

Bioinformatics

LAB 2

Prof.ssa Elisa Ficarra

Prof.ssa Santa Di Cataldo

Eng. Marta Lovino

Eng. Francesco Ponzio

Eng. Emanuele Parisi

Politecnico di Torino

DAUIN

Dept. of Control and Computer Engineering





Organization

Schedule

This week

LAB2 & LAB3

Wednesday, May 27th 11.30 – 13.00

Thursday, May 28th 11.30 – 13.00 + 13.00 – 14.30 (additional Q&A session)

Projects

Friday, May 29th 11.30 – 13.00 (genomics) + 13.00 – 14.30 (bioimaging)

Next week

LAB4 & LAB5

Wednesday, June 3rd 11.30 – 13.00 + 13.00 – 14.30 (additional Q&A session)

Thursday, June 4th 11.30 – 13.00 + 13.00 – 14.30 (additional Q&A session)

Please check the [Teaching Portal](#) and the [Telegram group](#) to be updated.



The background features a light blue DNA double helix on the left side. Scattered across the teal and white background are several chemical structures: a large polycyclic aromatic hydrocarbon (PAH) in the center-left, a benzene ring with a substituent at the bottom right, and several smaller molecules like water (H₂O) and diatomic molecules (O₂) in the upper right.

LAB 2 - Goals

GOALS

- Absolute and relative UNIX paths
- Basic bash commands
- Running a bash script
- Understand global and local alignment
- Use bwa aligner



Global alignment

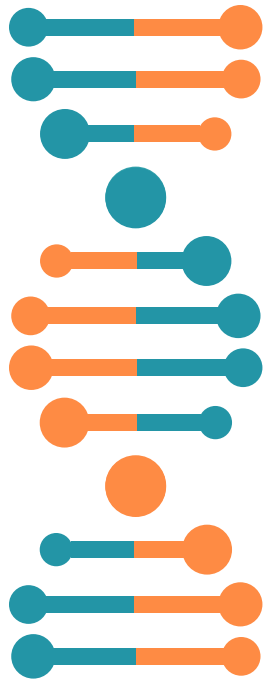
Definitions:

| | |
|-------------|---|
| $s(a, b)$ | the substitution score for aligning letters a and b |
| g | the gap score for aligning any letter to a null |
| X_i | the partial sequence consisting of the first i letters of $X \equiv x_1x_2 \dots x_m$ |
| Y_j | the partial sequence consisting of the first j letters of $Y \equiv y_1y_2 \dots y_n$ |
| $SIM(i, j)$ | the similarity of X_i and Y_j |

Consider the *last column* of an optimal alignment of X_i and Y_j . This column either aligns x_i to y_j , or x_i to a null, or y_j to a null. Because we do not allow “crossing”, there are no other possibilities. This observation yields the following recurrence:

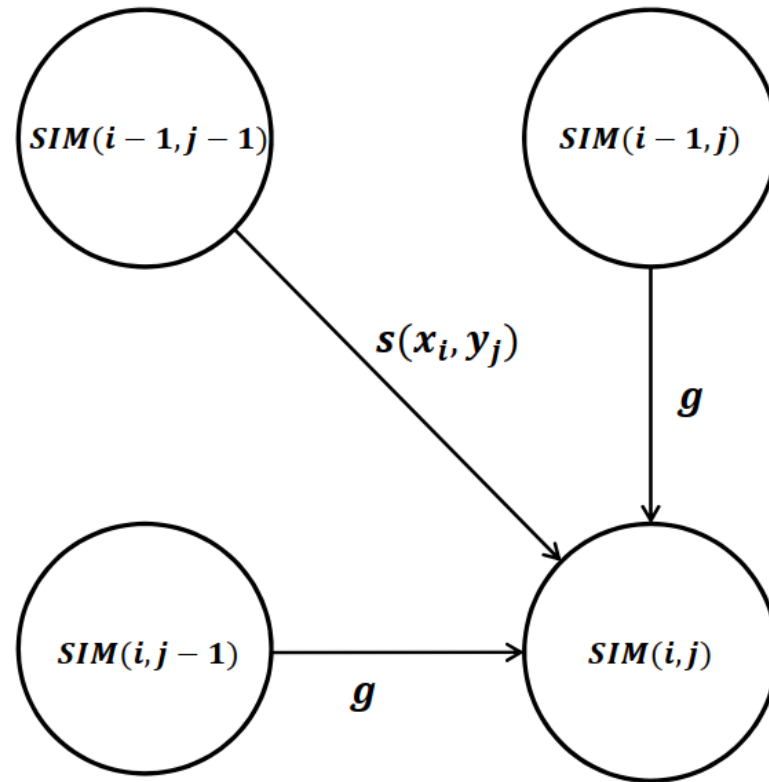
$$SIM(i, j) = \max \begin{cases} SIM(i-1, j-1) + s(x_i, y_j) & x_i \text{ and } y_j \text{ aligned} \\ SIM(i-1, j) + g & x_i \text{ aligned with a null} \\ SIM(i, j-1) + g & y_j \text{ aligned with a null} \end{cases}$$

