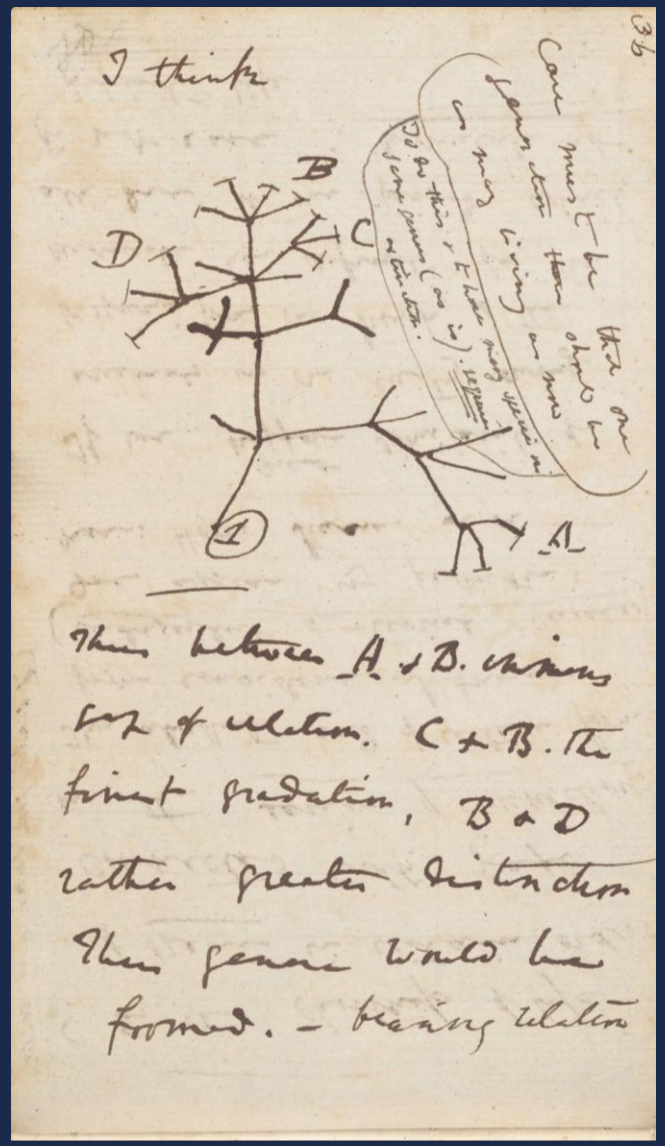
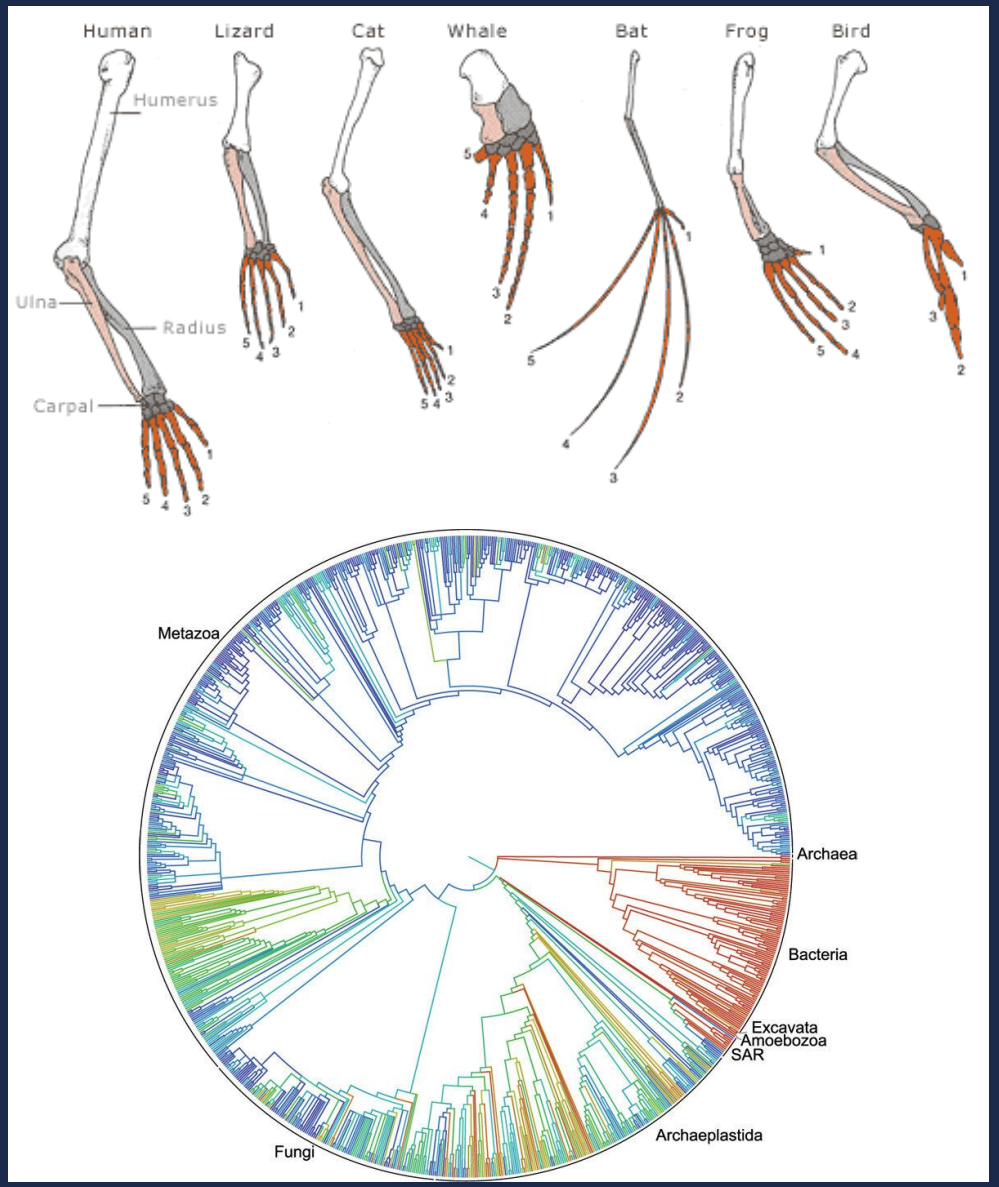
Aligning the Unseen

A Journey Through Sequence Alignment Algorithms

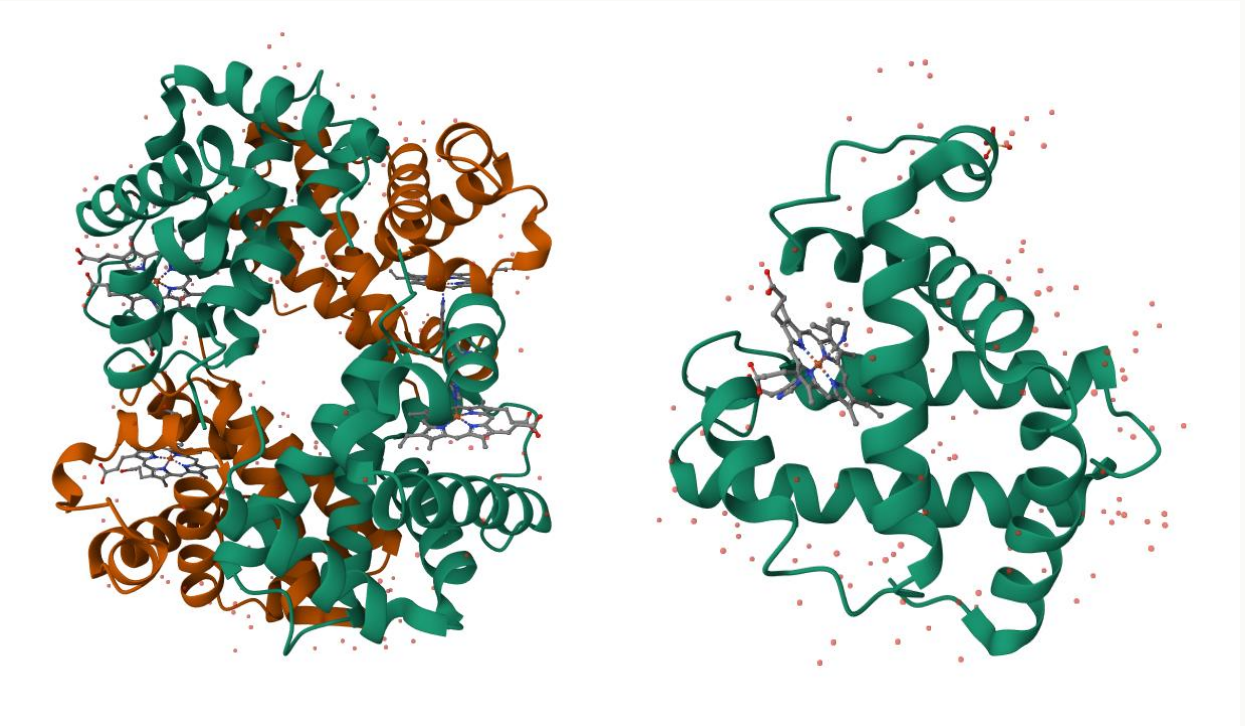
Mentor: Prof. Paolo Boldi


I think
B
C
D
A


Human
Lizard
Cat
Whale
Bat
Frog
Bird
Humerus
Ulna
Radius
Carpal
Metazoa
Archaea
Bacteria
Excavata
Amoebozoa
SAR
Archaeplastida
Fungi


137
XVIII, 2. 106. Leguminosae.
453. Pisum sativum L. Brech-Erbse.

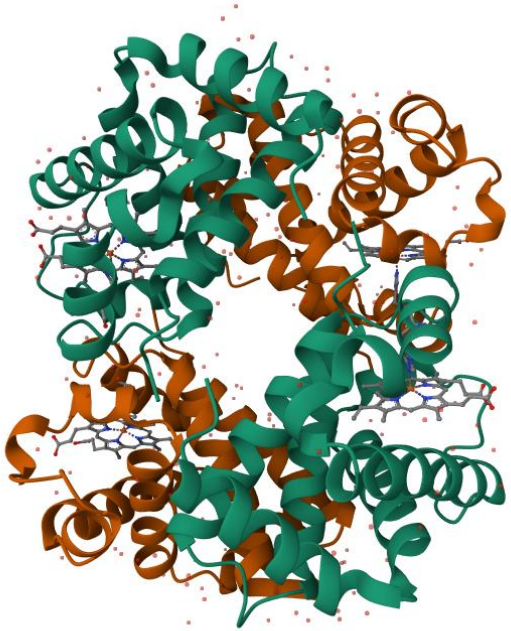
How do we compare sequences?



Human β-Hemoglobin

Whale Myoglobin

How do we compare sequences?



Human β-Hemoglobin

Whale Myoglobin

- Are these proteins related?

How do we compare sequences?

Human β-Hemoglobin

Whale Myoglobin

- Are these proteins related?
- Do they share a common ancestral gene?

How do we compare sequences?

Human β-Hemoglobin

Whale Myoglobin

- Are these proteins related?
- Do they share a common ancestral gene?
- How do we quantify their similarity?

How do we compare sequences?

Human β-Hemoglobin

Whale Myoglobin

- Are these proteins related?
- Do they share a common ancestral gene?
- How do we quantify their similarity?

We need an algorithm.

How do we compare sequences?

HBB	**MYG**
...	...
Trp	Trp
Gly	Ala
Lys	Lys
Val	Val
Asn	Glu
Val	Ala
Asp	Asp
Glu	Val
Val	Ala
Gly	Gly
...	...

- Are these proteins related?
- Do they share a common ancestral gene?
- How do we quantify their similarity?

We need an algorithm.

How do we compare sequences?

HBB	**MYG**
...	...
Trp	Trp
Gly	Ala
Lys	Lys
Val	Val
Asn	Glu
Val	Ala
Asp	Asp
Glu	Val
Val	Ala
Gly	Gly
...	...

- Are these sequences related?
- Do they share a common structure?
- How do we measure their similarity?

We need an algorithm.

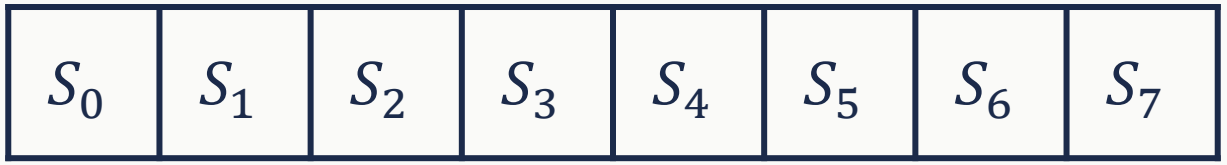
Fitch (1966)

Fitch (1966)

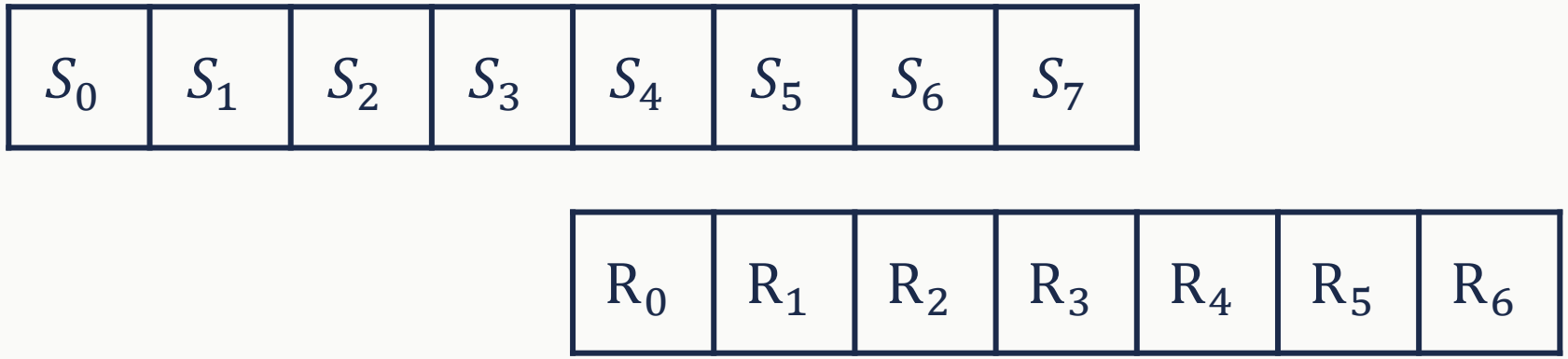
- Use the genetic code as a distance measure
- Minimum mutations required between amino acid pairs
- Sliding window + statistical threshold

