

## Lab 2 - Alignments in R

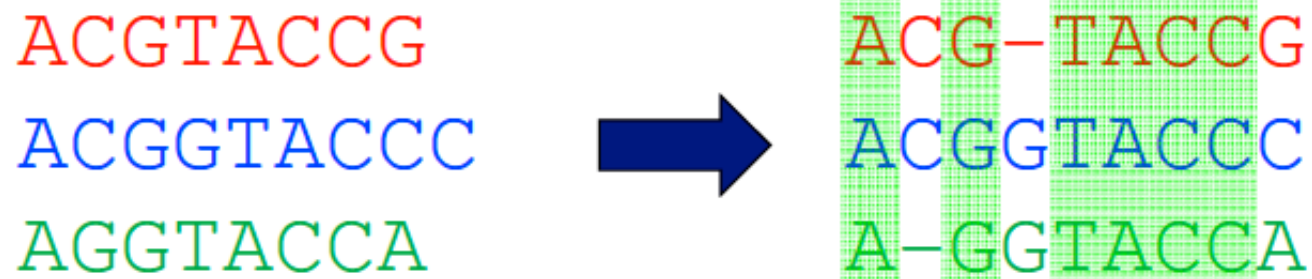
# Let's align things!

Open R-studio



# Align Multiple Sequences

- Want a global alignment over all the sequences
- Different scoring functions possible



# Sum of Pairwise Scores

ACGTACCG  
 ACGGTACCC  
 AGGTACCA

→

ACG-TACCG  
 ACGGTACCC  
 A-GGTACCA

ACG-TACCG  
 ACGGTACCC  
 A-GGTACCA  
 mxmgmmmmx

ACG-TACCG  
 ACGGTACCC  
 A-GGTACCA  
 mgmgmmmmx

ACG-TACCG  
 ACGGTACCC  
 A-GGTACCA  
 mgmxmmmmx

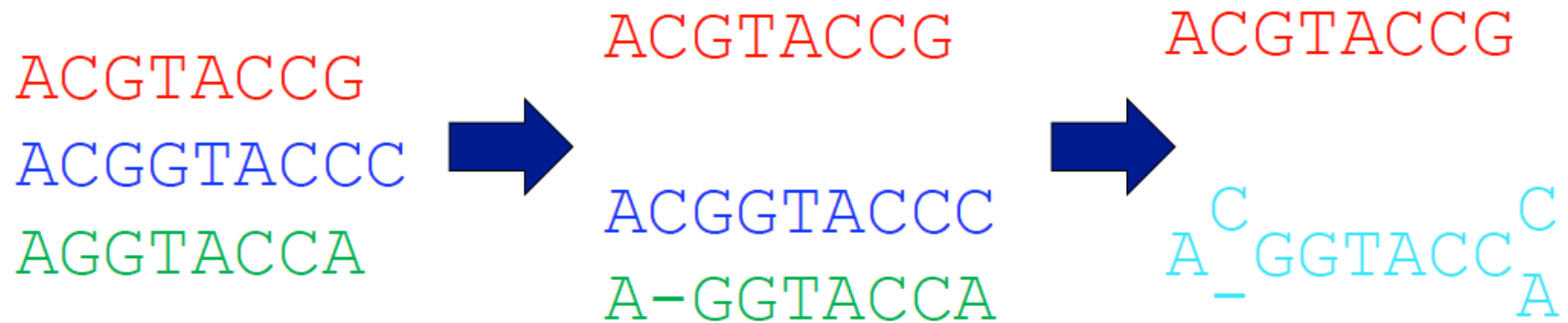
$$S = 18m + 5x + 4g = -12$$

match=+1, mismatch (x) = -2, gap = -5

# Algorithms for MSA

- *Computational intractable*
  - Time on the order of length of each sequence multiplied
  - For 100 sequence of 100 bp it would be  $100^{100}$
- Use heuristics to make more manageable
  - Not guaranteed to find highest scoring global alignments
  - Usually use progressive alignments

