



# Sequence alignment

The star alignment approach to  
multiple alignment

TGCTG

# Outline

- Heuristic methods for multiple sequence alignment
- Star alignment

# The need for heuristic approaches to multiple sequence alignment

- Dynamic programming algorithm
  - exists but is not efficient/practical
  - $O(n^k 2^k)$  time complexity
    - ( $k$  = # sequences,  $n$  = length of sequences)
- Heuristic algorithms
  - Fast but not guaranteed to find optimum
  - Most rely on computing pairwise alignments

# Common heuristic approaches

- Star alignment
  - Align all sequences pairwise with a “central” sequence
  - Combine pairwise alignments into a multiple alignment
- Progressive alignment
  - Align sequences or alignments pairwise according to a tree structure
- Iterative refinement
  - Pull out a sequence from a multiple alignment and then realign it using a pairwise algorithm
  - Iterate until convergence

# Star Alignment Approach

- Given:  $k$  sequences to be aligned and a scoring scheme (i.e., substitution matrix and gap/space scores)

$$x^1, \dots, x^k$$

- Algorithm:
  - pick one sequence  $x^c$  as the “center”
  - for each  $x^i \neq x^c$  determine an optimal alignment between  $x^i$  and  $x^c$
  - Aggregate pairwise alignments and return multiple alignment

# Picking the center in star alignments

Two possible approaches:

1. try each sequence as the center, return the best multiple alignment
2. compute all pairwise alignments and select the string  $x^c$  that maximizes:

$$\sum_{i \neq c} \text{sim}(x^i, x^c)$$

# Aligning to an existing partial alignment

- Need to treat each “partial alignment” as a single entity
  - Partial alignment should not be changed other than gap insertions
- Shift entire columns when incorporating gaps

