# Sequence alignment

杨建益

Email: yangjy@nankai.edu.cn

Webpage: http://yanglab.nankai.edu.cn/

Course: http://yanglab.nankai.edu.cn/teaching/bioinformatics/

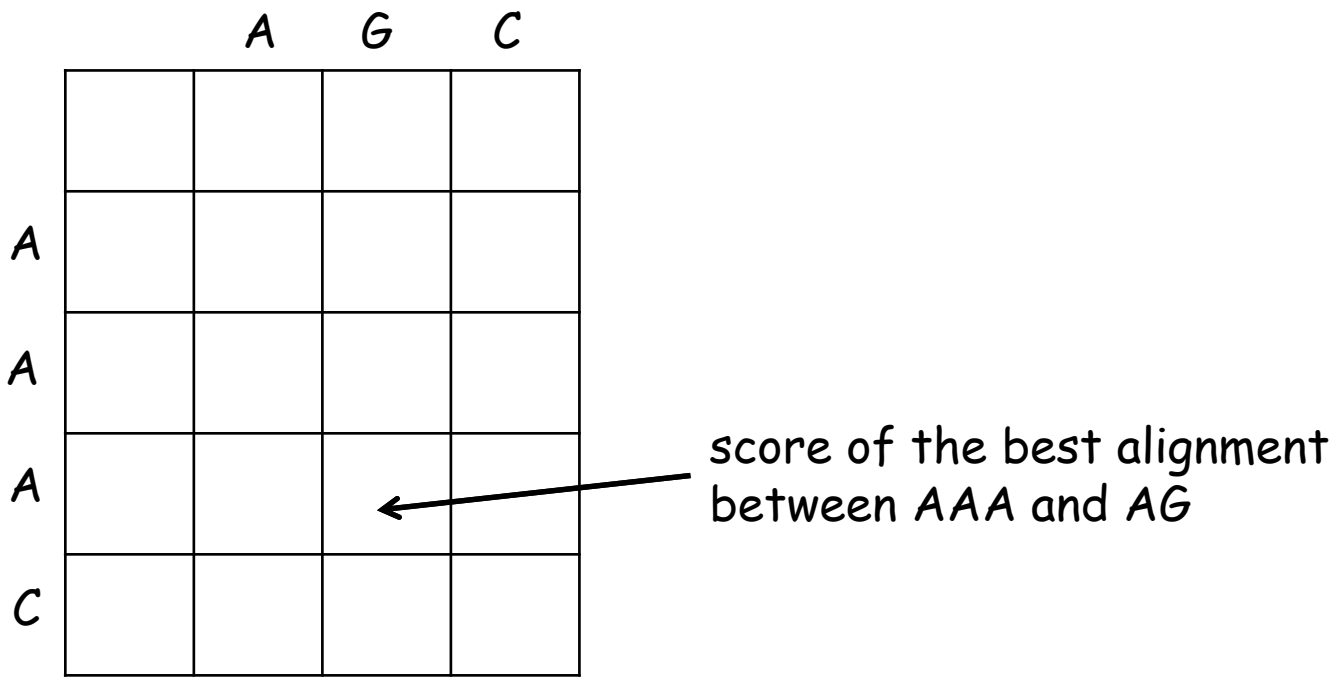
Office: 数学科学学院，419室

# Content

1. Why to make sequence alignment?
2. What is a sequence alignment?
3. How to derive a mutation matrix-PAM
4. How to derive a mutation matrix-BLOSUM
5. Gap penalty
6. Dynamic programming
   a. Global alignment: Needleman-Wunsch
   b. Local alignment: Smith-Waterman
7. Heuristic algorithms

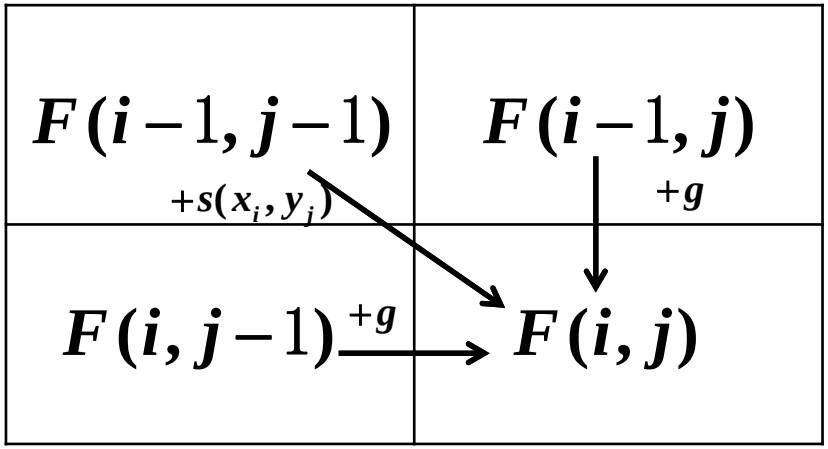
# Needleman-Wunsch's dynamic programming (DP) idea

