How does BLAST work

- Find very short exact matches
- Extend the short exact matches to longer regions
- Perform optimal alignment on the extended regions
BLAST+ command line

Download BLAST+

There are two versions: blast+ (new blast) and legacy blast (old blast)
Tutorials don’t always label them properly.

How to tell old/new blasts apart

• New blast+ → uses programs such as makeblastdb, blastn, blastp and has search tasks such as megablast

• Old blast → uses programs such as formatdb, blastall, megablast and has search strategies such as blastn, blastp

(If you find the above confusing you are not alone)

Picking the right tool can be complicated

| Program Selection for Nucleotide Queries |
|-------------------------------|-----------------|-----------------|-----------------|-----------------|
| Length | DATABASES | PURPOSE | PROGRAM | EXPLANATION |
| 20 bp or longer | Nucleotide Search | Identify the query sequence | makeblastdb, blastn, blastp | Learn more |
| 1 - 20 bp | Nucleotide Search | Find sequences similar to query sequence | makeblastdb, blastn, blastp | Learn more |
| 20 bp or longer | Nucleotide Search | Find similar sequence from the Trace archive | tracecl, blastn, blastp | Learn more |
| 1 - 20 bp | Nucleotide Search | Find similar sequence to translated query in a translated database | tracecl, blastn, blastp | Learn more |
| 20 bp or longer | Protein Search | Find similar sequence to translated query in a protein database | megablast, blastp, blastn | Learn more |
| 1 - 20 bp | Protein Search | Find primer binding sites or map short contiguous motifs | megablast, blastn, blastp | Learn more |

INDEX OF /blast/executables/blast+/LATEST/

Get the binary for your system

Download and unpack – then move the entire blast+ directory to the src folder you’ve set up before.
The list of blast+ tools

- blast_formatter
deltblast
rpblastn
blastdb aliastool dustmasker segmasker
blastdbcheck legacy Blast.pl tblastn
blastdbcmd makeblastdb tblastx
blastn makebindex update_blastdb.pl
blatp makeprofiledb windowmasker
blastx psiblast
convert2blastmask rpblast

Running blast+

- We need a query sequence that we search with
- We need a subject/target database that we search in
- We need a search strategy \( \rightarrow \) how are we looking for it

Prepare a target/subject database

- Score: normalized score, expressed in bits, that is meant to estimate the magnitude of the search space
- EValue: the number of distinct alignments, with a score equivalent to or better than S, that are expected to occur by chance in the search database
Output formats can be changed

Exploring blast alignments

With an editor change your sequence and rerun the alignments.

Make changes: mismatches, deletions, insertions

Rerun alignments and see what changed.

Increase the edits and observe what happens.

Homework 11

• Pick a 5000bp region from one Ebola genome

• Align this region against all the 99 genomes and sort the alignments by the percentage of identity of the alignment.

• What are the three most dissimilar genomes that you find?

• What is the alignment length and number of mismatches for each of your three worst alignments?