	A	C	E	F	G	H	I	L	M	N	O	P	Q	R	S	T	U	V	W	Y
Asp	0	2	2	2	1	1	2	1	3	1	1	2	2	2	2	3	2	1	2	1
Cys	2	0	2	1	3	2	3	2	3	2	1	2	2	1	1	1	2	1	2	1
Thr	2	2	0	2	2	2	1	1	1	1	2	1	2	1	1	2	2	2	1	2
Phe	2	1	2	0	3	2	3	2	2	2	1	2	2	2	1	2	1	1	1	2
Glu	1	3	2	3	0	2	1	1	2	2	2	2	1	2	2	2	2	1	3	1
His	1	2	2	2	2	0	2	2	3	1	1	1	1	1	2	3	1	2	2	2
Lys	2	3	1	3	1	2	0	2	1	1	2	2	1	1	2	2	2	2	2	2
Ala	1	2	1	2	1	2	2	0	2	2	2	1	2	2	1	2	2	1	2	1
Met	3	3	1	2	2	3	1	2	0	2	3	2	2	1	2	2	1	1	1	2
Asn	1	2	1	2	2	1	1	2	2	0	1	2	2	2	1	3	2	2	1	2
Tyr	1	1	2	1	2	1	2	2	3	1	0	2	1	2	1	2	2	2	2	2
Pro	2	2	1	2	2	1	2	1	2	2	2	0	1	1	1	2	1	2	2	2
Gln	2	2	2	2	1	1	1	2	2	2	1	1	0	1	1	1	1	2	3	2
Arg	2	1	1	2	2	1	1	2	1	2	2	1	1	0	1	1	1	2	2	1
Ser	2	1	1	1	2	2	2	1	2	1	1	1	1	1	0	1	1	2	1	1
Trp	3	1	2	2	2	3	2	2	2	3	2	2	1	1	1	0	1	2	3	1
Leu	2	2	2	1	2	1	2	2	1	2	2	1	1	1	1	1	0	1	1	1
Val	1	1	2	1	1	2	2	1	1	2	2	2	2	2	2	2	1	0	1	1
Ile	2	2	1	1	3	2	2	2	1	1	2	2	3	2	1	3	1	1	0	2
Gly	1	1	2	2	1	2	2	1	2	2	2	2	2	1	1	1	1	1	2	0

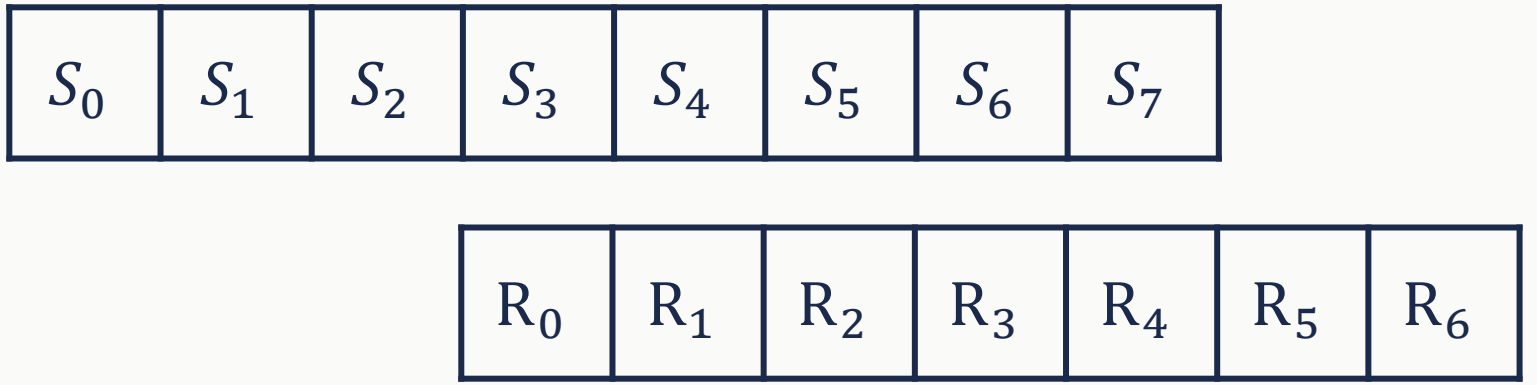
Needleman & Blair (1969)



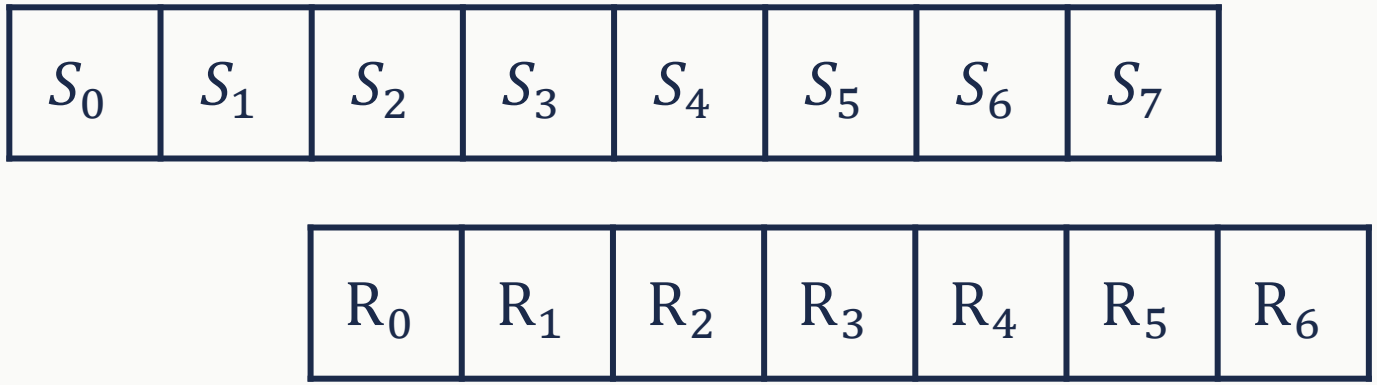
Needleman & Blair (1969)



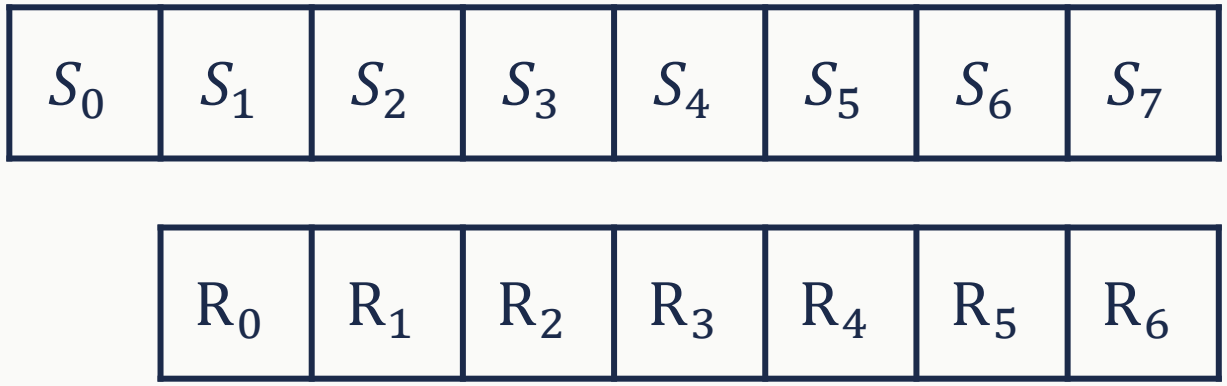
Needleman & Blair (1969)



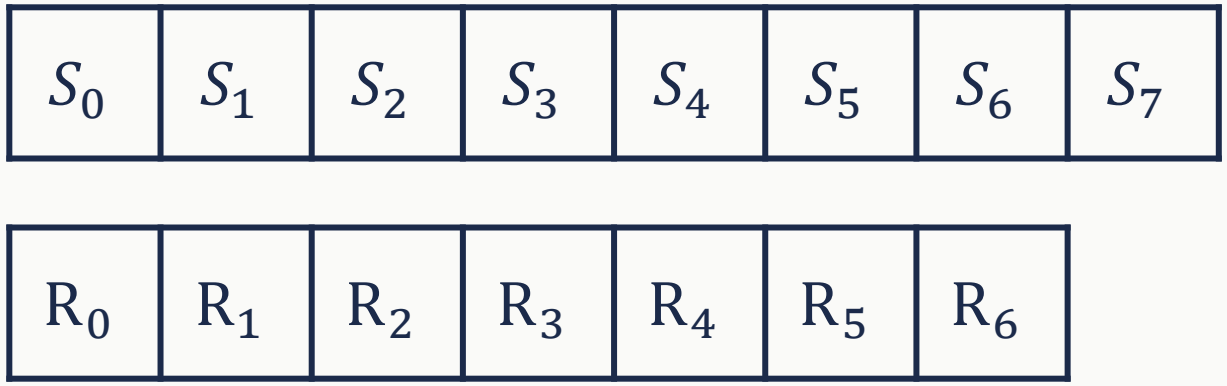
Needleman & Blair (1969)



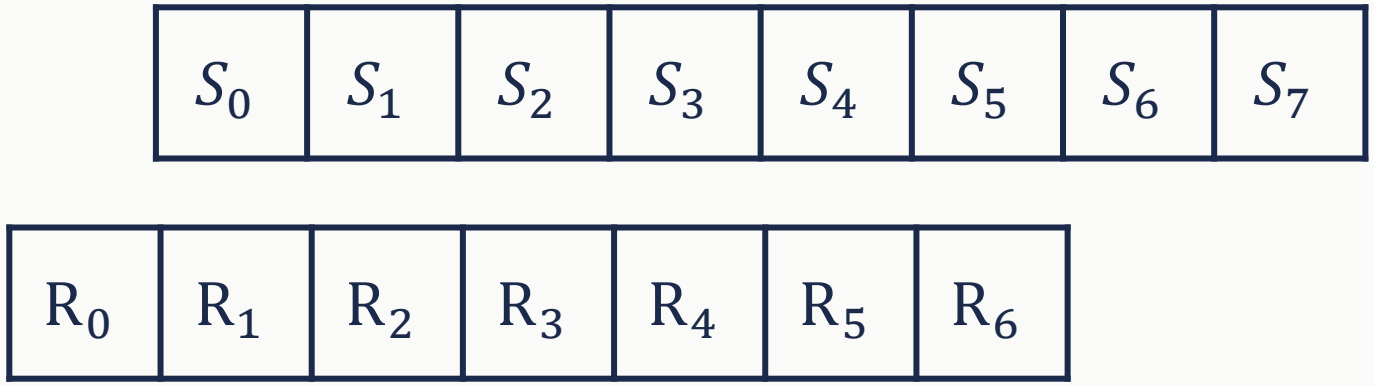
Needleman & Blair (1969)



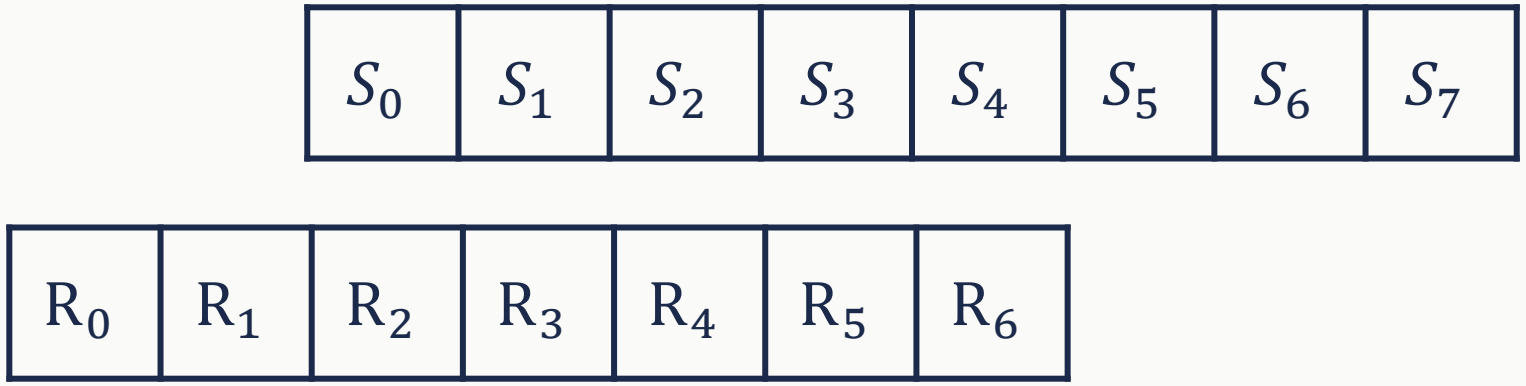
Needleman & Blair (1969)



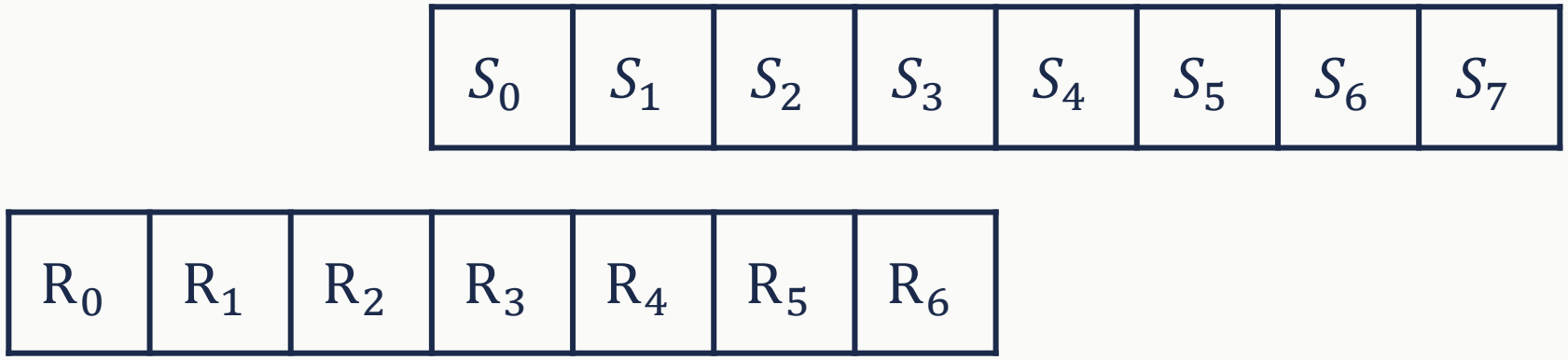
Needleman & Blair (1969)



