



## Get NCBI BLAST databases

The best way to obtain BLAST databases is to download them from NCBI or the cloud (currently from Google Cloud Platform – an experimental feature - see details below). These are the same databases available via the public BLAST Web Service (<https://blast.ncbi.nlm.nih.gov>), are updated regularly, and contain taxonomic information built into them. These can also be a source of biological sequence data (see below).

To download a preformatted NCBI BLAST database, run the `update_blastdb.pl` program followed by any relevant options and the name(s) of the BLAST databases to download. For example:

```
$ update_blastdb.pl --decompress nr [*]
```

This command will download the compressed nr BLAST database *from NCBI* to the current working directory and decompress it. **Any subsequent identical invocations of this script with the same parameters in that directory will only download any data if it has a different time stamp when compared to the data at NCBI.**

### Experimental:

If the `--source gcp` option is provided, the data will be downloaded from the Google Cloud!

If you would like to see what BLAST databases are available to download, please run:

```
$ update_blastdb.pl --showall [*]
```

For more information on available NCBI BLAST databases, please see <https://go.usa.gov/xPhky>. For a demo of this tool, please see <https://bit.ly/2UA7tYb> (external link).

For more details about what command line options this tool supports, please run:

```
$ update_blastdb.pl --help
```

If you need FASTA from these BLAST databases, you can obtain it as follows:

```
$ blastdbcmd -entry all -db nr -out nr.fsa
```

If you need FASTA for selected sequence(s) from these BLAST databases, you can obtain it as follows (the sequence of interest is identified by the accession u00001 in this example):

```
$ blastdbcmd -entry u00001 -db nr -out u00001.fsa
```

[\*] If you run into any problems with this invocation, please try the `-passive` option, which is enabled by default in BLAST+ 2.8.1 and following. The `--decompress` option is only needed if the source data comes from NCBI