

NAME

bl2seq, blast, blastall, blastall_old, blastcl3, blastpgp, impala, megablast, rpsblast, seedtop – Basic Local Alignment Search Tool

SYNOPSIS

bl2seq [-] [-A] [-D N] [-E N] [-F str] [-G N] [-I "start stop"] [-J "start stop"] [-M str] [-S N] [-T] [-U] [-V] [-W N] [-X N] [-Y X] [-a filename] [-d N] [-e X] [-g F] [-i filename] [-j filename] [-m] [-o filename] [-p str] [-q N] [-r N] [-t N]

blast [-] [-B N] [-D N] [-C x] [-E N] [-F str] [-G N] [-H] [-I "start stop"] [-J "start stop"] [-K N] [-L] [-M str] [-N] [-P X] [-Q N] [-R] [-S N] [-T N] [-W N] [-X N] [-Y X] [-Z N] [-a N] [-b N] [-c] [-d str] [-e X] [-f X] [-g] [-h N] [-i filename] [-j filename] [-k str] [-m N] [-n] [-o filename] [-p str] [-q N] [-r N] [-s] [-t N] [-u] [-v N] [-w N] [-y N] [-z N]

blastall [-] [-A N] [-B N] [-C x] [-D N] [-E N] [-F str] [-G N] [-I] [-J] [-K N] [-L start,stop] [-M str] [-O filename] [-P N] [-Q N] [-R filename] [-S] [-T] [-U] [-V] [-W N] [-X N] [-Y X] [-Z N] [-a N] [-b N] [-d str] [-e X] [-f X] [-g F] [-i filename] [-l str] [-m N] [-n] [-o filename] [-p str] [-q N] [-r N] [-s] [-t N] [-v N] [-w N] [-y X] [-z X]

blastall_old [-] [-A N] [-B N] [-C x] [-D N] [-E N] [-F str] [-G N] [-I] [-J] [-K N] [-L start,stop] [-M str] [-O filename] [-P N] [-Q N] [-R filename] [-S] [-T] [-U] [-W N] [-X N] [-Y X] [-Z N] [-a N] [-b N] [-d str] [-e X] [-f X] [-g F] [-i filename] [-l str] [-m N] [-n] [-o filename] [-p str] [-q N] [-r N] [-s] [-t N] [-v N] [-w N] [-y X] [-z X]

blastcl3 [-] [-A N] [-C x] [-D N] [-E N] [-F str] [-G N] [-I] [-J] [-K N] [-L start,stop] [-M str] [-O filename] [-Q N] [-R] [-S] [-T] [-U] [-W N] [-X N] [-Y X] [-Z N] [-a N] [-b N] [-d str] [-e X] [-f X] [-g F] [-i filename] [-m N] [-n] [-o filename] [-p str] [-q N] [-r N] [-s] [-t N] [-u str] [-v N] [-w N] [-y X] [-z X]

blastpgp [-] [-A N] [-B filename] [-C filename] [-E N] [-F T] [-G N] [-H N] [-I] [-J] [-K N] [-L N] [-M str] [-N X] [-O filename] [-P N] [-Q filename] [-R filename] [-S N] [-T] [-U] [-W N] [-X N] [-Y X] [-Z N] [-a N] [-b N] [-c N] [-d str] [-e X] [-f N] [-h X] [-i filename] [-j N] [-k filename] [-l str] [-m N] [-o filename] [-p str] [-q N] [-s] [-t N] [-u] [-v N] [-y X] [-z N]

impala [-] [-E N] [-F str] [-G N] [-H] [-I] [-J] [-M str] [-O filename] [-P filename] [-a N] [-b N] [-c N] [-d str] [-e X] [-h X] [-i filename] [-j N] [-m N] [-o filename] [-v N] [-y X] [-z N]

megablast [-] [-A N] [-D N] [-E N] [-F str] [-G N] [-H N] [-I] [-J] [-L start,stop] [-M N] [-N N] [-O filename] [-P N] [-Q filename] [-R] [-S N] [-T] [-U] [-V] [-W N] [-X N] [-Y X] [-Z N] [-a N] [-b N] [-d str] [-e X] [-f] [-g F] [-i filename] [-l str] [-m N] [-n] [-o filename] [-p X] [-q N] [-r N] [-t N] [-s N] [-v N] [-y N] [-z X]

