

Computational Biology

Introduction

Sequence Alignment

Global alignment

Local alignment

Exercise

Lecturers:

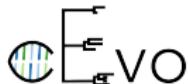
Tanja Stadler, Carsten Magnus & Tim Vaughan

Teaching Assistants:

Jūlija Pečerska, Jérémie Sciré,
Sarah Nadeau & Marc Manceau

Computational Evolution
Department of Biosystems Science and Engineering

HS 2019



Tutorial Setup

Introduction

Sequence Alignment

Global alignment

Local alignment

Exercise

► 12 tutorials

- Every Monday (Zürich) and Thursday (Basel)
30 Sept - 19 Dec
- **This week:**
 - Introduction of Homework 1
 - A pen and paper exercise to practice the algorithm of Homework 1
- Odd week number (next week):
 - Help session for Homework 1
- Any time:
 - Seek help on Moodle forums

Homework Setup

Introduction
Sequence Alignment
Global alignment
Local alignment
Exercise

► 5 homework assignments

- 25 % of your grade
- Hand out 30 Sept, 14 Oct, 28 Oct, 11 Nov, 25 Nov
- Hand in 2 weeks after (Monday at noon)
 - Submit on Moodle
 - Submission must contain code (*.R) and answers to theory questions (*.pdf)
 - Submissions can be updated until the deadline
 - Follow the structure of the skeleton
 - Points are attributed on a per-function basis
- Example homework assignment available on Moodle (with solution)
- **No late submissions**
- **Plagiarism will be detected**

Sequence alignment

Introduction

Sequence Alignment

Global alignment

Local alignment

Exercise

Aim: Find the best global/local alignment of sequences

$\text{seqA} = a_1 a_2 \dots a_{n_A}$ and $\text{seqB} = b_1 b_2 \dots b_{n_B}$.

Needleman-Wunsch for global alignment

Introduction
Sequence Alignment
Global alignment
Local alignment
Exercise

Task: Align TTGC and AGCAC using **Needleman-Wunsch**

Given: Gap penalty of -2, match score of +3, mismatch penalty of -1

Steps:

- ▶ **Initialization** of score and path matrices
- ▶ **Iterative calculation** of score and path matrices
- ▶ **Reconstruction** of the optimal global alignment

N-W initialization

- Initialize `score_matrix H` of size $n_A + 1 \times n_B + 1$

[Introduction](#)[Sequence Alignment](#)[Global alignment](#)[Local alignment](#)[Exercise](#)

i \ j	0	1 A	2 G	3 C	4 A	5 C
0						
1 T						
2 T						
3 G						
4 C						

N-W initialization

[Introduction](#)
[Sequence Alignment](#)
[Global alignment](#)
[Local alignment](#)
[Exercise](#)

- ▶ Initialize `score_matrix H` of size $n_A + 1 \times n_B + 1$
- ▶ Fill first row & column of `H` with increments of gap score

i \ j	0	1 A	2 G	3 C	4 A	5 C
0	0	-2	-4	-6	-8	-10
1 T	-2					
2 T	-4					
3 G	-6					
4 C	-8					

N-W initialization

[Introduction](#)
[Sequence Alignment](#)
[Global alignment](#)
[Local alignment](#)
[Exercise](#)

- ▶ Initialize `score_matrix H` of size $n_A + 1 \times n_B + 1$
- ▶ Fill first row & column of `H` with increments of gap score
- ▶ Fill first row of `path_matrix` with \leftarrow
- ▶ Fill first column of `path_matrix` with \uparrow

j \ i	0	1 A	2 G	3 C	4 A	5 C
0	0	-2	-4	-6	-8	-10
1 T	-2					
2 T	-4					
3 G	-6					
4 C	-8					

N-W iterative calculation

- ▶ Compute partial alignment scores $H[i, j]$ for $2 < i < n_A + 1$ and $2 < j < n_B + 1$

$$H[i, j] = \max \left\{ \begin{array}{ll} H[i - 1, j - 1] + 3 & \text{if } a_i = b_j \\ H[i - 1, j - 1] - 1 & \text{if } a_i \neq b_j \\ H[i - 1, j] - 2 & \text{gap in sequence seqB} \\ H[i, j - 1] - 2 & \text{gap in sequence seqA} \end{array} \right. \begin{array}{l} \text{match} \\ \text{mismatch} \\ \text{gap in sequence seqB} \\ \text{gap in sequence seqA} \end{array} \right\}$$