Next sequence to add to  
the multiple alignment

**TGTTAAC**

**TGT-AAC**

center sequence

**TGTTAAC**

**-TGT AAC**

**-TGT -AC**

**ATGT --C**

**ATGT GGC**

**-TGTTAAC**

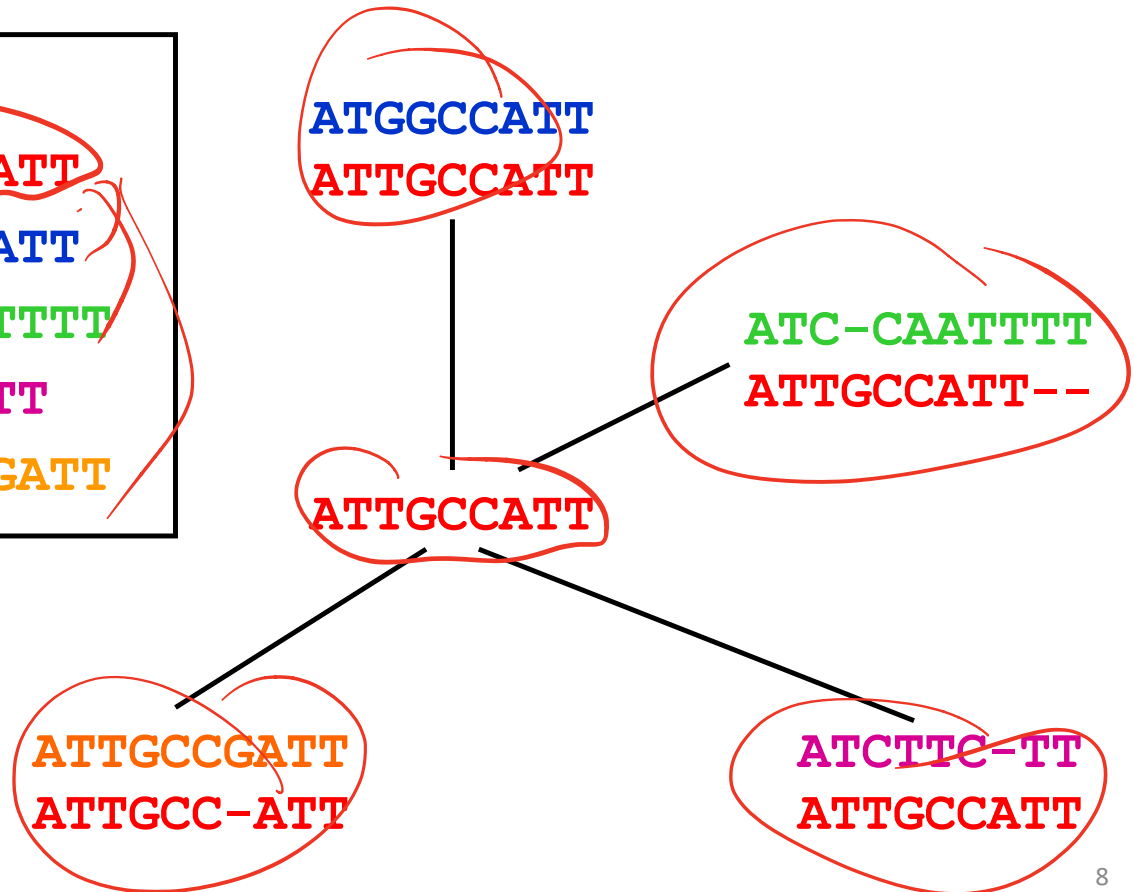
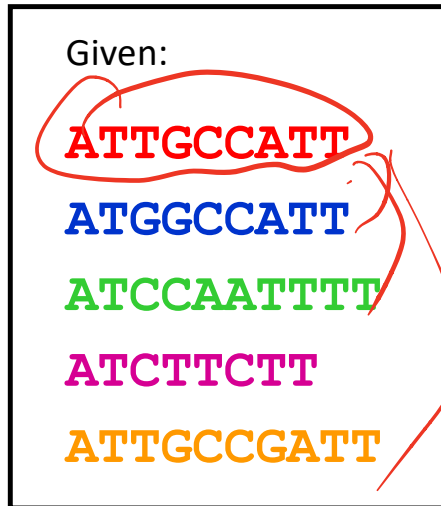
**-TGT-AAC**

**-TGT--AC**

**ATGT---C**

**ATGT-GGC**



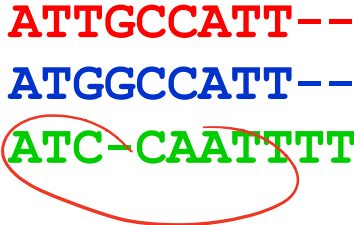
# Star Alignment Example





# Star Alignment Example

- Aggregate pairwise alignments

	present pair	Current multiple alignment
1.	 <p>ATGGCCATT ATTGCCATT</p>	<p>ATTGCCATT ATGGCCATT</p>
2.	 <p>ATC-CAATTTT ATTGCCATT--</p>	 <p>ATTGCCATT-- ATGGCCATT-- ATC-CAATTTT</p>

# Star Alignment Example

present pair

Current multiple alignment

3. **ATCTTC-TT**  
**ATTGCCATT**

**ATTGCCATT--**  
**ATGGCCATT--**  
**ATC-CAATTTT**  
**ATCTTC-TT--**

4. **ATTGCCGATT**  
**ATTGCC-ATT**

**ATTGCC- A TT--**  
**ATGGCC- A TT--**  
**ATC-CA- A TTTT**  
**ATCTTC- - TT--**  
**ATTGCCG A TT--**

shift entire columns  
when incorporating a gap



## Comments about Star alignment

- Conceptually simple
- Dependent only upon pairwise alignments
- Does not consider any position-specific information of the partial multiple sequence alignment while aligning a new sequence to it

# Summary

- Heuristic methods needed for practical multiple sequence alignment
- Star alignment is one of the simplest heuristic methods
  - builds a multiple alignment from a set of pairwise alignments to a single, center, sequence