

Spring 2024
BIOS 477/877
Bioinformatics and Molecular Evolution
Lecture 6

BIOS477/877 L6 - 1

1

TODAY'S TOPICS

- **Pairwise Alignment**
 - Dotplot
 - Simple Alignments (scoring scheme)
 - Dynamic Programming Algorithm
- **Assignment 3 (a short one!)**
Due: February 11 (Sun)

BIOS477/877 L6 - 2

2

Pairwise Sequence Comparison

- **What can we learn from comparing sequences?**
 - **Their relatedness:**
 - **Protein/gene family classification**
 - **Evolutionary relationships/history**
 - **Functional domains/motifs:**
 - They are shared among a group of molecules sharing the same functions
- ➔ **Gene/protein functions and structures**

BIOS477/877 L6 - 3

3

Pairwise Sequence Comparison

- **When we compare similar sequences:**

AATCTGC ATTCTC

↖ ↗

Current Sequences

BIOS477/877 L6 - 4

4

Pairwise Sequence Comparison

- **Hypothesis: common descent**

Ancestral Sequence

???????

↙ ↘

Evolution Evolution

AATCTGC ATTCTC

↖ ↗

Current Sequences

BIOS477/877 L6 - 5

5

Pairwise Sequence Comparison

- **Hypothesis: common descent**

Ancestral Sequence

ATTCTGC

Substitution T G

↙ ↘

A Deletion

AATCTGC ATTCTC

↖ ↗

Current Sequences

BIOS477/877 L6 - 6

6

Pairwise Sequence Comparison

➤ Alignment represents a **hypothesis** regarding the evolutionary history of the two sequences

(Usually unknown)
ATCTGC

ATCT-C
: : : : :
AATCTGC

AATCTGC ATTCTC

BIOS477/877 L6 - 7

7

Pairwise Sequence Comparison

➤ Insertion or Deletion? ← **Indel**

ATTCTGC
A T G
AATCTGC ATTCTC

ATTCTC
Insertion T
G A
AATCTGC ATTCTC

ATTCT-C
: : : : :
AATCTGC

gap

BIOS477/877 L6 - 8

8

Pairwise Sequence Comparison: Dot Plot

➤ **Dot Plot (Dot Matrix)** visualizes sequence similarity

S1 \ S2	A	A	T	C	T	G	C
A	•	•					
T			•				
T							
C							
T							
C							

Fill up the rest of the matrix!

Go to Dotplot slides

BIOS477/877 L6 - 9

9

Pairwise Sequence Comparison: Dot Plot

- Identical sequences
- Internal repeats
- Single nucleotide repeats

Diagonal line

Parallel to diagonal
or
Perpendicular to diagonal
(single: inverted sequence)
(multiple: inverted repeats)

Vertical or horizontal lines

BIOS477/877 L6 - 10

10

Pairwise Sequence Comparison: Dot Plot

➤ Pairwise sequence comparison by dotplot

- **Dotlet JS** <https://dotlet.vital-it.ch>
- **Dotlet** (old Java version) <https://myhits.isb-sib.ch/cgi-bin/dotlet>
[For security reasons, old Java programs are not available within UNL network]
- **DotMatcher** (A program in EMBOSS)
<http://emboss.bioinformatics.nl/cgi-bin/emboss/dotmatcher>
(see course Web for other EMBOSS links)
- **JDotter: Java Dot Plot Alignments**
<http://pgrc.ink-gatersleben.de/jdotter/>
- **YASS: Visualization of local pairwise alignments**
<https://bioinfo.univ-lille.fr/vass/vass.php>
- **Dotter (part of SeqTools)**
<http://www.sanger.ac.uk/science/tools/seqtools>

See the Course website for more programs

BIOS477/877 L6 - 11

11

Protein domain detection by dot plot

Dotlet JS <https://dotlet.vital-it.ch>

[Help] (from the old version)
https://myhits.sib.swiss/ntil/dotlet/doc/dotlet_help.html