rpsblast [-] [-F str] [-I] [-J] [-L start,stop] [-N X] [-O filename] [-P N] [-T] [-U] [-X N] [-Y X] [-Z N] [-a N] [-b N] [-d filename] [-e X] [-i filename] [-l filename] [-m N] [-o filename] [-p F] [-v N] [-y X] [-z N]

seedtop [-] [-C N] [-D N] [-E N] [-F] [-G N] [-I] [-J] [-M str] [-O filename] [-S N] [-X N] [-d str] [-e X] [-f] [-i filename] [-k filename] [-o filename] [-p str] [-q N] [-r N]

DESCRIPTION

This manual page documents briefly the commands **bl2seq**, **blast**, **blastall**, **blastcl3**, **blastpgp**, **impala**, **megablast**, **rpsblast**, and **seedtop**. These commands are documented together because they have a lot of common options.

bl2seq performs a comparison between two sequences using either the blastn or blastp algorithm. Both sequences must be either nucleotides or proteins.

blast compares a sequence against either a local database or a second sequence; it incorporates most of the functionality of both **bl2seq** and **blastall**, but uses a semi-experimental new internal engine.

blastall and **blastall_old** find the best matches in a local database for a sequence. **blastall** uses a newer

engine than **blastall_old** by default, but supports using the older engine as well (when invoked with the option **-V F**).

blastcl3 accesses the newest NCBI BLAST search engine (version 2.0). The software behind BLAST version 2.0 was written from scratch to allow BLAST to handle the new challenges posed by the sequence databases in the coming years. Updates to this software will continue in the coming years.

blastpgp performs gapped blastp searches and can be used to perform iterative searches in psi-blast and phi-blast mode.

impala searches a database of score matrices, prepared by **copymat**(1), producing BLAST-like output.

megablast uses the greedy algorithm of Webb Miller et al. for nucleotide sequence alignment search and concatenates many queries to save time spent scanning the database. This program is optimized for aligning sequences that differ slightly as a result of sequencing or other similar "errors". It is up to 10 times faster than more common sequence similarity programs and therefore can be used to swiftly compare two large sets of sequences against each other.

rpsblast (Reverse PSI-BLAST) searches a query sequence against a database of profiles. This is the opposite of PSI-BLAST that searches a profile against a database of sequences, hence the 'Reverse'. **rpsblast** uses a BLAST-like algorithm, finding single- or double-word hits and then performing an ungapped extension on these candidate matches. If a sufficiently high-scoring ungapped alignment is produced, a gapped extension is performed and those (gapped) alignments with sufficiently low expect value are reported. This procedure is in contrast to IMPALA that performs a Smith-Waterman calculation between the query and each profile, rather than using a word-hit approach to identify matches that should be extended.

seedtop answers two relatively simple questions:

1. Given a sequence and a database of patterns, which patterns occur in the sequence and where?
2. Given a pattern and a sequence database, which sequences contain the pattern and where?

Some of these commands support multiple types of comparison, governed by the **-p** ("program") flag:

blastp	compares an amino acid query sequence against a protein sequence database.
blastn	compares a nucleotide query sequence against a nucleotide sequence database.
blastx	compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database. For bl2seq , the nucleotide should be the first sequence given.
psitblastn	compares a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames (both strands) using a position specific matrix created by PSI-BLAST.
tblastn	compares a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames (both strands). For bl2seq , the nucleotide should be the second sequence given.
tblastx	compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

OPTIONS

A summary of options is included below.

- Print usage message
- A (bl2seq)
Input sequences in the form of accession.version
- A *N* (blastall, blastall_old, blastcl3, blastpgp, megablast)
Multiple Hits window size; generally defaults to 0 (for single-hit extensions), but defaults to 40 when using discontinuous templates.