- Given an n-character sequence x, and an m-character sequence y
- Construct an (n+1)x(m+1) matrix F(o…n, o…m)
- F(i,j)=score of the best alignment between x[1…i] and y[1…j]



# Iteration formula

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



# Two steps in Needleman-Wunsch algorithm

## Step 1: Fill in the matrix F iteratively

| | | 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 |

# Two steps in Needleman-Wunsch algorithm

## Step 2: Traceback to find the optimal alignment

| | | 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 |

x: AAAC

y: AG–C

# Two steps in Needleman-Wunsch algorithm

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

Gap penalty: g=-2

extension = opening

## Step 1: Fill in the matrix F iteratively

Draw an (n+1)x(m+1) matrix F(o…n, o…m) first

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

# Initialize the 1st column and 1st row

Gap penalty: $g=-2$

extension = opening

| | | A | G | C |
|---|---|---|---|---|
| | 0 | -2 | -4 | -6 |
| A | -2 | | | |
| A | -4 | | | |
| A | -6 | | | |
| C | -8 | | | |

# Begin filling in column-wise or row-wise order

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

Gap penalty: $w(k) = -2k$

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

# filling...

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

Gap penalty: -2

extension = opening

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

# finally

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

Gap penalty: -2

extension = opening

| | | 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 |

# Traceback

## Step 2: Traceback to find the optimal alignment

Starting from F(n,m) to F(0,0)

x: C
y: C

| | | 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 |

# Traceback

## Step 2: Traceback to find the optimal alignment

x: AC
y: −C

| | | 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 |

# Traceback

## Step 2: Traceback to find the optimal alignment

x: AAC
y: G–C

| | | 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 |

# Traceback

## Step 2: Traceback to find the optimal alignment

one optimal alignment

x: AAAC
y: AG–C

| | | 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 |

# Traceback

## Step 2: Traceback to find the optimal alignment

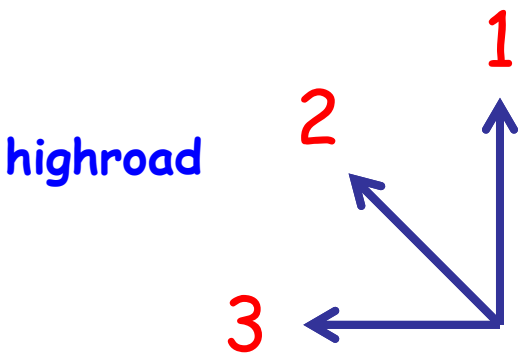
another optimal alignment

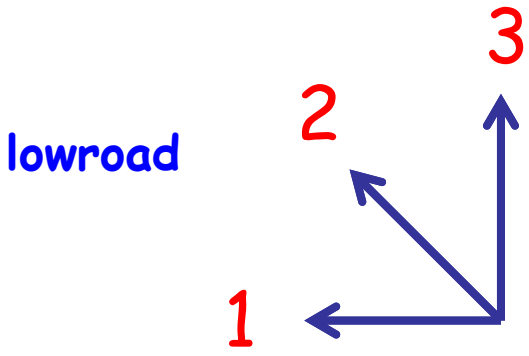
x: AAAC
y: −AGC

| | | 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 |

# Equally optimal alignments

can use preference ordering over paths when doing traceback



# Another example

$$s(x_i, y_j) = \begin{cases} 2, & \textbf{if } x_i = y_j \\ -3, & \textbf{otherwise} \end{cases}$$

Gap penalty: $g$=-2

extension = opening

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

# Another example

$$s(x_i, y_j) = \begin{cases} 2, & \text{if } x_i = y_j \\ -3, & \text{otherwise} \end{cases}$$

Gap penalty: $g$=-2

extension = opening

| | | A | C | T | G | A | T | T | C | A |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | -2 | -4 | -6 | -8 | -10 | -12 | -14 | -16 | -18 |
| A | -2 | 2 | | | | | | | | |
| C | -4 | 0 | | | | | | | | |
| G | -6 | -2 | | | | | | | | |
| C | -8 | -4 | | | | | | | | |
| A | -10 | -6 | | | | | | | | |
| T | -12 | -8 | | | | | | | | |
| C | -14 | -10 | | | | | | | | |
| A | -16 | -12 | | | | | | | | |

# Another example

$$s(x_i, y_j) = \begin{cases} 2, & \text{if } x_i = y_j \\ -3, & \text{otherwise} \end{cases}$$