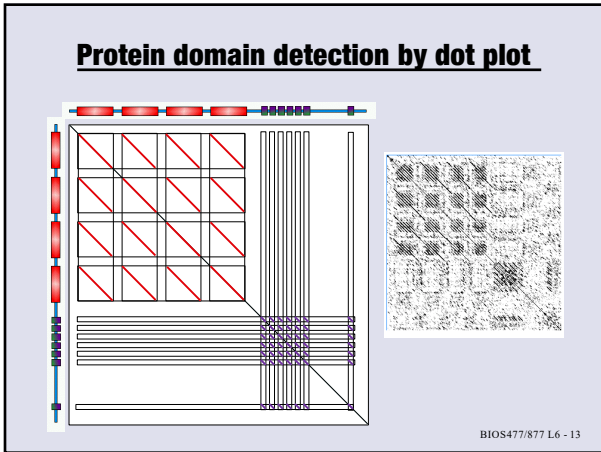
[Examples]
https://myhits.isb-sib.ch/ntil/dotlet/doc/dotlet_examples.html

Leading rich repeat EGF domain

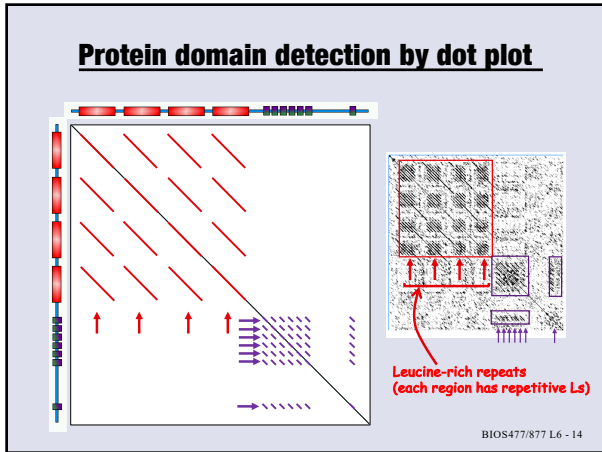
Dot plot can be used to compare:
- a sequence to itself to identify internal repeats, etc.
- two sequences to identify their similarity

BIOS477/877 L6 - 12

12



13



14

Simple Alignments

➤ Same length & no gap

```

AATCTATA
AAGATATA
    
```

→ Only one possible alignment

BIOS477/877 L6 - 15

15

Simple Alignments

➤ Varied length & no (internal) gap

```

AATCTATA
AAGATA
    
```

• If we only shift the shorter sequence, how many possible alignments?

BIOS477/877 L6 - 16

16

Simple Alignments

➤ Varied length & no (internal) gap

```

AATCTATA   AATCTATA   AATCTATA
AAGATA     AAGATA     AAGATA
    
```

→ Three possible alignments.

• Which alignment is the best?

BIOS477/877 L6 - 17

17

Simple Alignments

➤ Varied length & no (internal) gap

```

      ↓ ↓
AATCTATA
AAGATA
↑ ↑
    
```

Score

```

      ↓ ↓ ↓ ↓
AATCTATA
AAGATA
↑ ↑
    
```

Score

```

      ↓ ↓ ↓
AATCTATA
AAGATA
↑ ↑ ↑
    
```

Score

• Matches } Two elements
• Mismatches }

BIOS477/877 L6 - 18

18

Simple Alignments

➤ Varied length & no (internal) gap

Score ? Score ? Score ?

- Alignment Score =
 (match score) x (the number of **matched pairs**) +
 (mismatch score) x (the number of **mismatched pairs**)
 If **match score = 1**, **mismatch score = 0**

BIOS477/877 L6 - 19

19

Simple Alignments

➤ Varied length & no (internal) gap

Score ? Score ? Score ?

- Alignment Score =
 (match score) x (the number of **matched pairs**) +
 (mismatch score) x (the number of **mismatched pairs**)
 If **match score = 1**, **mismatch score = -2**

BIOS477/877 L6 - 21

21

Simple Alignments

➤ Varied length & gaps considered

• How many possible alignments?