- B *N*** (blast)
 - Produce on-the-fly output:
 - 0 none (default)
 - 1 table of offsets and quality values
 - 2 add sequence data
 - 3 text ASN.1
 - 4 binary ASN.1
- B *N*** (blastall, blastall_old)
 - Number of concatenated queries, in blastn or tblastn mode
- B *filename*** (blastpgp)
 - Input Alignment File for PSI-BLAST Restart
- C *X*** (blast, blastall, blastall_old, blastcl3)
 - Use composition-based statistics for blastp or tblastn:
 - D or d Default (equivalent to **T**)
 - 0, F, or f No composition-based statistics
 - 1, T, or t Composition-based statistics as in *NAR* 29:2994-3005, 2001
 - 2 Composition-based score adjustment as in *Bioinformatics* 21:902-911, 2005, conditioned on sequence properties
 - 3 Composition-based score adjustment as in *Bioinformatics* 21:902-911, 2005, unconditionally

When enabling statistics in blastall, blastall_old, or blastcl3 (*i.e.*, not blast), appending **u** (case-insensitive) to the mode enables use of unified p-values combining alignment and compositional p-values in round 1 only.
- C *filename*** (blastpgp)
 - Output File for PSI-BLAST Checkpointing
- C *N*** (seedtop)
 - Score only or not (default = 1)
- D *N*** (bl2seq)
 - Output format:
 - 0 traditional (default)
 - 1 tabular
- D *N*** (blast, blastall, blastall_old, blastcl3)
 - Translate sequences in the database according to genetic code *N* in /usr/share/ncbi/data/gc.prt (default is 1; only applies to tblast*)
- D *N*** (megablast)
 - Type of output:
 - 0 alignment endpoints and score
 - 1 all ungapped segments endpoints
 - 2 traditional BLAST output (default)
 - 3 tab-delimited one line format
 - 4 incremental text ASN.1
 - 5 incremental binary ASN.1
- D *N*** (seedtop)
 - Cost decline to align (default = 99999)
- E *N*** (bl2seq, blastcl3, megablast)
 - Extending a gap costs *N* (-1 invokes default behavior)

- E** *N* (blast, blastall, blastall_old)
Extending a gap costs *N* (-1 invokes default behavior: non-affine if greedy, 2 otherwise)
- E** *N* (blastpgp, impala, seedtop)
Extending a gap costs *N* (default is 1)
- F** *str* (bl2seq, blast, blastall, blastall_old, blastpgp, blastcl3, impala, megablast, rpsblast) Filter options for DUST or SEG; defaults to **T** for bl2seq, blast, blastall, blastall_old, blastcl3, and megablast, and to **F** for blastpgp, impala, and rpsblast.
- F** (seedtop)
Filter sequence with SEG.
- G** *N* (bl2seq, blastcl3, megablast)
Opening a gap costs *N* (-1 invokes default behavior)
- G** *N* (blast, blastall, blastall_old)
Opening a gap costs *N* (-1 invokes default behavior: non-affine if greedy, 5 if using dynamic programming)
- G** *N* (blastpgp, impala, seedtop)
Opening a gap costs *N* (default is 11)
- H** (blast)
Produce HTML output
- H** *N* (blastpgp)
End of required region in query (-1 indicates end of query)
- H** (impala)
Print help (different from usage message)
- H** *N* (megablast)
Maximal number of HSPs to save per database sequence (default is 0, unlimited)
- I** "*start stop*" (bl2seq, blast)
Location on first (query) sequence (applies only if file specified with **-i** contains a single sequence)
- I** (blastall, blastall_old, blastcl3, blastpgp, impala, megablast, rpsblast, seedtop) Show GIs in defines
- J** "*start stop*" (bl2seq, blast)
Location on second (subject) sequence (applies only if file specified with **-j** contains a single sequence)
- J** (blastall, blastall_old, blastcl3, blastpgp, impala, megablast, rpsblast, seedtop) Believe the query define
- K** *N* (blast, blastall, blastall_old, blastcl3, blastpgp)
Number of best hits from a region to keep (off by default, if used a value of 100 is recommended)
- L** (blast)
Use (classical Mega BLAST) lookup table with width 12
- L** *start,stop* (blastall, blastall_old, blastcl3, megablast, rpsblast) Location on query sequence (for rpsblast, only valid in blastp mode)
- M** *str* (bl2seq, blast, blastall, blastall_old, blastcl3, blastpgp, impala, seedtop) Use matrix *str* (default = BLOSUM62)
- M** *N* (megablast)
Maximal total length of queries for a single search (default = 5000000)
- N** (blast)
Show only accessions for sequence IDs in tabular output

