**Official BLAST® Command Line Applications User Manual by NCBI** <https://www.ncbi.nlm.nih.gov/books/NBK279690/>

**Directions based off of NCBI Website (Standalone BLAST Setup for Unix)** <https://www.ncbi.nlm.nih.gov/books/NBK52640/>

**Set Up of BLAST+ app**

**Download**

The executable BLAST package is available for a variety of computer platforms on <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST/>

For MacOSX version 10.5 or higher, download ncbi-blast-#.#.#+.dmg, and double click on ncbi-blast-#.#.#+.pkg.

For older versions of MacOSX, download ncbi-blast-#.#.#+--x64-macosx.tar.gz.

The archive for Mac OSX is a gzip-compressed tar file where the #.#.# represents the version number of the current release.

To download, click “Save link as…” from the popup menu and save the archive to a desired location on the local disk.

Open the MacOSX terminal app (this is the Mac version of Bash which is a Unix shell and command language) and downloading through the ftp client is shown below. Copy and paste the red underlined commands into the terminal app. Make sure to modify the download title of the zip file.

$ ftp ftp.ncbi.nlm.nih.gov  
Connected to ftp.wip.ncbi.nlm.nih.gov.  
220-  
Warning Notice!  
This is a U.S. Government computer system, which may be accessed and used   
[ ... extra warning message removed ... ]  
There is no right of privacy in this system.  
---  
Welcome to the NCBI ftp server! The official anonymous access URL is ftp://ftp.ncbi.nih.gov   
Public data may be downloaded by logging in as "anonymous" using your E-mail  
address as a password.  
Please see ftp://ftp.ncbi.nih.gov/README.ftp for hints on large file transfers  
220 FTP Server ready.  
Name (ftp.ncbi.nlm.nih.gov:tao): anonymous  
331 Anonymous login ok, send your complete email address as your password.  
Password: [*note: enter your email address at this prompt; encrypted so the text will not show up as you type*]  
230-Anonymous access granted, restrictions apply.  
Please read the file README.ftp  
230 it was last modified on Fri Mar 28 14:05:45 2008 - 716 days ago  
Remote system type is UNIX.  
Using binary mode to transfer files.  
ftp> cd blast/executables/LATEST/  
250 CWD command successful  
ftp> bin  
200 Type set to I  
ftp> get ncbi-blast-2.6.0+-x64-macosx.tar.gz  
local: ncbi-blast-2.6.0+-x64-macosx.tar.gz remote: ncbi-blast-2.6.0+-x64-macosx.tar.gz  
227 Entering Passive Mode (130,14,29,30,215,39)  
150 Opening BINARY mode data connection for ncbi-blast-2.6.0+-x64-macosx.tar.gz (158357911 bytes)  
226 Transfer complete  
158357911 bytes received in 2.88 secs (54996.76 Kbytes/sec)  
ftp> bye  
221 Goodbye.  
$

**Installation**

To install, extract the downloaded package after placing it under a desired directory. This can be accomplished by a combination of gunzip and tar commands.

$ gunzip -d ncbi-blast-2.6.0+-x64-macosx.tar.gz  
$ tar xvpf ncbi-blast-2.6.0+-x64-macosx.tar.gz

Successful execution of the above commands installs the package and generates a new ncbi-blast-#.#.#+ directory under the working directory selected. This new directory contains the bin and doc subdirectories, as well as the VERSION file. The bin subdirectory contains these programs:

| **Program** | **Function** |
| --- | --- |
| blastdbcheck | Checks the integrity of a [BLAST](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/blast/) database |
| blastdbcmd | Retrieves sequences or other information from a [BLAST](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/blast/) database |
| blastdb\_aliastool | Creates database alias (to tie volumes together for example) |
| Blastn | Searches a nucleotide [query](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/query/) against a nucleotide database |
| blastp | Searches a protein [query](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/query/) against a protein database |
| blastx | Searches a nucleotide [query](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/query/), dynamically translated in all six frames, against a protein database |
| blast\_formatter | Formats a blast result using its assigned request ID (RID) or its saved archive |
| convert2blastmask | Converts lowercase [masking](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/masking/) into makeblastdb readable data |
| deltablast | Searches a protein [query](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/query/) against a protein database, using a more sensitive [algorithm](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/algorithm/) |
| dustmasker | Masks the low complexity regions in the input nucleotide sequences |
| legacy\_blast.pl | Converts a legacy blast search command line into blast+ counterpart and execute it |
| makeblastdb | Formats input [FASTA](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/fasta/) file(s) into a [BLAST](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/blast/) database |
| makembindex | Indexes an existing nucleotide database for use with megablast |
| makeprofiledb | Creates a conserved [domain](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/domain/) database from a list of input position specific scoring matrix (scoremats) generated by psiblast |
| psiblast | Finds members of a protein family, identifies proteins distantly related to the [query](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/query/), or builds position specific scoring matrix for the [query](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/query/) |
| rpsblast | Searches a protein against a conserved [domain](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/domain/) database to identify functional domains present in the [query](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/query/) |
| rpstblastn | Searches a nucleotide [query](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/query/), by dynamically translating it in all six-frames first, against a conserved [domain](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/domain/) database |
| segmasker | Masks the low complexity regions in input protein sequences |
| tblastn | Searches a protein [query](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/query/) against a nucleotide database dynamically translated in all six frames |
| tblastx | Searches a nucleotide [query](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/query/), dynamically translated in all six frames, against a nucleotide database similarly translated |
| update\_blastdb.pl | Downloads preformatted blast databases from NCBI |
| windowmasker | Masks repeats found in input nucleotide sequences |

**Configuration**

Using the BLAST+ package without configuration can be cumbersome – it requires the installation path to be prefixed to the program call and database specification since the system doesn’t know where to look for the installed program and specified database when doing future calls. To streamline BLAST searches, two environments, PATH and BLASTDB, need to be modified and specified to point to the corresponding directories.

Under bash, the following command appends the path to the new BLAST bin directory to the existing PATH setting.

$ export PATH=$PATH:$HOME/ncbi-blast-2.6.0+/bin

The modified $PATH can be examined using the echo command and what the output will show with the new underlined extension. Everything before the underlined portion will be different but still similar.

$ echo $PATH

/Library/Frameworks/Python.framework/Versions/3.6/bin:/usr/bin:/bin:/usr/sbin:/sbin:/usr/local/bin:/usr/local/ncbi/blast/bin

To get an idea of what this step did, you can view this path in your Mac Finder by clicking on USER 🡪 ncbi-blast-#.#.#+ 🡪 bin.

To manage available BLAST databases, create a directory to store them:

$ mkdir $HOME/blastdb

To get an idea of what this step did, you can view this path in your Mac finder by clicking on USER 🡪 blastdb.

**Database Download**

BLAST database is a key component of any BLAST search. To fully test the BLAST+ package thus installed, a functional database is needed. The following demonstrates the process of downloading and installation of the refseq\_rna.00.tar.gz into the database you configured in the previous step. This is a volume of a pre-formatted database from NCBI. This same procedure can be used to download other BLAST databases. Other databases can be found at <ftp://ftp.ncbi.nlm.nih.gov/blast/db/>

$ cd $HOME/blastdb  
home/tao/blastdb$ ftp ftp.ncbi.nlm.nih.gov  
Connected to ftp.wip.ncbi.nlm.nih.gov.  
220-  
Warning Notice!  
[ ... Extra warning message removed for brevity ... ]  
Name (ftp.ncbi.nlm.nih.gov:tao): anonymous  
331 Anonymous login ok, send your complete email address as your password.  
Password: [*note: enter your email address at this prompt; encrypted so the text will not show up as you type*]

230-Anonymous access granted, restrictions apply.  
Please read the file README.ftp  
230 it was last modified on Fri Mar 28 14:05:45 2008 - 716 days ago  
Remote system type is UNIX.  
Using binary mode to transfer files.  
ftp> cd blast/db  
250 CWD command successful  
ftp> bin  
200 Type set to I  
ftp> get refseq\_rna.00.tar.gz

local: refseq\_rna.00.tar.gz remote: refseq\_rna.tar.gz

229 Entering Extended Passive Mode (|||50279|)  
150 Opening BINARY mode data connection for refseq\_rna.00.tar.gz (857150245 bytes)  
100% |\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*| 817 MB 21.48 MB/s 00:00 ETA  
226 Transfer complete.  
857150245 bytes received in 00:38 (21.48 MB/s)  
ftp> bye  
221 Goodbye.   
home/blastdb$

Inflating the compressed archive and extracting the tar file will regenerate the files for this database. The example tar command and its output are given below. Make sure the path in the terminal is directed into the blastdb. To be sure, you can enter the first command for future experiments.

