

- uniform
- power law
- should randomise the order of taxons for any non-uniform distribution

Community Profile

Reference Sequence

Simulate

Hi-C reads

WGS reads

A5 - produces a folder which will may wish to retain.

Assembly

Contigs

Map

Map

Map

Hi-C placements

Read placements

Contig placements

Parser

Graph

Clustering

Taxon Assignment

Scoring

f-score or a closely related variant.

Hi-C reads
read-length
inter-prob
num-frag
geom-prob

WGS reads
read-length
insert-mean
insert-sd
coverage
error-model

Community Profile
distribution
ref-seq

This could be made independent of ref-seq (taxon number) if functionally defined.

Contigs
assembler (A5)

Placements
mapper (BWA)
strategy dealing with degeneracy

Graph
edge weight calculation
pruning or other internal cleanup

Taxon Assignment
algorithm (MCL, SR-MCL, R-MCL, etc)
soft/hard clusters

Scoring
chosen metric (f-score)