In brief, we can solve for $SIM(m, n)$ by solving smaller versions of the problem first.

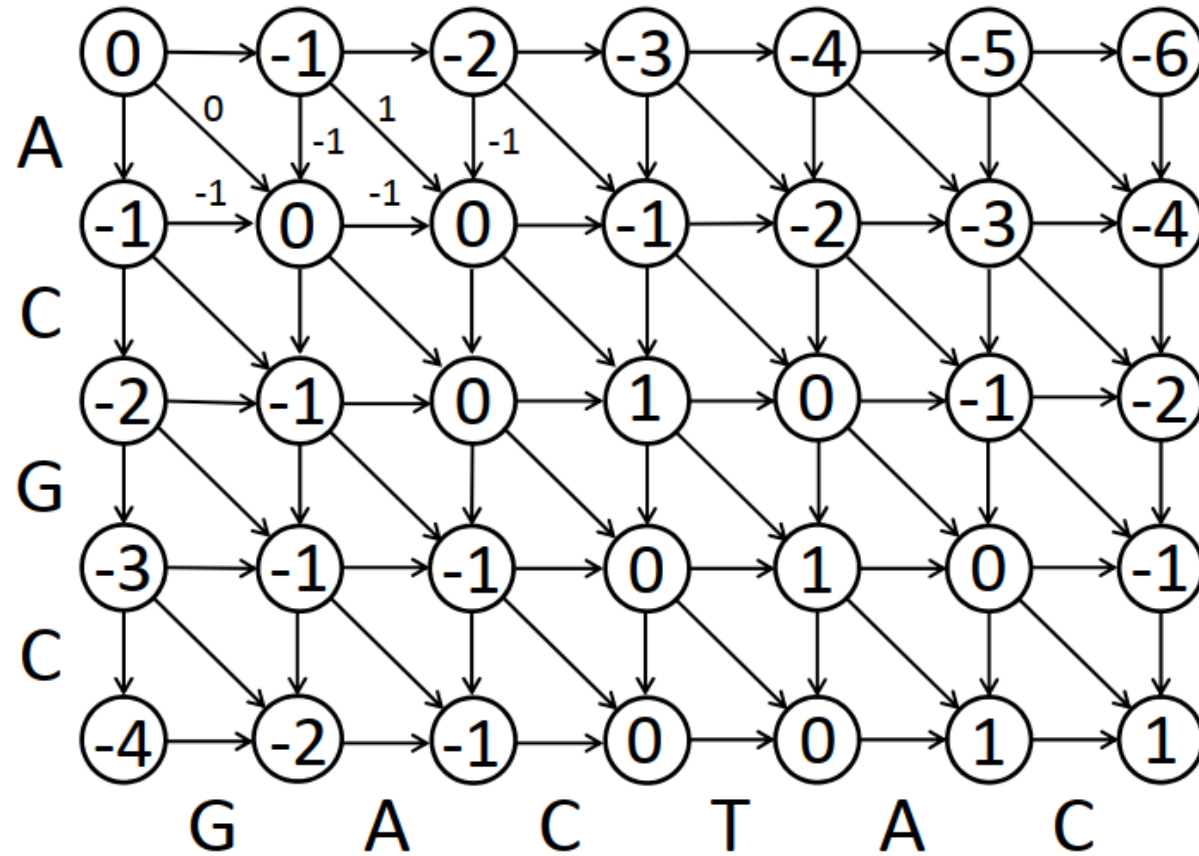


Global alignment

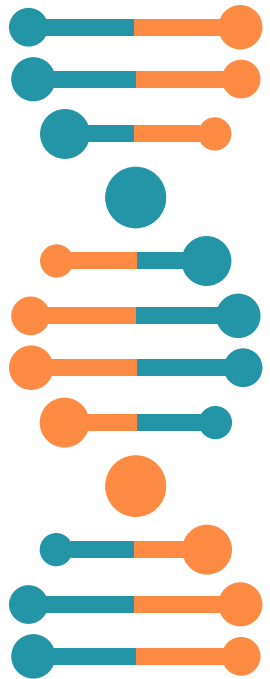
One may associate a partial similarity with each node of a path graph. If the values of $SIM(i-1, j-1)$, $SIM(i-1, j)$ and $SIM(i, j-1)$ are known, the value of $SIM(i, j)$ may be calculated.