Gap penalty: $g$=-2

extension = opening

| | | A | C | T | G | A | T | T | C | A |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | -2 | -4 | -6 | -8 | -10 | -12 | -14 | -16 | -18 |
| A | -2 | 2 | 0 | | | | | | | |
| C | -4 | 0 | 4 | | | | | | | |
| G | -6 | -2 | 2 | | | | | | | |
| C | -8 | -4 | 0 | | | | | | | |
| A | -10 | -6 | -2 | | | | | | | |
| T | -12 | -8 | -4 | | | | | | | |
| C | -14 | -10 | -6 | | | | | | | |
| A | -16 | -12 | -8 | | | | | | | |

# Another example

$$s(x_i, y_j) = \begin{cases} 2, & \text{if } x_i = y_j \\ -3, & \text{otherwise} \end{cases}$$

Gap penalty: $g$=-2

extension = opening

| | | A | C | T | G | A | T | T | C | A |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | -2 | -4 | -6 | -8 | -10 | -12 | -14 | -16 | -18 |
| A | -2 | 2 | 0 | -2 | -4 | | | | | |
| C | -4 | 0 | 4 | 2 | 0 | | | | | |
| G | -6 | -2 | 2 | 1 | 4 | | | | | |
| C | -8 | -4 | 0 | -1 | 2 | | | | | |
| A | -10 | -6 | -2 | -3 | 0 | | | | | |
| T | -12 | -8 | -4 | 0 | -2 | | | | | |
| C | -14 | -10 | -6 | -2 | -3 | | | | | |
| A | -16 | -12 | -8 | -4 | -5 | | | | | |

# Another example

$$s(x_i, y_j) = \begin{cases} 2, & \text{if } x_i = y_j \\ -3, & \text{otherwise} \end{cases}$$

Gap penalty: $g$=-2

extension = opening

| | | A | C | T | G | A | T | T | C | A |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | -2 | -4 | -6 | -8 | -10 | -12 | -14 | -16 | -18 |
| A | -2 | 2 | 0 | -2 | -4 | -6 | -8 | -10 | -12 | -14 |
| C | -4 | 0 | 4 | 2 | 0 | -2 | -4 | -6 | -8 | -10 |
| G | -6 | -2 | 2 | 1 | 4 | 2 | 0 | -2 | -4 | -6 |
| C | -8 | -4 | 0 | -1 | 2 | 1 | -1 | -3 | 0 | -2 |
| A | -10 | -6 | -2 | -3 | 0 | 4 | 2 | 0 | -2 | 2 |
| T | -12 | -8 | -4 | 0 | -2 | 2 | 6 | 4 | 2 | 0 |
| C | -14 | -10 | -6 | -2 | -3 | 0 | 4 | 3 | 6 | 4 |
| A | -16 | -12 | -8 | -4 | -5 | -2 | 2 | 1 | 4 | 8 |

# Another example

one optimal alignment

x: AC-GCA-TCA

y: ACTG-ATTCA

| | | A | C | T | G | A | T | T | C | A |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | -2 | -4 | -6 | -8 | -10 | -12 | -14 | -16 | -18 |
| A | -2 | 2 | 0 | -2 | -4 | -6 | -8 | -10 | -12 | -14 |
| C | -4 | 0 | 4 | 2 | 0 | -2 | -4 | -6 | -8 | -10 |
| G | -6 | -2 | 2 | 1 | 4 | 2 | 0 | -2 | -4 | -6 |
| C | -8 | -4 | 0 | -1 | 2 | 1 | -1 | -3 | 0 | -2 |
| A | -10 | -6 | -2 | -3 | 0 | 4 | 2 | 0 | -2 | 2 |
| T | -12 | -8 | -4 | 0 | -2 | 2 | 6 | 4 | 2 | 0 |
| C | -14 | -10 | -6 | -2 | -3 | 0 | 4 | 3 | 6 | 4 |
| A | -16 | -12 | -8 | -4 | -5 | -2 | 2 | 1 | 4 | 8 |

# Another example

another optimal alignment

x: AC-GCAT-CA

y: ACTG-ATTCA

| | | A | C | T | G | A | T | T | C | A |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | -2 | -4 | -6 | -8 | -10 | -12 | -14 | -16 | -18 |
| A | -2 | 2 | 0 | -2 | -4 | -6 | -8 | -10 | -12 | -14 |
| C | -4 | 0 | 4 | 2 | 0 | -2 | -4 | -6 | -8 | -10 |
| G | -6 | -2 | 2 | 1 | 4 | 2 | 0 | -2 | -4 | -6 |
| C | -8 | -4 | 0 | -1 | 2 | 1 | -1 | -3 | 0 | -2 |
| A | -10 | -6 | -2 | -3 | 0 | 4 | 2 | 0 | -2 | 2 |
| T | -12 | -8 | -4 | 0 | -2 | 2 | 6 | 4 | 2 | 0 |
| C | -14 | -10 | -6 | -2 | -3 | 0 | 4 | 3 | 6 | 4 |
| A | -16 | -12 | -8 | -4 | -5 | -2 | 2 | 1 | 4 | 8 |

# Exercise

sequence x: GAATTCAGTTA

sequence y: GGATCGA

Score matrix: $$s(x_i, y_j) = \begin{cases} 2, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

Gap penalty: g=-2

extension = opening

# Questions to think about

- How about local alignment?
- What happens if gap opening penalty and gap extension penalties are not equal?

