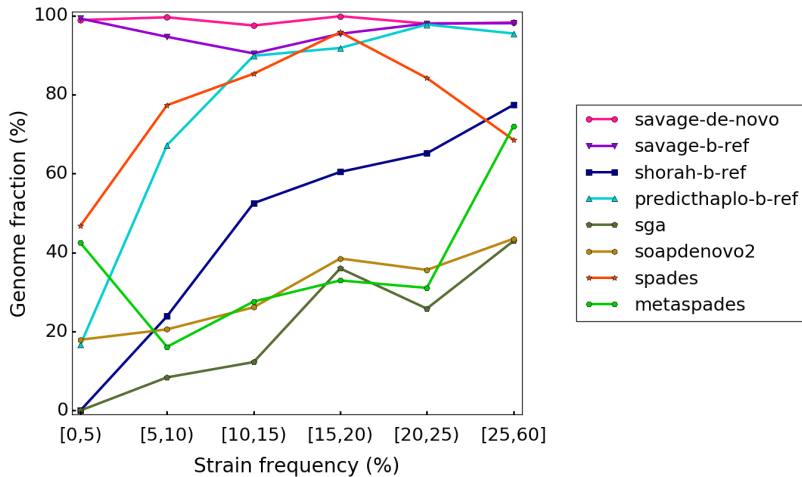
Computational challenge: $\# \text{ overlaps} = O(\text{depth}^2)$



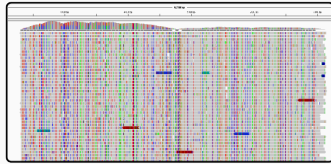
SAVAGE: *Strain Aware VirAI Genome assEmbly*

Benchmarking experiments

Overall genome fraction recovered per strain



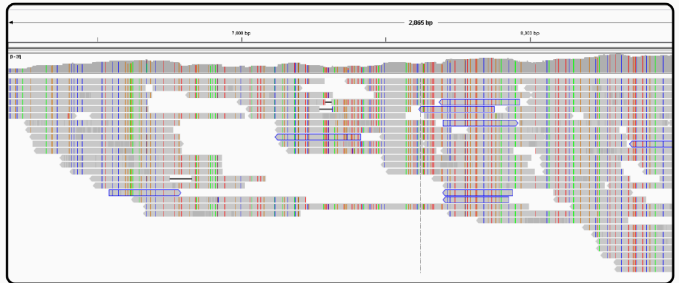
Hepatitis C virus



80.000x

SAVAGE

>16
strains



Future work

- Explore possibilities on **other species**, e.g. human genomes
- **Improve efficiency** of overlap graph construction
- Extend our algorithm to work for **other sequencing technologies**

J. Baaijens, A.Z. El Aabidine, E. Rivals, A. Schönhuth
De novo viral quasispecies assembly using overlap graphs
Genome Research, accepted for publication

Thanks for listening!

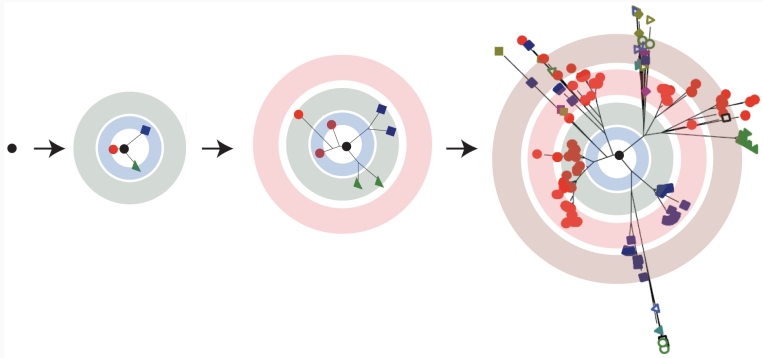
Future work

- Explore possibilities on **other species**, e.g. human genomes
- **Improve efficiency** of overlap graph construction
- Extend our algorithm to work for **other sequencing technologies**

J. Baaijens, A.Z. El Aabidine, E. Rivals, A. Schönhuth
De novo viral quasispecies assembly using overlap graphs
Genome Research, accepted for publication

Thanks for listening!

Viral quasispecies evolution



[Lauring and Andino, 2010]