The Biostar Handbook: A Beginner's Guide to Bioinformatics[4]

#course

blast

BLAST Glossary - BLAST® Help - NCBI Bookshelf
BLAST® Command Line Applications User Manual - NCBI Bookshelf
Linux系统中NCBI BLAST+本地化教程 · Wei Shen's Note

> BLAST stands for Basic Local Alignment Search Tool and is an algorithm for searching a sequence (called a query) against very large sequence databases (called a target).

> 1.Search may take place in nucleotide and/or protein space or translated spaces where nucleotides are translated into proteins.
> 2.Searches may implement search "strategies": optimizations to a certain task. Different search strategies will return different alignments.
> 3.Searches use alignments that rely on **scoring matrices**
> 4.Searches may be customized with many additional parameters. BLAST has many subtle functions that most users never need.

> **How to do blast?**

> 1.Prepare a **BLAST database** with makeblastdb. This only needs to be done once.
> 2.Pick a **blast tool**: blastn, blastp as appropriate (you may need to tune the parameters).
> 3.Run the tool and format **the output** as needed.

> "What are the blast tools?"

> **blastn, blastp, blastx, tblastn, tblastx**

Query sequence type	Database sequence type	Alignment level type	What the program should be called	What the program is actually called
nucleotide	nucleotide	nucleotide	**blastNN**	**blastn**
peptide	peptide	peptide	**blastPP**	**blastp**
nucleotide	peptide	peptide	**blastNP**	**blastx**
peptide	nucleotide	peptide	**blastPN**	**tblastn**
nucleotide	nucleotide	peptide	**blastNNP**	**tblastx**

About database?

makeblastdb creates blast databases.
blastdbcmd queries blast databases.
update_blastdb.pl updates prebuilt blast databases.

About -task?

blastn - finds more divergent sequences
megablast - finds less divergenent sequences
blastn-short - short queries

megablast is default

Expect value (E): a parameter that describes the number of hits one can expect to see by chance when searching a database of a particular size. It decreases exponentially as the score(S) of the match increases.

-evalue 科学计数法，比如说**1e3**，定义期望值阈值。**E**值表明在随机的情况下，其它序列与目标序列相似度要大于这条显示的序列的可能性。 与**S**值有关，**S**值表示两序列的同源性，分值越高表明它们之间相似的程度越大

blast