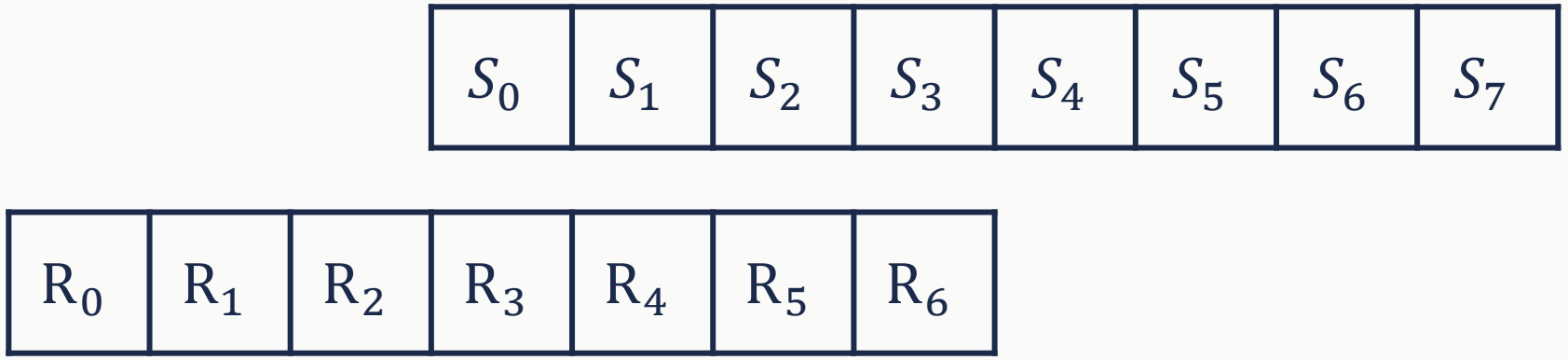
Needleman & Blair (1969)



Needleman & Blair (1969)

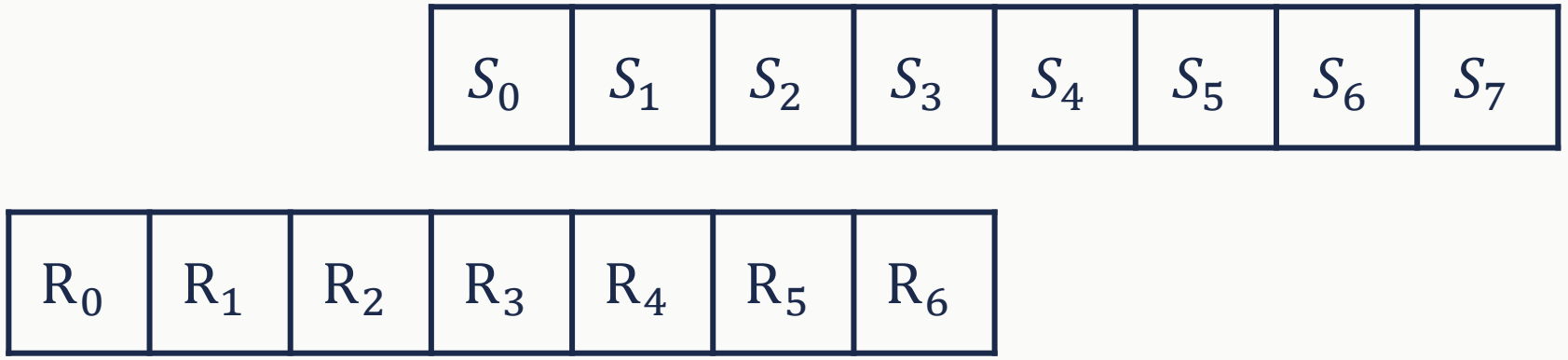


Needleman & Blair (1969)



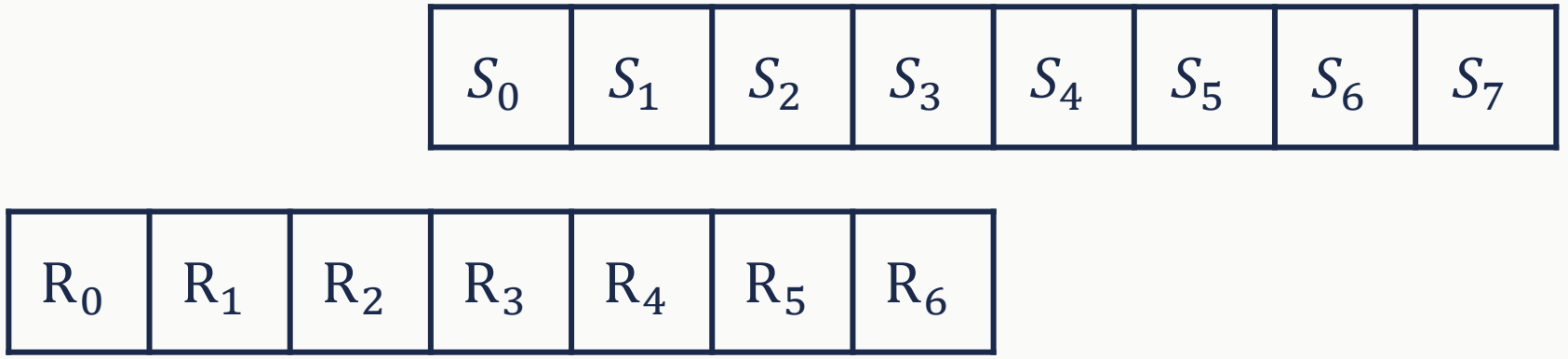
N_{aa}

Needleman & Blair (1969)



N_{aa} P_f

Needleman & Blair (1969)



N_{aa} P_f N_f

What's missing?

Enter: Dynamic Programming

Needleman-Wunsch (1970)

2 strings S & R

Needleman-Wunsch (1970)

2 strings S & R

ES: Match Score

Needleman-Wunsch (1970)

2 strings S & R

ES: Match Score

DS: Mismatch Score

Needleman-Wunsch (1970)

2 strings S & R

ES: Match Score

GP: Gap Penalty

DS: Mismatch Score

Needleman-Wunsch (1970)

$$SIM_{ij} = \begin{cases} ES & if \quad S_i = R_j \\ DS & if \quad S_i \neq R_j \end{cases}$$

Needleman-Wunsch (1970)

$$SIM_{ij} = \begin{cases} ES & if \quad S_i = R_j \\ DS & if \quad S_i \neq R_j \end{cases}$$

	R_0	R_1	R_2	R_3	R_4	R_5	R_6
S_0							
S_1							
S_2							
S_3							
S_4							
S_5							

Needleman-Wunsch (1970)

$$SIM_{ij} = \begin{cases} ES & if \quad S_i = R_j \\ DS & if \quad S_i \neq R_j \end{cases}$$

	R_0	R_1	R_2	R_3	R_4	R_5	R_6
S_0							
S_1							
S_2							
S_3							
S_4							
S_5							

$ES: 1$

$DS: 0$

$GP: -1$

Needleman-Wunsch (1970)

$$SIM_{ij} = \begin{cases} ES & if \quad S_i = R_j \\ DS & if \quad S_i \neq R_j \end{cases}$$

S: elixir

R: elicsir

ES: 1

DS: 0

GP: -1

	R_0	R_1	R_2	R_3	R_4	R_5	R_6
S_0							
S_1							
S_2							
S_3							
S_4							
S_5							

Needleman-Wunsch (1970)

$$SIM_{ij} = \begin{cases} ES & if \quad S_i = R_j \\ DS & if \quad S_i \neq R_j \end{cases}$$

S: elixir

R: elicsir

ES: 1

DS: 0

GP: -1

	e	l	i	c	s	i	r
e							
l							
i							
x							
i							
r							

Needleman-Wunsch (1970)

$$SIM_{ij} = \begin{cases} ES & if \quad S_i = R_j \\ DS & if \quad S_i \neq R_j \end{cases}$$

S: elixir

R: elicsir

ES: 1

DS: 0

GP: -1

	e	l	i	c	s	i	r
e	1						
l		1					
i			1			1	
x							
i			1			1	
r							1

Needleman-Wunsch (1970)

$$SIM_{ij} = \begin{cases} ES & if \quad S_i = R_j \\ DS & if \quad S_i \neq R_j \end{cases}$$

S: elixir

R: elicsir

ES: 1

DS: 0

GP: -1

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

Needleman-Wunsch (1970)

$$SIM_{ij} = \begin{cases} ES & if \quad S_i = R_j \\ DS & if \quad S_i \neq R_j \end{cases}$$

S: elixir

R: elicsir

ES: 1

DS: 0

GP: −1

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

Needleman-Wunsch (1970)

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

ES: 1

DS: 0

GP: -1

Needleman-Wunsch (1970)

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

ES: 1

DS: 0

GP: -1

MS

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								
S_1								
S_2								
S_3								
S_4								
S_5								

Needleman-Wunsch (1970)

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

MS

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								$6GP$
S_1								$5GP$
S_2								$4GP$
S_3								$3GP$
S_4								$2GP$
S_5								GP
	$7GP$	$6GP$	$5GP$	$4GP$	$3GP$	$2GP$	GP	0

Needleman-Wunsch (1970)

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

ES: 1

DS: 0

GP: -1

MS

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								$6GP$
S_1								$5GP$
S_2								$4GP$
S_3								$3GP$
S_4								$2GP$
S_5								GP
	$7GP$	$6GP$	$5GP$	$4GP$	$3GP$	$2GP$	GP	0

Needleman-Wunsch (1970)

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

MS

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								$6GP$
S_1								$5GP$
S_2								$4GP$
S_3								$3GP$
S_4								$2GP$
S_5								GP
	$7GP$	$6GP$	$5GP$	$4GP$	$3GP$	$2GP$	GP	0

Needleman-Wunsch (1970)

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

MS

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								$6GP$
S_1								$5GP$
S_2								$4GP$
S_3								$3GP$
S_4								$2GP$
S_5								GP
	$7GP$	$6GP$	$5GP$	$4GP$	$3GP$	$2GP$	GP	0

Needleman-Wunsch (1970)

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

MS

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								$6GP$
S_1								$5GP$
S_2								$4GP$
S_3								$3GP$
S_4								$2GP$
S_5								GP
	$7GP$	$6GP$	$5GP$	$4GP$	$3GP$	$2GP$	GP	0

Needleman-Wunsch (1970)

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

MS

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								$6GP$
S_1								$5GP$
S_2								$4GP$
S_3								$3GP$
S_4								$2GP$
S_5								GP
	$7GP$	$6GP$	$5GP$	$4GP$	$3GP$	$2GP$	GP	0

Needleman-Wunsch (1970)

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

MS

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								$6GP$
S_1								$5GP$
S_2								$4GP$
S_3								$3GP$
S_4								$2GP$
S_5								GP
	$7GP$	$6GP$	$5GP$	$4GP$	$3GP$	$2GP$	GP	0

Needleman-Wunsch (1970)

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

MS

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								$6GP$
S_1								$5GP$
S_2								$4GP$
S_3								$3GP$
S_4								$2GP$
S_5								GP
	$7GP$	$6GP$	$5GP$	$4GP$	$3GP$	$2GP$	GP	0

Needleman-Wunsch (1970)

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

MS

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								$6GP$
S_1								$5GP$
S_2								$4GP$
S_3								$3GP$
S_4								$2GP$
S_5								GP
	$7GP$	$6GP$	$5GP$	$4GP$	$3GP$	$2GP$	GP	0

Needleman-Wunsch (1970)

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

MS

S: elixir

R: elicsir

ES: 1

DS: 0

GP: -1

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								$6GP$
S_1								$5GP$
S_2								$4GP$
S_3								$3GP$
S_4								$2GP$
S_5								GP
	$7GP$	$6GP$	$5GP$	$4GP$	$3GP$	$2GP$	GP	0

Needleman-Wunsch (1970)

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$MS_{i,j}$ «*Delete R_i - insert gap in S*»

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

S: elixir

R: elicsir

ES: 1

DS: 0

GP: -1

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								$6GP$
S_1								$5GP$
S_2								$4GP$
S_3								$3GP$
S_4								$2GP$
S_5								GP
	$7GP$	$6GP$	$5GP$	$4GP$	$3GP$	$2GP$	GP	0

Needleman-Wunsch (1970)

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$MS_{i,j}$ «*Delete S_j - insert gap in R*»

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								6GP
S_1								5GP
S_2								4GP
S_3								3GP
S_4								2GP
S_5								GP
	7GP	6GP	5GP	4GP	3GP	2GP	GP	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$MS_{i,j}$ «*Align S_i with R_j - pay ES or DS*»

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								6GP
S_1								5GP
S_2								4GP
S_3								3GP
S_4								2GP
S_5								GP
	7GP	6GP	5GP	4GP	3GP	2GP	GP	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$MS_{i,j} = max\{RIGHT_{i,j}; DOWN_{i,j}; DIAG_{i,j}\}$$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								6GP
S_1								5GP
S_2								4GP
S_3								3GP
S_4								2GP
S_5								GP
	7GP	6GP	5GP	4GP	3GP	2GP	GP	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{RIGHT_{i,j}; DOWN_{i,j}; DIAG_{i,j}\}$$

MS

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								$6GP$
S_1								$5GP$
S_2								$4GP$
S_3								$3GP$
S_4								$2GP$
S_5								GP
	$7GP$	$6GP$	$5GP$	$4GP$	$3GP$	$2GP$	GP	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	R_0	R_1	R_2	R_3	R_4	R_5	R_6	
S_0								$6GP$
S_1								$5GP$
S_2								$4GP$
S_3								$3GP$
S_4								$2GP$
S_5								GP
	$7GP$	$6GP$	$5GP$	$4GP$	$3GP$	$2GP$	GP	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e								$6GP$
l								$5GP$
i								$4GP$
x								$3GP$
i								$2GP$
r								GP
	$7GP$	$6GP$	$5GP$	$4GP$	$3GP$	$2GP$	GP	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e								−6
l								−5
i								−4
x								−3
i								−2
r								−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e								−6
l								−5
i								−4
x								−3
i								−2
r							?	−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$RIGHT = -1 - 1$

$DOWN = -1 - 1$

$DIAG = 0 + 1$

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$RIGHT_{i,j} = MS_{i,j+1} + GP$

