

Bioinformatics Lab
Lab 3
Standalone BLAST

<https://www.ncbi.nlm.nih.gov/books/NBK279671/>
<https://www.ncbi.nlm.nih.gov/guide/howto/run-blast-local/>

Help for set up Linux/Unix blast

<http://www.ncbi.nlm.nih.gov/books/NBK52640/>

Help for running blast on Windows:

<http://www.ncbi.nlm.nih.gov/books/NBK52637/>

Video:

How to run a local blast on PC:

https://www.youtube.com/watch?v=d_WMep3Q9ec

Command line BLAST (Mac):

<https://www.youtube.com/watch?v=sU44ZtZlzo4>

Windows Command Line Tutorial

<https://www.bleepingcomputer.com/tutorials/windows-command-prompt-introduction/>

Windows Command Line Tutorial - 1 - Introduction to the Command Prompt

<https://www.youtube.com/watch?v=MBBWVgE0ewk>

LINUX MS-Dos basic commands:

http://proteomics.ysu.edu/courses/BIOL4800_6900/ExtraReadings/ShellIntro.pdf

SSH-Secure Shell:

http://proteomics.ysu.edu/courses/BIOL4800_6900/LAB/

Example – how to run:

```
Makeblastdb -in input_file -dbtype prot/nucl -out database_name
```

```
Blastn -query text_query.txt -db db_name -out output.txt
```

```
ca. Command Prompt
C:\Users\Public>cd TestDB
C:\Users\Public\TestDB>ls
'ls' is not recognized as an internal or external command,
operable program or batch file.
C:\Users\Public\TestDB>dir
Volume in drive C has no label.
Volume Serial Number is 5444-1ED5

Directory of C:\Users\Public\TestDB
01/19/2018  03:23 PM    <DIR>          .
01/19/2018  03:23 PM    <DIR>          ..
01/19/2018  03:23 PM                166,464 human_prot.txt
                1 File(s)                166,464 bytes
                2 Dir(s)          187,151,167,488 bytes free

C:\Users\Public\TestDB>\Program Files\NCBI\blast-2.2.30+bin\makeblastdb
'\Program' is not recognized as an internal or external command,
operable program or batch file.
C:\Users\Public\TestDB>"\Program Files\NCBI\blast-2.2.30+bin\makeblastdb.exe"
USAGE
    makeblastdb.exe [-hl] [-help] [-in input_file] [-input_type type]
                    [-dbtype molecule_type] [-title database_title] [-parse_seqs]
                    [-hash_index] [-mask_data mask_data_files] [-mask_id mask_algo_ids]
                    [-mask_desc mask_algo_descriptions] [-gi_mask]
                    [-gi_mask_name gi_based_mask_names] [-out database_name]
                    [-max_file_sz number_of_bytes] [-logfile File_Name] [-taxid TaxID]
                    [-taxid_map TaxIDMapFile] [-version]

DESCRIPTION
    Application to create BLAST databases, version 2.2.30+

Use '-help' to print detailed descriptions of command line arguments
=====
Error: Argument "dbtype". Mandatory value is missing: 'String, 'nucl', 'prot''
C:\Users\Public\TestDB>"\Program Files\NCBI\blast-2.2.30+bin\makeblastdb.exe"
-in human_prot.txt -dbtype
USAGE
    makeblastdb.exe [-hl] [-help] [-in input_file] [-input_type type]
                    [-dbtype molecule_type] [-title database_title] [-parse_seqs]
                    [-hash_index] [-mask_data mask_data_files] [-mask_id mask_algo_ids]
                    [-mask_desc mask_algo_descriptions] [-gi_mask]
                    [-gi_mask_name gi_based_mask_names] [-out database_name]
                    [-max_file_sz number_of_bytes] [-logfile File_Name] [-taxid TaxID]
                    [-taxid_map TaxIDMapFile] [-version]

DESCRIPTION
    Application to create BLAST databases, version 2.2.30+

Use '-help' to print detailed descriptions of command line arguments
=====
Error: Argument "-dbtype". Value is missing
C:\Users\Public\TestDB>"\Program Files\NCBI\blast-2.2.30+bin\makeblastdb.exe"
-in human_prot.txt -dbtype prot -out human_prot.db

Building a new DB, current time: 01/19/2018 15:30:31
New DB name:      human_prot.db
New DB title:    human_prot.txt
Sequence type:   Protein
Keep Linkouts:   T
Keep MBits:      T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 317 sequences in 0.117115 seconds.
C:\Users\Public\TestDB>"\Program Files\NCBI\blast-2.2.30+bin\makeblastdb.exe"
-in human_prot.txt -dbtype prot -out human_prot.db
```

```

Task to execute
Default = 'blastp'
-db <String>
  BLAST database name
  * Incompatible with: subject, subject_loc
-out <File_Out>
  Output file name
  Default = '-'
-evalue <Real>
  Expectation value (E) threshold for saving hits
  Default = '10'
-word_size <Integer, >=2>
  Word size for wordfinder algorithm
-gapopen <Integer>
  Cost to open a gap
-gapextend <Integer>
  Cost to extend a gap
-matrix <String>
  Scoring matrix name (normally BLOSUM62)
-threshold <Real, >=0>
  Minimum word score such that the word is added to the BLAST lookup table
-comp_based_stats <String>
  Use composition-based statistics:
    D or d: default (equivalent to 2 )
    0 or F or f: No composition-based statistics
    1: Composition-based statistics as in NAR 29:2994-3005, 2001
    2 or I or t : Composition-based score adjustment as in Bioinformatics
    21:902-911,
    2005, conditioned on sequence properties
    3: Composition-based score adjustment as in Bioinformatics 21:902-911,
    2005, unconditionally
  Default = '2'

*** BLAST-2-Sequences options
-subject <File_In>
  Subject sequence(s) to search
  * Incompatible with: db, gillist, seqidlist, negative_gillist,
  negative_seqidlist, taxids, taxidlist, negative_taxids, negative_taxidlist,
  ipglist, negative_ipglist, db_soft_mask, db_hard_mask
-subject_loc <String>
  Location on the subject sequence in 1-based offsets (Format: start-stop)
  * Incompatible with: db, gillist, seqidlist, negative_gillist,
  negative_seqidlist, taxids, taxidlist, negative_taxids, negative_taxidlist,
  ipglist, negative_ipglist, db_soft_mask, db_hard_mask, remote

*** Formatting options
-outfmt <String>
  alignment view options:
    0 = Pairwise,
    1 = Query-anchored showing identities,
    2 = Query-anchored no identities,
    3 = Flat query-anchored showing identities,
    4 = Flat query-anchored no identities,
    5 = BLAST XML,
    6 = Tabular,
    7 = Tabular with comment lines,
    8 = Seqalign (Text ASN.1),
    9 = Seqalign (Binary ASN.1),
    10 = Comma-separated values,
    11 = BLAST archive (ASN.1),
    12 = Seqalign (JSON),
    13 = Multiple-file BLAST JSON,
    14 = Multiple-file BLAST XML2,
    15 = Single-file BLAST JSON,
    16 = Single-file BLAST XML2,
    18 = Organism Report