/home/tao/blastdb$ tar zxvpf refseq\_rna.00.tar.gz

refseq\_rna.nal  
refseq\_rna.00.nhr  
refseq\_rna.00.nin  
refseq\_rna.00.nnd  
refseq\_rna.00.nni  
refseq\_rna.00.nog  
refseq\_rna.00.nsd  
refseq\_rna.00.nsi  
refseq\_rna.00.nsq

home/tao/blastdb$ ls -ltr refseq\_rna.00\*  
-rw-rw-r-- 1 tao sdesk 30757308 Dec 15 19:24 refseq\_rna.00.nin  
-rw-rw-r-- 1 tao sdesk 999999519 Dec 15 19:24 refseq\_rna.00.nsq  
-rw-rw-r-- 1 tao sdesk 80420 Dec 15 19:24 refseq\_rna.00.nni  
-rw-rw-r-- 1 tao sdesk 20575424 Dec 15 19:24 refseq\_rna.00.nnd  
-rw-rw-r-- 1 tao sdesk 373245118 Dec 15 19:24 refseq\_rna.00.nhr  
-rw-rw-r-- 1 tao sdesk 2260381 Dec 15 19:25 refseq\_rna.00.nsi  
-rw-rw-r-- 1 tao sdesk 108664178 Dec 15 19:25 refseq\_rna.00.nsd  
-rw-rw-r-- 1 tao sdesk 10252432 Dec 15 19:25 refseq\_rna.00.nog  
-rw-r--r-- 1 tao sdesk 1048449064 Apr 14 16:32 refseq\_rna.00.tar.gz  
/home/tao/blastdb$

For other experiments where you’ve downloaded many databases and wish to save disk space, the refseq\_rna.tar.gz can be removed after installation if you no longer need it.

/home/tao/blastdb$ rm refseq\_rna.00.tar.gz  
/home/tao/blastdb$

**Execution and validation**

With the above BLAST+ setup, BLAST programs installed under the “ncbi-blast-#.#.#+/bin” directory can be invoked by name from any directory. Type the command “blastn –help” without quotes, which displays the program parameters of blastn to the console.

**$** blastn -help  
USAGE  
 blastn [-h] [-help] [-import\_search\_strategy filename]  
 [-export\_search\_strategy filename] [-task task\_name] [-db database\_name]  
 [-dbsize num\_letters] [-gilist filename] [-seqidlist filename]  
 [-negative\_gilist 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]  
 [-perc\_identity float\_value] [-xdrop\_ungap float\_value]  
 [-xdrop\_gap float\_value] [-xdrop\_gap\_final float\_value]  
 [-searchsp int\_value] [-max\_hsps int\_value] [-sum\_statistics]  
 [-penalty penalty] [-reward reward] [-no\_greedy]  
 [-min\_raw\_gapped\_score int\_value] [-template\_type type]  
 [-template\_length int\_value] [-dust DUST\_options]  
 [-filtering\_db filtering\_database]  
 [-window\_masker\_taxid window\_masker\_taxid]  
 [-window\_masker\_db window\_masker\_db] [-soft\_masking soft\_masking]  
 [-ungapped] [-culling\_limit int\_value] [-best\_hit\_overhang float\_value]  
 [-best\_hit\_score\_edge float\_value] [-window\_size int\_value]  
 [-off\_diagonal\_range int\_value] [-use\_index boolean] [-index\_name string]  
 [-lcase\_masking] [-query\_loc range] [-strand strand] [-parse\_deflines]  
 [-outfmt format] [-show\_gis] [-num\_descriptions int\_value]  
 [-num\_alignments int\_value] [-html] [-max\_target\_seqs num\_sequences]  
 [-num\_threads int\_value] [-remote] [-version]  
  
DESCRIPTION  
 Nucleotide-Nucleotide BLAST 2.6.0+  
  
OPTIONAL ARGUMENTS

-h  
 Print USAGE and DESCRIPTION; ignore all other parameters  
 -help  
 Print USAGE, DESCRIPTION and ARGUMENTS; ignore all other parameters  
 -version  
 Print version number; ignore other arguments  
  
 \*\*\* Input query options  
 - query <File\_In>  
 Input file name  
 Default = `-'  
 -query\_loc <String>  
 Location on the query sequence in 1-based offsets (Format: start-stop)  
 -strand <String, `both', `minus', `plus'>  
 Query strand(s) to search against database/subject  
 Default = `both'  
...

