

BIOINFORMATICS INTERNSHIP

Session twenty two

FUN FACT

- Only four "learning" sessions left !!!
- Let's reflect on what we've learned
 - Add your thoughts in the shared google doc

BLAST

- Basic local alignment and search tool
- Query
- Database

TYPES OF BLASTS

- Nucleotide data?
 - BLAST_n
 - BLAST_x
 - tBLAST_x
- Protein data?
 - BLAST_p
 - tBLAST_n

HOW BLAST WORKS

1. Build a lookup table

Preprocess: Build a *lookup table* of size $|\Sigma|^w$ for all w -length words in D

$\Sigma = \{A, C, G, T\}$

$w = 2$

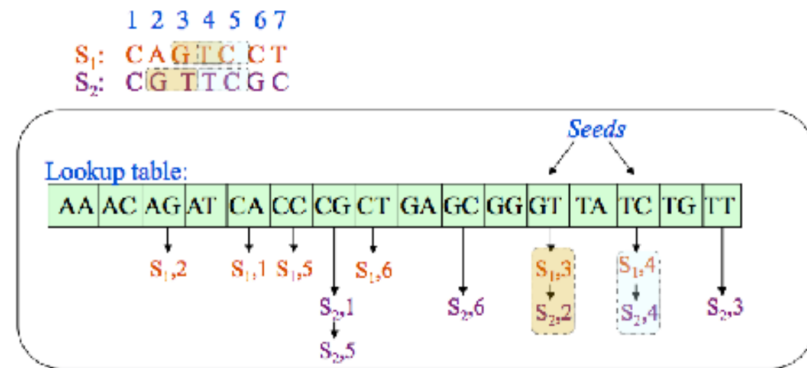
$\rightarrow 4^2 (=16)$ entries in lookup table

Lookup table:

AA	AC	AG	AT	CA	CC	CG	CT	GA	GC	GG	GT	TA	TC	TG	TT
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

CONT.

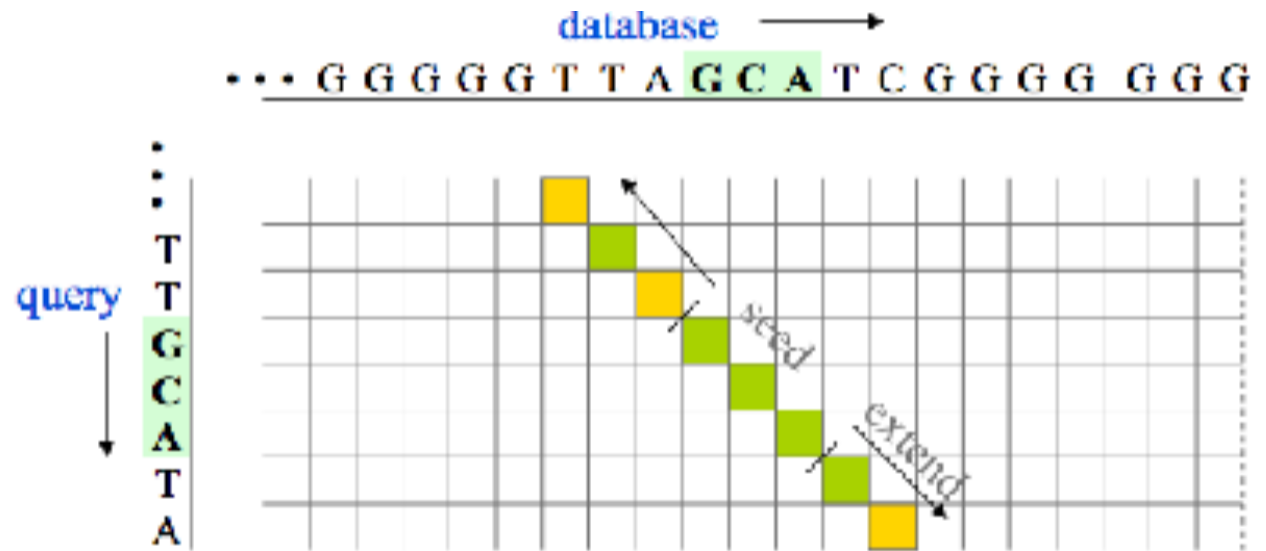
2. Filter low complexity and identify seeds



CONT.

3. Bidirectional extension (smith waterman algorithm)

Smith waterman- method of
determining regions of
sequence similarity



LETS DO IT!

- <https://blast.ncbi.nlm.nih.gov/Blast.cgi>