



Use blastdb_aliastool to manage the BLAST databases

Created: June 23, 2008; Updated: January 7, 2021.

Often, one needs to search multiple databases together or wishes to search a specific subset of sequences within an existing database. For these type of searches a convenient way to conduct them is by creating a virtual BLAST database. The blastdb_aliastool can perform three types of tasks to assist in that process. First, it can build an alias file to transparently combine searches of different databases. Second, it can build an alias file that limits a search based on a list of GIs (numerical IDs) or accessions. Finally, it can convert the list of GI's or accessions to a more efficient binary format.

Note: When combining BLAST databases, all the databases must be of the same molecule type. The following examples assume that the two databases as well as the GI file are in the current working directory. The binary format for accessions is only supported in the newer version 5 of the BLAST databases (BLAST+ 2.10.0 or newer suggested). Version 5 of the BLAST databases supports limiting a search natively by taxonomy, and only the relevant TAXIDs are needed.

Aggregate existing BLAST databases

To combine the two nematode nucleotide databases, named “nematode_mrna” and “nematode_genomic”, we use the following command line:

```
$ blastdb_aliastool -dblist "nematode_mrna nematode_genomic" -dbtype nucl \  
-out nematode_all -title "Nematode RefSeq mRNA + Genomic"
```

Create a subset of a BLAST database

The nematode_mrna database contains RefSeq mRNAs for several species of round worms. The best subset is from *C. elegans*. In most cases, we want to search this subset instead of the complete collection. Since the database entries are from NCBI nucleotide databases and the database is formatted with “-parse_seqsids”, we can use the “-gilist *c_elegans_mrna.gi*” parameter/value pair to limit the search to the subset of interest, alternatively, we can create a subset of the nematode_mrna database as follows:

```
$ blastdb_aliastool -db nematode_mrna -gilist c_elegans_mrna.gi -dbtype \  
nucl -out c_elegans_mrna -title "C. elegans refseq mRNA entries"
```

Note: one can also specify multiple databases using the -db parameter of blastdb_aliastool.

Convert a GI or accession list to binary format

The blastdb_aliastool can convert a GI or accession list to a binary format that is more efficient during the BLAST search. The example below converts a list of accessions to the binary format. The last two options shown (-seqid_db and -seqid_dbtype) are optional and limit the contents of the resulting accession list to accessions in

the specified database, in this case swissprot. This may result in a much smaller file and shorter run times, but BLAST will exit with an error if the specified database is not used. As mentioned earlier, binary accession lists are only supported with version 5 BLAST databases.

```
$ blastdb_aliastool -seqid_file_in myacc.acc -seqid_file_out myacc.bin.acc -seqid_db  
swissprot -seqid_dbtype prot
```