

Blastall

Blastall may be used to perform all five flavors of blast comparison. One may obtain the blastall options by executing 'blastall -' (note the dash). A typical use of blastall would be to perform a blastn search (nucl. vs. nucl.) of a file called QUERY would be:

```
blastall -p blastn -d nr -i QUERY -o out.QUERY
```

The output is placed into the output file out.QUERY and the search is performed against the 'nr' database. If a protein vs. protein search is desired, then 'blastn' should be replaced with 'blastp' etc.

Some of the most commonly used blastall options are:

blastall arguments:

-p Program Name [String]

Input should be one of "blastp", "blastn", "blastx", "tblastn", or "tblastx".

-d Database [String]

default = nr

The database specified must first be formatted with formatdb. Multiple database names (bracketed by quotations) will be accepted. An example would be

```
-d "nr est"
```

which will search both the nr and est databases, presenting the results as if one 'virtual' database consisting of all the entries from both were searched. The statistics are based on the 'virtual' database of nr and est.

-i Query File [File In]

default = stdin

The query should be in FASTA format. If multiple FASTA entries are in the input file, all queries will be searched.

-e Expectation value (E) [Real]

default = 10.0

-o BLAST report Output File [File Out] Optional

default = stdout

-F Filter query sequence (DUST with blastn, SEG with others) [String]

default = T

BLAST 2.0 and 2.1 uses the dust low-complexity filter for blastn and seg for the other programs. Both 'dust' and 'seg' are integral parts of the NCBI toolkit and are accessed automatically.

If one uses "-F T" then normal filtering by seg or dust (for blastn) occurs (likewise "-F F" means no filtering whatsoever).

This options also takes a string as an argument. One may use such a string to change the specific parameters of seg or invoke other filters. Please see the "Filtering Strings" section (below) for details.

-S Query strands to search against database (for blast[nx], and tblastx). 3 is both, 1 is top, 2 is bottom [Integer]

default = 3

-T Produce HTML output [T/F]

default = F

-l Restrict search of database to list of GI's [String] Optional

This option specifies that only a subset of the database should be searched, determined by the list of gi's (i.e., NCBI identifiers) in a file. One can obtain a list of gi's for a given Entrez query from <http://www.ncbi.nlm.nih.gov/Entrez/batch.html>. This file should be in the same directory as the database, or in the directory that BLAST is called from.

-U Use lower case filtering of FASTA sequence [T/F] Optional

default = F

This option specifies that any lower-case letters in the input FASTA file should be masked.

Enhancements:

A new option has been added to search multiple queries at once for the blastn and tblastn program options of blastall.

-B Number of concatenated queries, for blastn and tblastn [Integer]

Optional

default = 0

This new feature similar in principle, but different in implementation from the support for multiple queries already existing in megablast. The combination of ungapped search (-g F) and multiple queries (-B N) is not supported. The argument to -B option must be equal to the number of sequences in the FASTA input file.

Processing multiple query sequences in one run can be much faster than processing them with separate runs because the database is scanned only 1 time for the entire set of queries. When the -B option is used, the results may differ from the ones produced with individual queries. Usually results will be at least as good or better (in terms of score/evaluate) than the results of corresponding individual queries; exceptions occur due to the heuristic nature of BLAST. Additional alignments may appear. It is guaranteed that matching sequences will appear in the same order when they are tied in evaluate and are part of the output both with and without -B. When the -B option is used, the summary statistics at the bottom of the output are for the combined set of queries; at present, the summary statistics are not tabulated for the individual queries in a multiple-query input.