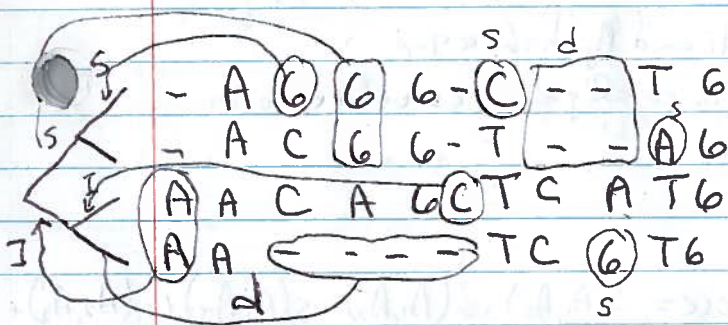
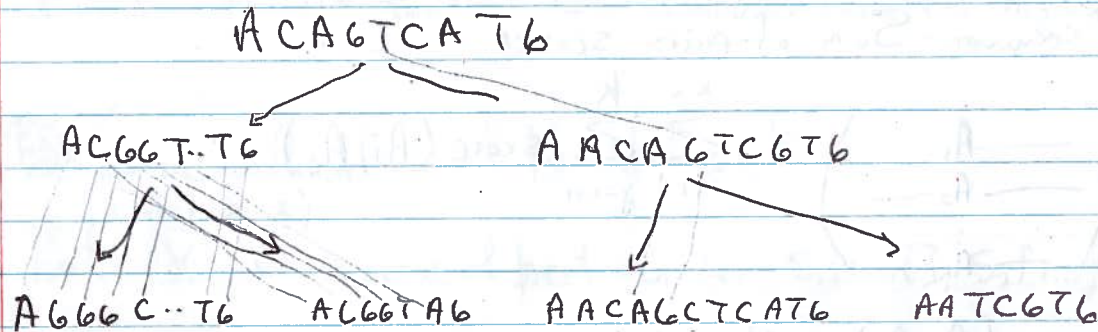


4522 Pontiac
→ 3452

Notes: Local alignment problem only makes sense
when $\sum M(a,b) < 0$
 $\begin{matrix} a & e & a & c & g \\ b & e & a & c & g \end{matrix}$

Multiple seq. alignment.



Biological problem: Given n sequences
Phylogenetic tree T , where each sequence
is assigned to a leaf

Find: Multiple alignment where all
nucleotides derived from the same
common ancestor are aligned to each other.

Scoring a given MSA

Goal: Find a ^{scoring} function score: MSA $\rightarrow \mathbb{R}$

such that $\text{score}(\text{MSA}) > \text{score}(\text{MSA}') \Rightarrow \text{MSA}$ is more likely to be biologically correct than

Many choices of scoring functions are possible

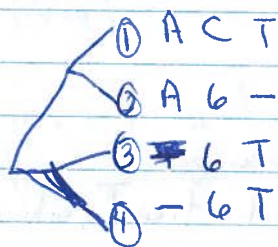
Most common: Sum-of-pairs score

$$\text{Score} \left(\begin{array}{c} \text{---} A_1 \text{---} \\ \text{---} A_2 \text{---} \\ \vdots \\ \text{---} A_k \text{---} \end{array} \right) = \sum_{i=1}^{k-1} \sum_{j=i+1}^k \text{score}(A_i, A_j)$$

where $\text{score}(A_i, A_j)$ is obtained by

- ① Removing positions where both A_i and A_j have a gap
- ② Scoring the remaining ~~MSA~~ pairwise alignment as before, with subst. Matrix M , gap penalty c .