The real test of this installation should be example searches. The work session shown below performs the following task:

Call blastdbcmd to extract the sequence of nm\_000122 from the installed database (refseq\_rna.00) to a text file (test\_query.txt)

$ blastdbcmd -db refseq\_rna.00 -entry nm\_000122 -out test\_query.fa

Run a test blastn search using the sequence in test\_query.txt as query against refseq\_rna.00 database

$ blastn -[query](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/query/) test\_query.fa -db refseq\_rna.00 -task blastn -dust no -outfmt "7 qseqid sseqid evalue bitscore" -max\_target\_seqs 2  
BLASTN 2.2.29+  
Query: gi|263191547|ref|NM\_000122.3| Homo sapiens mutL homolog 1 (MLH1), transcript variant 1, mRNA  
Database: refseq\_rna.00  
Fields: [query](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/query/) id, subject id, evalue, [bit score](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/bit-score/)  
2 hits found  
gi|263191547|ref|NM\_000122.3| gi|263191547|ref|NM\_000122.3| 0.0 4801  
gi|263191547|ref|NM\_000122.3| gi|332816398|ref|[XM\_001170433.2](https://www.ncbi.nlm.nih.gov/nuccore/332816398)| 0.0 4758  
# [BLAST](https://www.ncbi.nlm.nih.gov/books/n/helpblast/blast_glossary/def-item/blast/) processed 1 queries  
$

For a variety of other options for the command-line, the following will lead to other specific parameters for other BLAST applications, like blastn, tblastn, etc: <https://www.ncbi.nlm.nih.gov/books/NBK279675/>

**Steps for BLAST Protein-Protein in Batches**

The following steps can be altered based on specifics for different experiments.

**Downloading and Installing swissprot.tar.gz database from NCBI.**

$ cd $HOME/blastdb  
home/tao/blastdb$ ftp ftp.ncbi.nlm.nih.gov  
Connected to ftp.wip.ncbi.nlm.nih.gov.  
220-  
Warning Notice!  
[ ... Extra warning message removed for brevity ... ]  
Name (ftp.ncbi.nlm.nih.gov:tao): anonymous  
331 Anonymous login ok, send your complete email address as your password.  
Password: [*note: enter your email address at this prompt; encrypted so the text will not show up as you type*]

230-Anonymous access granted, restrictions apply.  
Please read the file README.ftp  
230 it was last modified on Fri Mar 28 14:05:45 2008 - 716 days ago  
Remote system type is UNIX.  
Using binary mode to transfer files.  
ftp> cd blast/db  
250 CWD command successful  
ftp> bin  
200 Type set to I  
ftp> get swissprot.tar.gz

local: swissprot.tar.gz remote: swissprot.tar.gz

229 Entering Extended Passive Mode (|||50279|)  
150 Opening BINARY mode data connection for swissprot.tar.gz (857150245 bytes)  
100% |\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*| 817 MB 21.48 MB/s 00:00 ETA  
226 Transfer complete.  
857150245 bytes received in 00:38 (21.48 MB/s)  
ftp> bye  
221 Goodbye.   
home/blastdb$

**Inflating compressed archive and extracting the tar file**

Same as before, go onto <ftp://ftp.ncbi.nlm.nih.gov/blast/db/>

Right-click on swissprot.tar.gz and taxdb.tar.gz to save into USER/blastdb folder.

dhcp-172-21-66-152:blastdb legalllab$ tar zxvpf swissprot.tar.gz

swissprot.pal

swissprot.00.pin

swissprot.00.phr

swissprot.00.psq

swissprot.00.pni

swissprot.00.pnd

swissprot.00.psi

swissprot.00.psd

swissprot.00.ppi

swissprot.00.ppd

swissprot.00.pog

dhcp-172-21-66-152:blastdb legalllab$ ls -ltr swissprot\*  
-rw-r--r-- 1 legalllab staff 176065572 May 11 12:00 swissprot.00.psq

