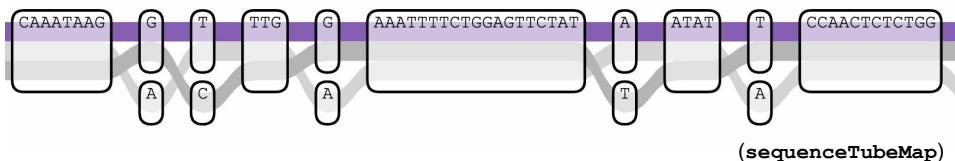
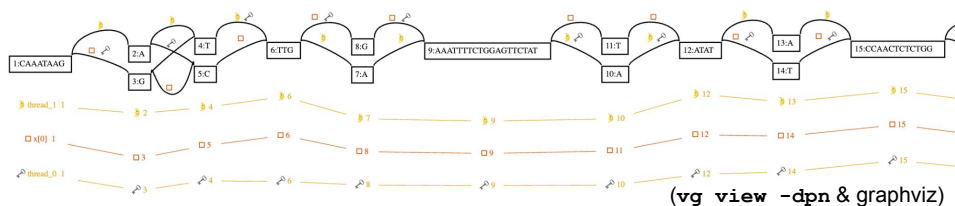


Variation Graph Visualization

William Gao

Background

- a *pangenome* represents genetic variation as a graph (with nodes and edges)
 - a linear sequence of the pangenome is represented as a *path*
 - a *variation graph* has a collection of nodes, edges, and paths



- a variation graph is nonplanar -- it requires at least 3 dimensions to draw without overlapping^[1]
- **goal:** prototype a new way to visualize paths in a variation graph

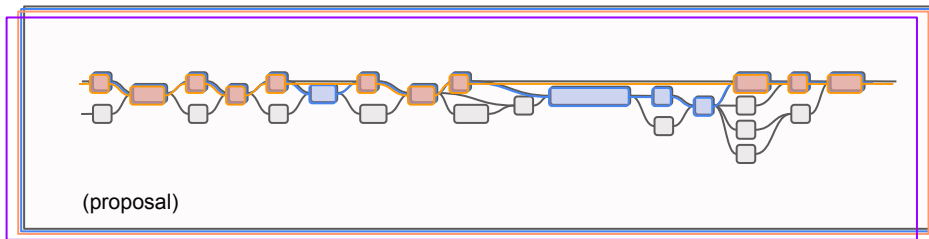
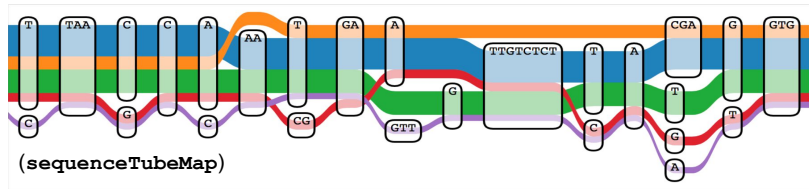
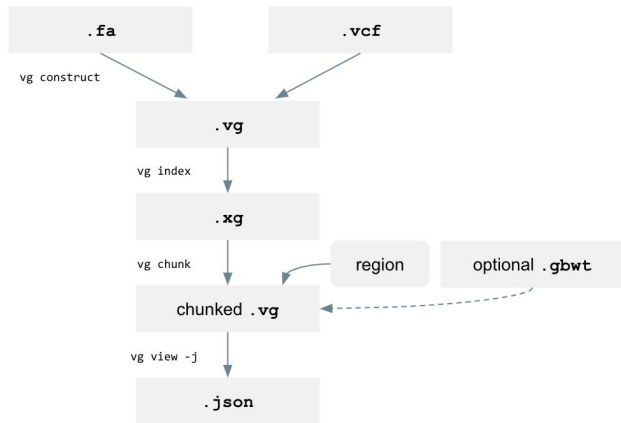
Methods

Data pre-processing

- take in **FASTA** and **VCF** files (reference + variants), a region, and a **gbwt** file (haplotypes)
- construct and encode graph as a **JSON** file for the UI

Layout

- put graph (nodes and edges) on a plane
- put haplotypes (paths) on separate “layers” that can be interactively selected



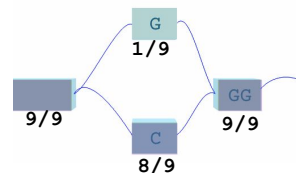
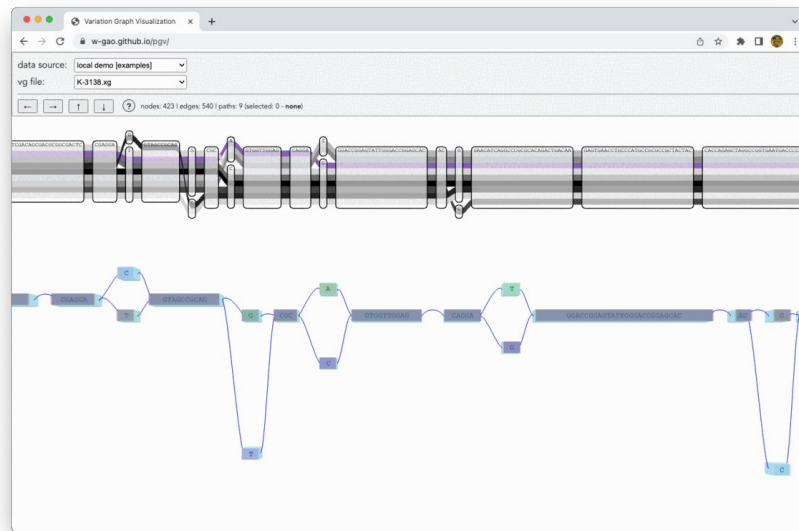
Results

Proof of concept

- <https://w-gao.github.io/pgv/>
- a web-based tool built with Three.js
- graphs are pre-processed and statically served

Features

- node coverage
 - paths are layered: the more paths that pass through a node, the darker the node is
 - hovering over a node displays the number of paths that visit the node
- path selection
 - a particular path of interest can be highlighted



Discussion

Future improvements

- support different zoom levels / resolution
 - currently doesn't scale well to larger graphs
- dynamic rendering
 - requires backend support
 - download / render graph based on viewport

Takeaway

- the goal is to display paths in a variation graph in an interactive layout
- there's no universal way to visualize graphs
 - this is intended to provide a platform to explore variation graphs in an accessible way