**Paper to read for the next class:**

- T F Smith & M S Waterman, Identification of common molecular subsequences. J Mol Biol (1981) 147, 195-197.
- O. Gotoh. An improved algorithm for matching biological sequences. Journal of Molecular Biology 162 705-708 1982.

# Content

1. Why to make sequence alignment?
2. What is a sequence alignment?
3. How to derive a mutation matrix-PAM
4. How to derive a mutation matrix-BLOSUM
5. Gap penalty
6. Dynamic programming
   a. Global alignment: Needleman-Wunsch
   b. Local alignment: Smith-Waterman
7. Heuristic algorithms

# Smith-Waterman algorithm

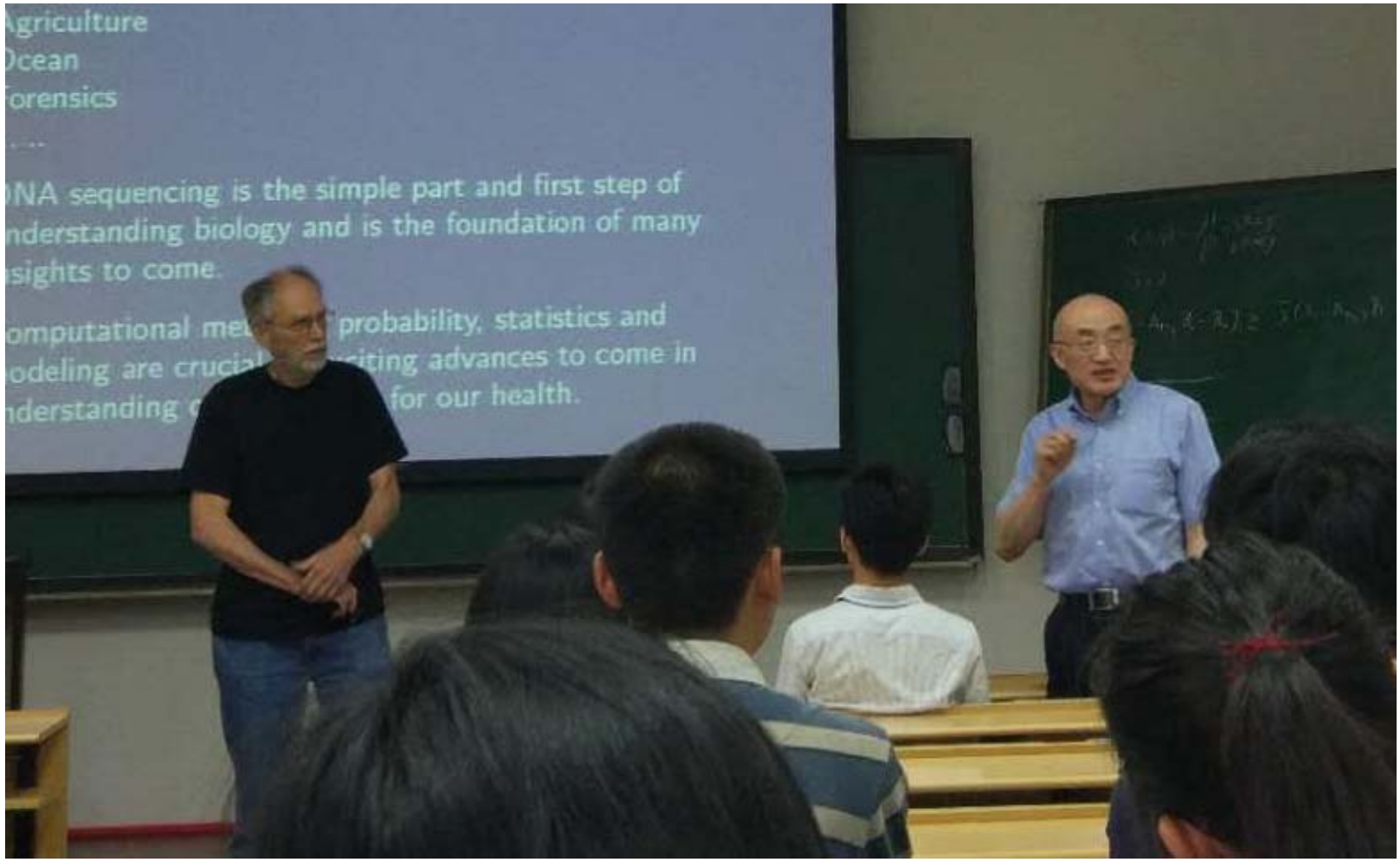
- For generating optimal local alignment?

