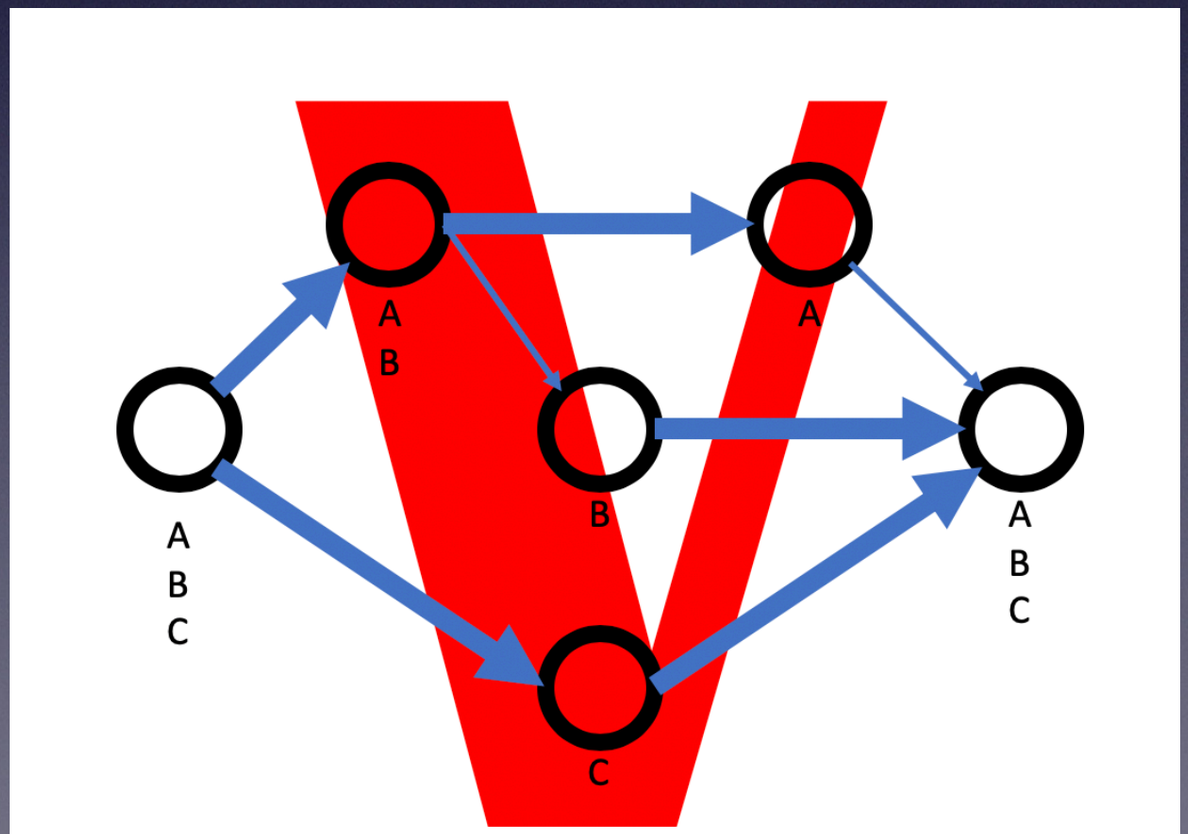
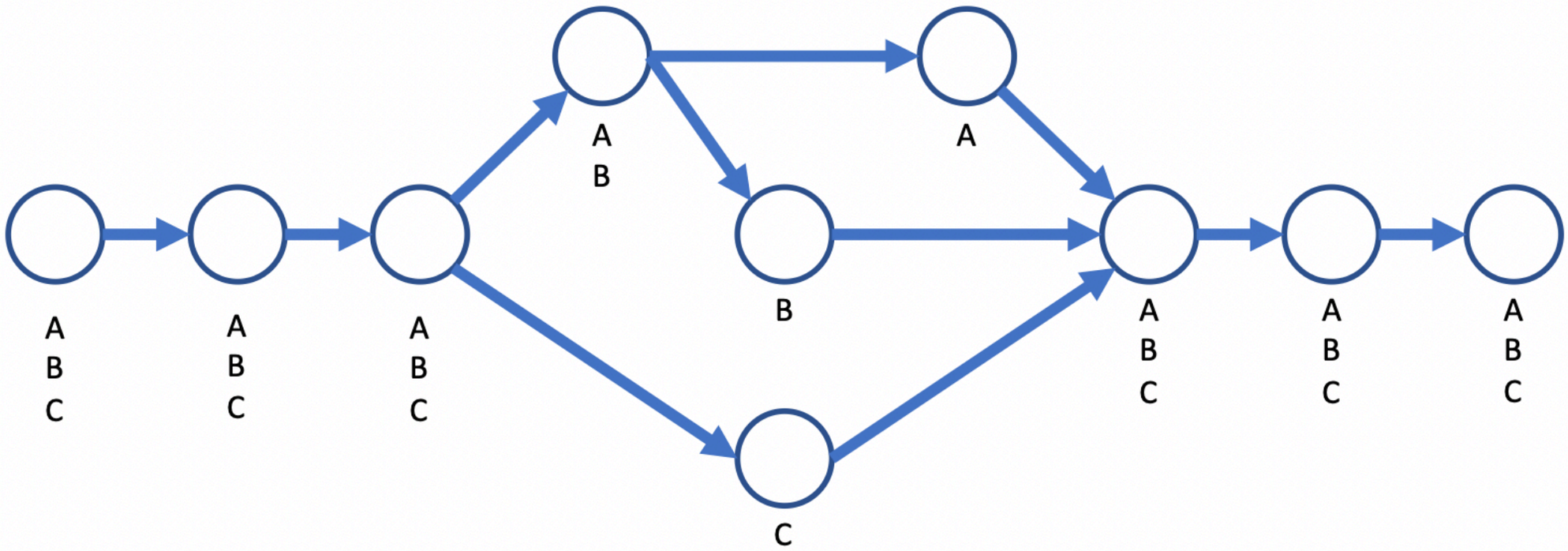
VirusGraph Team

