



Limiting a Search by taxonomy

In order to limit a BLAST search by taxonomy, one needs to provide the NCBI taxonomy ID(s) (taxid) for given organism(s). A taxid is simply a number that specifies a node in the taxonomic tree. For example, 9606 is the taxid for human, 9989 is the taxid for rodentia, and 2 is the taxid for all bacteria. Taxids are preferable to organism names as the latter can be ambiguous. For instance, bacteria is both a genus of insects as well as a superkingdom. **BLAST only accepts taxids that are at or below the species level.**

The BLAST+ command line tool suite provides the `get_species_taxids.sh` script to translate higher level taxids (e.g.: Enterobacterales) into a list of taxids that are at the appropriate level. This script can also be used to look up (and disambiguate) taxids based upon a taxonomic name. If unsure whether your taxid is at the species level or lower, it is safe to run `get_species_taxids.sh` and feed the output to BLAST.

Additionally, one may use the `-negative_taxids` and `-negative_taxidlist` options to exclude sequences by taxid.

Note: this script depends on EDirect and it only works on Linux and Mac. For EDirect installation instructions, please see <https://www.ncbi.nlm.nih.gov/books/NBK179288/>.

Limiting a BLAST search with a high-level taxonomic node

```
$ get_species_taxids.sh -n Enterobacterales
Taxid: 91347
rank: order
division: enterobacteria
scientific name: Enterobacterales
common name:
1 matches found
$ get_species_taxids.sh -t 91347 > 91347.txids
$ blastn -db nt -query QUERY -taxidlist 91347.txids -outfmt 7 -out OUTPUT.tab
```

Limiting a BLAST search with a species-level taxonomic node

```
$ blastn -db nt -query QUERY -taxids 9606 -outfmt 7 -out OUTPUT.tab
```