Sequence comparison

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

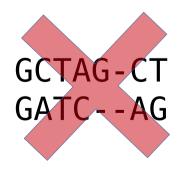
Global comparison

- ¹ Sequences of enarly equal length
- Goal: to learnWhere they are alike
 - ¹ Where they differ

Alignments

- Add spaces to sequences
- Until sequences have the same length
- ¹ Place one over the other
- Make sure no two spaces align
- Compute alignment score





Scoring system

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

Similarity

- Highest score among all alignments
 Notation:
- sim(s,t)

Algorithm

- Construct a matrix
- Vertical: s
- Horizontal: t
- ^{II} 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 O(mn) where m = |s| and n = |t|