Evan Biederstedt, Jan Buchmann, Alex Gener, Joan Martí-Carreras,
Harihara Muralidharan, Alexis Norris, Surya Saha, Mike Tisza, & Valerie Virta



An example of HIV inference obtained using Next Generation Sequencing data



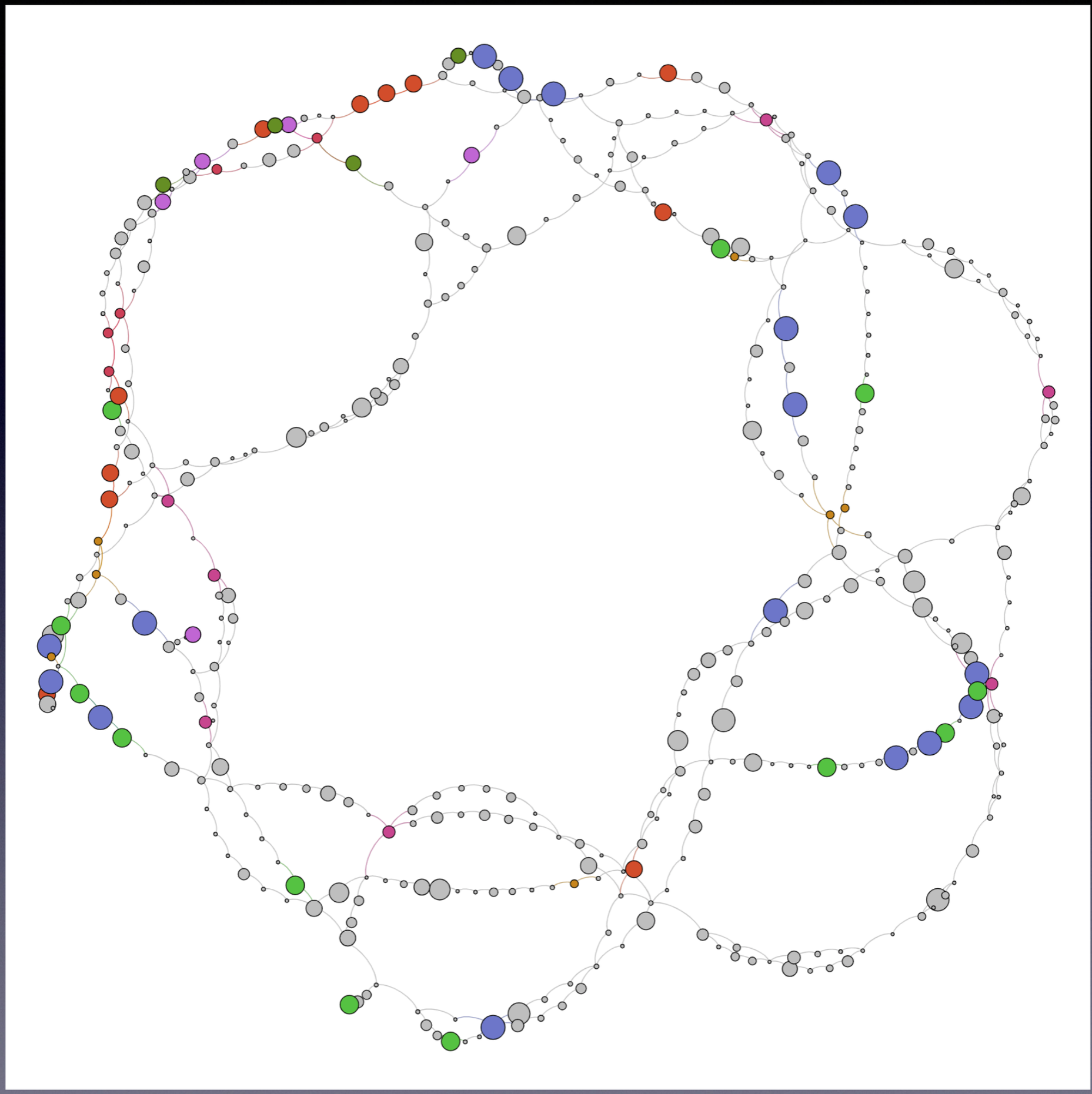