Introduction

Sequence Alignment

Global alignment

Local alignment

Exercise

N-W iterative calculation

- ▶ Compute partial alignment scores $H[i, j]$ for $2 < i < n_A + 1$ and $2 < j < n_B + 1$

$$H[i, j] = \max \left\{ \begin{array}{ll} H[i - 1, j - 1] + 3 & \text{if } a_i = b_j \\ H[i - 1, j - 1] - 1 & \text{if } a_i \neq b_j \\ H[i - 1, j] - 2 & \text{gap in sequence seqB} \\ H[i, j - 1] - 2 & \text{gap in sequence seqA} \end{array} \right. \begin{array}{l} \text{match} \\ \text{mismatch} \\ \text{gap in sequence seqB} \\ \text{gap in sequence seqA} \end{array} \right\}$$

i \ j	0	1 A	2 G	3 C	4 A	5 C
0	0	-2	-4	-6	-8	-10
1	T	-2	-1			
2	T	-4				
3	G	-6				
4	C	-8				

Introduction

Sequence Alignment

Global alignment

Local alignment

Exercise

N-W iterative calculation

- ▶ Compute partial alignment scores $H[i, j]$ for $2 < i < n_A + 1$ and $2 < j < n_B + 1$

$$H[i, j] = \max \left\{ \begin{array}{ll} H[i - 1, j - 1] + 3 & \text{if } a_i = b_j \\ H[i - 1, j - 1] - 1 & \text{if } a_i \neq b_j \\ H[i - 1, j] - 2 & \text{gap in sequence seqB} \\ H[i, j - 1] - 2 & \text{gap in sequence seqA} \end{array} \right. \begin{array}{l} \text{match} \\ \text{mismatch} \\ \\ \end{array} \right\}$$

Introduction

Sequence Alignment

Global alignment

Local alignment

Exercise

i \ j	0	1 A	2 G	3 C	4 A	5 C
0	0	-2	-4	-6	-8	-10
1	T	-2	-4			
2	T	-4				
3	G	-6				
4	C	-8				

N-W iterative calculation

- ▶ Compute partial alignment scores $H[i, j]$ for $2 < i < n_A + 1$ and $2 < j < n_B + 1$

$$H[i, j] = \max \left\{ \begin{array}{ll} H[i - 1, j - 1] + 3 & \text{if } a_i = b_j \\ H[i - 1, j - 1] - 1 & \text{if } a_i \neq b_j \\ H[i - 1, j] - 2 & \text{gap in sequence seqB} \\ H[i, j - 1] - 2 & \text{gap in sequence seqA} \end{array} \right. \begin{array}{l} \text{match} \\ \text{mismatch} \\ \text{gap in sequence seqB} \\ \text{gap in sequence seqA} \end{array} \right\}$$

Introduction

Sequence Alignment

Global alignment

Local alignment

Exercise

i \ j	0	1 A	2 G	3 C	4 A	5 C
0	0	-2	-4	-6	-8	-10
1 T	-2	-4				
2 T	-4					
3 G	-6					
4 C	-8					

N-W iterative calculation

- ▶ Compute partial alignment scores $H[i, j]$ for $2 < i < n_A + 1$ and $2 < j < n_B + 1$

$$H[i, j] = \max \left\{ \begin{array}{ll} H[i - 1, j - 1] + 3 & \text{if } a_i = b_j \\ H[i - 1, j - 1] - 1 & \text{if } a_i \neq b_j \\ H[i - 1, j] - 2 & \text{gap in sequence seqB} \\ H[i, j - 1] - 2 & \text{gap in sequence seqA} \end{array} \right. \begin{array}{l} \text{match} \\ \text{mismatch} \\ \text{gap in sequence seqB} \\ \text{gap in sequence seqA} \end{array} \right\}$$

i \ j	0	1 A	2 G	3 C	4 A	5 C
0	0	-2	-4	-6	-8	-10
1	T	-2	-1			
2	T	-4				
3	G	-6				
4	C	-8				

Introduction

Sequence Alignment

Global alignment

Local alignment

Exercise

N-W iterative calculation

- ▶ Compute partial alignment scores $H[i, j]$ for $2 < i < n_A + 1$ and $2 < j < n_B + 1$

$$H[i, j] = \max \left\{ \begin{array}{ll} H[i - 1, j - 1] + 3 & \text{if } a_i = b_j \\ H[i - 1, j - 1] - 1 & \text{if } a_i \neq b_j \\ H[i - 1, j] - 2 & \text{gap in sequence seqB} \\ H[i, j - 1] - 2 & \text{gap in sequence seqA} \end{array} \right. \begin{array}{l} \text{match} \\ \text{mismatch} \\ \text{gap in sequence seqB} \\ \text{gap in sequence seqA} \end{array} \right\}$$