$DOWN_{i,j} = MS_{i+1,j} + GP$

$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$

$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$

MS

	e	l	i	c	s	i	r	
e								−6
l								−5
i								−4
x								−3
i								−2
r							?	−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$RIGHT = -2$

$DOWN = -2$

$DIAG = 1$

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e								−6
l								−5
i								−4
x								−3
i								−2
r							?	−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$RIGHT = -3$

$DOWN = 0$

$DIAG = -1$

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e								−6
l								−5
i								−4
x								−3
i							?	−2
r							1	−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$RIGHT = 0$

$DOWN = -3$

$DIAG = -1$

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e								−6
l								−5
i								−4
x								−3
i							0	−2
r						?	1	−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e								−6
l								−5
i								−4
x							?	−3
i						?	0	−2
r					?	0	1	−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e								−6
l								−5
i								−4
x							-1	−3
i						2	0	−2
r					-1	0	1	−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e	4	2*	1*	0*	-1*	-2	-4	−6
l	2	3	2	1*	0*	-1	-3	−5
i	0	1	2	2	1*	0*	-2	−4
x	-2*	-1*	0*	1*	2	1	-1	−3
i	-3	-2	-1*	0	1	2	0	−2
r	-5	-4	-3	-2	-1	0	1	−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e	4	2*	1*	0*	-1*	-2	-4	−6
l	2	3	2	1*	0*	-1	-3	−5
i	0	1	2	2	1*	0*	-2	−4
x	-2*	-1*	0*	1*	2	1	-1	−3
i	-3	-2	-1*	0	1	2	0	−2
r	-5	-4	-3	-2	-1	0	1	−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$\bar{S}$: ε

$\bar{R}$: ε

S: elixir

R: elicsir

ES: 1

DS: 0

GP: -1

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e	4	2*	1*	0*	-1*	-2	-4	−6
l	2	3	2	1*	0*	-1	-3	−5
i	0	1	2	2	1*	0*	-2	−4
x	-2*	-1*	0*	1*	2	1	-1	−3
i	-3	-2	-1*	0	1	2	0	−2
r	-5	-4	-3	-2	-1	0	1	−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$\bar{S}$: e

$\bar{R}$: e

S: elixir

R: elicsir

ES: 1

DS: 0

GP: -1

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e	**4**		1*	0*	-1*	-2	-4	−6
l		**3**	2	1*	0*	-1	-3	−5
i	0	1	2	2	1*	0*	-2	−4
x	-2*	-1*	0*	1*	2	1	-1	−3
i	-3	-2	-1*	0	1	2	0	−2
r	-5	-4	-3	-2	-1	0	1	−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$\bar{S}$: el

$\bar{R}$: el

S: elixir

R: elicsir

ES: 1

DS: 0

GP: -1

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e	**4**			0*	-1*	-2	-4	−6
l		**3**		1*	0*	-1	-3	−5
i			**2**	2	1*	0*	-2	−4
x	-2*	-1*	0*	1*	2	1	-1	−3
i	-3	-2	-1*	0	1	2	0	−2
r	-5	-4	-3	-2	-1	0	1	−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$\bar{S}$: eli

$\bar{R}$: eli

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e	**4**				-1*	-2	-4	−6
l		**3**			0*	-1	-3	−5
i			**2**		1*	0*	-2	−4
x				**1***	2	1	-1	−3
i	-3	-2	-1*	0	1	2	0	−2
r	-5	-4	-3	-2	-1	0	1	−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$\bar{S}$: elix

$\bar{R}$: elic

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e	**4**					-2	-4	−6
l		**3**				-1	-3	−5
i			**2**			0*	-2	−4
x				**1***		1	-1	−3
i					**1**	2	0	−2
r	-5	-4	-3	-2	-1	0	1	−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$\bar{S}$: elix~

$\bar{R}$: elics

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e	**4**						-4	−6
l		**3**					-3	−5
i			**2**				-2	−4
x				**1***			-1	−3
i					**1**	**2**	0	−2
r	-5	-4	-3	-2	-1	0	1	−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$\bar{S}$: elix~i

$\bar{R}$: elicsi

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e	**4**							−6
l		**3**						−5
i			**2**					−4
x				**1***				−3
i					**1**	**2**		−2
r							**1**	−1
	−7	−6	−5	−4	−3	−2	−1	0

Needleman-Wunsch (1970)

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$\bar{S}$: elix~ir

$\bar{R}$: elicsir

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e	4							
l		3						
i			2					
x				1*				
i					1	2		
r							1	
								0

Needleman-Wunsch ~ LCS

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$\bar{S}$: elix~ir

$\bar{R}$: elicsir

S: elixir

R: elicsir

$ES: 1$

$DS: 0$

$GP: -1$

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e								-6
l								-5
i								-4
x								-3
i								-2
r								-1
	-7	-6	-5	-4	-3	-2	-1	0

Needleman-Wunsch ~ LCS

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$\bar{S}$: elix~ir

$\bar{R}$: elicsir

S: elixir

R: elicsir

ES: 1

DS: 0

GP: 0

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e								0
l								0
i								0
x								0
i								0
r								0
	0	0	0	0	0	0	0	0

Needleman-Wunsch ~ LCS

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$\bar{S}$: elix~ir

$\bar{R}$: elicsir

S: elixir

R: elicsir

ES: 1

DS: 0

GP: 0

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e	5	4	3	2**	2**	2	1	0
l	4	4	3	2**	2**	2	1	0
i	3	3	3	2**	2**	2*	1	0
x	2**	2**	2**	2**	2**	2	1	0
i	2	2	2*	2	2	2	1	0
r	1	1	1	1	1	1	1	0
	0	0	0	0	0	0	0	0

Needleman-Wunsch ~ LCS

SIM

	e	l	i	c	s	i	r
e	1	0	0	0	0	0	0
l	0	1	0	0	0	0	0
i	0	0	1	0	0	1	0
x	0	0	0	0	0	0	0
i	0	0	1	0	0	1	0
r	0	0	0	0	0	0	1

$\bar{S}$: elix~ir

$\bar{R}$: elicsir

S: elixir

R: elicsir

ES: 1

DS: 0

GP: 0

$$RIGHT_{i,j} = MS_{i,j+1} + GP$$

$$DOWN_{i,j} = MS_{i+1,j} + GP$$

$$DIAG_{i,j} = MS_{i+1,j+1} + SIM_{i,j}$$

$$MS_{i,j} = max\{R_{i,j}; D_{i,j}; D_{i,j}\}$$

MS

	e	l	i	c	s	i	r	
e	**5**	4	3	2**	2**	2	1	0
l	4	4	3	2**	2**	2	1	0
i	3	3	3	2**	2**	2*	1	0
x	2**	2**	2**	2**	2**	2	1	0
i	2	2	2*	2	2	2	1	0
r	1	1	1	1	1	1	1	0
	0	0	0	0	0	0	0	0

Needleman-Wunsch ~ LCS

2 strings S & R

$|S| = m \quad \& \quad |R| = n$

Needleman-Wunsch ~ LCS

2 strings S & R

$|S| = m \quad \& \quad |R| = n$

Time Complexity: $O(mn)$

Space Complexity: $O(mn)$

Can we do it in less space?

Linear Space LCS

Hirschberg (1975)

$L_{0\ldots m,0\ldots n}$

$S_{1\ldots m}$

$R_{1\ldots m}$

$ES: 1$

$DS: 0$

$GP: 0$

Hirschberg (1975)

$L_{0\ldots m,0\ldots n}$

$S_{1\ldots m}$

$R_{1\ldots m}$

$ES: 1$

$DS: 0$

$GP: 0$

	ε	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	R_{11}	R_{12}
ε	0	0	0	0	0	0	0	0	0	0	0	0	0
S_1	0												
S_2	0												
S_3	0												
S_4	0												
S_5	0												
S_6	0												
S_7	0												
S_8	0												
S_9	0												
S_{10}	0												

Hirschberg (1975)

$L_{0\ldots m,0\ldots n}$

S: fondazione

R: associazione

$ES: 1$

$DS: 0$

$GP: 0$

	ε	a	s	s	o	c	i	a	z	i	o	n	e
ε	0	0	0	0	0	0	0	0	0	0	0	0	0
f	0												
o	0												
n	0												
d	0												
a	0												
z	0												
i	0												
o	0												
n	0												
e	0												

Hirschberg (1975)

$L_{0\ldots m,0\ldots n}$

$i-1$
i

	ε	a	s	s	o	c	i	a	z	i	o	n	e
ε	0	0	0	0	0	0	0	0	0	0	0	0	0
f	0												
o	0												
n	0												
d	0												
a	0												
z	0												
i	0												
o	0												
n	0												
e	0												

S: fondazione

R: associazione

$ES: 1$

$DS: 0$

$GP: 0$

Hirschberg (1975)

$L_{0\ldots m,0\ldots n}$

$i-1$
i

		ε	a	s	s	o	c	i	a	z	i	o	n	e
	ε	0	0	0	0	0	0	0	0	0	0	0	0	0
	f	0	0	0	0	0	0	0	0	0	0	0	0	0
o		0												
n		0												
d		0												
a		0												
z		0												
i		0												
o		0												
n		0												
e		0												

S: fondazione

R: associazione

$ES: 1$

$DS: 0$

$GP: 0$

Hirschberg (1975)

$L_{0\ldots m,0\ldots n}$

$i-1$
i

S: fondazione

R: associazione

$ES: 1$

$DS: 0$

$GP: 0$

	ε	a	s	s	o	c	i	a	z	i	o	n	e
ε	0	0	0	0	0	0	0	0	0	0	0	0	0
f	0	0	0	0	0	0	0	0	0	0	0	0	0
o	0	0	0	0	**1**	1	1	1	1	1	**1**	1	1
n	0												
d	0												
a	0												
z	0												
i	0												
o	0												
n	0												
e	0												

Hirschberg (1975)

$L_{0\ldots m,0\ldots n}$

$i-1$
i

		ε	a	s	s	o	c	i	a	z	i	o	n	e
	ε	0	0	0	0	0	0	0	0	0	0	0	0	0
	f	0	0	0	0	0	0	0	0	0	0	0	0	0
	o	0	0	0	0	**1**	1	1	1	1	1	**1**	1	1
	n	0	0	0	0	1	1	1	1	1	1	1	**2**	2
	d	0												
	a	0												
	z	0												
	i	0												
	o	0												
	n	0												
	e	0												

S: fondazione

R: associazione

$ES: 1$

$DS: 0$

$GP: 0$

Hirschberg (1975)

$L_{0\ldots m,0\ldots n}$

S: fondazione

R: associazione

$ES: 1$

$DS: 0$

$GP: 0$

$i-1$
i

	ε	a	s	s	o	c	i	a	z	i	o	n	e
ε	0	0	0	0	0	0	0	0	0	0	0	0	0
f	0	0	0	0	0	0	0	0	0	0	0	0	0
o	0	0	0	0	**1**	1	1	1	1	1	**1**	1	1
n	0	0	0	0	1	1	1	1	1	1	1	**2**	2
d	0	0	0	0	1	1	1	1	1	1	1	2	2
a	0												
z	0												
i	0												
o	0												
n	0												
e	0												

Hirschberg (1975)

$L_{0\ldots m,0\ldots n}$

$L_{i,j} = |LCS(S_{1\ldots i}; R_{1\ldots j})|$

S: fondazione

R: associazione

ES: 1

DS: 0

GP: 0

	ε	a	s	s	o	c	i	a	z	i	o	n	e
ε	0	0	0	0	0	0	0	0	0	0	0	0	0
f	0	0	0	0	0	0	0	0	0	0	0	0	0
o	0	0	0	0	**1**	1	1	1	1	1	**1**	1	1
n	0	0	0	0	1	1	1	1	1	1	1	**2**	2
d	0	0	0	0	1	1	1	1	1	1	1	2	2
a	0	**1**	1	1	1	1	1	**2**	2	2	2	2	2
z	0												
i	0												
o	0												
n	0												
e	0												

$i - 1$

i

Hirschberg (1975)

$L_{0\ldots m,0\ldots n}$

$L_{i,j} = |LCS(S_{1\ldots i}; R_{1\ldots j})|$

S: fondazione

R: associazione

ES: 1

DS: 0

GP: 0

		a	s	s	o	c	i	a	z	i	o	n	e
	ε	a	s	s	o	c	i	a	z	i	o	n	e
ε	0	0	0	0	0	0	0	0	0	0	0	0	0
f	0	0	0	0	0	0	0	0	0	0	0	0	0
o	0	0	0	0	**1**	1	1	1	1	1	**1**	1	1
n	0	0	0	0	1	1	1	1	1	1	1	**2**	2
d	0	0	0	0	1	1	1	1	1	1	1	2	2
a	0	**1**	1	1	1	1	1	**2**	2	2	2	2	2
z	0	1	1	1	1	1	1	2	**3**	3	3	3	3
i	0												
o	0												
n	0												
e	0												

$i-1$

i

Hirschberg (1975)

$L_{0\ldots m,0\ldots n}$

$L_{i,j} = |LCS(S_{1\ldots i}; R_{1\ldots j})|$

S: fondazione

R: associazione

ES: 1

DS: 0

GP: 0

	ε	a	s	s	o	c	i	a	z	i	o	n	e
ε	0	0	0	0	0	0	0	0	0	0	0	0	0
f	0	0	0	0	0	0	0	0	0	0	0	0	0
o	0	0	0	0	**1**	1	1	1	1	1	**1**	1	1
n	0	0	0	0	1	1	1	1	1	1	1	**2**	2
d	0	0	0	0	1	1	1	1	1	1	1	2	2
a	0	**1**	1	1	1	1	1	**2**	2	2	2	2	2
z	0	1	1	1	1	1	1	2	**3**	3	3	3	3
i	0	1	1	1	1	1	**2**	2	3	**4**	4	4	4
o	0	1	1	1	**2**	2	2	2	3	4	**5**	5	5
n	0	1	1	1	2	2	2	2	3	4	5	**6**	6
e	0	1	1	1	2	2	2	2	3	4	5	6	**7**

$i - 1$

i

Hirschberg (1975)

$L_{0\ldots m,0\ldots n}$

$L_{i,j} = |LCS(S_{1\ldots i}; R_{1\ldots j})|$

S: fondazione

R: associazione

$ES: 1$

$DS: 0$

$GP: 0$

	ε	a	s	s	o	c	i	a	z	i	o	n	e
S	0	1	1	1	2	2	2	2	3	4	5	6	**7**

Hirschberg (1975)

$L_{0\ldots m,0\ldots n}$

$L_{i,j} = |LCS(S_{1\ldots i}; R_{1\ldots j})|$

$LL_{0\ldots n}$

$LL_j = L_{\mathrm{m},j}$

S: fondazione

R: associazione

$ES: 1$

$DS: 0$

$GP: 0$

	ε	a	s	s	o	c	i	a	z	i	o	n	e
S	0	1	1	1	2	2	2	2	3	4	5	6	7

Hirschberg (1975)

$L_{0\ldots m,0\ldots n}$

$L_{i,j} = |LCS(S_{1\ldots i}; R_{1\ldots j})|$

$LL_{0\ldots n}$

$LL_j = L_{\mathrm{m},j}$

S: fondazione

R: associazione

ES: 1

DS: 0

GP: 0

	ε	a	s	s	o	c	i	a	z	i	o	n	e
S	0	1	1	1	2	2	2	2	3	4	5	6	7

Reconstructing the actual subsequence seems impossible.