BIOS477/877 L6 - 23

23

Simple Alignments

➤ Varied length & gaps considered

And many more ...

- If two gaps are allowed in the shorter sequence ...
 → 28 alignments possible!

BIOS477/877 L6 - 24

24

Simple Alignments

➤ Varied length & gaps considered

- **Matches**
- **Mismatches**
- **Gaps**

}

Three elements

BIOS477/877 L6 - 25

25

Simple Alignments

➤ Varied length & gaps considered

Score 3 Score 3 Score 1

- Alignment Score =
 (match score) x (the number of **matched pairs**) +
 (mismatch score) x (the number of **mismatched pairs**) +
 (gap penalty) x (the number of **gaps**)
 If **match score = 1**, **mismatch score = 0**, **gap penalty = -1**

BIOS477/877 L6 - 26

26

Pairwise Alignment Problem

- To find the best alignment(s) between two sequences
 - **Exhaustive search method**
 - Examine all possible alignments
 - Develop a method to score each alignment (Match score, mismatch score, gap penalty)
 - Choose the best (optimal) alignment based on the score
 - **Not feasible** for long sequences
(e.g., 100bp x 95bp with 5 gaps only in the shorter sequence would require testing ~ 55 million possible alignments!)

BIOS477/877 L6 - 27

27

Dynamic Programming Algorithm

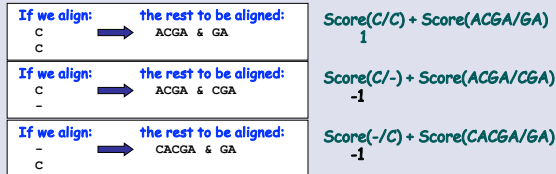
- What is **Dynamic Programming**?
 - An algorithm in which an optimization problem is solved by saving the optimal scores for the solution of every sub-problem instead of recalculating them repetitively.
- A problem is broken apart into reasonably sized **sub-problems**.
- These partial results are used to compute the final answer.

BIOS477/877 L6 - 28

28

Dynamic Programming Algorithm

- Needleman & Wunsch (1970) applied the Dynamic Programming approach to the sequence alignment problem.
- Problem: Align **CACGA** and **CGA**
(Match: 1, mismatch: 0, gap penalty: -1)

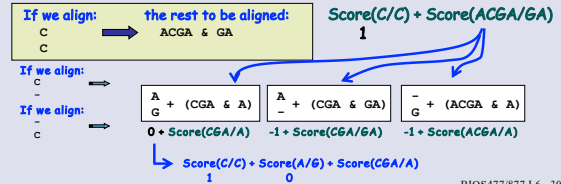


BIOS477/877 L6 - 29

29

Dynamic Programming Algorithm

- Needleman & Wunsch (1970) applied the Dynamic Programming approach to the sequence alignment problem.
- Problem: Align **CACGA** and **CGA**
(Match: 1, mismatch: 0, gap penalty: -1)

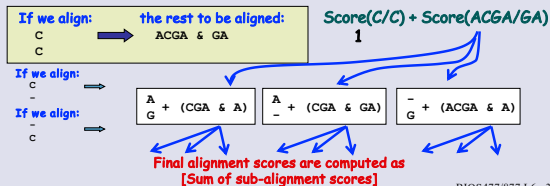


BIOS477/877 L6 - 30

30

Dynamic Programming Algorithm

- Needleman & Wunsch (1970) applied the Dynamic Programming approach to the sequence alignment problem.
- Problem: Align **CACGA** and **CGA**
(Match: 1, mismatch: 0, gap penalty: -1)



BIOS477/877 L6 - 31

31

Dynamic Programming Algorithm

- Three steps for sequence alignment (Eddy, 2004)

1. Initialization
2. Matrix filling (Scoring)
3. Traceback (Alignment)

- Sequences a and b
- Simple scoring scheme:

- $s(a_i, b_j) = 2$ ($a_i = b_j$, match)
- $s(a_i, b_j) = -1$ ($a_i \neq b_j$, mismatch)
- $w = -2$ (gap penalty)

