



Performing a DELTA-BLAST search

DELTA-BLAST searches a protein sequence database using a PSSM constructed from conserved domains matching a query. It first searches the NCBI CDD database to construct the PSSM.

Download the `cdd_delta` database

Obtain this database from <ftp://ftp.ncbi.nlm.nih.gov/blast/db> using the `update_blastdb.pl` tool (provided as part of the BLAST+ package). Note that the `cdd_delta` database must be downloaded and installed to the standard BLAST database directory (see [Configuring BLAST](#)) or in the current working directory.

Execute the `deltablast` search

```
$ deltablast -query query.fsa -db pataa
```

Indexed megaBLAST search

The indexed megaBLAST search requires both BLAST databases as well as an index of the words found in the database. The index of words may be produced with `makembindex`. The example below demonstrates how to produce the index as well as perform an indexed megaBLAST search. This example assumes that the `nt.00` BLAST database has been placed in the current directory (before `makembindex` is run) and that `QUERY` is a file containing a nucleotide query. Results will appear in `OUTPUT`. See tables C2 and C11 for information on command-line options.

```
$ makembindex -input nt.00 -ifformat blastdb -old_style_index false  
$ blastn -db ./nt.00 -query QUERY -use_index true -out OUTPUT
```

The BLAST databases may contain filtering (or masking) information for the database sequences. `Makembindex` can access this information and exclude the masked regions of the database from the index. This is demonstrated below. The first command shows how to discover the masking “Algorithm ID” from the BLAST database using `blastdbcmd`. In this case, the ID is 30. The second command demonstrates how to build an index that excludes the masked regions. Once the index has been built, it can be used as shown above. In the example below, the `ref_contig` BLAST database had been placed in the directory before `makembindex` was run.

```
$ blastdbcmd -db ref_contig -info  
Database: ref_contig  
364 sequences; 2,938,626,560 total bases  
  
Date: Oct 7, 2011 10:34 AM Longest sequence: 115,591,997 bases  
  
Available filtering algorithms applied to database sequences:
```

Algorithm ID	Algorithm name	Algorithm options
30	windowmasker	default options used

```
$ makembindex -input ref_contig -ifformat blastdb -old_style_index false -db_mask 30
creating /export/home/madden/INDEX_TEMP/ref_contig.00.idx...done
creating /export/home/madden/INDEX_TEMP/ref_contig.01.idx...done
creating /export/home/madden/INDEX_TEMP/ref_contig.02.idx...removed (empty)
```

BLAST+ remote service

The BLAST+ applications can perform a search on the NCBI servers if invoked with the “-remote” flag. All other command-line options are the same as for a stand-alone search.

The box below shows an example BLAST+ remote search using the blastn application. First, blastn searches the query against the nt database and produces a standard BLAST report. The query file (nt.u00001) contains the sequence for accession u00001 as FASTA. Second, the UNIX grep utility is used to find the RID for the search. Note that the RID can simply be found near the top of the BLAST report. Third, the RID is then used with blast_formatter to print out the results as a tabular report. Finally, the results are formatted as XML. The RID is only printed as an example and is no longer valid.

```
$ blastn -db nt -query nt.u00001 -out test.out -remote
$ grep RID test.out
RID: X3R7GAUS014

$blast_formatter -rid X3R7GAUS014 -out test.tab -outfmt 7

$blast_formatter -rid X3R7GAUS014 -out test.xml -outfmt 5
```