- N** *X* (blastpgp, rpsblast)
Number of bits to trigger gapping (default = 22.0)
- N** *N* (megablast)
Type of a discontinuous word template:
 - 0 coding (default)
 - 1 optimal
 - 2 two simultaneous
- O** *filename* (blastall, blastall_old, blastcl3, blastpgp, impala, megablast, rpsblast, seedtop) Write (ASN.1) sequence alignments to *filename*; only valid for blastpgp, impala, rpsblast, and seedtop with **-J**, and only valid for megablast with **-D2**.
- P** *X* (blast)
Identity percentage cut-off
- P** *N* (blastall, blastall_old, blastcl3, blastpgp, rpsblast)
Set to 1 for single-hit mode or 0 for multiple-hit mode (default). Does not apply to blastn.
- P** *filename* (impala)
Read matrix profiles from database *filename*
- P** *N* (megablast)
Maximal number of positions for a hash value (set to 0 [default] to ignore)
- Q** *N* (blast, blastall, blastall_old, blastcl3)
Translate query according to genetic code *N* in /usr/share/ncbi/data/gc.prt (default is 1)
- Q** *filename* (blastpgp)
Output File for PSI-BLAST Matrix in ASCII
- Q** *filename* (megablast)
Masked query output; requires **-D 2**
- R** (blast)
Compute locally optimal Smith-Waterman alignments. (This option is only available for gapped tblastn.)
- R** *filename* (blastall, blastall_old)
Read PSI-TBLASTN checkpoint file *filename*
- R** (blastcl3)
RPS Blast search
- R** *filename* (blastpgp)
Input File for PSI-BLAST Restart
- R** (megablast)
Report the log information at the end of output
- S** *N* (bl2seq, blast, blastall, blastall_old, blastcl3, megablast) Query strands to search against database for blastn, blastx, tblastx:
 - 1 top
 - 2 bottom
 - 3 both (default)
- S** *N* (blastpgp)
Start of required region in query (default = 1)
- S** *N* (seedtop)
Cutoff cost (default = 30)

- T (bl2seq, blastall, blastall_old, blastcl3, blastpgp, megablast, rpsblast) Produce HTML output
- T N (blast)
 - Type of a discontinuous word template:
 - 0 coding (default)
 - 1 optimal
 - 2 two simultaneous
- U (bl2seq, blastall, blastall_old, blastcl3, blastpgp, megablast, rpsblast) Use lower case filtering for the query sequence
- V (bl2seq, blastall, megablast)
 - Force use of legacy engine
- V (blast)
 - Use variable word size approach to database scanning
- W N (bl2seq, blast, blastall, blastall_old, blastcl3, blastpgp, megablast, rpsblast) Use words of size *N* (length of best perfect match; zero invokes default behavior, except with megablast, which defaults to 28, and blastpgp, which defaults to 3. The default values for the other commands vary with "program": 11 for blastn, 28 for megablast, and 3 for everything else.)
- X N (bl2seq, blast, blastall, blastall_old, blastcl3, blastpgp, megablast, rpsblast, seedtop) *X* dropoff value for gapped alignment (in bits) (zero invokes default behavior, except with megablast, which defaults to 20, and rpsblast and seedtop, which default to 15. The default values for the other commands vary with "program": 30 for blastn, 20 for megablast, 0 for tblastx, and 15 for everything else.)
- Y X (bl2seq, blast, blastall, blastall_old, blastcl3, blastpgp, megablast, rpsblast) Effective length of the search space (use zero for the real size)
- Z N (blast, blastall, blastall_old, blastcl3, blastpgp, megablast, rpsblast) *X* dropoff value for final [dynamic programming?] gapped alignment in bits (default is 50 for blastn and megablast, 0 for tblastx, 25 for others)
- a *filename* (bl2seq)
 - Write text ASN.1 output to *filename*
- a N (blast, blastall, blastall_old, blastcl3, blastpgp, impala, megablast, rpsblast) Number of threads to use (default is one)
- b N (blast, blastall, blastall_old, blastcl3, blastpgp, impala, megablast, rpsblast) Number of database sequences to show alignments for (B) (default is 250)
- c (blast)
 - Mask lower case
- c N (blastpgp, impala)
 - Constant in pseudocounts for multipass version (default is 9)
- d N (bl2seq)
 - Use theoretical DB size of *N* (zero stands for the real size)
- d *str* (blast, blastall, blastall_old, blastcl3, blastpgp, impala, megablast, seedtop) Database to use (default is nr for all executables except blast, which requires a second FASTA sequence if this is not set)
- d *filename* (rpsblast)
 - RPS BLAST Database
- e X Expectation value (E) (default = 10.0)