Materials & Methods

- Sequence files:
  - T4 virus
  - HIV with references
- Genome graphing software programs:
  - vg —
  - NovoGraph
  - SWIGG

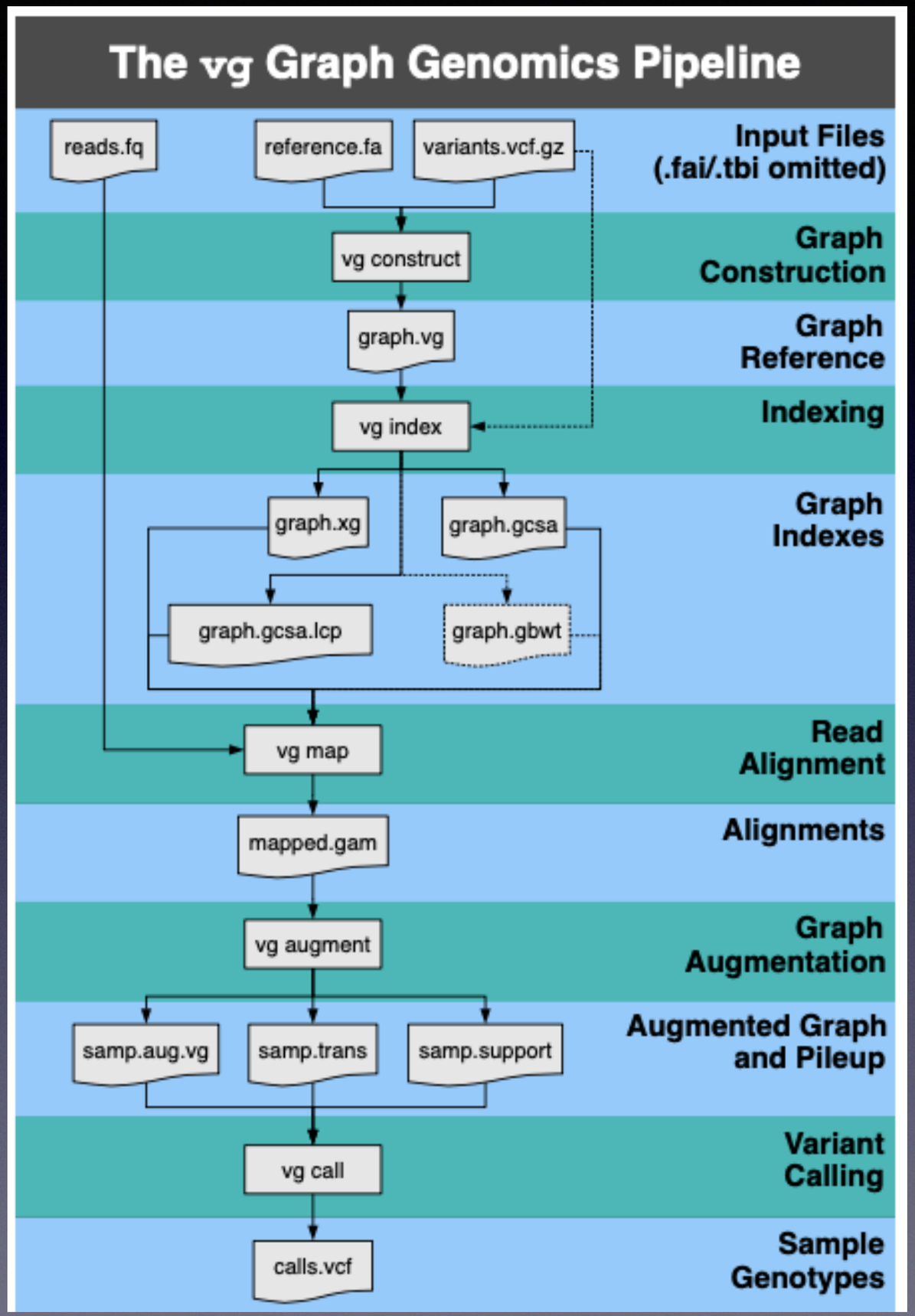
171 HIV genomes run on SWIGG with annotated kmer/nodes

