

Application of genome graph in AMR profiling

AMROMIC^{1,*}

¹Vietnam-Australia-USA

* To whom correspondence should be addressed. E-mails: aba@xyz.com

This manuscript was compiled on **November 20, 2019 22:24**

5

Abstract: Genome graph has been proposed and applied in genomics as the solution for a better reference system to increase mapping quality and variant calling performance. Even at early stage, this direction shows promising results in human genomics and metagenomics. In this study, we are focus on the application of this technique to build a pangenome suitable for resistomes. The results are expected to presented more comprehensive insights into structural variants (SV) that are correlated to the resistance behaviour of the fast-growing superbugs.

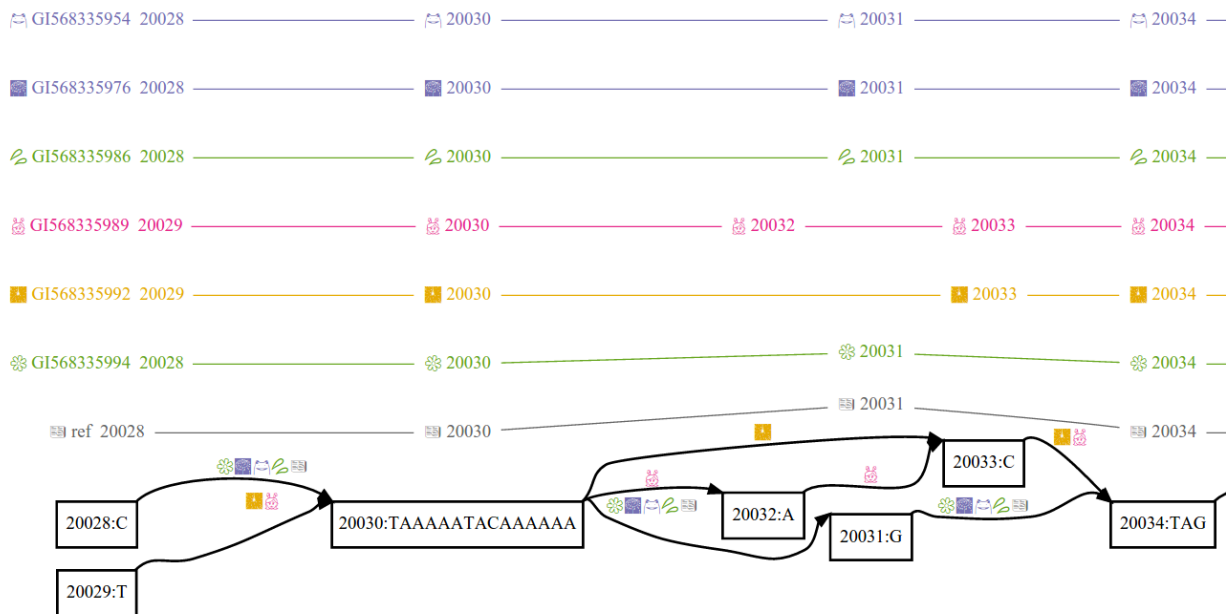
Keywords: microbial genomics, pangenome, genome graph, antibiotic resistance

Background

The idea of pangenome graph had been proposed long ago as a proof-of-concept for a comprehensive but compact reference model. After having a genome graph, we can align a query sequence directly to the graph for variant calling. There were improvement in terms of mapping rate **vg** tool sets (<https://github.com/vgteam/vg>) had been developed for this purpose and currently the most active project working on this field.

10

Figure 1: Example of a genome graph, steal from **vg**.



To date, there are two practical implementations for pangenome graph construction to considered:

- <https://github.com/lh3/minigraph>: *reference model* that focus on a stable, linear coordiante system of the reference genome thus restraint the graph structure. The graph can be build progressively.
- <https://github.com/ekg/seqwish>: more flexible graph using a *path model* that can incorporate multiple references, created by all-to-all pairwise alignment.

15

Resources to considered:

- Tools on github: **vg**, **minigraph**, **seqwish**, **cactus**
- Sites: <https://www.sevenbridges.com/graph/>, <https://ekg.github.io/>, <https://lh3.github.io/>
- Debate: <https://github.com/lh3/gfatools/issues/1>

20

Technical focus

Aims

1. Construct genome graph for AMR strains, e.g. *K. pneumoniae*
2. Identify structural variants of the pangenome by studying the graph topology (sequence segments and their order)
- 5 3. ...

Approaches

One way to construct the genome graph is to integrate a variant catalog (VCF file) into a reference. This is a common practice for human genome when we have a stable reference and a set of known variants output in a VCF file ...

10 Another option is to build the graph on the assembly sequences from the beginning. One problem addressed by researcher when constructing a genome graph progressively is the unavoidable *reference bias*. **seqwish** try to avoid this by aligning all pairs of sequences together, sacrificing the ability to integrate genomes into the graph one-by-one from the very beginning.

Results