CS 68: BIOINFORMATICS

Prof. Sara Mathieson
Swarthmore College
Spring 2018

Outline: Feb 5

- Global sequence alignment (Needleman-Wunsch)
- Dynamic Programming table + traceback

Notes:

- Office hours today 3-5pm

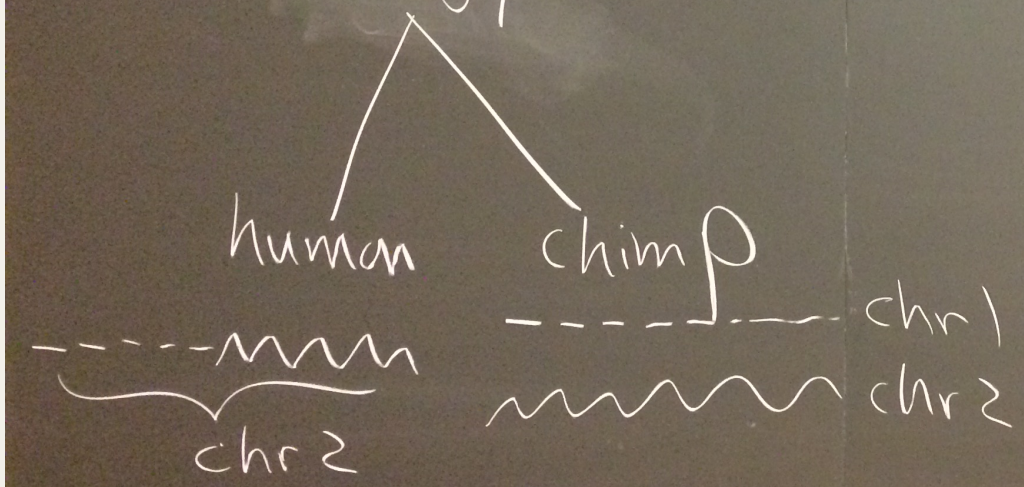
Sequence Alginment

Sequence Alignment

Why? understand evolution

↳ natural selection

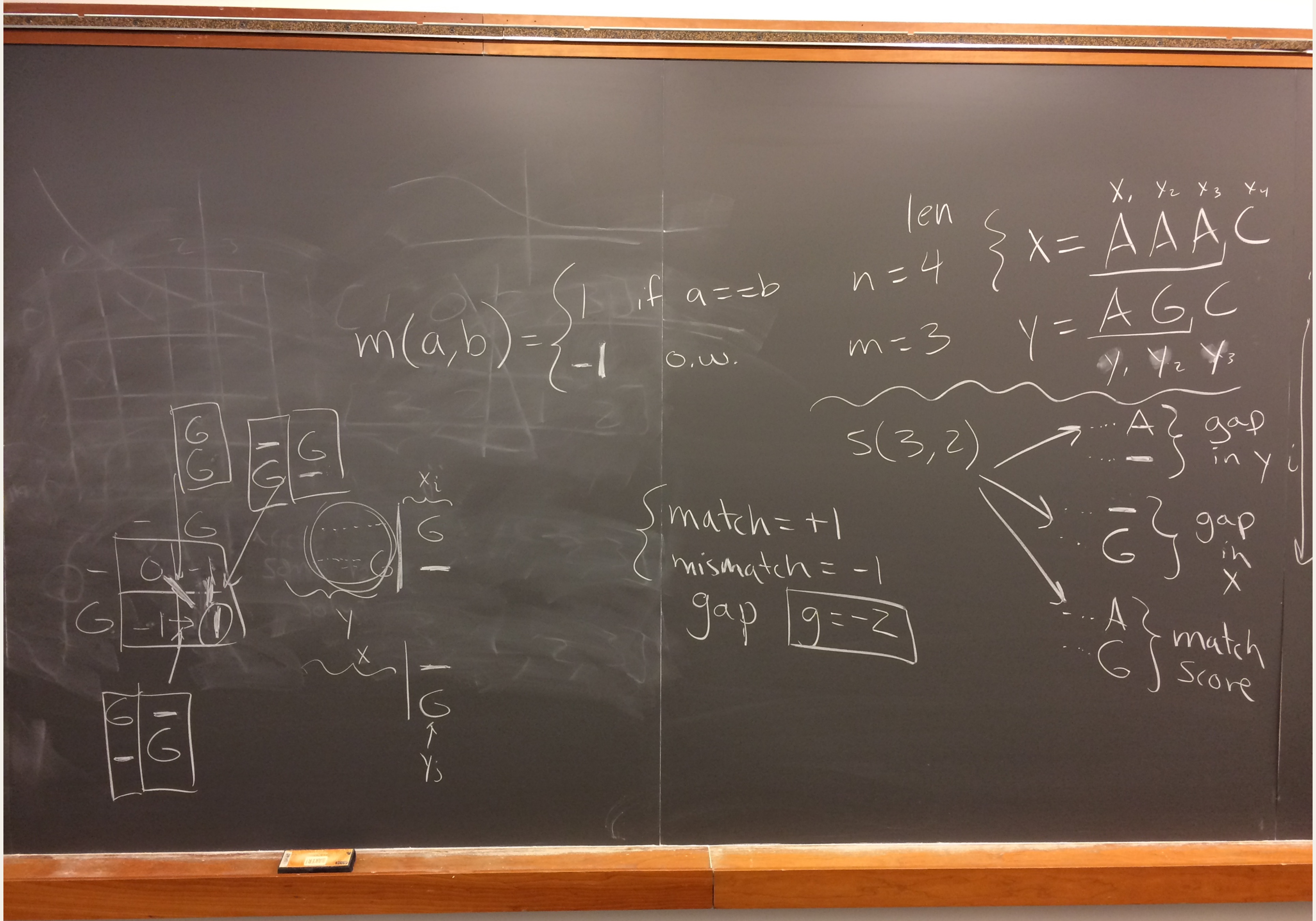
↳ homology



ways seq can change

1. substitution / mutation
2. insertion
3. deletions

} gaps



$$m(a,b) = \begin{cases} 1 & \text{if } a==b \\ -1 & \text{o.w.} \end{cases}$$

len

$$\left\{ \begin{array}{ll} n=4 & x = \underline{\overset{x_1}{A}\overset{y_2}{A}\overset{x_3}{A}}\,\overset{x_4}{C} \\ m=3 & y = \underline{\underset{y_1}{A}\underset{y_2}{G}}\,\underset{y_3}{C} \end{array} \right.$$

$$\left\{ \begin{array}{l} \text{match} = +1 \\ \text{mismatch} = -1 \end{array} \right.$$

gap $\boxed{g=-2}$

$$S(3,2) \begin{cases} \xrightarrow{} \begin{matrix} \dots A \\ \dots - \end{matrix} \Big\} \text{ gap in } y \\ \xrightarrow{} \begin{matrix} \dots - \\ \dots G \end{matrix} \Big\} \text{ gap in } x \\ \xrightarrow{} \begin{matrix} \dots A \\ \dots G \end{matrix} \Big\} \text{ match score} \end{cases}$$