- f** *X* (blast, blastall, blastall_old, blastcl3)
Threshold for extending hits, default if zero: 0 for blastn and megablast, 11 for blastp, 12 for blastx, and 13 for tblastn and tblastx.
- f** *N* (blastpgp)
Threshold for extending hits (default 11)
- f** (megablast)
Show full IDs in the output (default - only GIs or accessions)
- f** (seedtop)
Force searching for patterns even if they are too likely
- g** **F** (bl2seq, blastall, blastall_old, blastcl3)
Do not perform gapped alignment (N/A for tblastx)
- g** (blast)
Use greedy algorithm for gapped extensions
- g** **F** (megablast)
Make discontinuous megablast generate words for every base of the database (mandatory with the current BLAST engine)
- h** *N* (blast)
Frame shift penalty for out-of-frame gapping (blastx, tblastn only; default is zero)
- h** *X* (blastpgp, impala)
e-value threshold for inclusion in multipass model (default = 0.002 for blastpgp, 0.005 for impala)
- i** *filename*
Read (first, query) sequence or set from *filename* (default is stdin; not needed for blastpgp if restarting from scoremat)
- j** *filename* (bl2seq, blast)
Read second (subject) sequence or set from *filename*
- j** *N* (blastpgp)
Maximum number of passes to use in multipass version (default = 1)
- k** *str* (blast)
Pattern for PHI-BLAST
- k** *filename* (blastpgp, seedtop)
Input hit file for PHI-BLAST (default = hit_file)
- l** *str* (blastall, blastall_old, blastpgp, megablast)
Restrict search of database to list of GI's [String]
- l** *filename* (rpsblast)
Log messages to *filename* rather than standard error.
- m** (bl2seq)
Use Mega Blast for search
- m** *N* (blast, blastall, blastall_old, blastcl3, blastpgp, impala, megablast, rpsblast) alignment view options:
 - 0 pairwise (default)
 - 1 query-anchored showing identities
 - 2 query-anchored, no identities
 - 3 flat query-anchored, show identities
 - 4 flat query-anchored, no identities
 - 5 query-anchored, no identities and blunt ends
 - 6 flat query-anchored, no identities and blunt ends
 - 7 XML Blast output (not available for impala)