i \ j	0	1 A	2 G	3 C	4 A	5 C
0	0	-2	-4	-6	-8	-10
1	T	-2	-1	-3		
2	T	-4				
3	G	-6				
4	C	-8				

Introduction

Sequence Alignment

Global alignment

Local alignment

Exercise

N-W iterative calculation

- ▶ Compute partial alignment scores $H[i, j]$ for $2 < i < n_A + 1$ and $2 < j < n_B + 1$

$$H[i, j] = \max \left\{ \begin{array}{ll} H[i - 1, j - 1] + 3 & \text{if } a_i = b_j \\ H[i - 1, j - 1] - 1 & \text{if } a_i \neq b_j \\ H[i - 1, j] - 2 & \text{gap in sequence seqB} \\ H[i, j - 1] - 2 & \text{gap in sequence seqA} \end{array} \right. \begin{array}{l} \text{match} \\ \text{mismatch} \\ \\ \end{array} \right\}$$

i \ j	0	1 A	2 G	3 C	4 A	5 C
0	0	-2	-4	-6	-8	-10
1 T	-2	-1	-6			
2 T	-4					
3 G	-6					
4 C	-8					

Introduction

Sequence Alignment

Global alignment

Local alignment

Exercise

N-W iterative calculation

- ▶ Compute partial alignment scores $H[i, j]$ for $2 < i < n_A + 1$ and $2 < j < n_B + 1$

$$H[i, j] = \max \left\{ \begin{array}{ll} H[i - 1, j - 1] + 3 & \text{if } a_i = b_j \\ H[i - 1, j - 1] - 1 & \text{if } a_i \neq b_j \\ H[i - 1, j] - 2 & \text{gap in sequence seqB} \\ H[i, j - 1] - 2 & \text{gap in sequence seqA} \end{array} \right. \begin{array}{l} \text{match} \\ \text{mismatch} \\ \text{gap in sequence seqB} \\ \text{gap in sequence seqA} \end{array} \right\}$$

i \ j	0	1 A	2 G	3 C	4 A	5 C
0	0	-2	-4	-6	-8	-10
1 T	-2	-1	-3			
2 T	-4					
3 G	-6					
4 C	-8					

Introduction

Sequence Alignment

Global alignment

Local alignment

Exercise

N-W iterative calculation

- ▶ Compute partial alignment scores $H[i, j]$ for $2 < i < n_A + 1$ and $2 < j < n_B + 1$

$$H[i, j] = \max \left\{ \begin{array}{ll} H[i - 1, j - 1] + 3 & \text{if } a_i = b_j \\ H[i - 1, j - 1] - 1 & \text{if } a_i \neq b_j \\ H[i - 1, j] - 2 & \text{gap in sequence seqB} \\ H[i, j - 1] - 2 & \text{gap in sequence seqA} \end{array} \right. \begin{array}{l} \text{match} \\ \text{mismatch} \\ \text{gap in sequence seqB} \\ \text{gap in sequence seqA} \end{array} \right\}$$

i \ j	0	1 A	2 G	3 C	4 A	5 C
0	0	-2	-4	-6	-8	-10
1	T	-2	-1	-3		
2	T	-4				
3	G	-6				
4	C	-8				

Introduction

Sequence Alignment

Global alignment

Local alignment

Exercise

N-W iterative calculation

- ▶ Compute partial alignment scores $H[i, j]$ for $2 < i < n_A + 1$ and $2 < j < n_B + 1$

$$H[i, j] = \max \left\{ \begin{array}{ll} H[i - 1, j - 1] + 3 & \text{if } a_i = b_j \\ H[i - 1, j - 1] - 1 & \text{if } a_i \neq b_j \\ H[i - 1, j] - 2 & \text{gap in sequence seqB} \\ H[i, j - 1] - 2 & \text{gap in sequence seqA} \end{array} \right. \begin{array}{l} \text{match} \\ \text{mismatch} \\ \text{gap in sequence seqB} \\ \text{gap in sequence seqA} \end{array} \right\}$$

i \ j	0	1 A	2 G	3 C	4 A	5 C	
0	0	-2	-4	-6	-8	-10	
1	T	-2	-1	-3	-5	-7	-9
2	T	-4	-3	-2	-4	-6	-8
3	G	-6	-5	0	-2	-4	-6
4	C	-8	-7	-2	3	1	-1

