

Genome Directory

Define the number of genomes to pick (default = 10) and number of reads to generate per file (default = 100,000)

Execute **simulate** mode

PD on
concatenated tree

Determines PD contributions for taxa present in concatenated guide tree in PhyloSift marker directory

Select Taxa

Two separate approaches used:

1. Select some number of taxa that contribute to PD (user input, default = 10 taxa)
2. Sample taxa uniformly without replacement

Compute metrics
between target and
remaining taxa

Calculated metrics include: the distance to nearest neighbors, connecting branch lengths, and the number of sampled nodes within various PD units of connecting nodes.

Knockout
Swaths of Taxa

Workflow plugs into updateDB to remove genomes which have been used to simulate metagenome data, as well as a swath of related taxa.

Generated
Simulated Reads

Grinder algorithm randomly generates reads from selected genomes, outputs simulated PE-Illumina and 454 datasets

Simulation
Marker Directory

A new marker directory is created, where simulated genomes have been knocked out from marker packages.

