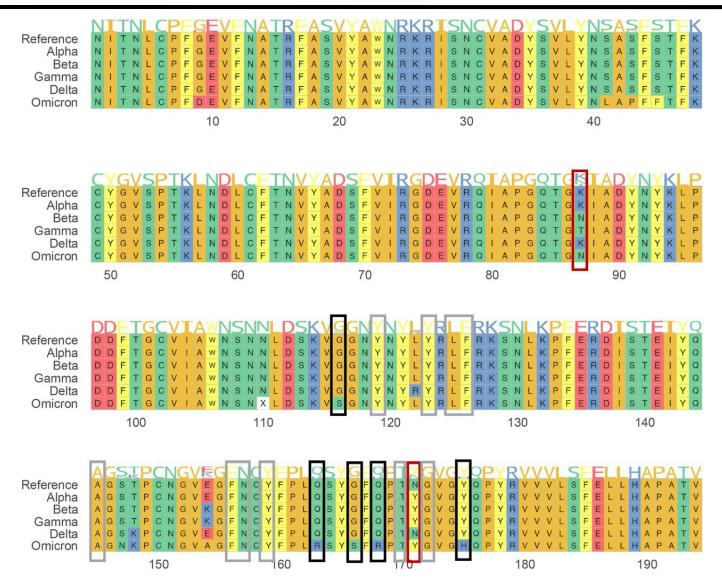
Multiple Sequence Alignment

Multiple Alignment Example



Contacting residues between SARS-CoV-2 and ACE2. Boxes denote the contacting residues. Black boxes denote mutations unique to omicron, red boxes denote mutations occurring in multiple variants, and grey boxes denote no mutations in any variant.

Image credit: Li et al. doi: https://doi.org/10.1101/202 1.12.08.471688

From pairwise to multiple

 A multiple sequence alignment of k sequences is an insertion of gaps in the positions of the sequences, just like a pairwise alignment.

```
H.s. Wee1 409-457
                       QVGRGLRYIMSMS-LVMMDIMPSMIFISRTSIPNAASEEGDEDDWASNK----
H.s. Ttk 614-659
S.c. Ste7 313-358
                       GVLNGLDHLYRQYKIIHRDIKPSNVLINSK----GQIKLC
S.c. Mkk1 332-376
                       AVLRGLSYLHEKK-VI
                                                                      FGVSGEAV----
S.p. Byr1 168-213
                       SMVKGLIYLYNVLHIIHRDLKPSNVVVNSR----GEIKLC
                                                                      FGVSGELV----
S.c. St20 722-767
                       ETLSGLEFLHSKG-VLHRDIKSDNILLSME
                                                                     FGFCAQINE---
S.c. Cc15 129-172
                       QTLLGLKYLHGEG-VIHRDIKAAN
                                                                      FGVSTIV----
S.p. Byr2 505-553
                       QTLKGLEYLHSRG-IVHRDIKGAN
                                                                      EGISKKLELNST
                       QILTAIKYIHSMG-ISHROLKPONILIEQDD--PVLVKIT
S.c. Spk1 302-348
S.p. Kin1 249-293
                       QIGSALSYLHQNS-VVHRDLKIENILISKT----GDIKII
                                                                      FGLSNLYR----
S.p. Cdr1 111-156
                       QILDAVAHCHRFR-FRHRDLKLENILIKVN---EQQIKIA
M.m. K6a1 507-556
                       TISKTVEYLHSQG-VVHRDLKPSNILYVDESGNPECLRIC
R.n. Kpbh 136-180
                       SLLEAVNELHVNN-IVHRDLKPENILLDDN----MQIRLS
H.s. Erk2 132-176
                       QILRGLKYIHSAM-VLHRDLKPSNLLLNTT---CLSCKIC
S.c. Kss1 137-182
@ 1999-2004 New Science Press
```

 "Two homologous sequences whisper, a multiple alignment shouts loudly" -- Arthur M. Lesk

Heuristic Algorithm for Multiple Alignment

- There is a simple algorithm that can merge pairwise alignments to a multiple alignment.
- The algorithm does not guarantee the optimality of the result. But runs relatively fast.