Hirschberg (1975)

$LL_{0\ldots n}$

$LL_j = L_{\mathrm{m},j}$

	ε	a	s	s	o	c	i	a	z	i	o	n	e
S	0	1	1	1	2	2	2	2	3	4	5	6	7

$L_{0\ldots m,0\ldots n}$

$L_{i,j} = |LCS(S_{1\ldots i}; R_{1\ldots j})|$

S: fondazione

R: associazione

Hirschberg (1975)

$LL_{0\ldots n}$

$LL_j = L_{\mathrm{m},j}$

	ε	a	s	s	o	c	i	a	z	i	o	n	e
S	0	1	1	1	2	2	2	2	3	4	5	6	7

$L_{0\ldots m,0\ldots n}$

$L_{i,j} = |LCS(S_{1\ldots i}; R_{1\ldots j})|$

$$L^*_{i,j} = |LCS(S_{m\ldots i+1}; R_{n\ldots j+1})|$$

S: fondazione

R: associazione

Hirschberg (1975)

$LL_{0\ldots n}$

$LL_j = L_{\mathrm{m},j}$

	ε	a	s	s	o	c	i	a	z	i	o	n	e
S	0	1	1	1	2	2	2	2	3	4	5	6	7

$L_{0\ldots m,0\ldots n}$

$L_{i,j} = |LCS(S_{1\ldots i}; R_{1\ldots j})|$

$L^*_{i,j} = |LCS(S_{m\ldots i+1}; R_{n\ldots j+1})|$

S: fondazione

R: associazione

$$M(i) = \max_{j\in[0;n]}\{L_{i,j} + L^*_{i,j}\}$$

Hirschberg (1975)

$LL_{0\dots n}$

$LL_j = L_{\mathrm{m},j}$

	ε	a	s	s	o	c	i	a	z	i	o	n	e
S	0	1	1	1	2	2	2	2	3	4	5	6	7

$L_{0\dots m,0\dots n}$

$L_{i,j} = |LCS(S_{1\dots i}; R_{1\dots j})|$

$L^*_{i,j} = |LCS(S_{m\dots i+1}; R_{n\dots j+1})|$

S: fondazione

R: associazione

$$M(i) = \max_{j\in[0;n]}\{L_{i,j} + L^*_{i,j}\}$$

$$Thm: M(i) = L_{m,n}\ \forall i \in [0;m]$$

Hirschberg (1975)

$LL_{0\dots n}$

$LL_j = L_{\mathrm{m},j}$

	ε	a	s	s	o	c	i	a	z	i	o	n	e
S	0	1	1	1	2	2	2	2	3	4	5	6	7

$L_{0\dots m,0\dots n}$

$L_{i,j} = |LCS(S_{1\dots i}; R_{1\dots j})|$

$L^*_{i,j} = |LCS(S_{m\dots i+1}; R_{n\dots j+1})|$

S: fondazione

R: associazione

$$M(i) = \max_{j\in[0;n]}\{L_{i,j} + L^*_{i,j}\}$$

$$\textit{Thm}: M(i) = L_{m,n}\ \forall i \in [0;m]$$

Hirschberg (1975)

$LL_{0\dots n}$

$LL_j = L_{\mathrm{m},j}$

	ε	a	s	s	o	c	i	a	z	i	o	n	e
S	0	1	1	1	2	2	2	2	3	4	5	6	7

$$M(i) = \max_{j\in[0;n]}\{L_{i,j} + L^*_{i,j}\}$$

Thm: $M(i) = L_{m,n}\ \forall i \in [0;m]$

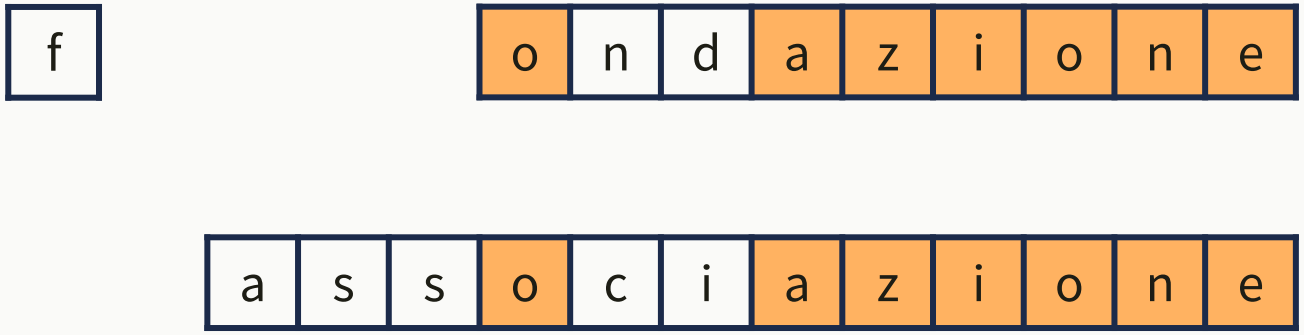
$L_{0\dots m,0\dots n}$

$L_{i,j} = |LCS(S_{1\dots i}; R_{1\dots j})|$

$L^*_{i,j} = |LCS(S_{m\dots i+1}; R_{n\dots j+1})|$

S: fondazione

R: associazione



Hirschberg (1975)

$LL_{0\dots n}$

$LL_j = L_{\mathrm{m},j}$

	ε	a	s	s	o	c	i	a	z	i	o	n	e
S	0	1	1	1	2	2	2	2	3	4	5	6	7

$$M(i) = \max_{j\in[0;n]}\{L_{i,j} + L^*_{i,j}\}$$

$$\textit{Thm}: M(i) = L_{m,n}\ \forall i \in [0;m]$$

$L_{0\dots m,0\dots n}$

$L_{i,j} = |LCS(S_{1\dots i}; R_{1\dots j})|$

$L^*_{i,j} = |LCS(S_{m\dots i+1}; R_{n\dots j+1})|$

S: fondazione

R: associazione

f	o

n	d	a	z	i	o	n	e

a	s	s	o

c	i	a	z	i	o	n	e

Hirschberg (1975)

$LL_{0\dots n}$

$LL_j = L_{\mathrm{m},j}$

	ε	a	s	s	o	c	i	a	z	i	o	n	e
S	0	1	1	1	2	2	2	2	3	4	5	6	7

$$M(i) = \max_{j\in[0;n]}\{L_{i,j} + L^*_{i,j}\}$$

$$Thm: M(i) = L_{m,n}\ \forall i \in [0;m]$$

$L_{0\dots m,0\dots n}$

$L_{i,j} = |LCS(S_{1\dots i}; R_{1\dots j})|$

$L^*_{i,j} = |LCS(S_{m\dots i+1}; R_{n\dots j+1})|$

S: fondazione

R: associazione

f	o	n

d	a	z	i	o	n	e

a	s	s	o

c	i	a	z	i	o	n	e

Hirschberg (1975)

$LL_{0\dots n}$

$LL_j = L_{\mathrm{m},j}$

	ε	a	s	s	o	c	i	a	z	i	o	n	e
S	0	1	1	1	2	2	2	2	3	4	5	6	7

$$M(i) = \max_{j\in[0;n]}\{L_{i,j} + L^*_{i,j}\}$$

$$Thm: M(i) = L_{m,n}\ \forall i \in [0;m]$$

$L_{0\dots m,0\dots n}$

$L_{i,j} = |LCS(S_{1\dots i}; R_{1\dots j})|$

$L^*_{i,j} = |LCS(S_{m\dots i+1}; R_{n\dots j+1})|$

S: fondazione

R: associazione

f	o	n	d

a	z	i	o	n	e

a	s	s	o

c	i	a	z	i	o	n	e

Hirschberg (1975)

$LL_{0\dots n}$

$LL_j = L_{\mathrm{m},j}$

$|LCS(S;R)| = 7$

$L_{0\dots m,0\dots n}$

$L_{i,j} = |LCS(S_{1\dots i};R_{1\dots j})|$

$L^*_{i,j} = |LCS(S_{m\dots i+1};R_{n\dots j+1})|$

S: fondazione

R: associazione

$$M(i) = \max_{j\in[0;n]}\{L_{i,j} + L^*_{i,j}\}$$

$$Thm: M(i) = L_{m,n}\ \forall i \in [0;m]$$

Hirschberg (1975)

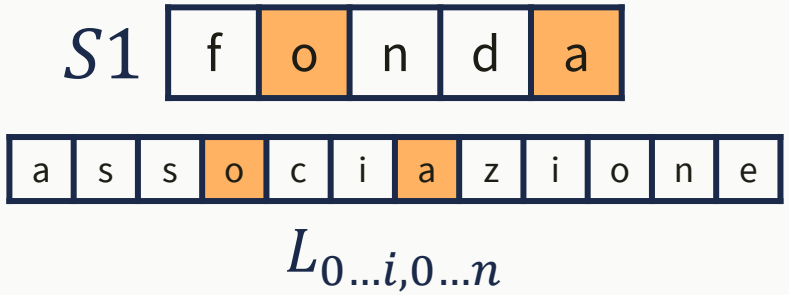
$|LCS(S;R)| = 7$

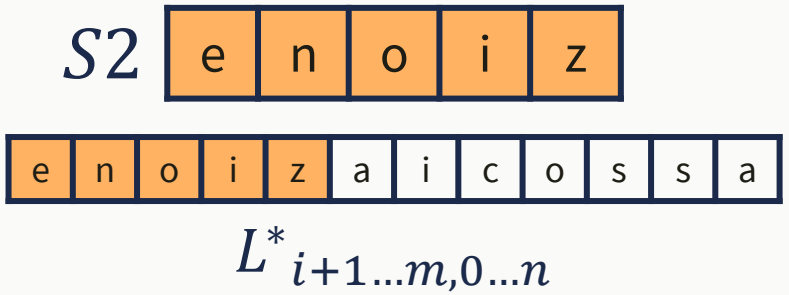
$L_{i,j} = |LCS(S_{1\dots i}; R_{1\dots j})|$

$L^*_{i,j} = |LCS(S_{m\dots i+1}; R_{n\dots j+1})|$

$$M(i) = \max_{j\in[0;n]}\{L_{i,j} + L^*_{i,j}\}$$

$Thm: M(i) = L_{m,n}\ \forall i \in [0;m]$





Hirschberg (1975)

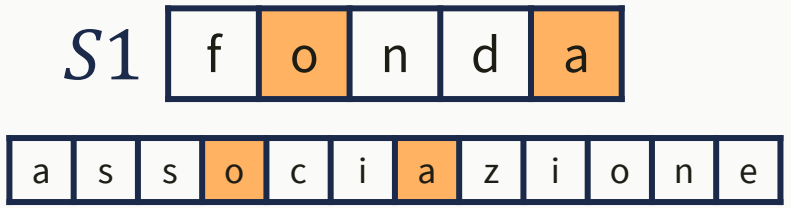
$|LCS(S;R)| = 7$

$L_{i,j} = |LCS(S_{1\dots i}; R_{1\dots j})|$

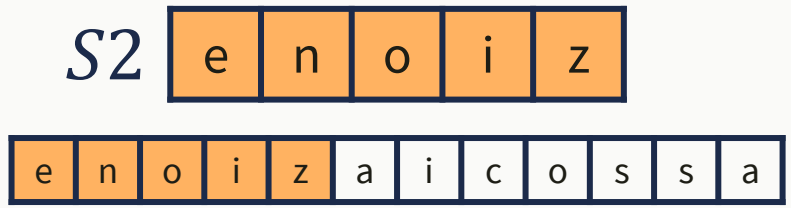
$L^*_{i,j} = |LCS(S_{m\dots i+1}; R_{n\dots j+1})|$

$$M(i) = \max_{j\in[0;n]}\{L_{i,j} + L^*_{i,j}\}$$

$Thm: M(i) = L_{m,n}\ \forall i \in [0;m]$



$L_{0\dots i,0\dots n}$



$L^*_{i+1\dots m,0\dots n}$

LL

j	0	1	2	3	4	5	6	7	8	9	10	11	12
R	ε	a	s	s	o	c	i	a	z	i	o	n	e
S1	0	1	1	1	1	1	1	2	2	2	2	2	2

j	12	11	10	9	8	7	6	5	4	3	2	1	0
R	ε	e	n	o	i	z	a	i	c	o	s	s	a
S2	0	1	2	3	4	5	5	5	5	5	5	5	5

Hirschberg (1975)