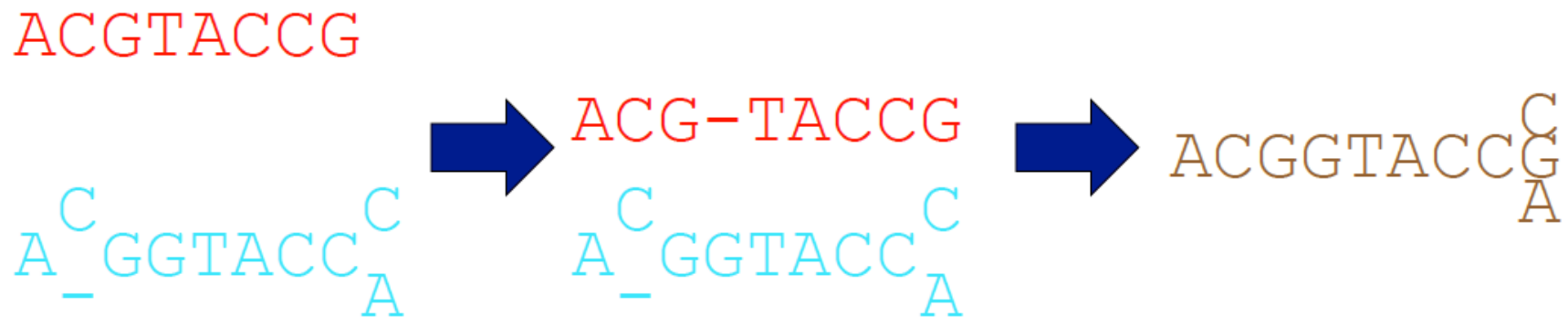
# Progressive Alignments



Find the best-scoring pair and align

Replace the pair with the *consensus sequence*

# Progressive Alignments



Find the best-scoring pair and align

Replace the pair with the *consensus sequence*



## Progressive Alignments

- Backtracking through the progressive alignment gives the MSA

ACG-TACCG  
ACGGTACCC  
A-GGTACCA

- We can also get a phylogenetic tree



Let's do some MSAs



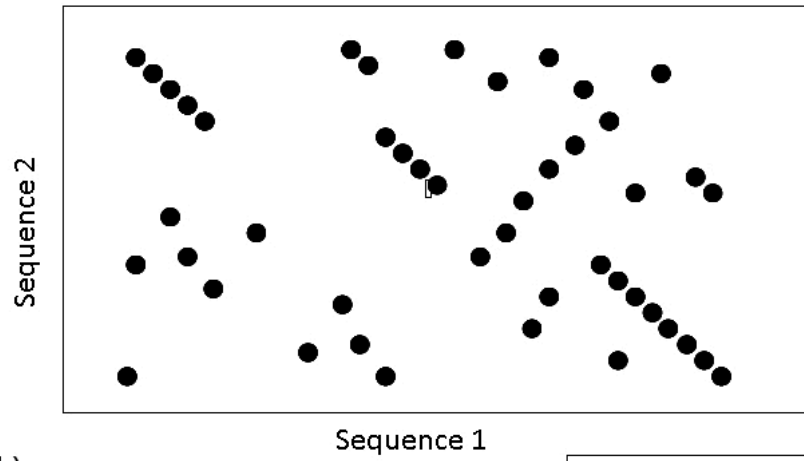
# Basic Local Alignment Search Tool (BLAST)

- Fast search through database of sequences
- Uses heuristics to make it fast
  - Not always “best” alignment

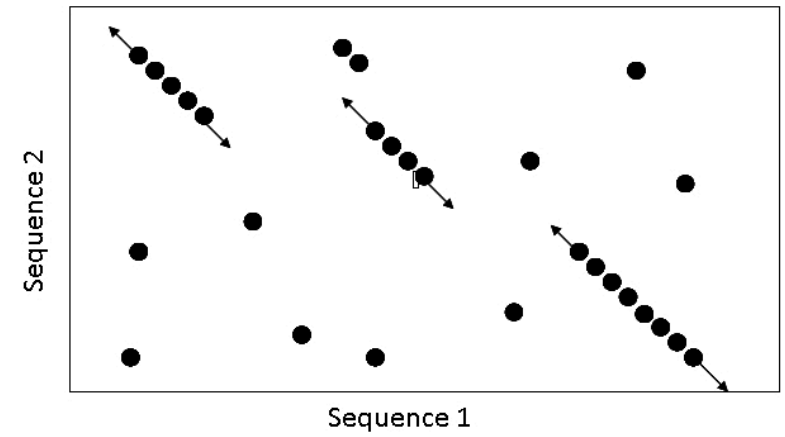
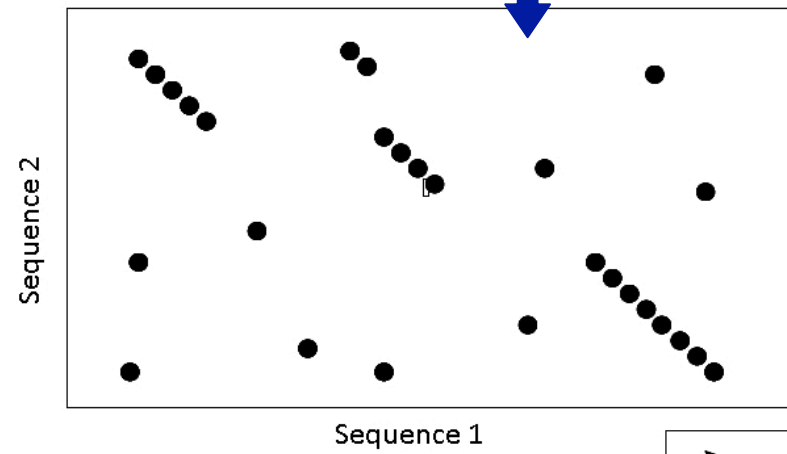
sequence      GCGTTGAGTGC  
                 ATGCGTTGAGT  
ATGCGTTGAGTGCGACGTAGCTAGCCGATCGATGCTAGCTAGCTAG  
                 TCGTTGAGTG  
                 CGTTGAGTGCG

1. Eleven letter “words” are made from the query sequence
2. List of neighborhood words are generated (small changes in sequence)
3. This list of actual and close words are compared to each sequence in data base looking for exact matches
4. When an exact match is found, it is extended both directions using a local alignment algorithm until the score reaches threshold
5. These local alignments are pooled and combined to make maximally extended alignments

a)



b)



Questions about alignments or BLAST?