- 8 tabular (not available for impala)
- 9 tabular with comment lines (not available for impala)
- 10 ASN.1 text (not available for impala or rpsblast)
- 11 ASN.1 binary (not available for impala or rpsblast)
- n** (blast)
 Show GIs in sequence IDs
- n** (blastall, blastall_old, blastcl3)
 MegaBlast search
- n** (megablast)
 Use non-greedy (dynamic programming) extension for affine gap scores
- o** *filename*
 Write final alignment report to *filename* rather than stdout
- p** *str* (bl2seq, blast, blastall, blastall_old, blastcl3)
 Use the "program" (comparison type) *str*. The **DESCRIPTION** section covers this option in more detail.
- p** *str* (blastpgp)
 program option for PHI-BLAST (default = blastpgp)
- p** *X* (megablast)
 Identity percentage cut-off (default = 0)
- p** **F** (rpsblast)
 Query sequence is nucleotide, not protein
- p** *str* (seedtop)
 program name:
 patmatchp indicates which patterns occur in a sequence
 patternp indicates which sequences contain a pattern
- q** *N* (bl2seq, blast, blastall, blastall_old, blastcl3,
 megablast, seedtop) Penalty for a nucleotide mismatch (blastn only) (default = -10 for seedtop, -3
 for everything else)
- q** *N* (blastpgp)
 ASN.1 Scoremat input of checkpoint data:
 0 no scoremat input (default)
 1 restart from ASCII scoremat checkpoint file
 2 restart from binary scoremat checkpoint file
- r** *N* (bl2seq, blast, blastall, blastall_old, blastcl3,
 megablast, seedtop) Reward for a nucleotide match (blastn only) (default = 10 for seedtop, -10 for
 everything else)
- s** (blast)
 No-op (formerly requested generating words for every base of the database)
- s** (blastall, blastall_old, blastcl3, blastpgp)
 Compute locally optimal Smith-Waterman alignments. For blastall, blastall_old, and blastcl3, this
 is only available in gapped tblastn mode.
- s** *N* (megablast)
 Minimal hit score to report (0 for default behavior)
- t** *N* (bl2seq, blast, blastall, blastall_old, blastcl3)
 Length of a discontinuous word template (the largest intron allowed in a translated nucleotide
 sequence when linking multiple distinct assignments; default = 0; negative values disable linking
 for blastall, blastall_old, and blastcl3.)

-t *N* [*u*] (blastpgp)

Composition-based statistics mode. The first character is interpreted as follows:

0, F, or f

no composition-based statistics

1, T, or t

composition-based statistics as in *NAR* 29:2994–3005, 2001

2 composition-based score adjustment as in *Bioinformatics* 21:902-911, 2005, conditioned on sequence properties in round 1

3 composition-based score adjustment as in *Bioinformatics* 21:902-911, 2005, unconditionally in round 1

When composition-based statistics are in use, appending **u** (case-insensitive) to the argument requests unified p-value combining alignment p-value and compositional p-value in round 1 only.

-t *N* (megablast)

Length of a discontiguous word template (contiguous word if 0 [default])

-u (blast)

Do only ungapped alignment (always TRUE for tblastx)

-u *str* (blastcl3)

Restrict search of database to results of Entrez2 lookup

-u *N* (blastpgp)

ASN.1 Scoremat output of checkpoint data:

0 no scoremat output (default)

1 output ASCII scoremat checkpoint file (requires **-J**)

2 output binary scoremat checkpoint file (requires **-J**)

-v *N* (blast, blastall, blastall_old, blastcl3, blastpgp,

impala, megablast, rpsblast) Number of one-line descriptions to show (**V**) (default = 500)

-w *N* (blast)

Window size (max. allowed distance between a pair of initial hits; 0 invokes default behavior, -1 turns off multiple hits)

-w *N* (blastall, blastall_old, blastcl3)

Frame shift penalty (OOF algorithm for blastx)

-y *X* (blast, blastall, blastall_old, blastcl3, blastpgp,

impala, rpsblast) **X** dropoff for ungapped extensions in bits (0.0 invokes default behavior: 20 for blastn, 10 for megablast, and 7 for all others.)

-y *N* (megablast)

X dropoff value for ungapped extension (default is 10)

-z *N* (blast)

Longest intron length for uneven gap HSP linking (tblastn only; default is 0)

-z *N* (blastall, blastall_old, blastcl3, blastpgp, impala,

megablast, rpsblast) Effective length of the database (use zero for the real size)

BUGS

This manual page is long and confusing; individual pages might be better.

AUTHOR

The National Center for Biotechnology Information.

SEE ALSO

blastclust(1), **copymat(1)**, **fastacmd(1)**, **formatdb(1)**, **formatrpsdb(1)**, **makemat(1)**, **blast.html**, **seed-top.html**, <http://www.ncbi.nlm.nih.gov/BLAST/>.