

Sequence comparison

- Comparing two sequences
 - Global comparison
 - Local comparison
 - Semi-local alignment
 - Prefix-suffix or containment
 - Application: fragment assembly

Global comparison

- Sequences of nearly equal length
- Goal: to learn
 - Where they are alike
 - Where they differ

Alignments

- Add spaces to sequences
- Until sequences have the same length
- Place one over the other
- Make sure no two spaces align
- Compute alignment score

GCTAGCT
GAT - CAG



~~GCTAG - CT
GATC - - AG~~

Scoring system

- Match: usually a positive value
- Mismatch: usually a negative value
- Space: usually a negative value, worse than mismatch
- Example:
 - Match = +1
 - Mismatch = -1
 - Space = -2

Similarity

- Highest score among all alignments
- Notation:
-
- $\text{sim}(s,t)$

Algorithm

- Construct a matrix
- Vertical: s
- Horizontal: t
- Leave room for empty prefixes
- Meaning of $a[i,j]$: similarity of $s[1,i]$, $t[1,j]$
- Filling matrix: spreadsheet
- Retrieving optimal alignments
- $O(mn)$ where $m = |s|$ and $n = |t|$