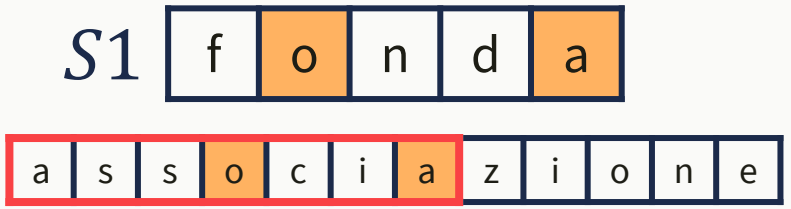
$$|LCS(S;R)| = 7$$

$$L_{i,j} = |LCS(S_{1\dots i}; R_{1\dots j})|$$

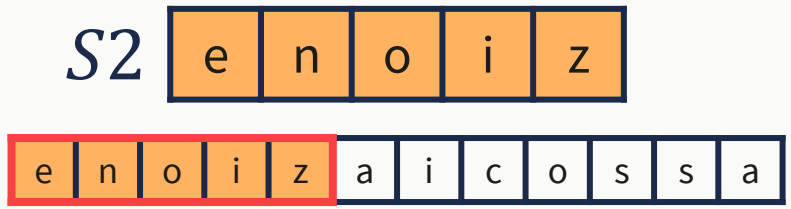
$$L^*_{i,j} = |LCS(S_{m\dots i+1}; R_{n\dots j+1})|$$

$$M(i) = \max_{j\in[0;n]}\{L_{i,j} + L^*_{i,j}\}$$

$$Thm: M(i) = L_{m,n}\ \forall i \in [0;m]$$



$L_{0\dots i,0\dots n}$



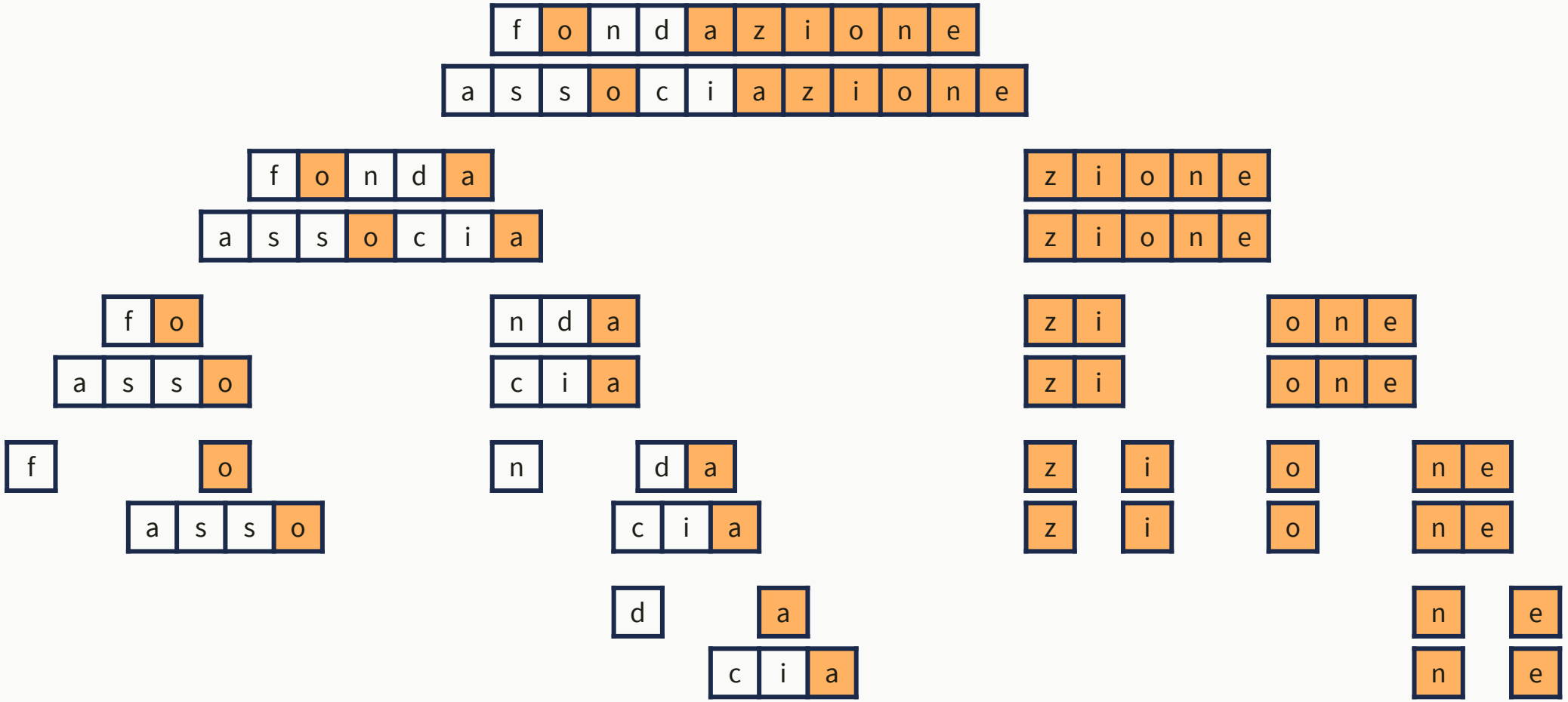
$L^*_{i+1\dots m,0\dots n}$

LL

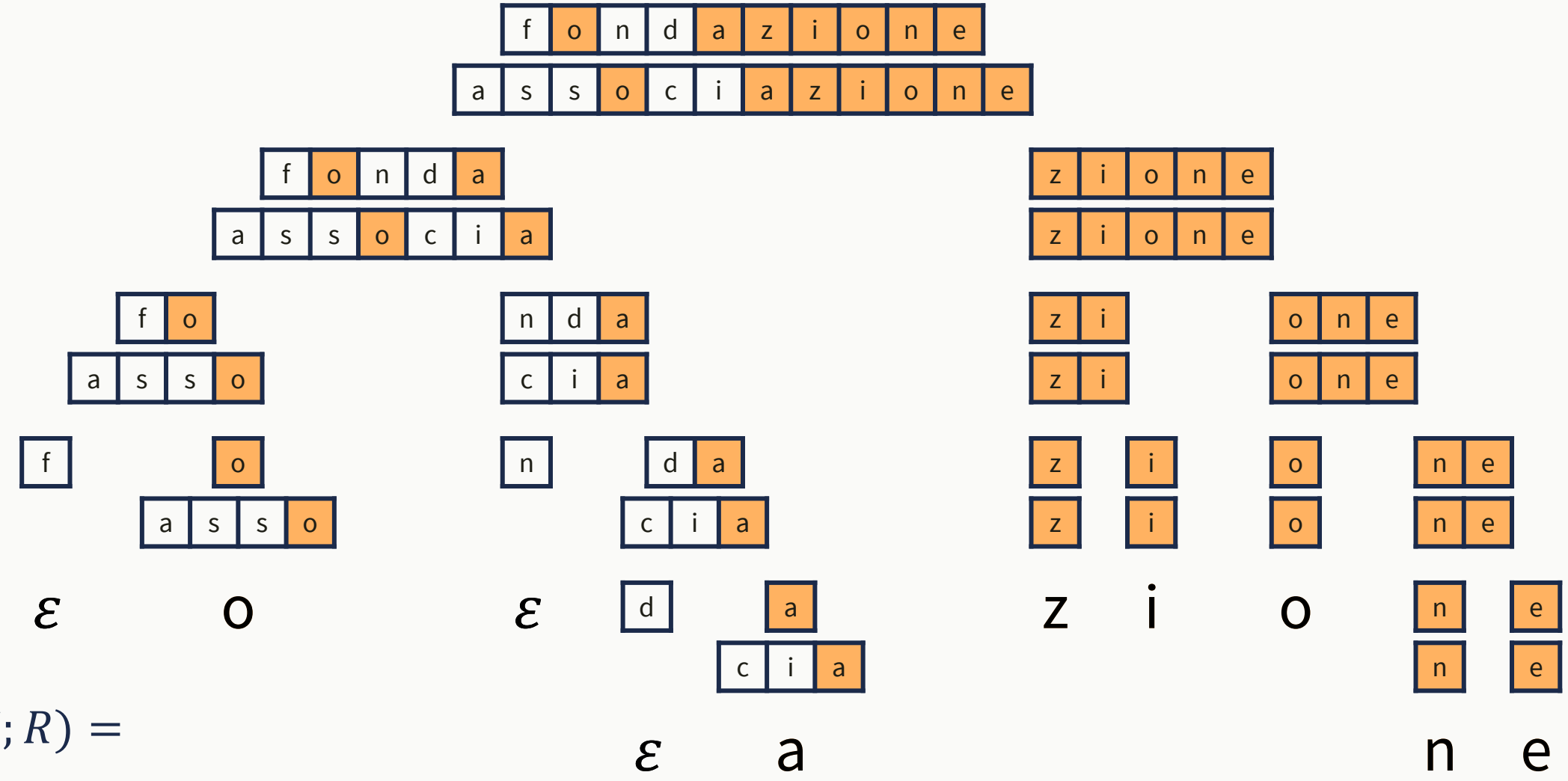
j	0	1	2	3	4	5	6	7	8	9	10	11	12
R	ε	a	s	s	o	c	i	a	z	i	o	n	e
S1	0	1	1	1	1	1	1	2	2	2	2	2	2

j	12	11	10	9	8	7	6	5	4	3	2	1	0
R	ε	e	n	o	i	z	a	i	c	o	s	s	a
S2	0	1	2	3	4	5	5	5	5	5	5	5	5

Hirschberg (1975)

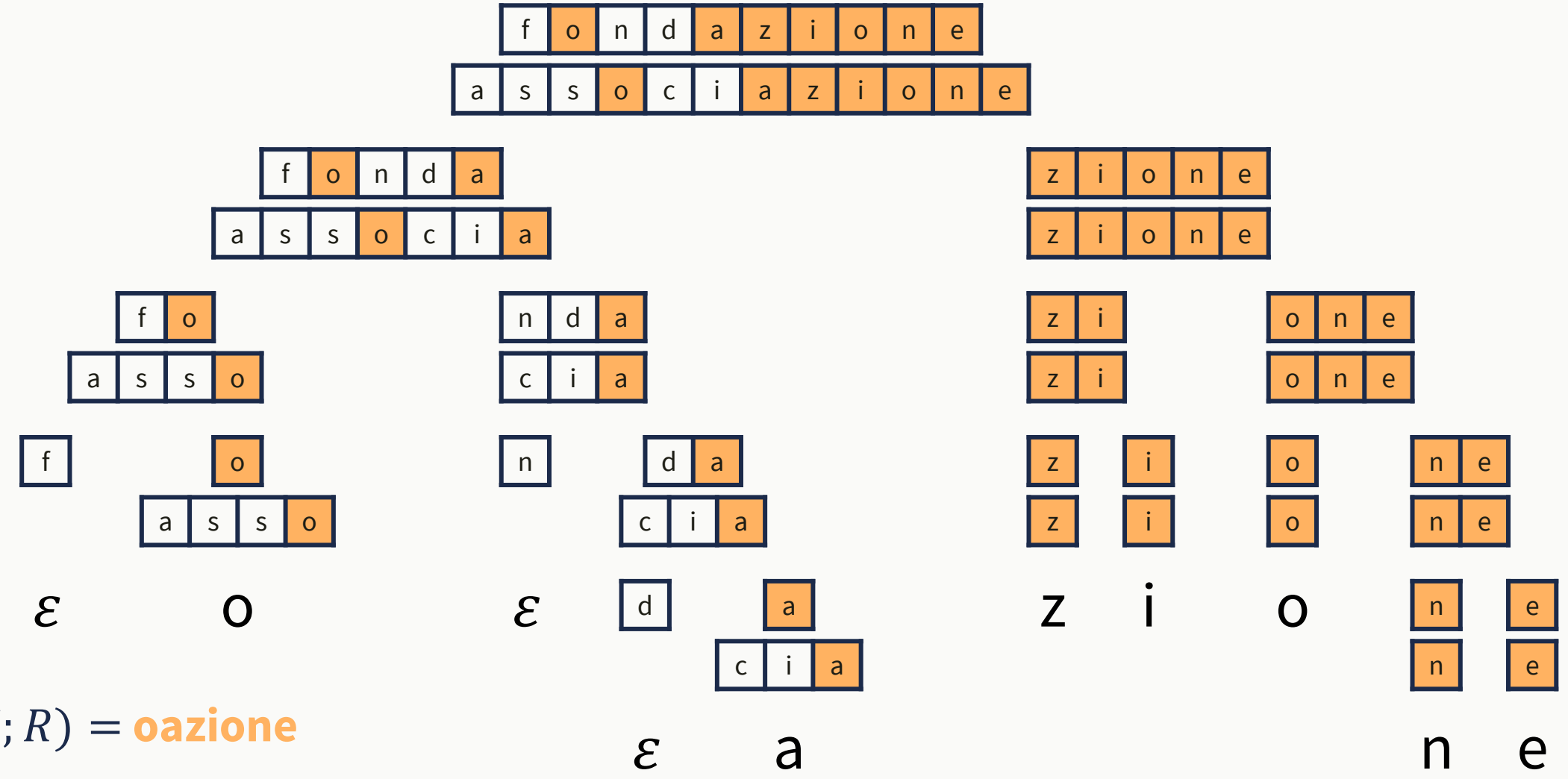


Hirschberg (1975)



$$LCS(S;R) =$$

Hirschberg (1975)



$LCS(S;R) =$ **oazione**

Hirschberg (1975)

2 strings S & R

$|S| = m \quad \& \quad |R| = n$

Time Complexity: $O(mn)$

Space Complexity: $O(m + n)$

What if we don't need to align everything?

