

Example

Linear gap model
Gap = -1
Match = 1
Mismatch = 0

-	A	G	G	T	C	A	G	T	A	T	A	
-	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
A	-1	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
T	-2	0	1	0	0	-1	-2	-3	-4	-5	-6	-7
G	-3	-1	1	2	1	0	-1	-1	-2	-3	-4	-5
A	-4	-2	0	1	2	1	1	0	-1	-1	-2	-3
G	-5	-3	-1	1	1	2	1	2	1	0	-1	-2
T	-6	-4	-2	0	2	1	2	1	3	2	1	0
A	-7	-5	-3	-1	1	2	2	2	2	4	3	2

41

Lecture outline

- Sequence alignment
 - Why do we need to align sequences?
 - Evolutionary relationships
- Gaps and scoring matrices
- Dynamic programming
 - Global alignment (Needleman & Wunsch)
 - Local alignment (Smith & Waterman)
- Database searches
 - BLAST
 - FASTA

53

BLAST

- Basic Local Alignment Search Tool
 - Altschul *et al.* 1990,1994,1997
- Heuristic method for local alignment
- Designed specifically for database searches
- Idea: good alignments contain short lengths of exact matches

54

Steps of BLAST

1. *Query words* of length 4 (for proteins) or 11 (for DNA) are created from query sequence using a sliding window

```
MEFPGLGSLGTSEPLPQFVDPALVSS
MEFP
EFPGL
FPGL
PGLG
GLGS
```

2. Filter low complexity regions (optional)

55

Steps of BLAST

3. Scan each database sequence for an exact match to *query words**. Each match is a *seed* for an ungapped alignment.

*Blast actually uses a list of *high scoring words* created from words *similar* to query words (High Scoring Pairs – HSPs).

56

Steps of BLAST

4. (*Original BLAST*) extend matching words to the left and right using ungapped alignments. Extension continues as long as score does not fall below a given threshold. This is a HSP (high scoring pair).

(*BLAST2*) Matches along the same diagonal (think dot plot) within a distance A of each other are joined and then the longer sequence is extended as before.

57

Steps of BLAST

5. Using a cutoff score S , keep only the extended matches that have a score at least S .
6. Determine statistical significance of each remaining match.
7.
 - (*Original BLAST*) Only ungapped alignments; sometimes combined together
 - (*BLAST2*) Extend the HSPs using gapped alignment

58

Summarizing BLAST

- One of the few algorithms to make it as a verb
 - Interesting article at:
<http://www.the-scientist.com/article/display/15682/>
- Extension is the most time-consuming step
- BLAST2 reported to be 3 times faster than the original version at the same quality

59

Example BLAST run

- BLAST website:
 - <http://www.ncbi.nlm.nih.gov/BLAST/>

60