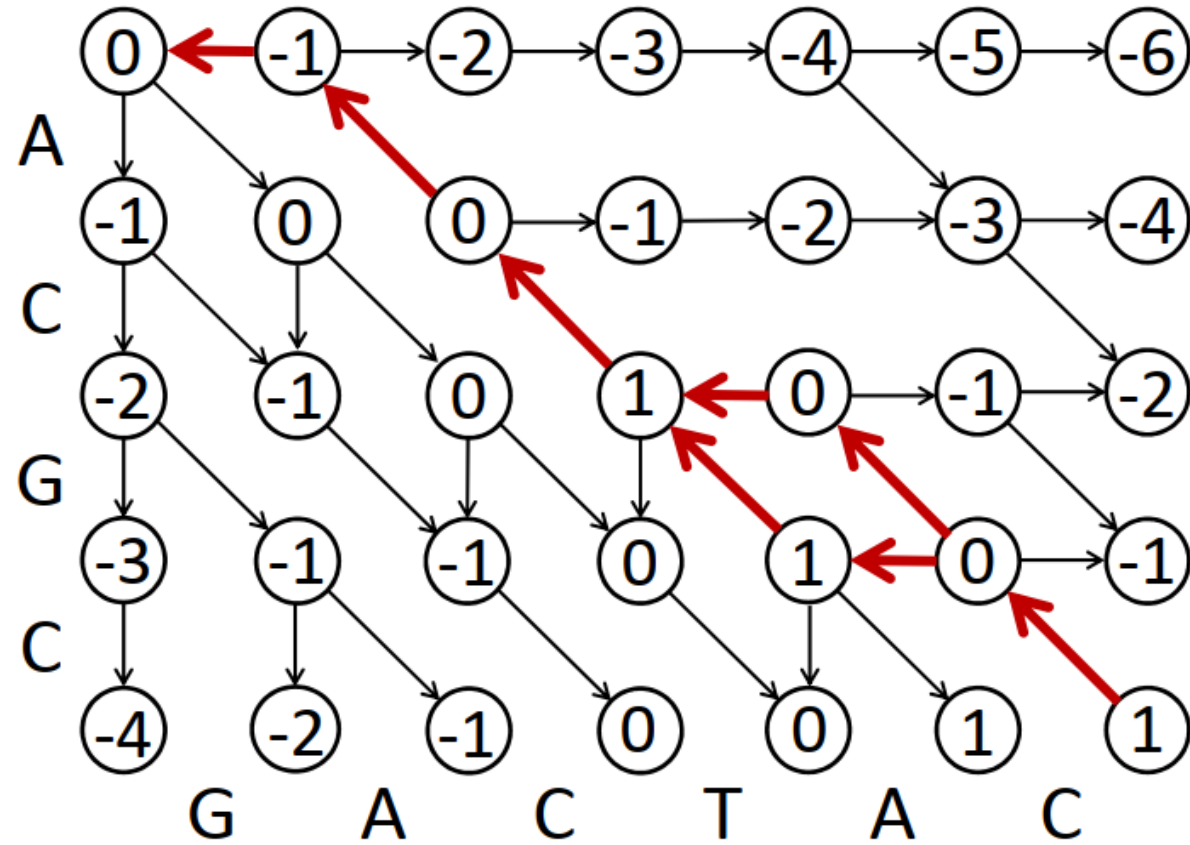
Global alignment



Scores: Match +1 Mismatch 0 Gap -1



Global alignment

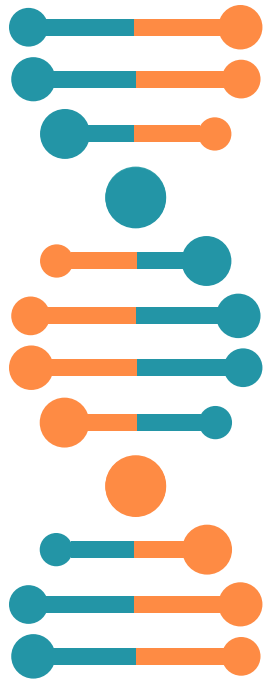


Optimal alignments:

-ACG-C
GACTAC

and

-AC-GC
GACTAC



Global alignment

Scores:

Match +1

Mismatch 0

Gap -1

| | | G | A | C | T | A | C |
|---|--|---|---|---|---|---|---|
| | | | | | | | |
| A | | | | | | | |
| C | | | | | | | |
| G | | | | | | | |
| C | | | | | | | |

Global alignment

Scores:

Match +1

Mismatch 0

Gap -1

| | | G | A | C | T | A | C |
|---|---|---|---|---|---|---|---|
| | 0 $\xrightarrow{-1}$ -1 $\xrightarrow{-1}$ -2 $\xrightarrow{-1}$ -3 $\xrightarrow{-1}$ -4 $\xrightarrow{-1}$ -5 $\xrightarrow{-1}$ -6 | | | | | | |
| A | -1 $\downarrow -1$ | | | | | | |
| C | -2 $\downarrow -1$ | | | | | | |
| G | -3 $\downarrow -1$ | | | | | | |
| C | -4 $\downarrow -1$ | | | | | | |

Global alignment

Scores:

Match +1

Mismatch 0

Gap -1

| | | G | A | C | T | A | C |
|---|----|----|----|----|----|----|----|
| | 0 | -1 | -2 | -3 | -4 | -5 | -6 |
| A | -1 | | | | | | |
| C | -2 | | | | | | |
| G | -3 | | | | | | |
| C | -4 | | | | | | |

Global alignment

Scores:

Match +1

Mismatch 0

Gap -1

| | | G | A | C | T | A | C |
|---|----|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | 0 | $\xrightarrow{-1}$ -1 | $\xrightarrow{-1}$ -2 | $\xrightarrow{-1}$ -3 | $\xrightarrow{-1}$ -4 | $\xrightarrow{-1}$ -5 | $\xrightarrow{-1}$ -6 |
| A | -1 | $\xrightarrow{-1}$ 0 | | | | | |
| C | -2 | | | | | | |
| G | -3 | | | | | | |
| C | -4 | | | | | | |

Global alignment

Scores:

Match +1

Mismatch 0

Gap -1

| | | G | A | C | T | A | C | |
|---|----|----|----|----|----|----|----|----|
| | 0 | -1 | -1 | -2 | -3 | -4 | -5 | -6 |
| A | -1 | 0 | -1 | | | | | |
| C | -2 | | | | | | | |
| G | -3 | | | | | | | |
| C | -4 | | | | | | | |

Global alignment

Scores:

Match +1

Mismatch 0

Gap -1

| | | G | A | C | T | A | C |
|---|----|----|----|----|----|----|----|
| | 0 | -1 | -2 | -3 | -4 | -5 | -6 |
| A | -1 | 0 | 0 | | | | |
| C | -2 | | | | | | |
| G | -3 | | | | | | |
| C | -4 | | | | | | |

Global alignment

Scores:

Match +1

Mismatch 0

Gap -1

| | | G | A | C | T | A | C | |
|---|----|----|----|----|----|----|----|----|
| | 0 | -1 | -1 | -2 | -3 | -4 | -5 | -6 |
| A | -1 | 0 | 0 | -1 | -2 | | | |
| C | -2 | | | | | | | |
| G | -3 | | | | | | | |
| C | -4 | | | | | | | |

Global alignment

Scores:

Match +1

Mismatch 0

Gap -1

| | | G | A | C | T | A | C | |
|---|----|----|----|----|----|----|----|----|
| | 0 | -1 | -1 | -2 | -3 | -4 | -5 | -6 |
| A | -1 | 0 | 0 | -1 | -2 | -3 | | |
| C | -2 | | | | | | | |
| G | -3 | | | | | | | |
| C | -4 | | | | | | | |

Global alignment

Scores:

Match +1

Mismatch 0

Gap -1

| | | G | A | C | T | A | C | |
|---|----|----|----|----|----|----|----|----|
| | 0 | -1 | -1 | -2 | -3 | -4 | -5 | -6 |
| A | -1 | 0 | 0 | -1 | -2 | -3 | -4 | |
| C | -2 | -1 | 0 | 1 | 0 | -1 | -2 | |
| G | -3 | -1 | -1 | 0 | 1 | 0 | -1 | |
| C | -4 | -2 | -1 | 0 | 0 | 1 | 1 | |

Global alignment

Scores:

Match +1

Mismatch 0

Gap -1

| | | G | A | C | T | A | C |
|---|----|----|----|----|----|----|----|
| | 0 | -1 | -2 | -3 | -4 | -5 | -6 |
| A | -1 | 0 | 0 | -1 | -2 | -3 | -4 |
| C | -2 | -1 | 0 | 1 | 0 | -1 | -2 |
| G | -3 | -1 | -1 | 0 | 1 | 0 | -1 |
| C | -4 | -2 | -1 | 0 | 0 | 1 | 1 |

C
C

Global alignment

Scores:

Match +1

Mismatch 0

Gap -1

| | | G | A | C | T | A | C |
|---|----|----|----|----|----|----|----|
| | 0 | -1 | -2 | -3 | -4 | -5 | -6 |
| A | -1 | 0 | 0 | -1 | -2 | -3 | -4 |
| C | -2 | -1 | 0 | 1 | 0 | -1 | -2 |
| G | -3 | -1 | -1 | 0 | 1 | 0 | -1 |
| C | -4 | -2 | -1 | 0 | 0 | 1 | 1 |

GC

AC

-C

AC

Global alignment

Scores:

Match +1

Mismatch 0

Gap -1

| | | G | A | C | T | A | C |
|---|----|----|----|----|----|----|----|
| | 0 | -1 | -2 | -3 | -4 | -5 | -6 |
| A | -1 | 0 | 0 | -1 | -2 | -3 | -4 |
| C | -2 | -1 | 0 | 1 | 0 | -1 | -2 |
| G | -3 | -1 | -1 | 0 | 1 | 0 | -1 |
| C | -4 | -2 | -1 | 0 | 0 | 1 | 1 |

