

usage: perl mkLTG.pl [-options] -in INPUT_FILE -taxonomy TAXONOMY -blast_db BLASTDB
-outdir OUTDIR -ltg_params PARAMETER_FILE

arguments:

-in name of the input fasta or TSV file containing the sequences to be assigned; the TSV file must have "sequence" as a heading for one of the columns
-taxonomy tsv file with the following tab separated columns:
taxid parent_taxid taxlevel taxname merged_taxid taxlevel_index
-ncbitax_dir directory of ncbi taxonomy dmp files (https://ftp.ncbi.nih.gov/pub/taxonomy/new_taxdump/)
only necessary if no -taxonomy file is provided
taxonomy file can be created from these files if all sequences in the

blast_db have ncbi taxids

-blast_db name of the blast database
-ltg_params tsv file with the following tab separated columns:
pid pcov phit taxn seqn refres ltgres
-outdir name of the output directory
-out_name alpha-numeric string for naming output files
-blastout if empty, run blast; otherwise this file is used to make ltg
-delete_tmp 0/1; if 1 delete temporary files after the run
-windows 0/1, set to one if running on windows

BLAST parameters

-task megablast/blastn
-blast_e maximum e-value
-dust yes/no
-max_target_seqs maximum number of blast hits per query
-num_threads number of threads
-qcov_hsp_perc minimum query coverage in blast
-batch_size batch size for blast