	[a]	A	A	T	C	T	G	C
[b]								
A								
T								
T								
C								
T								
C								

BIOS477/877 L6 - 32

32

Dynamic Programming Algorithm

Initialization step

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2							
T	-4							
T	-6							
C	-8							
T	-10							
C	-12							

Fill the first row and column with multiples of the gap penalty (-2)

BIOS477/877 L6 - 33

33

Dynamic Programming Algorithm

Matrix filling step

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2							
T	-4							
T	-6							
C	-8							
T	-10							
C	-12							

Start from this cell

BIOS477/877 L6 - 34

34

Dynamic Programming Algorithm

Matrix filling step

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A
[b]	0	-2
A	-2	

Start from the corner,
3 ways to reach the current cell

3 alignment possibilities

BIOS477/877 L6 - 35

35

Dynamic Programming Algorithm

Matrix filling step

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A
[b]	0	-2
A	-2	

3 alignment possibilities

1. A
A Aligned
2. -
A Gap in a
3. A
- Gap in b

Vertical or horizontal movement indicates adding a gap

BIOS477/877 L6 - 36

36

Dynamic Programming Algorithm

Matrix filling step

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

$j \backslash i$	[a]	A
[b]	$S_{i,j-1}$ 0	$S_{i,j-1}$ -2
A	$S_{i,j}$ -2	$S_{i,j}$

3 alignment possibilities

1. A
A Aligned (match)
 $S_{i,j} = S_{i,j-1} + s(a_i, b_j)$
2. -
A Gap in a
 $S_{i,j} = S_{i,j-1} + w$
3. A
- Gap in b
 $S_{i,j} = S_{i,j} + w$

Cell score

- 1. $S_{i,j} = S_{i,j-1} + s(a_i, b_j)$
- 2. $S_{i,j} = S_{i,j-1} + w$
- 3. $S_{i,j} = S_{i,j} + w$

Score in the previous cell

BIOS477/877 L6 - 37

37

Dynamic Programming Algorithm

Matrix filling step

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

$j \backslash i$	[a]	A
[b]	$S_{i,j-1}$ 0	$S_{i,j-1}$ -2
A	$S_{i,j}$ -2	$S_{i,j}$

3 alignment possibilities

1. A
A Aligned (match)
 $S_{i,j} = 0 + s(a_i, b_j)$
2. -
A Gap in a
 $S_{i,j} = -2 + w$
3. A
- Gap in b
 $S_{i,j} = -2 + w$

Cell score

- 1. $S_{i,j} = S_{i,j-1} + s(a_i, b_j)$
- 2. $S_{i,j} = S_{i,j-1} + w$
- 3. $S_{i,j} = S_{i,j} + w$

Score in the previous cell

BIOS477/877 L6 - 38

38

Dynamic Programming Algorithm

➤ Matrix filling step

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	$j \backslash i$	[a]	A	
[b]	$S_{i,j-1}$	0	$S_{i,j-1}$	-2
A	$S_{i-1,j}$	-2	$S_{i,j}$	$S_{i,j}$

3 alignment possibilities

1. A Aligned (match)
 $S_{i,j} = 0 + 2 = 2$ ← Maximum
2. - Gap in a
 $S_{i,j} = -2 - 2 = -4$
3. A Gap in b
 $S_{i,j} = -2 - 2 = -4$

Cell score

1. $S_{i,j} = S_{i-1,j-1} + s(a,b)$
2. $S_{i,j} = S_{i,j-1} + w$
3. $S_{i,j} = S_{i-1,j} + w$

Score in the previous cell

BIOS477/877 L6 - 39

39

Dynamic Programming Algorithm

➤ Matrix filling step

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	$j \backslash i$	[a]	A	
[b]	$S_{i,j-1}$	0	$S_{i,j-1}$	-2
A	$S_{i-1,j}$	-2	2	$S_{i,j}$