Local Alignment

Smith-Waterman (1981)

$H_{0\ldots m,0\ldots n}$

$S_{1\ldots m}$

$R_{1\ldots m}$

ES

DS

GP

	ε	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
ε	0	0	0	0	0	0	0	0	0
S_1	0								
S_2	0								
S_3	0								
S_4	0								
S_5	0								
S_6	0								
S_7	0								
S_8	0								

Smith-Waterman (1981)

$H_{0\ldots m,0\ldots n}$

$S_{1\ldots m}$

$R_{1\ldots m}$

ES

DS

$W_O \quad W_E$

$W_K = W_O + W_E \cdot k$

$W_O + W_E \geq DS$

	ε	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
ε	0	0	0	0	0	0	0	0	0
S_1	0								
S_2	0								
S_3	0								
S_4	0								
S_5	0								
S_6	0								
S_7	0								
S_8	0								

Smith-Waterman (1981)

$H_{0\ldots m,0\ldots n}$

$S_{1\ldots m}$

$R_{1\ldots m}$

ES

DS

$W_O \quad W_E$

$W_K = W_O + W_E \cdot k$

$W_O + W_E \geq DS$

H

	ε	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
ε	0	0	0	0	0	0	0	0	0
S_1	0								
S_2	0								
S_3	0								
S_4	0								
S_5	0								
S_6	0								
S_7	0								
S_8	0								

$$H_{i,j} = max\{LEFT_{i,j}; UP_{i,j}; DIAG_{i,j}\}$$

Smith-Waterman (1981)

$H_{0\dots m,0\dots n}$

$S_{1\dots m}$

$R_{1\dots m}$

ES

DS

$W_O \quad W_E$

$W_K = W_O + W_E \cdot k$

$W_O + W_E \geq DS$

H

	ε	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
ε	0	0	0	0	0	0	0	0	0
S_1	0								
S_2	0								
S_3	0								
S_4	0								
S_5	0								
S_6	0								
S_7	0								
S_8	0								

$$H_{i,j} = max\{0; LEFT_{i,j}; UP_{i,j}; DIAG_{i,j}\}$$

Smith-Waterman (1981)

$H_{0\ldots m,0\ldots n}$

$S_{1\ldots m}$

$R_{1\ldots m}$

ES

DS

$W_O \quad W_E$

$W_K = W_O + W_E \cdot k$

$W_O + W_E \geq DS$

H

	ε	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
ε	0	0	0	0	0	0	0	0	0
S_1	0								
S_2	0								
S_3	0								
S_4	0								
S_5	0								
S_6	0								
S_7	0								
S_8	0								

$$H_{i,j} = max\{0; L_{i,j}; U_{i,j}; D_{i,j}\}$$

Smith-Waterman (1981)

$H_{0\dots m,0\dots n}$

$S_{1\dots m}$

$R_{1\dots m}$

ES

DS

$W_O \quad W_E$

$W_K = W_O + W_E \cdot k$

$W_O + W_E \geq DS$

$$H_{i,j} = max\{0; L_{i,j}; U_{i,j}; D_{i,j}\}$$

$$LEFT_{i,j} = \max_k\{H_{i,j-k} - W_K\}$$

$$UP_{i,j} = \max_k\{H_{i-k,j} - W_K\}$$

$$DIAG_{i,j} = H_{i-1,j-1} + ES|DS$$

H

	ε	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
ε	0	0	0	0	0	0	0	0	0
S_1	0								
S_2	0								
S_3	0								
S_4	0								
S_5	0								
S_6	0								
S_7	0								
S_8	0								

Smith-Waterman (1981)

$H_{0\ldots m,0\ldots n}$

$S_{1\ldots m}$

$R_{1\ldots m}$

ES

DS

$W_O \quad W_E$

$W_K = W_O + W_E \cdot k$

$W_O + W_E \geq DS$

$$H_{i,j} = max\{0; L_{i,j}; U_{i,j}; D_{i,j}\}$$

$$LEFT_{i,j} = \max_k\{H_{i,j-k} - W_K\}$$

$$UP_{i,j} = \max_k\{H_{i-k,j} - W_K\}$$

$$DIAG_{i,j} = H_{i-1,j-1} + ES|DS$$

H

	ε	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
ε	0	0	0	0	0	0	0	0	0
S_1	0								
S_2	0								
S_3	0								
S_4	0								
S_5	0								
S_6	0								
S_7	0								
S_8	0								

Smith-Waterman (1981)

$H_{0\dots m,0\dots n}$

$S_{1\dots m}$

$R_{1\dots m}$

ES

DS

$W_O \quad W_E$

$W_K = W_O + W_E \cdot k$

$W_O + W_E \geq DS$

$$H_{i,j} = max\{0; L_{i,j}; U_{i,j}; D_{i,j}\}$$

$$LEFT_{i,j} = \max_k\{H_{i,j-k} - W_K\}$$

$$UP_{i,j} = \max_k\{H_{i-k,j} - W_K\}$$

$$DIAG_{i,j} = H_{i-1,j-1} + ES|DS$$

H

	ε	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
ε	0	0	0	0	0	0	0	0	0
S_1	0								
S_2	0								
S_3	0								
S_4	0								
S_5	0								
S_6	0								
S_7	0								
S_8	0								

TB

	ε	R_1	R_2	R_3	
ε	None	None	None	None	
S_1	None	None	None	None	
S_2	None	None	None	None	
S_3	None	None	None	None	

Smith-Waterman (1981)

$H_{0\ldots m,0\ldots n}$

$S_{1\ldots m}$

$R_{1\ldots m}$

ES

DS

$W_O \quad W_E$

$W_K = W_O + W_E \cdot k$

$W_O + W_E \geq DS$

$$H_{i,j} = max\{0; L_{i,j}; U_{i,j}; D_{i,j}\}$$

$$LEFT_{i,j} = \max_k\{H_{i,j-k} - W_K\}$$

$$UP_{i,j} = \max_k\{H_{i-k,j} - W_K\}$$

$$DIAG_{i,j} = H_{i-1,j-1} + ES|DS$$

$$TB_{i,j} = \begin{cases} ('Stop', 0, 0) \\ ('Horiz', k, 0) \\ ('Vert', 0, k) \\ ('Diag', 1, 1) \end{cases}$$

H

	ε	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
ε	0	0	0	0	0	0	0	0	0
S_1	0								
S_2	0								
S_3	0								
S_4	0								
S_5	0								
S_6	0								
S_7	0								
S_8	0								

TB

	ε	R_1	R_2	R_3	
ε	None	None	None	None	
S_1	None	None	None	None	
S_2	None	None	None	None	
S_3	None	None	None	None	

Smith-Waterman (1981)

$H_{0\ldots m,0\ldots n}$

$S_{1\ldots m}$

$R_{1\ldots m}$

ES

DS

$W_O \quad W_E$

$W_K = W_O + W_E \cdot k$

$W_O + W_E \geq DS$

$$H_{i,j} = max\{0; L_{i,j}; U_{i,j}; D_{i,j}\}$$

$$LEFT_{i,j} = \max_k\{H_{i,j-k} - W_K\}$$

$$UP_{i,j} = \max_k\{H_{i-k,j} - W_K\}$$

$$DIAG_{i,j} = H_{i-1,j-1} + ES|DS$$

$$TB_{i,j} = \begin{cases} ('Stop', 0, 0) \\ ('Horiz', k, 0) \\ ('Vert', 0, k) \\ ('Diag', 1, 1) \end{cases}$$

H

	ε	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8
ε	0	0	0	0	0	0	0	0	0
S_1	0								
S_2	0								
S_3	0								
S_4	0								
S_5	0								
S_6	0								
S_7	0								
S_8	0								

TB

	ε	R_1	R_2	R_3	
ε	None	None	None	None	
S_1	None	None	None	None	
S_2	None	None	None	None	
S_3	None	None	None	None	

Smith-Waterman (1981)

2 strings S & R

$|S| = m \quad \& \quad |R| = n$

Time Complexity: $O\big(mn \cdot (m+n)\big)$

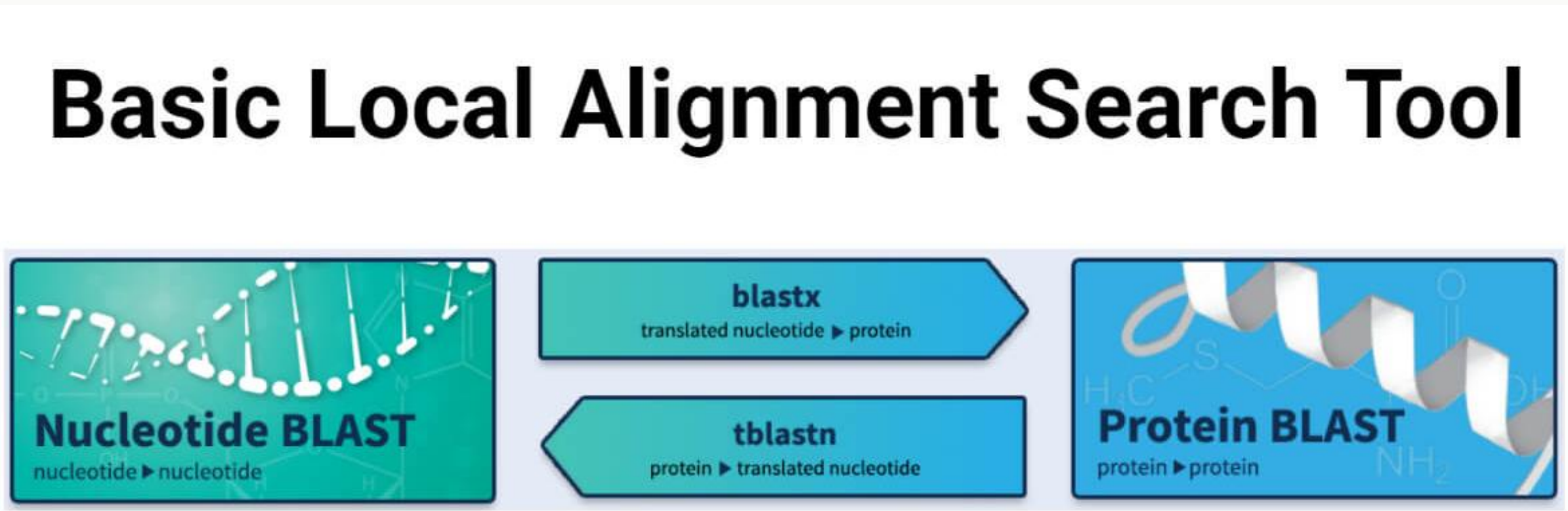
Space Complexity: $O(mn)$

When exact isn't fast enough...

Heuristic, but an order of magnitude faster

BLAST (1990)

Altschul, Gish, Miller, Myers, Lipman



BLAST (1990)

Altschul, Gish, Miller, Myers, Lipman

Query: **U F E S U U T M W T P N I Q L N A N A Y**

Target: **R S Y P R L R Y P S V S S W T P N I Q L N A W T C V R S R Q**

BLAST (1990)

Altschul, Gish, Miller, Myers, Lipman

Query: U F E S U U T M W T P N I Q L N A N A Y

Hit!

Target: R S Y P R L R Y P S V S S W T P N I Q L N A W T C V R S R Q

BLAST (1990)

Altschul, Gish, Miller, Myers, Lipman

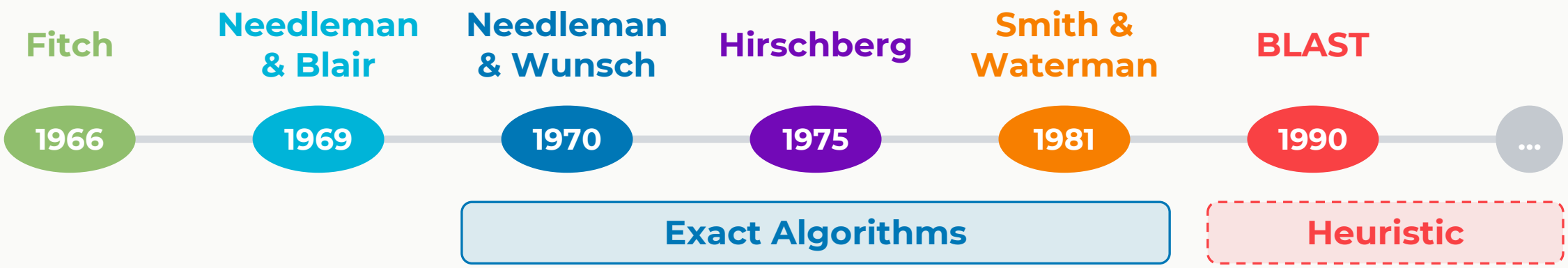
$10x\ faster!$

Query: U F E S U U T M **W T P N I Q L N A** N A Y

Hit!

Target: R S Y P R L R Y P S V S S **W T P N I Q L N A** W T C V R S R Q

The Roadmap


Fitch
1966
Needleman & Blair
1969
Needleman & Wunsch
1970
Hirschberg
1975
Smith & Waterman
1981
BLAST
1990
...
Exact Algorithms
Heuristic

References

- Fitch (1966) - "An improved method of testing for evolutionary homology" - *J. Mol. Biol.* 16(1):9-16
- Needleman & Blair (1969) - "Homology of Pseudomonas cytochrome c-551 with eukaryotic c-cytochromes" - *PNAS* 63(4):1227-1233
- Needleman & Wunsch (1970) - "A general method applicable to the search for similarities in the amino acid sequence of two proteins" - *J. Mol. Biol.* 48(3):443-453
- Hirschberg (1975) - "A linear space algorithm for computing maximal common subsequences" - *Commun. ACM* 18(6):341-343
- Smith & Waterman (1981) - "Identification of common molecular subsequences" - *J. Mol. Biol.* 147(1):195-197
- Altschul et al. (1990) - "Basic local alignment search tool" - *J. Mol. Biol.* 215(3):403-410

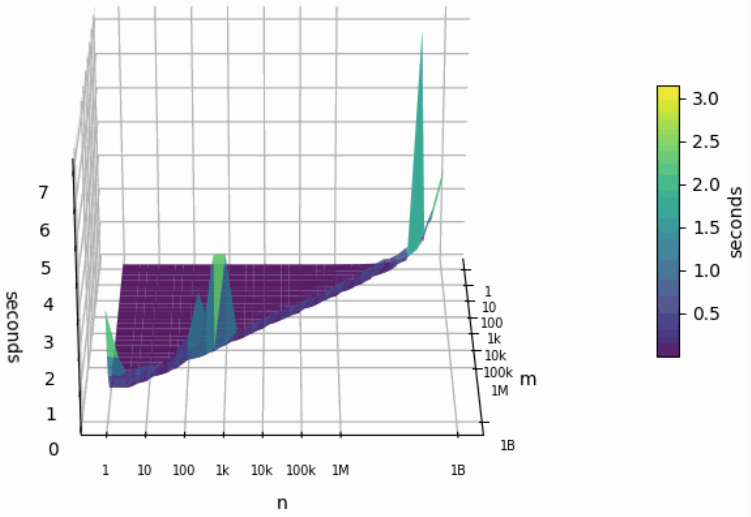
Images & Data

- «Big Bang», «First Galaxies», «The Earth in its Early Stages", "Primordial Soup", "First Life", "Earth's Biodiversity Through Time" – generated with Sora
- Darwin's species notebooks: 'I think . . .' – https://darwin-online.org.uk/content/frameset?viewtype=side&itemID=CUL-DAR121.-&pageseq=38
- Vertebrate forelimbs - https://open.lib.umn.edu/evolutionbiology/chapter/how-do-we-know-evolution-has-occurred-comparative-anatomy-2/
- Tree of Life - https://biology.duke.edu/news-events/news/%E2%80%98tree-life%E2%80%99-23-million-species-released
- Pisum Sativum - http://www.biolib.de/thome3/thome_flora_von_deutschland_3.pdf
- Human β-Hemoglobin image and sequence - https://www.uniprot.org/uniprotkb/P68871/entry
- Whale Myoglobin image and sequence - https://www.uniprot.org/uniprotkb/P02185/entry
- BLAST - https://microbenotes.com/blast-bioinformatics/

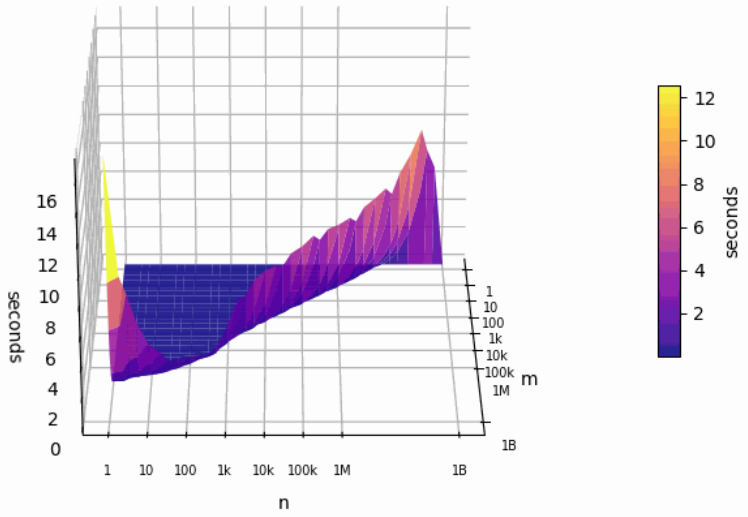
Questions?

