Build Kraken Database

Build a Kraken database from a genomic library or shrink a Kraken database.

Parameters in GUI

Parameter	Description	Default value
Mode	Select "Build" to create a new database from a genomic library (--build). Select "Shrink" to shrink an existing database to have only specified number of k-mers (--shrink).	Build
Database	Name of the output Kraken database (corresponds to --db that is used with --build, and to --new-db that is used with --shrink).	
Genomic library	Genomes that should be used to build the database. The genomes should be specified in FASTA format. The sequence IDs must contain either a GI number or a taxonomy ID.	
K-mer length	K-mer length in bp (--kmer-len).	31
Minimizer length	Minimizer length in bp (--minimizer-len). The minimizers serve to keep k-mers that are adjacent in query sequences close to each other in the database, which allows Kraken to exploit the CPU cache. Changing the value of the parameter can significantly affect the speed of Kraken, and neither increasing nor decreasing of the value will guarantee faster or slower speed.	15
Maximum database size	By default, a full database build is done. To shrink the database before the full build, input the size of the database in Mb (this corresponds to the --max-db-size parameter, but Mb is used instead of Gb). The size is specified together for the database and the index.	No limit
Clean	Remove unneeded files from a built database to reduce the disk usage (--clean).	True
Work on disk	Performs most operations on disk rather than in RAM (this will slow down build in most cases).	False
Jellyfiah hash size	The "kraken-build" tool uses the "jellyfish" tool. This parameter specifies the hash size for Jellyfish. Supply a smaller hash size to Jellyfish, if you encounter problems with allocating enough memory during the build process (--jellyfish-hash-size). By default, the parameter is not used.	Skip
Number of threads	Use multiple threads (--threads).	8

Parameters in Workflow File

Type: kraken-build

Parameter	Parameter in the GUI	Type
mode	**Mode**	*string*
database	**Database**	*string*
genomic-library	**Genomic library**	*url-datasets*
k-mer-length	**K-mer length**	*number*
minimizer-length	**Minimizer length**	*number*
maximum-database-size	**Maximum database size**	*number*
clean	**Clean**	*bool*
work-on-disk	**Work on disk**	*bool*
jellyfish-hash-size	**Jellyfiah hash size**	*number*
threads	**Number of threads**	*number*

Input/Output Ports

The element has 1 *output port*:

Name in GUI: Output Kraken database

Name in Workflow File: out

Slots:

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