-GC
TAC

G-C
TAC

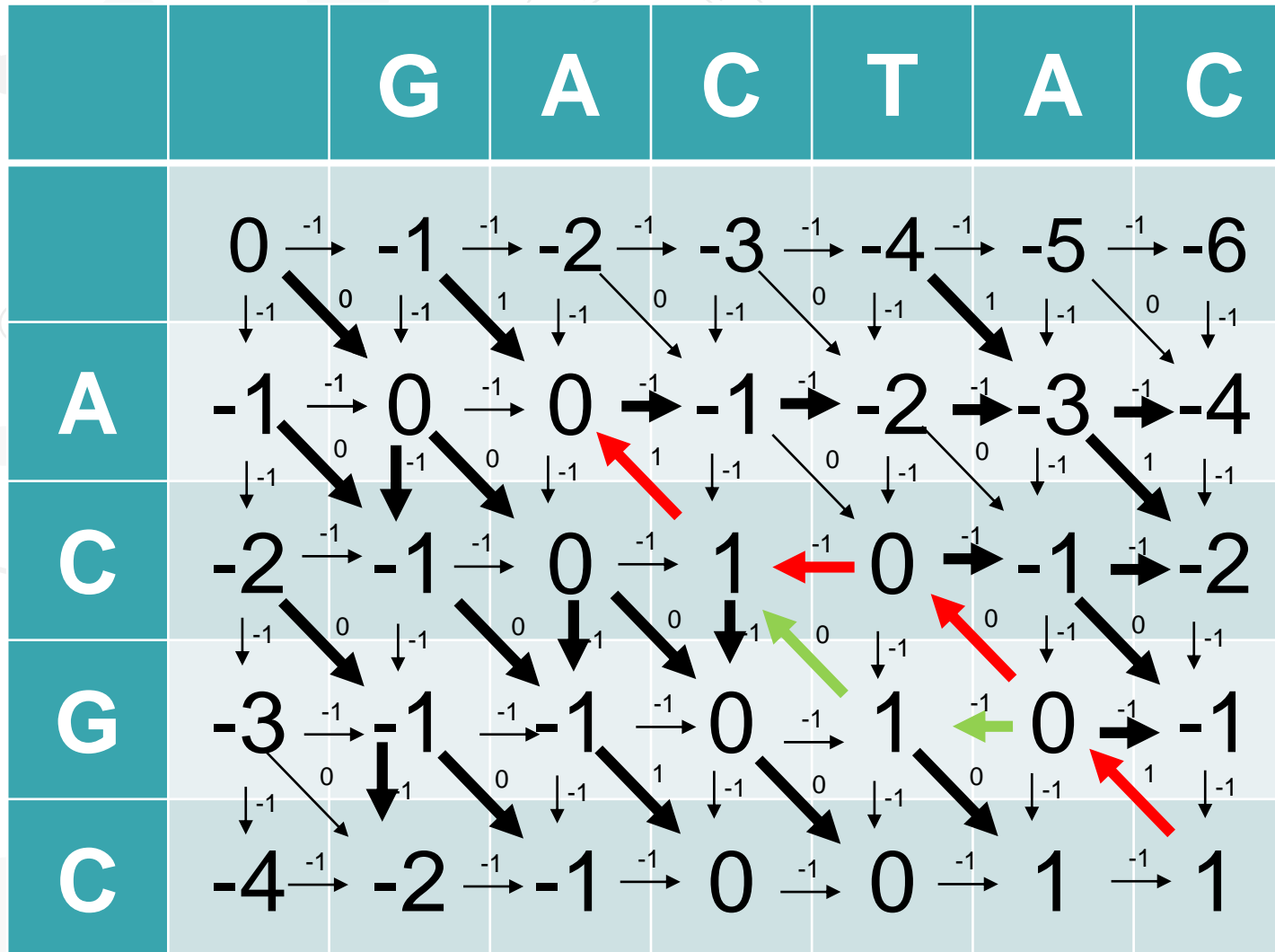
Global alignment

Scores:

Match +1

Mismatch 0

Gap -1



C-GC

CTAC

CG-C

CTAC

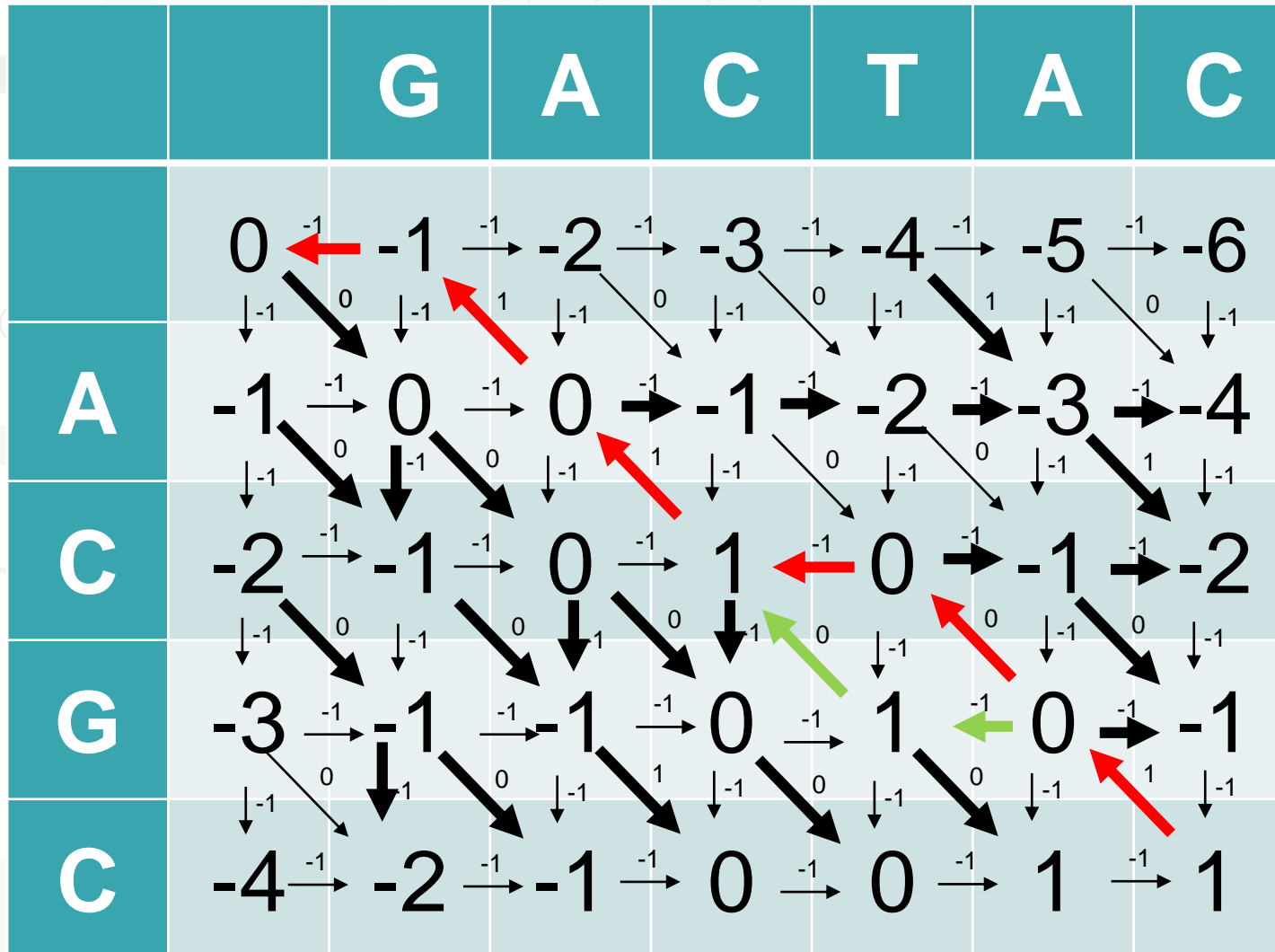
Global alignment

Scores:

Match +1

Mismatch 0

Gap -1



-AC-GC

GACTAC

-ACG-C

GACTAC

Global alignment

Similarity(X,Y):

For $i = 0, \dots, m$: $SIM[i,0] = i * g$

For $j = 1, \dots, n$: $SIM[0,j] = j * g$

For $i = 1, \dots, m$:

For $j = 1, \dots, n$:

$SIM[i,j] = \max(\$
 $SIM[i-1,j-1] + s(X[i],Y[j]),$
 $SIM[i-1,j] + g,$
 $SIM[i,j-1] + g$
 $)$