Introduction

Sequence Alignment

Global alignment

Local alignment

Exercise

N-W reconstruction

- ▶ Start at the bottom-right corner of the `score_matrix`

[Introduction](#)
[Sequence Alignment](#)
[Global alignment](#)
[Local alignment](#)
[Exercise](#)

i \ j	0	1	2	3	4	5	
0	0	-2	-4	-6	-8	-10	
1	T	-2	-1	-3	-5	-7	-9
2	T	-4	-3	-2	-4	-6	-8
3	G	-6	-5	0	-2	-4	-6
4	C	-8	-7	-2	3	1	-1

N-W reconstruction

[Introduction](#)

[Sequence Alignment](#)

[Global alignment](#)

[Local alignment](#)

[Exercise](#)

- ▶ Start at the bottom-right corner of the `score_matrix`
- ▶ Trace back until you reach the top-left corner

i \ j	0	1 A	2 G	3 C	4 A	5 C
0	0	-2	-4	-6	-8	-10
1 T	-2	-1	-3	-5	-7	-9
2 T	-4	-3	-2	-4	-6	-8
3 G	-6	-5	0	-2	-4	-6
4 C	-8	-7	-2	3	1	-1

N-W reconstruction

[Introduction](#)

[Sequence Alignment](#)

[Global alignment](#)

[Local alignment](#)

[Exercise](#)

- ▶ Start at the bottom-right corner of the `score_matrix`
- ▶ Trace back until you reach the top-left corner

i \ j	0	1 A	2 G	3 C	4 A	5 C
0	0	-2	-4	-6	-8	-10
1 T	-2	-1	-3	-5	-7	-9
2 T	-4	-3	-2	-4	-6	-8
3 G	-6	-5	0	-2	-4	-6
4 C	-8	-7	-2	3	1	-1

Optimal alignments with score -1:

N-W reconstruction

[Introduction](#)

[Sequence Alignment](#)

[Global alignment](#)

[Local alignment](#)

[Exercise](#)

- ▶ Start at the bottom-right corner of the `score_matrix`
- ▶ Trace back until you reach the top-left corner

i \ j	0	1 A	2 G	3 C	4 A	5 C
0	0	-2	-4	-6	-8	-10
1 T	-2	-1	-3	-5	-7	-9
2 T	-4	-3	-2	-4	-6	-8
3 G	-6	-5	0	-2	-4	-6
4 C	-8	-7	-2	3	1	-1

Optimal alignments with score -1:

TTGC--
-AGCAC;

N-W reconstruction

[Introduction](#)

[Sequence Alignment](#)

[Global alignment](#)

[Local alignment](#)

[Exercise](#)

- ▶ Start at the bottom-right corner of the `score_matrix`
- ▶ Trace back until you reach the top-left corner

i \ j	0	1 A	2 G	3 C	4 A	5 C
0	0	-2	-4	-6	-8	-10
1 T	-2	-1	-3	-5	-7	-9
2 T	-4	-3	-2	-4	-6	-8
3 G	-6	-5	0	-2	-4	-6
4 C	-8	-7	-2	3	1	-1

Optimal alignments with score -1:

TTGC--
-AGCAC;

TTGC--
A-GCAC;

N-W reconstruction

[Introduction](#)

[Sequence Alignment](#)

[Global alignment](#)

[Local alignment](#)

[Exercise](#)

- ▶ Start at the bottom-right corner of the `score_matrix`
- ▶ Trace back until you reach the top-left corner

i \ j	0	1 A	2 G	3 C	4 A	5 C
0	0	-2	-4	-6	-8	-10
1 T	-2	-1	-3	-5	-7	-9
2 T	-4	-3	-2	-4	-6	-8
3 G	-6	-5	0	-2	-4	-6
4 C	-8	-7	-2	3	1	-1

Optimal alignments with score -1:

TTGC--

-AGCAC;

TTG--C

-AGCAC;

TTGC--

A-GCAC;

N-W reconstruction

[Introduction](#)
[Sequence Alignment](#)
[Global alignment](#)
[Local alignment](#)
[Exercise](#)

- ▶ Start at the bottom-right corner of the `score_matrix`
- ▶ Trace back until you reach the top-left corner

i \ j	0	1 A	2 G	3 C	4 A	5 C
0	0	-2	-4	-6	-8	-10
1 T	-2	-1	-3	-5	-7	-9
2 T	-4	-3	-2	-4	-6	-8
3 G	-6	-5	0	-2	-4	-6
4 C	-8	-7	-2	3	1	-1

Optimal alignments with score -1:

TTGC--	TTG--C
-AGCAC;	-AGCAC;
TTGC--	TTG--C
A-GCAC;	A-GCAC.