Example:



$$\begin{aligned} \text{Score} &= s(A_1, A_2) + s(A_1, A_3) + s(A_1, A_4) + s(A_2, A_3) + s(A_2, A_4) + s(A_3, A_4) \\ &= -3 + -6 + -6 + -3 + -3 + 2 \\ &= -19 \end{aligned}$$

Optimal MSA problem

Given: Sequences $S_1 \dots S_n$
Subst. matrix M
Gap penalty c

Find: MSA for $S_1 \dots S_n$ that achieves the highest possible

~~Generalization~~ (Generalization of N-W algorithm)
(For $n=3$)

Let $X_{i,j,k}$ = score of best align for $S_1[1..i], S_2[1..j], S_3[1..k]$

$$X_{i,j,k} = \max \begin{cases} X_{i-1,j-1,k-1} + M[S_1[i], S_2[j]] + M[S_1[i], S_3[k]] + M[S_2[j], S_3[k]] \\ X_{i-1,j-1,k} + M[S_1[i], S_2[j]] + 2c \\ \vdots \\ X_{i,j,k-1} + 2c \end{cases}$$

For 3 seq. of length l

Dyn Prog table has $l \times l \times l$ entries

For n seq of length l : $O(l^n \cdot 2^n)$

↓
Exponential time algorithm

→ Heuristic, with no guarantee of optimality

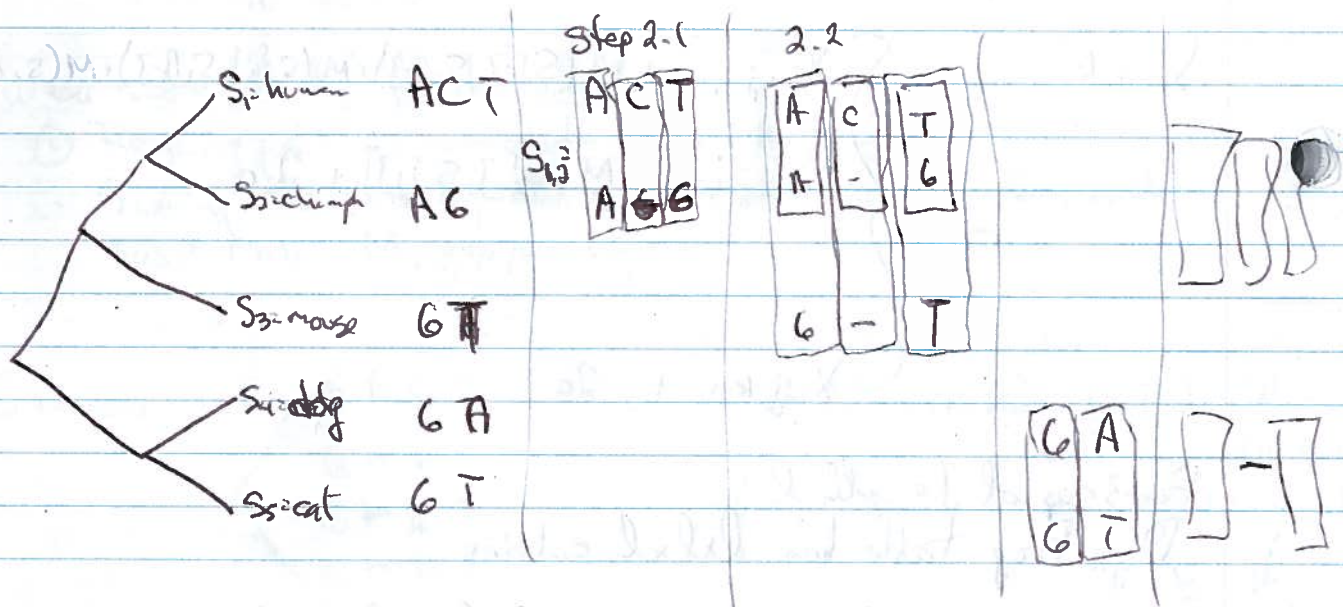
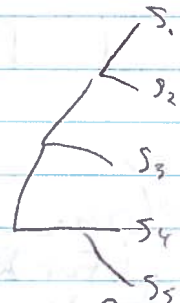
Progressive alignment Algo.

Input: $\begin{cases} S_1 \dots S_n \\ M \\ c \end{cases}$

① Guess phylogenetic tree for $S_1 \dots S_n$

② For each internal node in ~~pre~~ⁱⁿ order traversal
(from leaves to root)

~ Find optimal alignment btw pairs of sequences at two children



2.1. Align S_1 vs S_2 to obtain $S_{1,2}$

2.2 Align $S_{1,2}$ vs S_3 to obtain $S_{1,2,3}$

2.3 Align S_4 vs S_5 to obtain $S_{4,5}$