The journey continues...

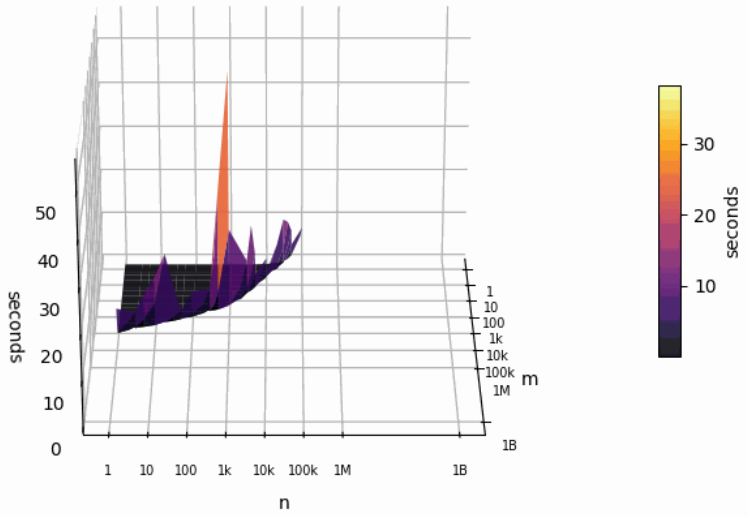
Time Complexity



$O(mn)$

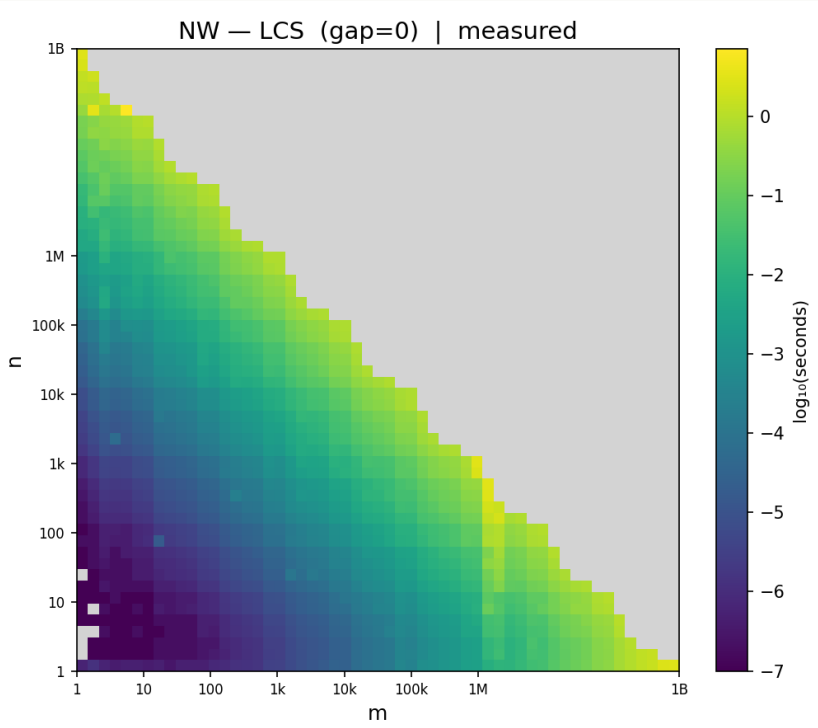


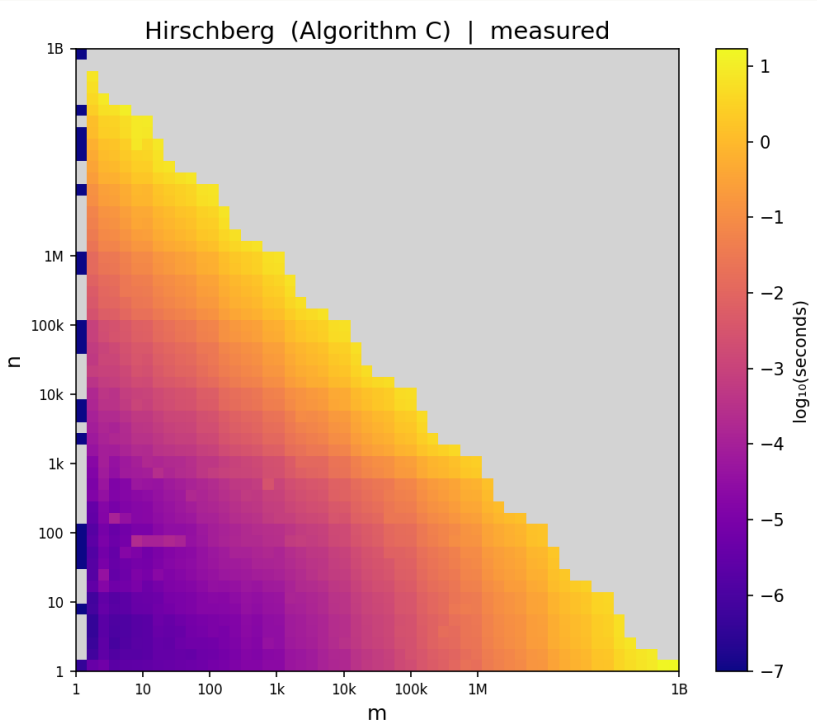
$O(mn)$

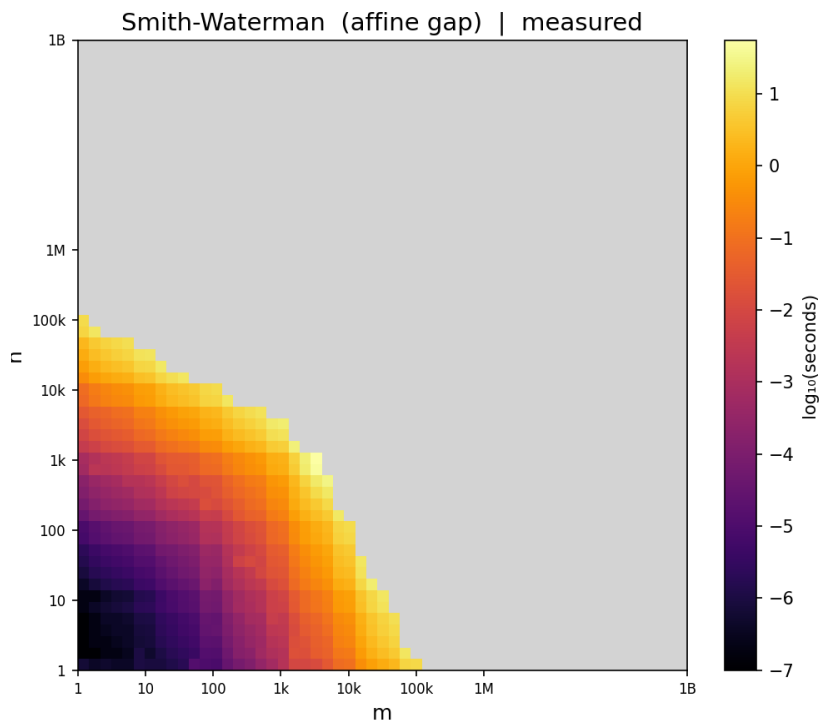


$O\big(mn \cdot (m+n)\big)$

Time Complexity





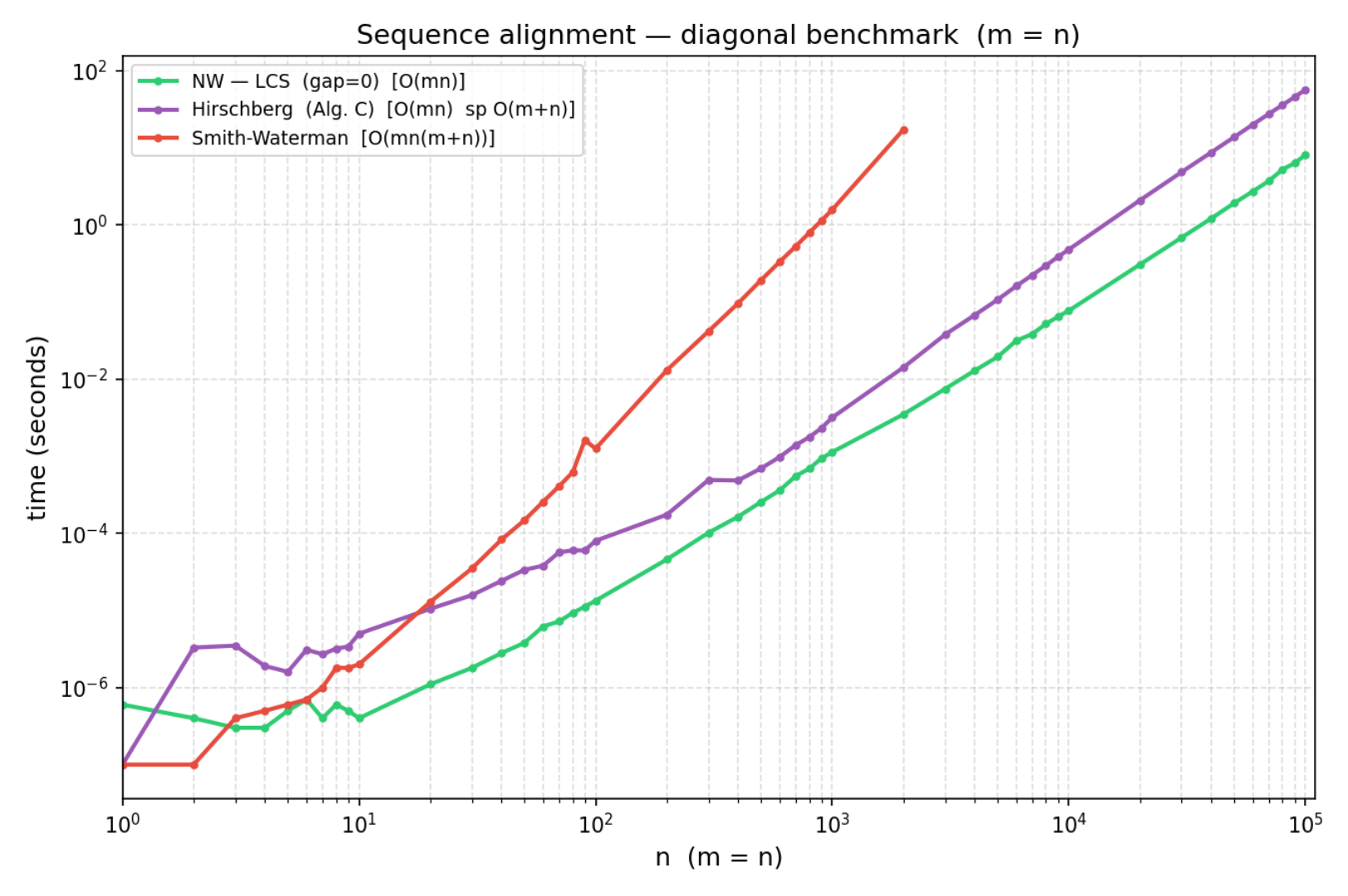


$O(mn)$

$O(mn)$

$O\big(mn \cdot (m+n)\big)$

Time Complexity



Can we do better than $O(n^2)$?

The Four Russians method improves NW by precomputing all possible subblocks of size $log(n) \times log(n)$.

Time: $O\left(\frac{n^2}{log(n)}\right)$

A real improvement, but only polylogarithmic.

Masek & Paterson (1980)

"A faster algorithm computing string edit distances"

J. Comput. System Sci., 20(1):18-31

Can we do better than $O(n^2)$?

Under SETH, no algorithm can solve LCS in $O(n^{2-\varepsilon})$ for any $\varepsilon > 0$.

Abboud, Backurs & Vassilevska Williams (2015)

"Tight Hardness Results for LCS and Other Sequence Similarity Measures"

FOCS, IEEE 56th Symposium on Foundations of Computer Science, pp. 59-78

$Thm: M(i) = L_{m,n}\ \forall i \in [0; m]$

Recall

$L_{i,j} = |LCS(S_{1\dots i}; R_{1\dots j})|$

$L^*_{i,j} = |LCS(S_{m\dots i+1}; R_{n\dots j+1})|$

$M(i) = \max_{j\in[0;n]}\{L_{i,j} + L^*_{i,j}\}$

$Proof.\ Let\ OPT\ = |LCS(S, R)|$

$(\leq)\ \forall j, concatenating\ an\ LCS\ of\ s_1 \dots s_i / r_1 \dots r_j\ with\ one\ of\ s_{i+1} \dots s_m / r_{j+1} \dots r_n\ yelds$
$a\ common\ subsequence\ of\ S, R,\ so\ L_{i,j} + L^*_{i,j} \leq OPT\ \forall j, hence\ M(i) \leq OPT.$

$(\leq)\ Let\ C\ be\ an\ optimal\ LCS.\ Split\ C\ at\ row\ i:\ let\ p\ pairs\ have\ S{-}index \leq i,$
$and\ set\ j^*\ to\ the\ R{-}index\ of\ the\ last\ such\ pair.$
$Then\ L_{i,j^*} \geq p\ and\ L^*_{i,j^*} \geq\ OPT - p, so\ L_{i,j^*} + L^*_{i,j^*} \geq OPT, hence\ M(i) \leq OPT.$

$$M(i) \leq OPT \wedge M(i) \geq OPT \Rightarrow M(i) = OPT$$