Same color = same taxonomical distribution of kmer (node)

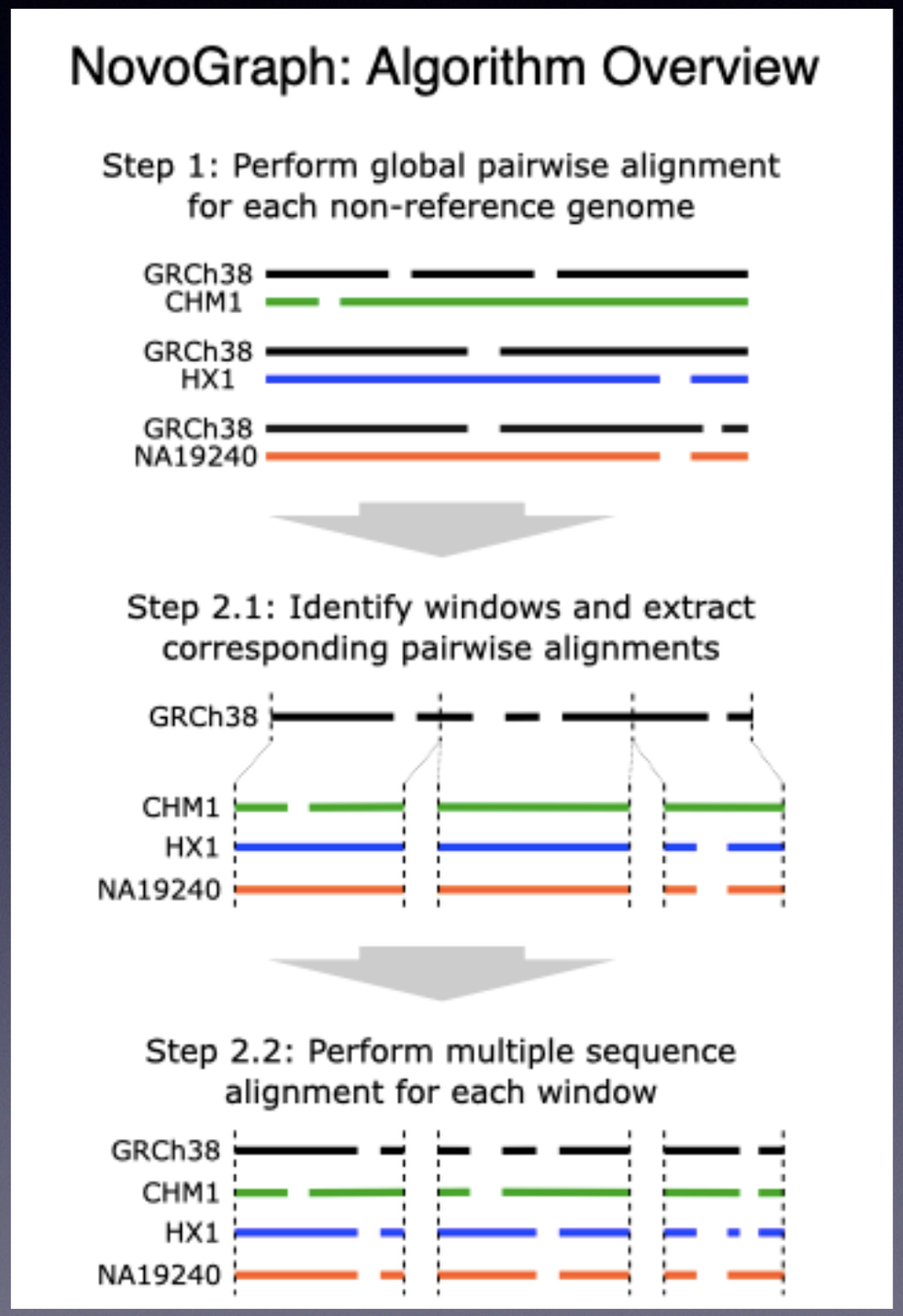
Size of node is proportional to occurrence of taxonomical category