Needleman-Wunsch for global alignment

Introduction
Sequence Alignment
Global alignment
Local alignment
Exercise

Steps:

1. Initialization:

- ▶ Initialize score_matrix H of size $n_A + 1 \times n_B + 1$
- ▶ Fill first row & column of H with increments of gap score
- ▶ Fill first row of path_matrix with \leftarrow
- ▶ Fill first column of path_matrix with \uparrow

2. Iterative calculation

of partial alignment scores $H[i, j]$ for
 $2 < i < n_A + 1$ and $2 < j < n_B + 1$;

3. Reconstruction

of the optimal global alignment:

- ▶ Start at the bottom right corner of the score_matrix;
- ▶ Trace back until you reach the top-left corner.

Smith-Waterman for local alignment

Introduction
Sequence Alignment
Global alignment
Local alignment
Exercise

Differences wrt. Needleman-Wunsch:

1. No negative scores allowed in $H_{SW}[i, j]$.
2. Reconstruction starts at the position with the highest score, rather than at $H_{SW}[n_A + 1, n_B + 1]$.
3. Reconstruction ends when a position with $H_{SW}[i, j] = 0$ is reached, rather than at $H_{SW}[1, 1]$.

Pen and paper exercise

[Introduction](#)
[Sequence Alignment](#)
[Global alignment](#)
[Local alignment](#)
[Exercise](#)

Task: Align AGCACACA and TCACACTA

A) Using the Needleman-Wunsch algorithm (global alignment)

$$H[i, j] = \max \left\{ \begin{array}{ll} H[i - 1, j - 1] + 3 & \text{if } a_i = b_j \\ H[i - 1, j - 1] - 1 & \text{if } a_i \neq b_j \\ H[i - 1, j] - 2 & \text{gap in sequence seqB} \\ H[i, j - 1] - 2 & \text{gap in sequence seqA} \end{array} \right. \begin{array}{l} \text{match} \\ \text{mismatch} \\ \text{gap in sequence seqB} \\ \text{gap in sequence seqA} \end{array} \right\}$$

B) Using the Smith-Waterman algorithm (local alignment)

$$H[i, j] = \max \left\{ \begin{array}{ll} 0 & \\ H[i - 1, j - 1] + 3 & \text{if } a_i = b_j \\ H[i - 1, j - 1] - 1 & \text{if } a_i \neq b_j \\ H[i - 1, j] - 2 & \text{gap in sequence seqB} \\ H[i, j - 1] - 2 & \text{gap in sequence seqA} \end{array} \right. \begin{array}{l} \text{match} \\ \text{mismatch} \\ \text{gap in sequence seqB} \\ \text{gap in sequence seqA} \end{array} \right\}$$

Solution: Needleman-Wunsch for global alignment

Task: Align AGCACACA and TCACACTA

i \ j	0	1	2	3	4	5	6	7	8	
0	0	-2	-4	-6	-8	-10	-12	-14	-16	
1	A	-2	-1	-3	-1	-3	-5	-7	-9	-11
2	G	-4	-3	-2	-3	-2	-4	-6	-8	-10
3	C	-6	-5	0	-2	0	-2	-1	-3	-5
4	A	-8	-7	-2	3	1	3	1	-1	0
5	C	-10	-9	-4	1	6	4	6	4	2
6	A	-12	-11	-6	-1	4	9	7	5	7
7	C	-14	-13	-8	-3	2	7	12	10	8
8	A	-16	-15	-10	-5	0	5	10	11	13

Optimal alignments:

AGCACAC-A
-TCACACTA

OR

AGCACAC-A
T-CACACTA

Score = 13

Introduction

Sequence Alignment

Global alignment

Local alignment

Exercise

Solution: Smith-Waterman for local alignment

Task: Align AGCACACCA and TCACACTA

i \ j	0	1	2	3	4	5	6	7	8	
0	0	0	0	0	0	0	0	0	0	
1	A	0	0	0	3	1	3	1	0	3
2	G	0	0	0	1	2	1	2	0	1
3	C	0	0	3	1	4	2	4	2	0
4	A	0	0	1	6	4	7	5	3	5
5	C	0	0	3	4	9	7	10	8	6
6	A	0	0	1	6	7	12	10	9	11
7	C	0	0	3	4	9	10	15	13	11
8	A	0	0	1	6	7	12	13	14	16

Optimal
alignment:

CACAC-A
CACACTA

Score = 16

Introduction
Sequence Alignment
Global alignment
Local alignment
Exercise