



## Reformat BLAST reports with blast\_formatter

It may be helpful to view the same BLAST results in different formats. A user may first parse the tabular format looking for matches meeting a certain criteria, then go back and examine the relevant alignments in the full BLAST report. He may also first look at pair-wise alignments, then decide to use a query-anchored view. Viewing a BLAST report in different formats has been possible on the NCBI BLAST web site since 2000, but has not been possible with stand-alone BLAST runs. The blast\_formatter allows this, if the original search produced blast archive format using the `-outfmt 11` switch. The query sequence, the BLAST options, the masking information, the name of the database, and the alignment are written out as ASN.1 (a structured format similar to XML). The `-max_target_seqs` option should be used to control the number of matches recorded in the alignment. The blast\_formatter reads this information and formats a report. The BLAST database used for the original search must be available, or the sequences need to be fetched from the NCBI, assuming the database contains sequences in the public dataset. The box below illustrates the procedure. A blastn run first produces the BLAST archive format, and the blast\_formatter then reads the file and produces tabular output.

blast\_formatter will format stand-alone searches performed with an earlier version of a database if both the search and formatting databases are prepared so that fetching by sequence ID is possible. To enable fetching by sequence ID use the `-parse_seqids` flag when running makeblastdb, or (if available) download preformatted BLAST databases from <ftp://ftp.ncbi.nlm.nih.gov/blast/db/> using [update\\_blastdb.pl](#) (provided as part of the BLAST+ package). Currently the blast archive format and blast\_formatter do not work with database free searches (i.e., `-subject` rather than `-db` was used for the original search).

```
$ echo 1786181 | blastn -db ecoli -outfmt 11 -out out.1786181.asn
$ blast_formatter -archive out.1786181.asn -outfmt "7 qacc sacc evaluate
qstart qend sstart send"
# BLASTN 2.2.24+
# Query: gi|1786181|gb|AE000111.1|AE000111 Escherichia coli K-12 MG1655
section 1 of 400
# Database: ecoli
# Fields: query acc., subject acc., evaluate, q. start, q. end,
s. start, s. end
# 85 hits found
AE000111      AE000111      0.0      1      10596      1      10596
AE000111      AE000174      8e-30     5565     5671      6928     6821
AE000111      AE000394      1e-27     5587     5671      135      219
AE000111      AE000425      6e-26     5587     5671      8552     8468
AE000111      AE000171      3e-24     5587     5671      2214     2130
AE000111      AE000171      1e-23     5587     5670      10559    10642
AE000111      AE000376      1e-22     5587     5675      129      42
AE000111      AE000268      1e-22     5587     5671      6174     6090
AE000111      AE000112      1e-22     10539    10596     1        58
AE000111      AE000447      5e-22     5587     5670      681      598
AE000111      AE000344      6e-21     5587     5671      4112     4196
```

|          |          |       |      |      |       |       |
|----------|----------|-------|------|------|-------|-------|
| AE000111 | AE000490 | 2e-20 | 5584 | 5671 | 4921  | 4835  |
| AE000111 | AE000280 | 2e-20 | 5587 | 5670 | 12930 | 12847 |