```

C:\Program Files\NCBI\blast-2.9.0+\bin>blastp -in

USAGE

```

blastp [-hl [-help] [-import_search_strategy filename]
[-export_search_strategy filename] [-task task_name] [-db database_name]
[-dbsize num_letters] [-gillist filename] [-seqidlist filename]
[-negative_gillist filename] [-negative_seqidlist filename]
[-taxids taxids] [-negative_taxids taxids] [-taxidlist filename]
[-negative_taxidlist filename] [-ipglist filename]
[-negative_ipglist filename] [-entrez_query entrez_query]
[-db_soft_mask filtering_algorithm] [-db_hard_mask filtering_algorithm]
[-subject subject_input_file] [-subject_loc range] [-query input_file]
[-out output_file] [-evalue evalue] [-word_size int_value]
[-gapopen open_penalty] [-gapextend extend_penalty]
[-qcov_hsp_perc float_value] [-max_hsps int_value]
[-xdrop_ungap float_value] [-xdrop_gap float_value]
[-xdrop_gap_final float_value] [-searchsp int_value] [-seg SEG_options]
[-soft_masking soft_masking] [-matrix matrix_name]
[-threshold float_value] [-culling_limit int_value]
[-best_hit_overhang float_value] [-best_hit_score_edge float_value]
[-subject_besthit] [-window_size int_value] [-lcase_masking]
[-query_loc range] [-parse_deflines] [-outfmt format] [-show_gis]
[-num_descriptions int_value] [-num_alignments int_value]
[-line_length line_length] [-html] [-sorthits sort_hits]
[-sorthsps sort_hsps] [-max_target_seqs num_sequences]
[-num_threads int_value] [-ungapped] [-remote] [-comp_based_stats compo]
[-use_sw_tback] [-version]

```

DESCRIPTION

Protein-Protein BLAST 2.9.0+

Use '-help' to print detailed descriptions of command line arguments

=====