Mapping File
(Accession, NCBI taxon ID)

PD cutoff

Alignment File
(Marker sequences in FASTA format)

NOTE: New marker packages are named according to input filenames (e.g. MarkerAlignment.fasta). Core marker data will be overwritten during new marker builds if input files do not have unique names compared to existing PhyloSift markers.

Execute build_marker mode

HMMbuild

Unalign sequences and create profile HMM out of input data

Generate unique IDs for input sequences

FastTree

Build tree and collapse topology according to a user-specified PD cutoff (e.g. 99%)

HB: Creepy monkey (top) or non creepy monkey (below)?? Or both (since they fill out the white space nicely)?

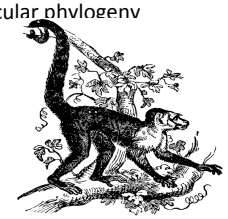
Sequence accessions mapped to phylogeny via a dummy pplacer file

HB: Want to mention the dummy pplacer step here but don't want to go into too much detail (might confuse readers). Is this appropriate?

Tree Reconciliation

Reconcile NCBI taxonomy IDs with phylogenetic topology

Quantitative metric (minimum hamming distance) used to match edges between NCBI taxon tree and molecular phylogeny



Clean and package new marker genes

Built Marker Packages

New marker gene packages placed into shared PhyloSift marker directory

Execute index mode

Index Marker Database

Indexes the marker databases needed for LAST and Bowtie

Locally indexed marker packages will not interfere with automatic updates to PhyloSift core markers

Built PhyloSift Marker package

Tree

Hmm profile

Taxon map

Representative sequences

Alignment