Chapter 1

Genotype Filters

Some of the detected trajectories or genotypes do not represent real mutations, and need to be removed from the analysis.

Common situations where a trajectory or genotype should be removed:

1. A mutation appears before and after a genotype sweeps and removes all pre-existing variation. Since the mutation should have been removed when the genotype fixed

2. A mutation fixes immediately during the same timepoint it is detected (there are no intermediate values), then becomes undetected again. This is likely due to a sampling error.

3. Similar to above, a mutation is detected at a single timepoint. These mutation do not provide much information and complicate the analysis, so they are filtered out. `--disable-filter-single` disables both this and the above filter.

4. The first timepoint shows that a mutation has already fixed. This contradicts the known biology of the experiment. `--disable-filter-startsfixed` disables this filter.

5. A mutation is relatively constant throughout the experiment and is not affected by the movement of other mutations. This filter can be disabled by setting `--filter-constant` to 0.

An exception has been made for the case where a mutation appears both before and after a genotype sweeps, but is undetected at the timepoint where this occurs. This represents a mutation arising, being removed during a genotype sweep, then arising again. Note that when a genotype is removed, all of the trajectories that comprise that genotype are removed and the clustering algorithm is re-run on the remaining dataset.