Filter by Classification

The filter takes files with NGS reads or contigs, classified by one of the tools: Kraken, CLARK, DIAMOND, WEVOTE.

For each input file, it outputs a file with unspecific sequences (i.e. sequences not classified by the tools, taxID = 0) and/or one or several files with sequences that belong to the specific taxonomic group(s).

Parameters in GUI

Parameter	Description	Defaultvalue
Input data	To filter single-end (SE) reads or contigs, received by reads de novo assembly, set this parameter to "SE reads or contigs". Use the "Input URL 1" slot of the input port. To filter paired-end (PE) reads, set the value to "PE reads". Use the ""Input URL 1" and "Input URL 2" slots of the input port to input the NGS reads data. Also, input the classification data, received from Kraken, CLARK, or DIAMOND, to the "Taxonomy classification data" input slot. Either one or two slots of the output port are used depending on the input data.	SE reads or contigs
Save unspecific sequences	Select "True" to put all unspecific input sequences (i. e. sequences with tax ID = 0) into a separate file. Select "False" to skip unspecific sequences. At least one specific taxon should be selected in the "Save sequences with taxID" parameter in this case.	True
Save sequences with taxID	Select a taxID to put all sequences that belong to this taxonomic group (i. e. the specified taxID and all children in the taxonomy tree) into a separate file.	

Parameters in Workflow File

Type: classification-filter

Parameter	Parameter in the GUI	Type
sequencing-reads	**Input data**	*string*
save-unspecific-sequences	**Save unspecific sequences**	*bool*
tax-ids	**Save sequences with taxID**	*string*

Input/Output Ports

The element has 1 *input port*:

Name in GUI: Input sequences and tax IDs:

The following input should be provided:

- URL(s) to FASTQ or FASTA file(s).
- Corresponding taxonomy classification of sequences in the files.

To process single-end reads or contigs, pass the URL(s) to the "Input URL 1" slot.

To process paired-end reads, pass the URL(s) to files with the "left" and "right" reads to the "Input URL 1" and "Input URL 2" slots correspondingly.

The taxonomy classification data are received by one of the classification tools (Kraken, CLARK, or DIAMOND) and should correspond to the input files.

Name in Workflow File: in

Slots:

SlotInGUI	Slot in Workflow File	Type
Input URL	**url**	*string*
Taxonomy data	**tax-data**	*tax-classification*

The element has 1 *output port*:

Name in GUI: Output file(s):

The port outputs URLs to files with NGS reads, classified by taxon IDs: one file per each specified taxon ID per each input file (or the pair of files in case of PE reads).

Either one (for SE reads or contigs) or two (for PE reads) output slots are used depending on the input data. See also the "Input data" parameter of the element.

Name in Workflow File: out

Slots:

SlotInGUI	Slot in Workflow File	Type
Output URL 1	**url**	*string*
Output URL 2	**url**	*string*