T F Smith & M S Waterman, Identification of common molecular subsequences. J Mol Biol (1981) 147, 195-197.

Temple F. Smith (1939-)
Boston University

Michael S. Waterman (1942-)
University of Southern California (USC)

# Michael S. Waterman (1942-)



# Two differences between SW and NW

- 1. Non-negative scores

$$F(i,j) = \max \begin{cases} F(i-1,j-1) + s(x_i, y_j) \\ F(i-1,j) + g \\ F(i,j-1) + g \\ 0 \end{cases}$$

| | |
|---|---|
| $F(i-1,j-1)$ $+s(x_i,y_j)$ | $F(i-1,j)$ $+g$ |
| $F(i,j-1)$ $+g$ | $F(i,j)$ |

# Two differences between SW and NW

- 2. Traceback starts at the highest scoring matrix cell and proceeds until a cell with score zero is encountered.

# Example

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

Gap penalty: g=-2

extension = opening

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

# Example

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

Gap penalty: g=-2

extension = opening

| | | A | A | G | A |
|---|---|---|---|---|---|
| | 0 | 0 | 0 | 0 | 0 |
| T | 0 | | | | |
| T | 0 | | | | |
| A | 0 | | | | |
| A | 0 | | | | |
| G | 0 | | | | |

Initialization

# Example

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

Gap penalty: g=-2

extension = opening

| | | A | A | G | A |
|---|---|---|---|---|---|
| | 0 | 0 | 0 | 0 | 0 |
| T | 0 | 0 | | | |
| T | 0 | 0 | | | |
| A | 0 | 1 | | | |
| A | 0 | 1 | | | |
| G | 0 | 0 | | | |

Filling....

# Example

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

Gap penalty: g=-2

extension = opening

| | | A | A | G | A |
|---|---|---|---|---|---|
| | 0 | 0 | 0 | 0 | 0 |
| T | 0 | 0 | 0 | | |
| T | 0 | 0 | 0 | | |
| A | 0 | 1 | 1 | | |
| A | 0 | 1 | 2 | | |
| G | 0 | 0 | 0 | | |

Filling....

# Example

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

Gap penalty: g=-2

extension = opening

| | | A | A | G | A |
|---|---|---|---|---|---|
| | 0 | 0 | 0 | 0 | 0 |
| T | 0 | 0 | 0 | 0 | |
| T | 0 | 0 | 0 | 0 | |
| A | 0 | 1 | 1 | 0 | |
| A | 0 | 1 | 2 | 0 | |
| G | 0 | 0 | 0 | 3 | |

Filling....

# Example

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

Gap penalty: g=-2

extension = opening

| | | A | A | G | A |
|---|---|---|---|---|---|
| | 0 | 0 | 0 | 0 | 0 |
| T | 0 | 0 | 0 | 0 | 0 |
| T | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 1 | 1 | 0 | 1 |
| A | 0 | 1 | 2 | 0 | 1 |
| G | 0 | 0 | 0 | 3 | 1 |

Filling done

# Example

x : G

y : G

| | | A | A | G | A |
|---|---|---|---|---|---|
| | 0 | 0 | 0 | 0 | 0 |
| T | 0 | 0 | 0 | 0 | 0 |
| T | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 1 | 1 | 0 | 1 |
| A | 0 | 1 | 2 | 0 | 1 |
| G | 0 | 0 | 0 | 3 | 1 |

Traceback…

# Example

x : AG

y : AG

| | | A | A | G | A |
|---|---|---|---|---|---|
| | 0 | 0 | 0 | 0 | 0 |
| T | 0 | 0 | 0 | 0 | 0 |
| T | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 1 | 1 | 0 | 1 |
| A | 0 | 1 | 2 | 0 | 1 |
| G | 0 | 0 | 0 | 3 | 1 |

Traceback…

# Example

x : AAG

y : AAG

| | | A | A | G | A |
|---|---|---|---|---|---|
| | 0 | 0 | 0 | 0 | 0 |
| T | 0 | 0 | 0 | 0 | 0 |
| T | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 1 | 1 | 0 | 1 |
| A | 0 | 1 | 2 | 0 | 1 |
| G | 0 | 0 | 0 | 3 | 1 |

Traceback done

# Affine Gap penalty

```
MVLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRVKHLKTEAEMKASEDLK
SLEWMVNWAMVNWAAVV--------DDFYQELFKAHPEYQNKFGFFKAHPEYQNKFGFKGVALG
```

Gap opening  Gap extension

- Gap penalty: $w(k) = a + b \times (k-1) \quad (k \geq 1;\ \ a, b < 0)$
  - k: length of continous gaps
  - a: gap opening penalty
  - b: gap entension penalty

- Linear gap penalty if a=b
- Affine gap penalty if a!=b

# DP for affine gap penalty case

- O. Gotoh. An improved algorithm for matching biological sequences. Journal of Molecular Biology 162 705-708 1982.

