Molecular epidemiology is increasingly important for limiting the spread of infections, defining risk factors, and understanding the development of disease to improve treatment and palliative care. Next generation sequencing technologies are rapidly reshaping the fields of medical diagnostics, molecular epidemiology, and sequence analysis. Galaxy provides a platform explicitly upholding the core tenets of transparency, accessibility, and reproducibility of analytical workflows.

**Map and Error Correct**

HIV samples are sequenced using sequencing technology. The sequenced reads are error corrected with qfilt, a tool that uses simple heuristics to determine high-quality reads. The error-corrected reads are then aligned with bealign, part of the BioExt python package. This tool aligns the reads using a codon-aware alignment algorithm against one of several built-in HIV reference genomes, HXB2 protease and reverse transcriptase in

**Compute Distances**

Distances between pairs are computed using the Tamura-Nei\(^1\) model implemented in the t9n3 package, enabling subsequent inference of transmission between individuals.

**Infer Transmission**

Computed distances are then used to cluster reads by inferred transmission using hivnetworkcsv. Unexpanded clusters are shown as circles, with the size indicating number of members in each group. In the expanded views, arrows indicate the transmission direction, if the software was given enough data to infer that.

**Further Analyses**

The HIV-TRACE pipeline supports a number of data visualizations, such as the ancestral sequence structural viewer, which combines a phylogenetic tree view with a protein rendering, showing the ancestral sequence changes related to specific portions of a protein.

**Availability**

All of the software mentioned here is or will soon be available on usegalaxy.org, the main Galaxy ToolShed, and on github.

**Additional Options**

There are far more ways to analyze and visualize the data generated in this pipeline than one poster can cover, what is shown here is intended only as a representative sampling.