Classify Sequences with MetaPhlAn2 Element

MetaPhlAn2 (METAgenomic PHyLogenetic ANalysis) is a tool for profiling the composition of microbial communities (bacteria, archaea, eukaryotes, and viruses) from whole-metagenome shotgun sequencing data.

Element type: metaphlan2-classify

Parameters

Parameter	Description	Defaultvalue	Parameter in Workflow File	Type
Input data	To classify single-end (SE) reads or contigs, received by reads de novo assembly, set this parameter to "SE reads or contigs". To classify paired-end (PE) reads, set the value to "PE reads".	SE reads or contigs	input-data	*string*
Input file format	Set type of an input file (--input-type). Each input file will usually contain a lot of sequences that should be classified.	FASTA	input-format	*string*
Database	A path to a folder with MetaPhlAn2 database: BowTie2 index files, built from reference genomes, and *.pkl file (--mpa-pkl, --bowtie2db). By default, "mpa_v20_m200" database is provided (if it has been downloaded). The database was built on ~1M unique clade-specific marker genes identified from ~17,000 reference genomes (~13,500 bacterial and archaeal, ~3,500 viral, and ~110 eukaryotic).		database	*string*
Number of threads	The number of CPUs to use for parallelizing the mapping (--nproc).	8	threads	*number*
Analysis type	Specify the type of analysis to perform: • Relative abundance - profiling of metagenomes in terms of relative abundances (corresponds to "-t rel_ab") • Relative abundance with reads statistics - profiling of metagenomes in terms of relative abundances and estimate the number of reads coming from each clade ("-t rel_ab_w_read_stats") • Reads mapping - mapping from reads to clades, the output contains reads that hit a marker only ("-t reads_map") • Clade profiles - normalized marker counts for clades with at least a non - null marker("-t clade_profiles") • Marker abundance table - normalized marker counts: only when > 0.0 and optionally normalized by metagenome size ("-t marker_ab_table"), see also "Normalize by metagenome size" parameter • Marker presence table - list of markers present in the sample ("-t marker_pres_table"), see also "Presence threshold" parameter	Relative abundance	analysis-type	*string*
Tax level	The taxonomic level for the relative abundance output: all, kingdoms (Bacteria and Archaea) only, phyla only, etc. (--tax_lev).	All	tax-level	*string*
Bowtie2 output file	The file for saving the output of BowTie2 (--bowtie2out). In the case of PE reads one file is created per each pair of files.	Auto	bowtie2-output-url	*string*
Output file	MetaPhlAn2 output depends on the "Analysis type" parameter. By default, it is a tab-delimited file with the predicted taxon relative abundances.	Auto	output-url	*string*
Normalize by metagenome size	The parameter is present only when "Analysis type" is equal to "Marker abundance table". It is a combo box with values "Skip" (default) and "Normalize". If "Normalize" is selected, the total number of reads in the original metagenome is taken into account for normalization: UGENE calculates the number of reads in an input FASTA/FASTQ file and passes "--nreads" parameter to MetaPhlAn2.		normalize-by-size	*boolean*
Presence threshold	The parameter is present only when "Analysis type" is equal to the "Marker presence table". It is an INT value >= 0. The default value is 1. Specify a threshold for calling a marker.		presence-threshold	*number*

Input/Output Ports

The element has 1 *input port*:

Name in GUI: Input sequences:

URL(s) to FASTQ or FASTA file(s) should be provided. In the case of SE reads or contigs use the "Input URL 1" slot only. In case of PE reads input "left" reads to "Input URL 1", "right" reads to "Input URL 2".See also the "Input data" parameter of the element

Name in Workflow File: in

Slots:

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