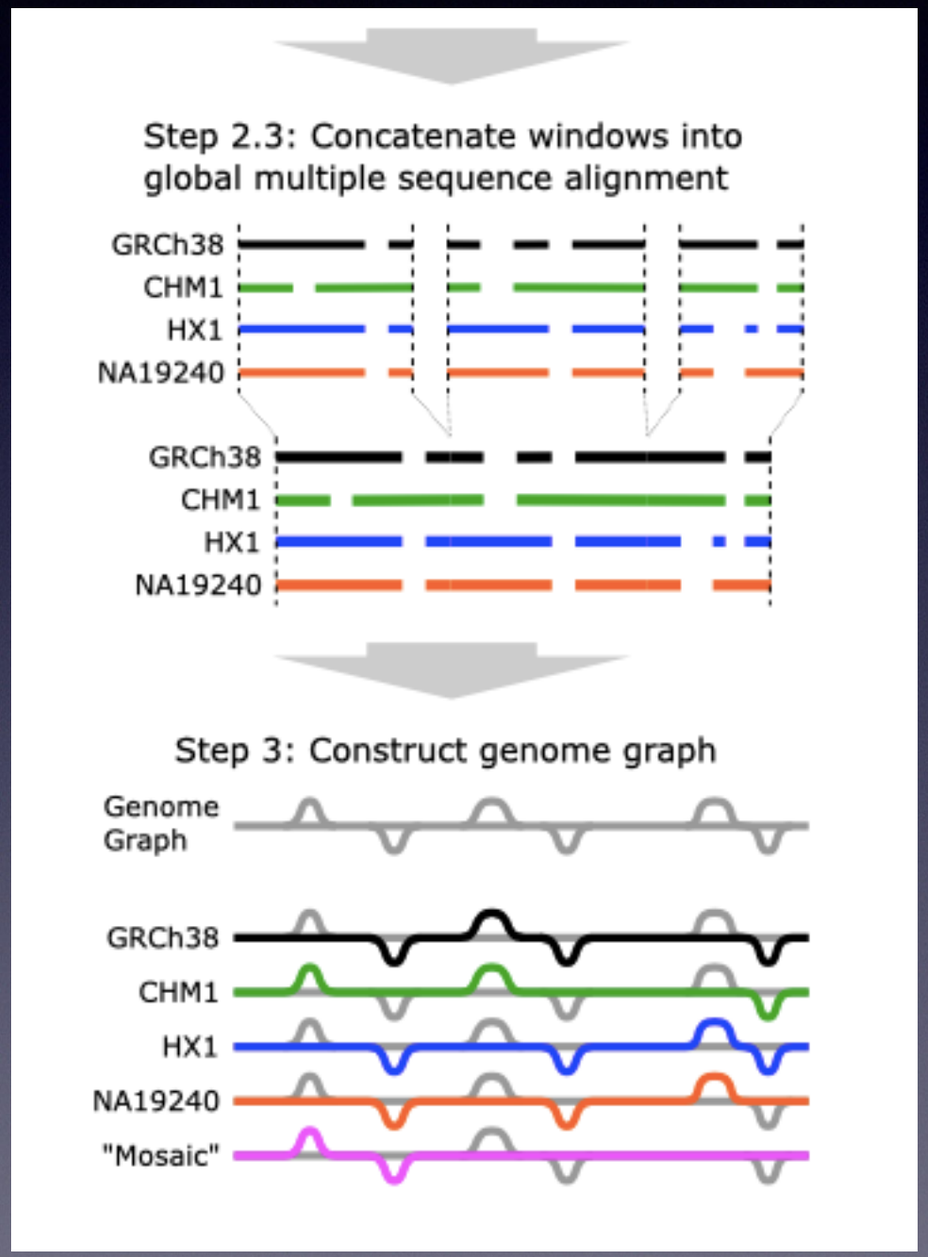


How vg works

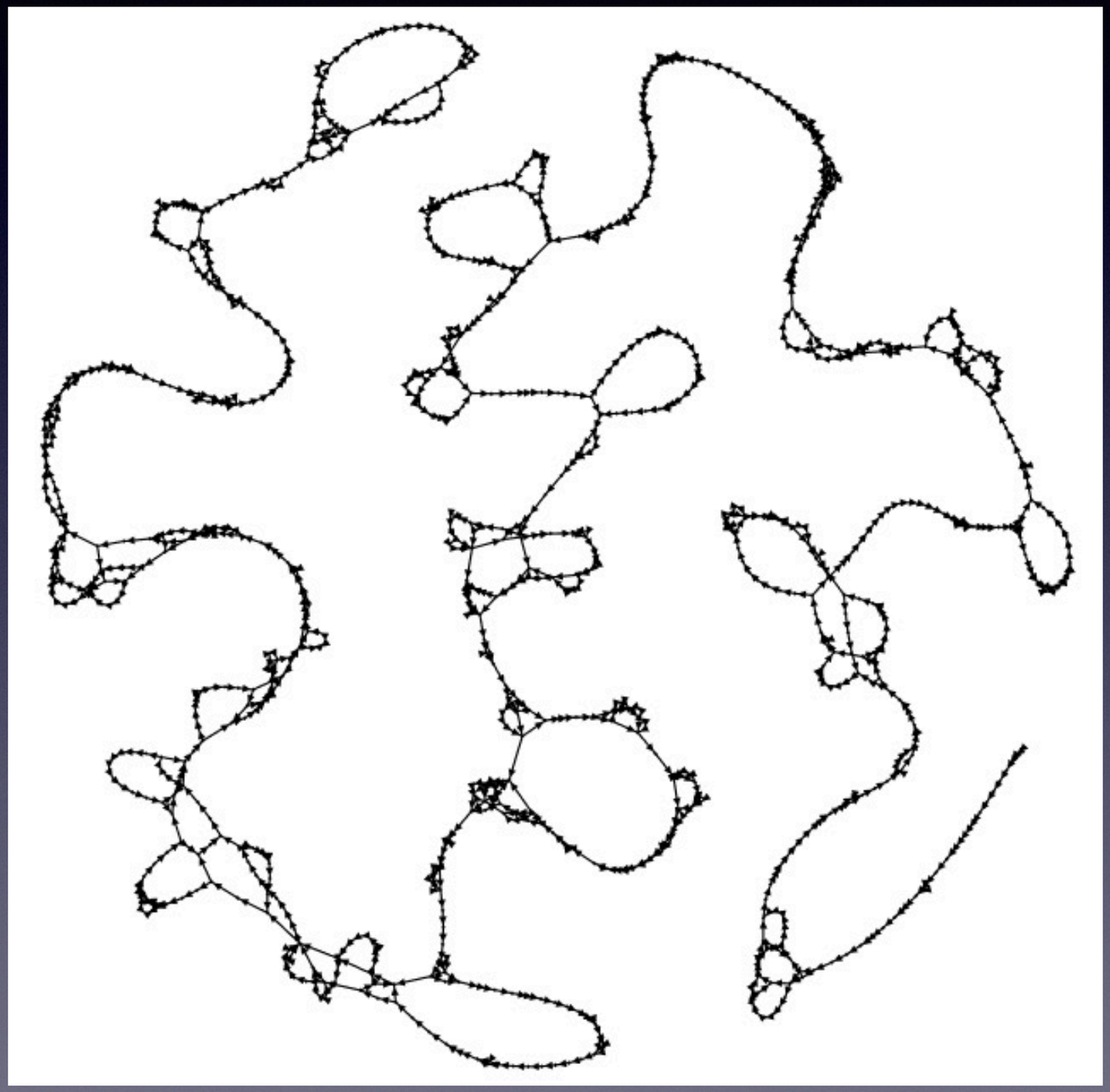

The vg Graph Genomics Pipeline
reads.fq
reference.fa
variants.vcf.gz
Input Files
(.fai/.tbi omitted)
vg construct
Graph
Construction
graph.vg
Graph
Reference
vg index
Indexing
graph.xg
graph.gcsa
Graph
Indexes
graph.gcsa.lcp
graph.gbwt
vg map
Read
Alignment
mapped.gam
Alignments
vg augment
Graph
Augmentation
samp.aug.vg
samp.trans
samp.support
Augmented Graph
and Pileup
vg call
Variant
Calling
calls.vcf
Sample
Genotypes