-rw-r--r-- 1 legalllab staff 3733400 May 11 12:00 swissprot.00.pin

-rw-r--r-- 1 legalllab staff 122615307 May 11 12:00 swissprot.00.phr

-rw-r--r-- 1 legalllab staff 14628 May 11 12:00 swissprot.00.ppi

-rw-r--r-- 1 legalllab staff 3733288 May 11 12:00 swissprot.00.ppd

-rw-r--r-- 1 legalllab staff 17372 May 11 12:00 swissprot.00.pni

-rw-r--r-- 1 legalllab staff 4433928 May 11 12:00 swissprot.00.pnd

-rw-r--r-- 1 legalllab staff 614544 May 11 12:00 swissprot.00.psi

-rw-r--r-- 1 legalllab staff 26271614 May 11 12:00 swissprot.00.psd

-rw-r--r-- 1 legalllab staff 1866676 May 11 12:00 swissprot.00.pog

-rw-r--r-- 1 legalllab staff 31 May 11 12:00 swissprot.pal

-rw-r--r-- 1 legalllab staff 124071726 May 12 00:01 swissprot.tar.gz  
dhcp-172-21-66-152:blastdb legalllab$

Repeat the two procedures with taxdb.

In the Mac Finder, create a folder in USER/blastdb and name it NCBI. Drag the swissprot and tax databases into the NCBI folder.

**Handling FASTA File**

Convert the Word Excel file of your local sequences that you wish to study with identifiers into a FASTA file. Create a new folder in USER/blastdb and name it sequences. Save the FASTA file into this folder.

**Building a BLAST database with local sequences**

<https://www.ncbi.nlm.nih.gov/books/NBK279688/>

The makeblastdb application produces BLAST databases from FASTA files The FASTA definition lines are not parsed by makeblastdb and may be completely unstructured. The text in the definition line will be stored in the BLAST database and displayed in the BLAST report, but it will not be possible to fetch individual sequences using blastdbcmd or to limit the search with the –seqidlist option. Use the –parse\_seqids flag when invoking makeblastdb to enable retrieval of sequences based upon sequence identifiers. Each sequence must have a unique identifier, and that identifier must have a specific format. It should begin right after the “>” sign on the definition line and contain no spaces.

$ cat sequences/sequence2.fasta

>1

AIDWLTG

>2

AQTYSPS

>3

AFRKFLPL

Makeblastdb can be invoked for this file as below.

$ makeblastdb -in sequences/sequence2.fasta -parse\_seqids -dbtype prot

Building a new DB, current time: 05/12/2017 14:08:59

New DB name: /Users/legalllab/blastdb/sequence2.fasta

New DB title: sequence2.fasta

Sequence type: Protein

Deleted existing Protein BLAST database named /Users/legalllab/blastdb/sequence2.fasta

Keep MBits: T

Maximum file size: 1000000000B

Adding sequences from FASTA; added 19 sequences in 0.00102186 seconds.

dhcp-172-21-66-152:blastdb legalllab$ cat sequence2.fasta

$

**Execution**

To execute the protein-protein comparison, invoke the following command into the terminal app.

$ blastp -task blastp-short -db NCBI/swissprot -query sequences/sequence2.fasta -out results/sequence2results.xls -outfmt “7 qseqid qseq qlen sseqid sacc slen score ppos”

$ blastp -task blastp-short -db swissprot –remote –entrez\_query “Homo sapiens [Organism]” -query sequences/sequence2.fasta -out results/sequence2results.xls -outfmt “7 qseqid qseq qlen sseqid sacc slen score ppos”

\*Sometimes, it is suggested to type this out instead of copying and pasting because the Terminal might reformat the next line and it might read it incorrectly. This is if you get an error saying “Too many positional arguments (1)” after copying and pasting. This is just a side note.

Explanation of command line

-task An option with blastp-short optimized for query sequences shorter than 30 residues

-db Refers to the NCBI database, swissprot in this experiment

-query Refers to the subject that will be compared to NCBI database = FASTA file with local sequences

-out Refers to the output and in this case, it will output as an Excel file that will show up in the blastdb folder

-outfmt Refers to customized format of output; 7 is output as tabular with comment line, qseqid is query sequence id, qseq is query sequence, qlen is query length, sacc is subject accession, score is raw score, ppos is percentage of positive-scoring matches”

**Results**

Insert a column between columns B and C. Type in the formula “=len(cell)” and for cell, replace that with the referenced cell with the query sequence.

This will allow you to see whether the query sequence is the same length as column D, that outputs query length.

It is also useful to look at sequences that output 100% for positive-scoring matches.

These are some of the ways that you can analyze the data you are given in order to calculate for accuracy in accession numbers.