DP table

	"–"	A	G	C
"–"	0	-2	-4	-6
A	-2	-1	-1	-3
A	-4	-1	0	-2
A	-6	-3	-2	-1
C	-8	-5	-4	-1

base case

base case

$$s(i,j) = \max \begin{cases} s(i-1,j) + g \\ s(i,j-1) + g \\ s(i-1,j-1) + m(x_i, y_j) \end{cases}$$

recursive

$$s(n,m) = s(x,y)$$

overall best alignment score

$$s(2,2) = \max \begin{cases} s(1, \\ s(2, \\ s(1, \end{cases}$$

$s(i-1, j) + g$

$s(i, j-1) + g$

$s(i-1, j-1) + m(x_i, y_j)$

$s(x, y)$

$$s(2,2) = \max \begin{cases} s(1,2) + g = -1 - 2 = -3 \\ s(2,1) + g = -1 - 2 = -3 \\ s(1,1) + m(A, G) = 1 + (-1) = 0 \end{cases}$$

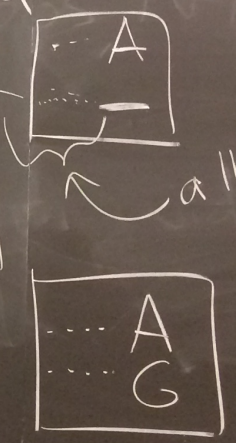
align x_i with gap

align y_j with gap

<u>Interpretation</u>

* Score is positive:
  → seqs are related

* Score is negative
  → seqs are unrelated



all of y

find alignment

- start at $s(n,m)$
- go back to cell(s) that produced best alignment

ll of
Y

```
AAAC
-AGC
```

```
AAAC
AG-C
```

```
AAAC
A-GC
```