Cell score

1. $S_{i,j} = S_{i-1,j-1} + s(a,b)$
2. $S_{i,j} = S_{i,j-1} + w$
3. $S_{i,j} = S_{i-1,j} + w$

Choose the maximum score and its path

BIOS477/877 L6 - 40

40

Dynamic Programming Algorithm

➤ Matrix filling step

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2	2						
T	-4							
T	-6							
C	-8							
T	-10							
C	-12							

Keep the maximum score path

BIOS477/877 L6 - 41

41

Dynamic Programming Algorithm

➤ Matrix filling step

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2	2	-2					
T	-4							
T	-6							
C	-8							
T	-10							
C	-12							

Calculate the next cell score

BIOS477/877 L6 - 42

42

Dynamic Programming Algorithm

➤ Matrix filling step

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2	2	0	-2				
T	-4							
T	-6							
C	-8							
T	-10							
C	-12							

BIOS477/877 L6 - 43

43

Dynamic Programming Algorithm

➤ Matrix filling step

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2	2	0	-2	-2			
T	-4							
T	-6							
C	-8							
T	-10							
C	-12							

BIOS477/877 L6 - 44

44

Dynamic Programming Algorithm

➤ Matrix filling step

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2	2	0	-2	-4	-6	-8	-10
T	-4							
T	-6							
C	-8							
T	-10							
C	-12							

BIOS477/877 L6 - 45

45

Dynamic Programming Algorithm

➤ Matrix filling step

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2	2	0	-2	-4	-6	-8	-10
T	-4							
T	-6							
C	-8							
T	-10							
C	-12							

BIOS477/877 L6 - 46

46

Dynamic Programming Algorithm

➤ Matrix filling step

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2	2	0	-2	-4	-6	-8	-10
T	-4							
T	-6							
C	-8							
T	-10							
C	-12							

BIOS477/877 L6 - 47

47

Dynamic Programming Algorithm

➤ Matrix filling step

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2	2	0	-2	-4	-6	-8	-10
T	-4	0	1	2	0	-2	-4	-6
T	-6	-2	-1	3	1	2	0	-2
C	-8	-4	-3	1	5	3	1	2
T	-10	-6	-5	-1	3	7	5	3
C	-12	-8	-7	-3	1	5	6	7

BIOS477/877 L6 - 48

48

Dynamic Programming Algorithm

➤ Traceback

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2	2	0	-2	-4	-6	-8	-10
T	-4	0	1	2	0	-2	-4	-6
T	-6	-2	-1	3	1	2	0	-2
C	-8	-4	-3	1	5	3	1	2
T	-10	-6	-5	-1	3	7	5	3
C	-12	-8	-7	-3	1	5	6	7

Trace back the arrows

Start from here

BIOS477/877 L6 - 49

49

Dynamic Programming Algorithm

➤ Traceback

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2	2	0	-2	-4	-6	-8	-10
T	-4	0	1	2	0	-2	-4	-6
T	-6	-2	-1	3	1	2	0	-2
C	-8	-4	-3	1	5	3	1	2
T	-10	-6	-5	-1	3	7	5	3
C	-12	-8	-7	-3	1	5	6	7

Trace back the arrows

Start from here

BIOS477/877 L6 - 50

50

Dynamic Programming Algorithm

➤ Traceback

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2	2	0	-2				
T	-4	0	1	2				
T	-6	-2	-1	3	1	2	0	-2
C	-8	-4	-3	1	5	3	1	2
T	-10	-6	-5	-1	3	7	5	3
C	-12	-8	-7	-3	1	5	6	7

Trace back the arrows

Start from here

Follow the arrows from the right-bottom corner to the left-top corner

BIOS477/877 L6 - 51

51

Dynamic Programming Algorithm

➤ Traceback

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2	2	0	-2				
T	-4	0	1	2				
T	-6	-2	-1	3	1	2	0	-2
C	-8	-4	-3	1	5	3	1	2
T	-10	-6	-5	-1	3	7	5	3
C	-12	-8	-7	-3	1	5	6	7

