

1 Notes for installing a local blast+ instance of NCBI BLAST

2 F. J. Pineda

3 09/25/2017

4  
5 see:

6 NCBI Blast tutorial, <http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.h>

7 BLAST, I. Korf, M. Yandell, & J. Bedell, O'Reilly, 2003

8  
9  
10  
11 The current most up-to-date version of ncbi-blast is 'blast+' which is at ve

12 To install blast+ locally please start here:

13  
14 <https://www.ncbi.nlm.nih.gov/guide/howto/run-blast-local/>

15  
16 Installation instructions are here

17  
18 <https://www.ncbi.nlm.nih.gov/books/NBK1762/>

19  
20 #####

21 # Step 1: install your own local instance of blast

22 #####

23  
24 These instructions are based on the ncbi blast installation pages by Tao

25  
26 <https://www.ncbi.nlm.nih.gov/books/NBK52640/>

27  
28  
29 #####

30 # Go to the ncbi blast download server and look at the latest available

31 #####

32  
33 > curl ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/

34  
35 # You should see a listing of distributions to download. What you downlo

36 # on your machine's OS and hardware, e.g. MAC, Windows, Linux, 64-bit, e

37 # It will also depend on whether you want to compile from source code. H

38 # download binaries that will run on the cluster

39 #

40 # cd into you home directory and download the latest blast for 64-bit li

41  
42 > cd

43 > curl -O ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/\n  
44 ncbi-blast-2.6.0+-x64-linux.tar.gz

