

Feng-Doolittle

Motivation

- Needleman-Wunsch $n=N$ needs $O(n^N)$ space and runtime complexity!
- A more efficient algorithm for multiple sequence alignment would be better.

The algorithm

- Compute pairwise alignments of all sequences
- Use similarities instead of distances
- Build tree with these values —> UPGMA/WPGMA
- Use the tree to generate the alignments

How to align?

- Bottom-Up: Take the sequences which have the same parent node and replace every gap with X
- Possible cases:
 - Sequence with sequence: best possible alignment e.g. with Needleman-Wunsch
 - group with sequence: every sequences of the group with the sequence —> take best alignment
 - group with group: every sequences of the group0 with every sequence of group1 —> take best alignment

Example

- Sequence A with ACTG
- Sequence B with AT
- Sequence C with ACG
- score: 1 if match, -1 else

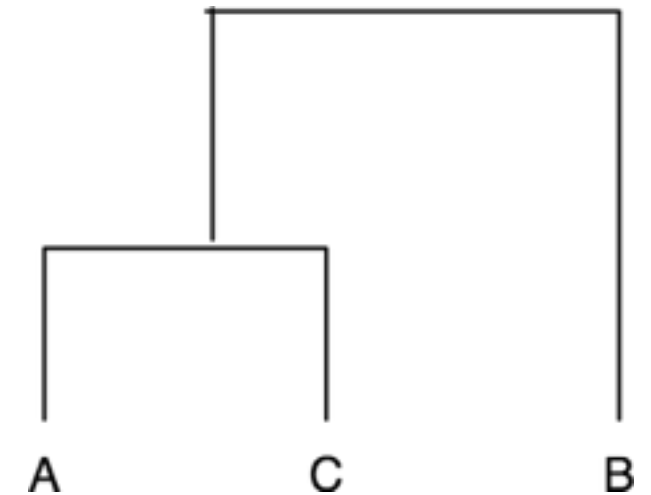
Example

- $s(A,B) = 0$

A	C	T	G
A	-	T	-

- $s(A,C) = 2$

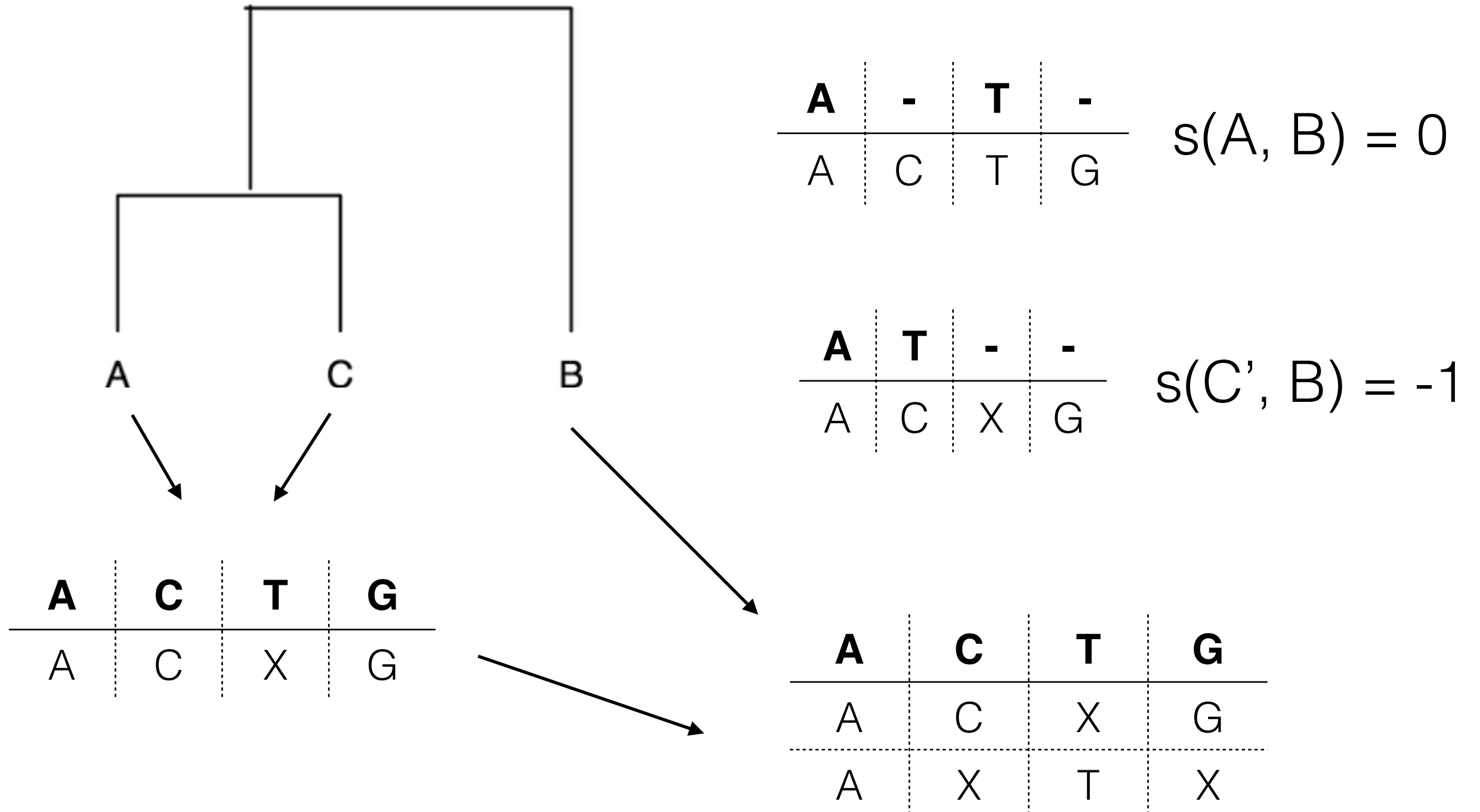
A	C	T	G
A	C	-	G



- $s(B,C) = -1$

A	T	-
A	C	G

Example



List of references

- Lecture “Multiples Sequence Alignment - Basics”
Bioinformatics I, Prof. Backofen

URL: http://www.bioinf.uni-freiburg.de//Lehre/Courses/2014_SS/V_Bioinformatik_1/multiple-alignment.pdf, visited: 13/11/2014