Trace back the arrows

Start from here

Follow the arrows from the right-bottom corner to the left-top corner

BIOS477/877 L6 - 52

52

Dynamic Programming Algorithm

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2	2	0	-2	-4	-6	-8	-10
T	-4	0	1	2	0	-2	-4	-6
T	-6	-2	-1	3	1	2	0	-2
C	-8	-4	-3	1	5	3	1	2
T	-10	-6	-5	-1	3	7	5	3
C	-12	-8	-7	-3	1	5	6	7

AATCTG Horizontal movement → a gap in b
ATTCT Vertical movement → a gap in a

BIOS477/877 L6 - 53

53

Dynamic Programming Algorithm

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2	2	0	-2	-4	-6	-8	-10
T	-4	0	1	2	0	-2	-4	-6
T	-6	-2	-1	3	1	2	0	-2
C	-8	-4	-3	1	5	3	1	2
T	-10	-6	-5	-1	3	7	5	3
C	-12	-8	-7	-3	1	5	6	7

AATCTGC
ATTCT-C

BIOS477/877 L6 - 54

54

Dynamic Programming Algorithm

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2	2	0	-2	-4	-6	-8	-10
T	-4	0	1	2	0	-2	-4	-6
T	-6	-2	-1	3	1	2	0	-2
C	-8	-4	-3	1	5	3	1	2
T	-10	-6	-5	-1	3	7	5	3
C	-12	-8	-7	-3	1	5	6	7

AATCTGC Match (+2) × 5 + Alignment Score
ATTCT-C Mismatch (-1) × 1 + Alignment Score
 Gap (-2) × 1 = 7

BIOS477/877 L6 - 55

55

Dynamic Programming Algorithm

➤ Initialization step [Gap penalty: -2]

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

Fill the first row and column with multiples of the gap penalty (-2)

BIOS477/877 L6 - 56

56

Dynamic Programming Algorithm

Initialization step

[Gap penalty: -2]

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

Equivalent to adding only gaps in the 1st or 2nd sequence

BIOS477/877 L6 - 57

57

Terminal Gaps

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

Gaps in the 2nd sequence

AATCTGC

BIOS477/877 L6 - 58

58

Terminal Gaps

	A	A	T	C	T	G	C	
	0	-2	-4	-6	-8	-10	-12	-14
A	-2							-16
T	-4							-18
T	-6							-20
C	-8							-22
T	-10							-24
C	-12							-26

Gaps in the 1st sequence

AATCTGC-----
-----ATTCTC

BIOS477/877 L6 - 59

59

Terminal Gaps

	A	A	T	C	T	G	C	
	0	-2	-4	-6	-8	-10	-12	-14
A	-2							
T	-4							
T	-6							
C	-8							
T	-10							
C	-12	-14	-16	-18	-20	-22	-24	-26

Gaps in the 1st sequence

Gaps in the 2nd sequence

AATCTGC-----
-----ATTCTC -----AATCTGC
ATTCTC-----

BIOS477/877 L6 - 60

60

Dynamic Programming Algorithm

Traceback

$s(a,b) = 2$ ($a=b$, match)
 $s(a,b) = -1$ ($a \neq b$, mismatch)
 $w = -2$ (gap penalty)

	[a]	A	A	T	C	T	G	C
[b]	0	-2	-4	-6	-8	-10	-12	-14
A	-2	2	0	-2	-4	-6	-8	-10
T	-4	0	1	2	0	-2	-4	-6
T	-6	-2	-1	3	1	2	0	-2
C	-8	-4	-3	1	5	3	1	2
T	-10	-6	-5	-1	3	7	5	3
C	-12	-8	-7	-3	1	5	6	7

BIOS477/877 L6 - 61

61

Dynamic Programming Exercise

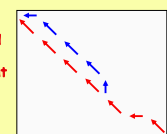
Assignment 3

Perform a small example of pairwise alignment using the dynamic programming algorithm

Due: February 11 (Sun)

We will go over it next Tuesday

Note: There could be more than one optimum alignment paths!



BIOS477/877 L6 - 62

62