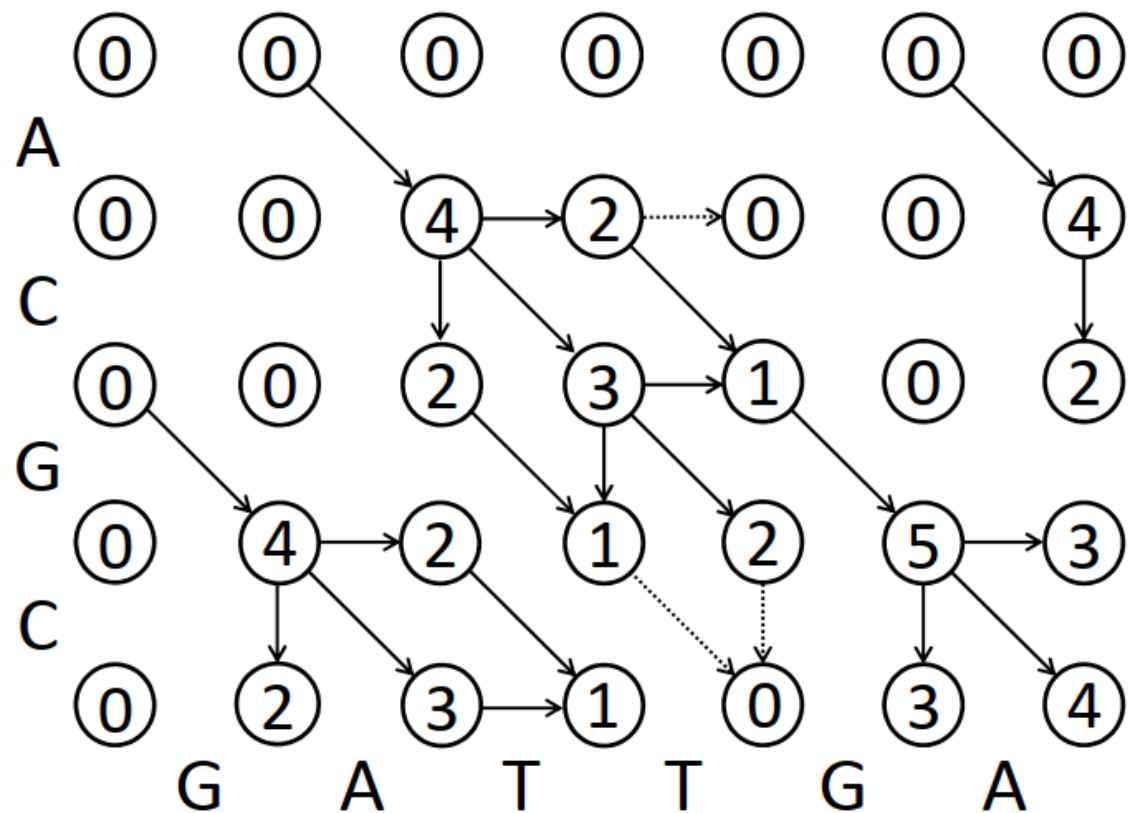
EndFor

EndFor

Return $SIM[m,n]$



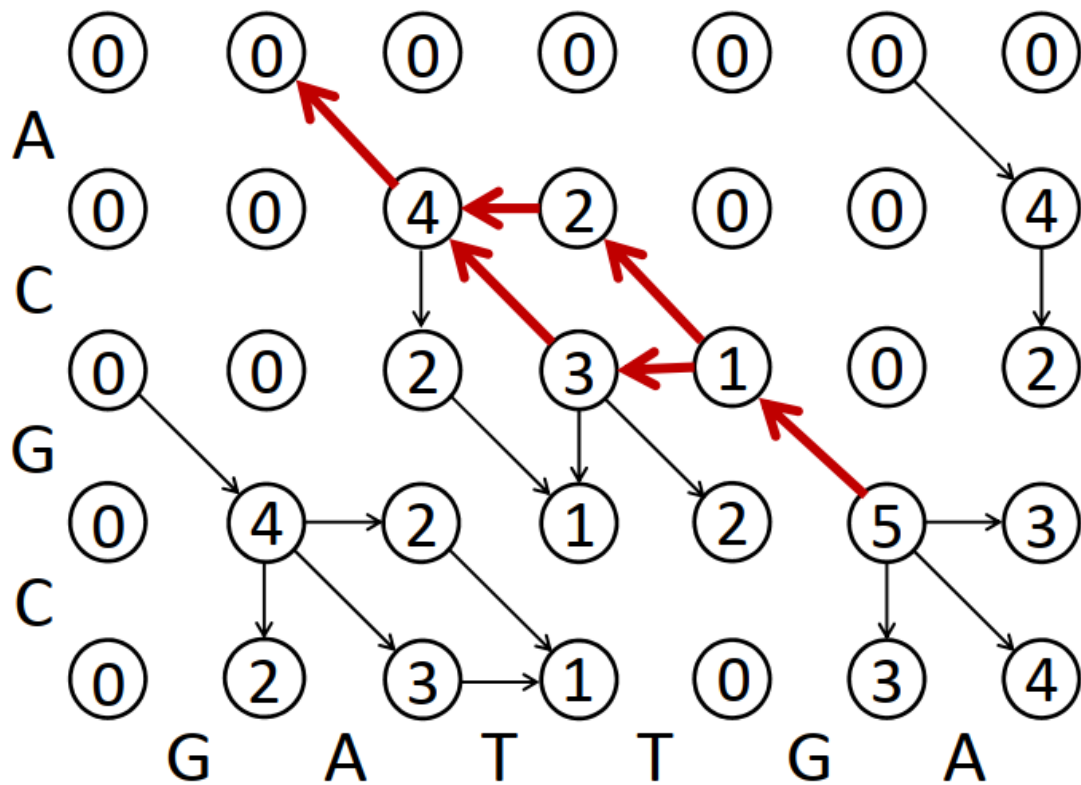
Local alignment



Scores: Match +4 Mismatch -1 Gap -2



Local alignment



Scores: Match +4 Mismatch -1 Gap -2



Local alignment

Local_Similarity(X,Y):

 $S=0$

For $i = 0, \dots, m$: $SIM[i,0] = i * g$

For $j = 1, \dots, n$: $SIM[0, j] = j * g$

For $i = 1, \dots, m$:

For $j = 1, \dots, n$:

$$\text{SIM}[i,j] = \max($$

0,

$$\text{SIM}[i-1,j-1] + s(X[i],Y[j]),$$
$$\text{SIM}[i-1,j]+g,$$
$$\text{SIM}[i,j-1] + g$$

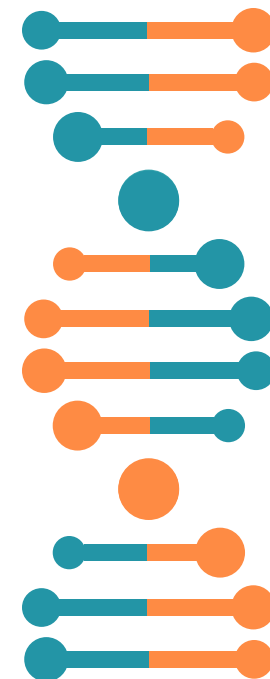
)

$$S = \max(S, \text{SIM}[i, j])$$

EndFor

EndFor

Return S





LAB 2 - Assignments

LAB2 - Assignments

- Global alignment
- Local alignment
- BWA
- Bash script file with BWA





Questions?

Remember:
no question is
stupid



**LET'S START
PRACTICING!!**