Time complexity $O(mn)$

# NW-DP with affine gap penalty

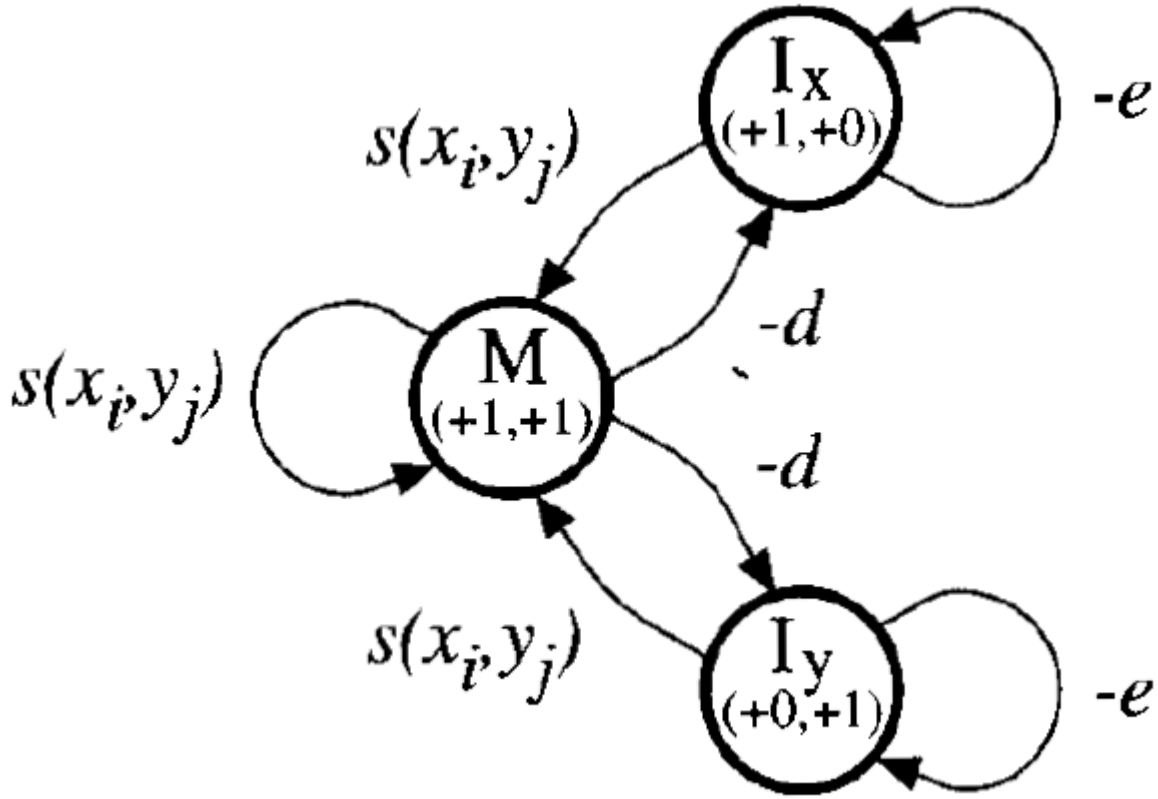
- Need 3 matrices instead of 1

$M(i, j)$ best score given that *x[i]* is aligned to *y[j]*

$I_x(i, j)$ best score given that *x[i]* is aligned to *a gap*

$I_y(i, j)$ best score given that *y[j]* is aligned to *a gap*

# NW-DP with affine gap penalty



# NW-DP with affine gap penalty

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

$$I_x(i,j) = \max \begin{cases} M(i-1,j) + a \\ I_x(i-1,j) + b \end{cases}$$



$$I_y(i,j) = \max \begin{cases} M(i,j-1) + a \\ I_y(i,j-1) + b \end{cases}$$

# NW-DP with affine gap penalty

- Initialization

$$\begin{cases} M(0,0) = 0; \\ M(i,0) = -\infty, M(0,j) = -\infty \quad (i, j \neq 0) \end{cases}$$

$$\begin{cases} I_x(0,0) = 0 \\ I_x(i,0) = a + b \times (i-1), \quad (0 < i \leq m) \\ I_x(0,j) = -\infty, \quad (0 < j \leq n) \end{cases}$$

$$\begin{cases} I_y(0,0) = 0 \\ I_y(0,j) = a + b \times (j-1), \quad (0 < j \leq n) \\ I_y(i,0) = -\infty, \quad (0 < i \leq m) \end{cases}$$

# NW-DP with affine gap penalty

- Traceback
  - Start at the largest of $M(m,n)$ , $I_x(m,n)$, $I_y(m,n)$
  - Stop at any of $M(0,0)$ , $I_x(0,0)$, $I_y(0,0)$