Merging Two Pairwise Alignments

- Suppose we have the following two pairwise alignments
- t: A-GAGC
- s1: ATGAGC
- and
- t: AGA-GC
- s2: AGTTGC
- Our main idea is to use the shared sequence t to construct a multiple alignment of s1, t, and s2.

Merging Two Pairwise Alignments

A-GA-GC A-GAGC t: A-GA-GC s1: ATGA-GC ATGAGC s1: ATGA-GC and and s2: A-GTTGC t: A-GA-GC AGA-GC A-GTTGC s2: s2: AGTTGC parrowise a hymners

induced by the multiple

ments, so that the

- Algorithm: Insert gaps to the two alignments, so that the superstrings for t become the same in the two alignments.
 Then put the two alignments together.
- Property: Maintains the pairwise alignment unchanged if ignoring the all-gap columns of the pairwise alignment.

Merging Two Pairwise Alignments

```
A-GA-GC
     A-GA-GC
 t:
                                                A-GA-GC
                     s1:
                          ATGA-GC
s1:
     ATGA-GC
                                               ATGA-GC \leftarrow
                                          s1:
                     s2: A-GTTGC
s2:
     A-GTTGC
                                          s2: A-GTTGC
                                          s3: A-TA--C \leftarrow
                          A-GA-GC
     AGA/GC
s3:
    ATAHC
                     s3: A-TA--C
```

- The process can be continued to merge two multiple alignments together as a bigger one.
- Note that the obtained multiple alignment may not be optimal.

Heuristic Algorithm for Multiple Alignment

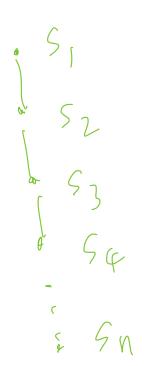
- Input: s₁,s₂,...,s_n.
- Algorithm:
 - Let A = pairwise alignment of s1 and s2.
 - For i from 3 to n.
 - Construct pairwise alignment P between s1 and si.
 - Let A = merge(A, P, s1), i.e., merging A and P using s1 as the template.
 - Return A.

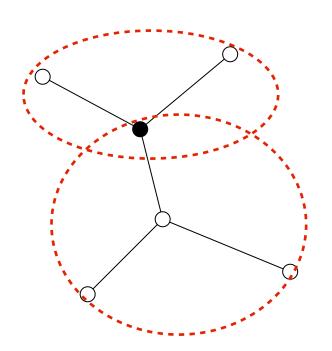
Progressive Alignment

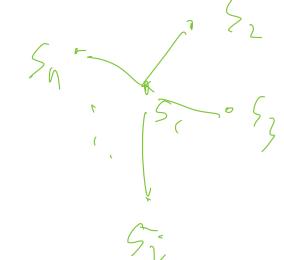
- The order of the merging is important to get good (but not optimal) multiple alignment.
- · We want to merge similar sequences first.

One way is to construct a minimum spanning tree, and then

merge using the shared vertices.





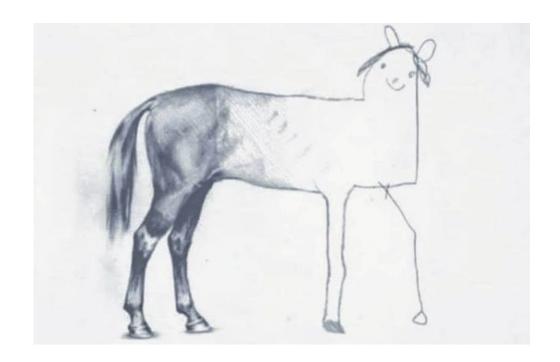


Heuristic Algorithm

- The exact algorithm for multiple alignment is super-polynomial. Before we study it, let's examine a heuristic algorithm.
- A heuristic algorithm is an algorithm that gives up quality for speed. It usually does not offer any performance guarantee in terms of quality.
- Well... We already sacrificed the quality because of a simple scoring function any way.
- If we cannot afford exponential time, the best we can ask for is a suboptimal solution.
- But practically, it might work sometimes.

Heuristic Algorithm

• We do not want to spend super-polynomial (e.g. exponential) time.



"You get what you pay for".

Exact Algorithm for Multiple Alignment

• When the optimal alignment is needed. There is an exact algorihtm as well.