

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.

Element type: clark-build

Parameters

Parameter	Description	Default value	Parameter in Workflow File	Type
Database	A folder that should be used to store the database files.		database	<i>string</i>
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	<i>url-datasets</i>
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	taxonomy-rank	<i>number</i>

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	<i>string</i>