# Example

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

a=-3,b=-1

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

# Example

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

a=-3,b=-1

Filling…

| | | A | C | A | C | T |
|---|---|---|---|---|---|---|
| | M:0 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_x$:-3 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_y$:-3 | -4 | -5 | -6 | -7 | -8 |
| | -∞ | | | | | |
| A | -4 | | | | | |
| | -∞ | | | | | |
| | -∞ | | | | | |
| A | -5 | | | | | |
| | -∞ | | | | | |
| | -∞ | | | | | |
| T | -6 | | | | | |
| | -∞ | | | | | |

# Example

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

a=-3,b=-1

Filling…

| | | A | C | A | C | T |
|---|---|---|---|---|---|---|
| | M:0 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_x$:-3 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_y$:-3 | -4 | -5 | -6 | -7 | -8 |
| | -∞ | 1 | | | | |
| A | -4 | -∞ | | | | |
| | -∞ | -∞ | | | | |
| | -∞ | -3 | | | | |
| A | -5 | -2 | | | | |
| | -∞ | -∞ | | | | |
| | -∞ | -6 | | | | |
| T | -6 | -3 | | | | |
| | -∞ | -∞ | | | | |

# Example

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

a=-3,b=-1

Filling...

| | | A | C | A | C | T |
|---|---|---|---|---|---|---|
| | M:0 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_x$:-3 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_y$:-3 | -4 | -5 | -6 | -7 | -8 |
| | -∞ | 1 | -5 | | | |
| A | -4 | -∞ | -∞ | | | |
| | -∞ | -∞ | -2 | | | |
| | -∞ | -3 | 0 | | | |
| A | -5 | -2 | -8 | | | |
| | -∞ | -∞ | -6 | | | |
| | -∞ | -6 | -3 | | | |
| T | -6 | -3 | -3 | | | |
| | -∞ | -∞ | -9 | | | |

# Example

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

a=-3,b=-1

Filling…

| | | A | C | A | C | T |
|---|---|---|---|---|---|---|
| | M:0 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_x$:-3 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_y$:-3 | -4 | -5 | -6 | -7 | -8 |
| | -∞ | 1 | -5 | -4 | | |
| A | -4 | -∞ | -∞ | -∞ | | |
| | -∞ | -∞ | -2 | -3 | | |
| | -∞ | -3 | 0 | -1 | | |
| A | -5 | -2 | -8 | -7 | | |
| | -∞ | -∞ | -6 | -3 | | |
| | -∞ | -6 | -3 | -1 | | |
| T | -6 | -3 | -3 | -4 | | |
| | -∞ | -∞ | -9 | -6 | | |

# Example

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

a=-3,b=-1

Filling…

| | | A | C | A | C | T |
|---|---|---|---|---|---|---|
| | M:0 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_x$:-3 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_y$:-3 | -4 | -5 | -6 | -7 | -8 |
| | -∞ | 1 | -5 | -4 | -7 | |
| A | -4 | -∞ | -∞ | -∞ | -∞ | |
| | -∞ | -∞ | -2 | -3 | -4 | |
| | -∞ | -3 | 0 | -1 | -4 | |
| A | -5 | -2 | -8 | -7 | -10 | |
| | -∞ | -∞ | -6 | -3 | -4 | |
| | -∞ | -6 | -3 | -1 | -2 | |
| T | -6 | -3 | -3 | -4 | -7 | |
| | -∞ | -∞ | -9 | -6 | -4 | |

# Example

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

a=-3,b=-1

Filling done

| | | A | C | A | C | T |
|---|---|---|---|---|---|---|
| | M:0 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_x$:-3 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_y$:-3 | -4 | -5 | -6 | -7 | -8 |
| | -∞ | 1 | -5 | -4 | -7 | -8 |
| A | -4 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | -∞ | -∞ | -2 | -3 | -4 | -5 |
| | -∞ | -3 | 0 | -1 | -4 | -5 |
| A | -5 | -2 | -8 | -7 | -10 | -11 |
| | -∞ | -∞ | -6 | -3 | -4 | -5 |
| | -∞ | -6 | -3 | -1 | -2 | -3 |
| T | -6 | -3 | -3 | -4 | -7 | -8 |
| | -∞ | -∞ | -9 | -6 | -4 | -5 |

# Example

x:A--AT
y:ACACT

Traceback

| | | A | C | A | C | T |
|---|---|---|---|---|---|---|
| | M:0 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_x$:-3 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_y$:-3 | -4 | -5 | -6 | -7 | -8 |
| | -∞ | 1 | -5 | -4 | -7 | -8 |
| A | -4 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | -∞ | -∞ | -2 | -3 | -4 | -5 |
| | -∞ | -3 | 0 | -1 | -4 | -5 |
| A | -5 | -2 | -8 | -7 | -10 | -11 |
| | -∞ | -∞ | -6 | -3 | -4 | -5 |
| | -∞ | -6 | -3 | -1 | -2 | -3 |
| T | -6 | -3 | -3 | -4 | -7 | -8 |
| | -∞ | -∞ | -9 | -6 | -4 | -5 |

