Build CLARK Database

Build a CLARK database from a set of reference sequences ("targets"). NCBI taxonomy data are used to map the accession number found in each reference sequence to its taxonomy ID.

Parameters in GUI

Parameter	Description	Default value
Database	A folder that should be used to store the database files.	
Genomic library	Genomes that should be used to build the database ("targets"). The genomes should be specified in FASTA format. There should be one FASTA file per reference sequence. A sequence header must contain an accession number (i.e., >accession.number ... or >gi|number|ref|accession.number| ...).	
Taxonomy rank	Set the taxonomy rank for the database. CLARK classifies metagenomic samples by using only one taxonomy rank. So as a general rule, consider first the genus or species rank, then if a high proportion of reads cannot be classified, reset your targets definition at a higher taxonomy rank (e.g., family or phylum).	Species

Parameters in Workflow File
Type: clark-build

Parameter	Parameter in the GUI	Type
database	**Database**	*string*
taxonomy	**Genomic library**	*url-datasets*
taxonomy-rank	**Taxonomy rank**	*number*

Input/Output Ports
The element has 1 *output port*:

Name in GUI: Output CLARK database

Name in Workflow File: out

Slots:

SlotInGUI	Slot in Workflow File	Type
Output URL	**url**	*string*