NovoGraph — novel approach for graphing long-read sequence assemblies

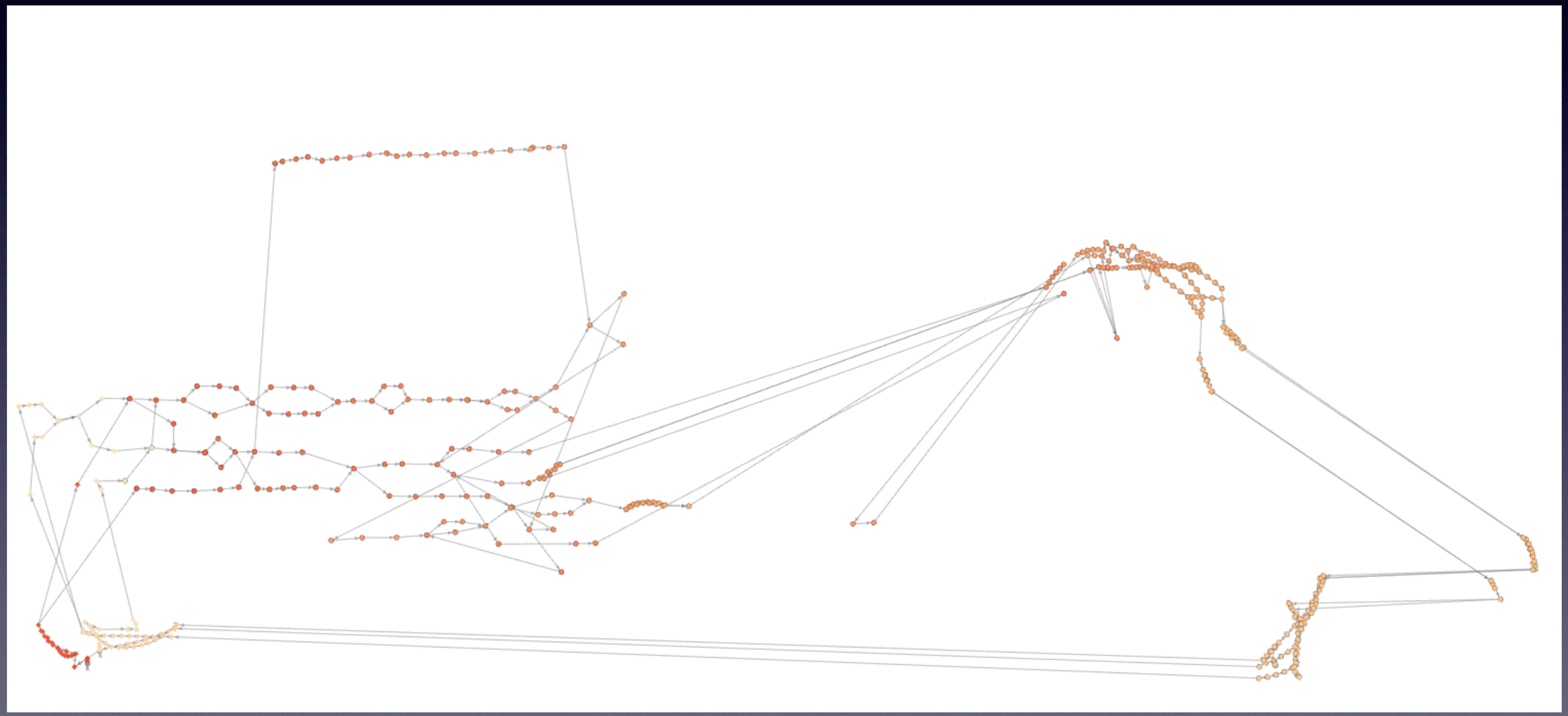




SWIGG — Automated pipeline builds graphs quickly using k-mer approach



SWIGG output into Gephi of HIV sequence



HIV Clades A-J, K=41, threshold >=2