Another alignment

x : AA−−T

y : ACACT

# Example

Traceback

| | | A | C | A | C | T |
|---|---|---|---|---|---|---|
| | M:0 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_x$:-3 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_y$:-3 | -4 | -5 | -6 | -7 | -8 |
| | -∞ | 1 | -5 | -4 | -7 | -8 |
| A | -4 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | -∞ | -∞ | -2 | -3 | -4 | -5 |
| | -∞ | -3 | 0 | -1 | -4 | -5 |
| A | -5 | -2 | -8 | -7 | -10 | -11 |
| | -∞ | -∞ | -6 | -3 | -4 | -5 |
| | -∞ | -6 | -3 | -1 | -2 | -3 |
| T | -6 | -3 | -3 | -4 | -7 | -8 |
| | -∞ | -∞ | -9 | -6 | -4 | -5 |

Another alignment

# Example

x : A–A–T

y : ACACT

Traceback

| | | A | C | A | C | T |
|---|---|---|---|---|---|---|
| | M:0 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_x$:-3 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | $I_y$:-3 | -4 | -5 | -6 | -7 | -8 |
| | -∞ | 1 | -5 | -4 | -7 | -8 |
| A | -4 | -∞ | -∞ | -∞ | -∞ | -∞ |
| | -∞ | -∞ | -2 | -3 | -4 | -5 |
| | -∞ | -3 | 0 | -1 | -4 | -5 |
| A | -5 | -2 | -8 | -7 | -10 | -11 |
| | -∞ | -∞ | -6 | -3 | -4 | -5 |
| | -∞ | -6 | -3 | -1 | -2 | -3 |
| T | -6 | -3 | -3 | -4 | -7 | -8 |
| | -∞ | -∞ | -9 | -6 | -4 | -5 |

Another alignment

# Example

$$s(x_i, y_j) = \begin{cases} 1, & \textbf{if } x_i = y_j \\ -1, & \textbf{otherwise} \end{cases}$$

a=-3,b=-1

```
x:A--AT
y:ACACT

x:AA--T
y:ACACT

x:A-A-T
y:ACACT
```

# SW-DP with affine gap penalty

$$M(i,j) = \max \begin{cases} M(i-1,j-1) + s(x_i, y_j) \\ I_x(i-1,j-1) + s(x_i, y_j) \\ I_y(i-1,j-1) + s(x_i, y_j) \\ 0 \end{cases}$$

$$I_x(i,j) = \max \begin{cases} M(i-1,j) + a \\ I_x(i-1,j) + b \end{cases}$$

$$I_y(i,j) = \max \begin{cases} M(i,j-1) + a \\ I_y(i,j-1) + b \end{cases}$$

# SW-DP with affine gap penalty

- Initialization

$$M(i,0) = M(0,j) = 0$$

$$I_x(i,0) = I_x(0,j) = -\infty$$

$$I_y(i,0) = I_y(0,j) = -\infty$$

# SW-DP with affine gap penalty

- Traceback
  - Start at the largest of M(i,j)
  - Stop at M(i,j)=0

# Project 2

1. Please find the best global alignment of following two sequences

- AGTTGC
- CAGA

Score matrix

| | A | T | G | C |
|---|---|---|---|---|
| A | 2 | 1 | -1 | -1 |
| T | 1 | 2 | -1 | -1 |
| G | -1 | -1 | 2 | 1 |
| C | -1 | -1 | 1 | 2 |

Gap penalty: open=extension=-2

# Project 2

2. Please find the best global alignment of following two sequences

- AGTTGC
- CAGA

Score matrix

| | A | T | G | C |
|---|---|---|---|---|
| A | 2 | 1 | -1 | -1 |
| T | 1 | 2 | -1 | -1 |
| G | -1 | -1 | 2 | 1 |
| C | -1 | -1 | 1 | 2 |

Gap penalty: open=-2, extension=-1

# Project 2

Your answers to 1,2 should include the following

1. Alignment matrix
2. Trace back path
3. Alignment result

in a similar format as the examples given in the class.

# Project 2

3. Write a program to align any two protein sequences with BLOSUM62 matrix, available at:
http://yanglab.nankai.edu.cn/teaching/bioinformatics/BLOSUM62.txt

Gap opening=-11
Gap extension=-1

To examine whether your program is correct, you can compare your result with the program at
http://zhanglab.ccmb.med.umich.edu/NW-align/

Please submit your reports to:
Ms Yajun Dai, 2120180074@mail.nankai.edu.cn
Deadline: 24:00, 04/25/2021