45

```
46 # This might take a long time so you can copy a previously downloaded v
47 > cp /users/sph140636/shared/code_examples/blast/ncbi-blast-2.6.0+-x64-l
48
49 # now expand the compressed tarball:
50 > tar -xzvf ncbi-blast-2.6.0+-x64-linux.tar.gz
51
52 # you should now have a directory named 'ncbi-blast-2.6.0+'
53 # If it worked, clean up by deleting the compressed tar file
54
55 rm ncbi-blast-2.6.0+-x64-linux.tar.gz
56
57
58 # finally to make your installation version independent, create a link:
59
60 ln -s ~/ncbi-blast-2.6.0+ ~/blast
61
62 # you can check that ~/blast links to ~/ncbi-blast-2.6.0+ by executing
63 ls -ld ~/blast
64
65 # Finally, we will need a directory where we keep the blast formatted da
66
67 mkdir ~/blastdb
68
69 #####
70 # Next we need to update the PATH environment variable so the shell looks fo
71 # blast-related commands where we just installed them. We also need to set t
72 # environment variable, so that blast knows to look in ~/blastdb for the da
73 #
74 #####
75
76 # Add the following two lines to your .bashrc file
77
78     export BLASTDB=$HOME/blastdb
79     export PATH=$PATH:$HOME/blast/bin
80
81 # You can either do this by editing .bashrc and adding the lines to the end
82 # or you can directly appending the lines to the end of you .bashrc by execu
83 # following commands
84
85     echo 'export BLASTDB=$HOME/blastdb' >> ~/.bashrc
86     echo ' export PATH=$PATH:$HOME/blast/bin' >> ~/.bashrc
87
88 # The commands in .bashrc don't update your environment as you can verify
89 # 'echoing' the PATH and BLASTDB environment variables:
90
```

```
91     echo $PATH
92     echo $BLASTDB
93
94 # Recall that .bashrc is only executed when you start a login shell. So your
95 # would have PATH and BLASTDB set on your next login. Alternately you can ex
96 # .bashrc by 'sourcing' it:
97
98     source ~/.bashrc
99
100 # Now check to see that your environment variables have actually been update
101
102     echo $PATH
103     echo $BLASTDB
104
105 #####
106 # Step 2: Download a preformatted database so you can test
107 # your blast installation
108 #####
109
110 # download a pre-formatted blast database into your blastdb directory:
111
112 > cd ~/blastdb
113 > curl -O ftp://ftp.ncbi.nlm.nih.gov/blast/db/refseq_rna.00.tar.gz
114
115 # This might take some time, so you can copy the database from the
116 # blast directory shared directory (/users/sph140636/shared/code_examples/bl
117
118 # expand the compressed tarball
119
120 > tar -xzvf refseq_rna.00.tar.gz
121
122 # you should see the following files:
123
124 refseq_rna.00.nnd
125 refseq_rna.00.nni
126 refseq_rna.00.nog
127 refseq_rna.00.nsd
128 refseq_rna.00.nsi
129 refseq_rna.00.nsq
130
131 # delete the tar file
132
133 > rm refseq_rna.00.tar.gz
134
135 #####
```

```
136 # Step 3: Test the installation
137 #####
138 # create a scratch directory to serve a temporary workspace
139 # and cd into it
140
141 mkdir ~/scratch
142 cd ~/scratch
143
144 # Call blastdbcmd to extract the sequence of nm_000122 from
145 # the installed database (refseq_rna.00) to a text file (test_query.txt)
146
147 > blastdbcmd -db refseq_rna.00 -entry nm_000122 -out test_query.fa
148
149 # and blast it against the database
150
151 blastn -query test_query.fa -db refseq_rna.00 -task blastn \
152 -dust no -outfmt "7 qseqid sseqid evalue bitscore" \
153 -max_target_seqs 2
154
155 # If everything is OK, you should see this:
156
157 # BLASTN 2.6.0+
158 # Query: NM_000122.1 Homo sapiens ERCC excision repair 3, TFIIH core complex
159 # Database: refseq_rna.00
160 # Fields: query id, subject id, evalue, bit score
161 # 2 hits found
162 NM_000122.1 gi|4557562|ref|NM_000122.1| 0.0 4962
163 NM_000122.1 gi|197101764|ref|NM_001132325.1| 0.0 4569
164 # BLAST processed 1 queries
165
166 #####
167 # Step 4: Create some "blastable" databases from fasta files
168 #####
169
170 #####
171 # make a temporary directory in your home directory and cd into it
172 #####
173
174 mkdir ~/scratch
175 cd ~/scratch
176
177 #####
178 # Download the most recent release of the Plasmodium 3D7
179 # annotated coding regions from PlasmoDB
180
```

```
181 #####
182
183 curl -O http://plasmodb.org/common/downloads/Current_Release/\
184 Pfalciparum3D7/fasta/data/PlasmoDB-34_Pfalciparum3D7_AnnotatedCDSs.fasta
185
186 #####
187 # Use 'makeblastdb' to format a blast database
188 #####
189
190     makeblastdb -in PlasmoDB-34_Pfalciparum3D7_AnnotatedCDSs.fasta -dbtype
191
192 # this formats the sequences in the fasta file and creates index files
193 # that speed up the alignment algorithms. If it all worked you will see this
194
195 Building a new DB, current time: 09/25/2017 13:04:54
196 New DB name: /users/fernando/scratch/PlasmoDB-34_Pfalciparum3D7_AnnotatedC
197 New DB title: PlasmoDB-34_Pfalciparum3D7_AnnotatedCDSs.fasta
198 Sequence type: Nucleotide
199 Keep MBits: T
200 Maximum file size: 1000000000B
201 Adding sequences from FASTA; added 5734 sequences in 0.353846 seconds.
202
203 # you can see the three database files:
204
205 > ls -l PlasmoDB-34_Pfalciparum3D7_AnnotatedCDSs.fasta.n*
206
207 # move the blast database files from your scratch directory into the blast d
208
209     mv PlasmoDB-34_Pfalciparum3D7_AnnotatedCDSs.fasta.n* $BLASTDB/.
210
211 # check that the files are in the right place
212
213     ls -ls $BLASTDB
214
215 #####
216 # clean up after yourself by deleting the scratch directory.
217 # warning: double check the line you type below. 'rm -r' is one of those com
218 # if you execute it in the wrong place, could delete valuable file!
219 #####
220
221     rm -r ~/scratch
222
223 #####
224 # Step 3: Do blast searches
225 #####
```

```
226
227 #####
228 # get some test sequences
229 #####
230
231     cp /users/sph140636/shared/code_examples/blast/example_queries.fasta .
232
233 #####
234 # test your installation
235 #####
236 blastn -query example_queries.fasta -db PlasmoDB-34_Pfalciparum3D7_Annotate
237 -task blastn -dust no -outfmt "7 qseqid sseqid evalue bitscore" \
238 -max_target_seqs 2
239
240
241 #####
242 # You can use NCBI Blast
243 #####
244
245 http://blast.ncbi.nlm.nih.gov/Blast.cgi
246
247 e.g. select 'nucleotide blast'
248 and try one of the following sequences
249
250 >NC_000521-644|NC_000521|1037656|1037762|3.27297995|0.34579439|-41.20
251 cactatatgtgtggtataaccacttgtgtggtatcactatatatgtgttattaccacgtatgtgtgttattacc
252 tgtgttgtataaccacatatgtgtg
253
254 >NC_000521-681|NC_000521|1035928|1036048|3.53349682|0.27272727|-35.74
255 agaatacgtattccttgaacataaaaatatataagagcatatcgtattctctcaacatatatacaagagcata
256 ctctgaacatatataacataaaggcatatcgtattct
257
258
```