

Pairwise Alignment

Anything can align to anything...

BIOL 435/535: Bioinformatics

January 25, 2021

**With enough gaps, any sequence can
be aligned to any other sequence**

Seq1 — ABCNJROCLCRPM

Seq2 — AJCJNRCKCRBP

**With enough gaps, any sequence can
be aligned to any other sequence**

Seq1 — ABCNJROCLCR—PM

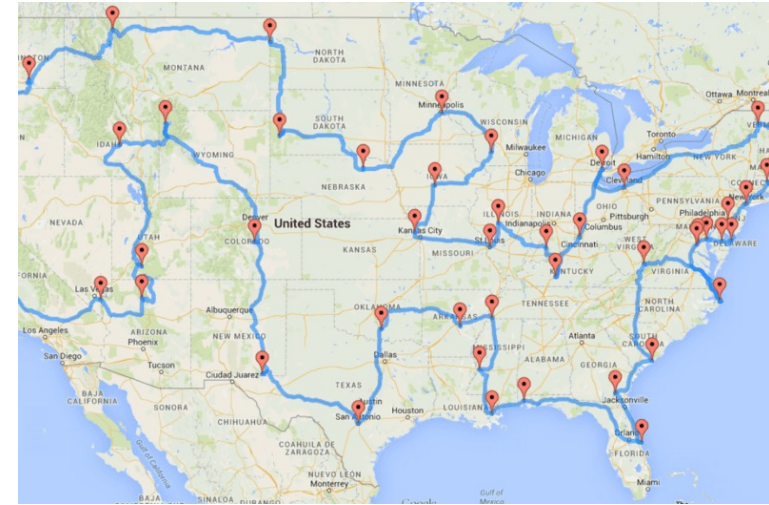
Seq2 — AJCJNR—CKCRBP—

Seems reasonable...
but not systematic/repeatable

How best to align two sequences

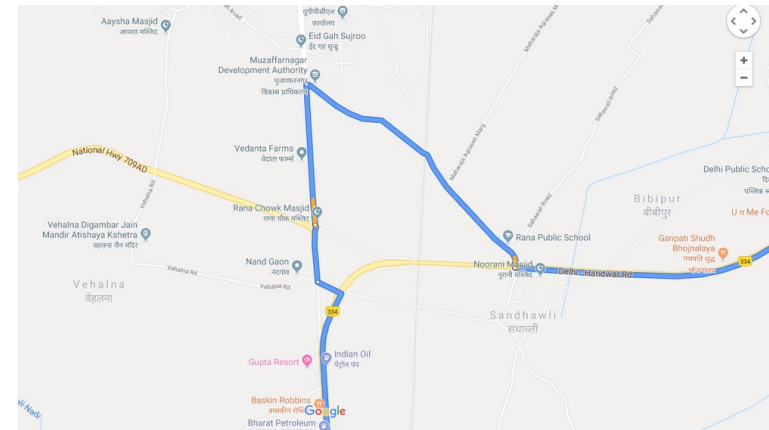
- Global alignment –

- Best overall alignment from end-to-end



- Local alignment –

- Short regions of the sequences that have homology



Scoring alignments

- Match
- Mismatch
- Gap
- Gap extension
 - One long gap is better than many short gaps

Needleman & Wunsch Global Pairwise Alignment

- Construct scoring matrix starting at origin
- Trace path through scoring matrix by selecting local maximum (start at bottom right corner, end at origin)

Needleman & Wunsch Global Pairwise Alignment Matrix construction

Rules:

Start at origin at 0

Match (diagonal) = +1

Mismatch (diagonal) = 0

Gap (right or down) = -1

Each cell gets highest possible score from three adjacent cells

[illegible]

Needleman & Wunsch Global Pairwise Alignment Matrix construction

Rules:

Start at origin at 0

Match (diagonal) = +1

Mismatch (diagonal) = 0

Gap (right or down) = -1

Each cell gets highest possible score from three adjacent cells

[illegible]

Needleman & Wunsch Global Pairwise Alignment Matrix construction

Rules:

Start at origin at 0

Match (diagonal) = +1

Mismatch (diagonal) = 0

Gap (right or down) = -1

Each cell gets highest possible score from three adjacent cells

[illegible]

Needleman & Wunsch Global Pairwise Alignment Matrix construction

Rules:

Start at origin at 0

Match (diagonal) = +1

Mismatch (diagonal) = 0

Gap (right or down) = -1

Each cell gets highest possible score from three adjacent cells

[illegible]

Needleman & Wunsch Global Pairwise Alignment Matrix construction

Rules:

Start at origin at 0

Match (diagonal) = +1

Mismatch (diagonal) = 0

Gap (right or down) = -1

Each cell gets highest possible score from three adjacent cells

[illegible]

Needleman & Wunsch Global Pairwise Alignment Matrix construction

Rules:

Start at origin at 0

Match (diagonal) = +1

Mismatch (diagonal) = 0

Gap (right or down) = -1

Each cell gets highest possible score from three adjacent cells

[illegible]

Needleman & Wunsch Global Pairwise Alignment Matrix construction

Rules:

Start at origin at 0

Match (diagonal) = +1

Mismatch (diagonal) = 0

Gap (right or down) = -1

Each cell gets highest possible score from three adjacent cells

[illegible]

Needleman & Wunsch Global Pairwise Alignment Matrix construction

Rules:

Start at origin at 0

Match (diagonal) = +1

Mismatch (diagonal) = 0

Gap (right or down) = -1


Each cell gets highest possible score from three adjacent cells

		A	B	C	N	J	R	O	C	L	C	R	P	M
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13
A	-1	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
J	-2	0	1	0	-1	-1	-2	-3	-4	-5	-6	-7	-8	-9
C	-3	-1	0	2	1	0	-1	-2	-2	-3	-4	-5	-6	-7
J	-4	-2	-1	1	2	2	1	0	-1	-2	-3	-4	-5	-6
N	-5	-3	-2	0	2	2	2	1	0	-1	-2	-3	-4	-5
R	-6	-4	-3	-1	1	2	3	2	1	0	-1	-1	-2	-3
C	-7	-5	-4	-2	0	1	2	3	3	2	1	0	-1	-2
K	-8	-6	-5	-3	-1	0	1	2	3	3	2	1	0	-1
C	-9	-7	-6	-4	-2	-1	0	1	3	3	4	3	2	1
R	-10	-8	-7	-5	-3	-2	0	0	2	3	3	5	3	2
B	-11	-9	-7	-6	-4	-3	-1	0	1	2	3	4	4	3
P	-12	-10	-8	-7	-5	-4	-2	-1	0	1	2	3	5	4

Needleman & Wunsch Global Pairwise Alignment Traceback

Rules:

Start at bottom right corner,
proceed through highest
scoring path at each cell,
(follow the arrows)

		A	B	C	N	J	R	O	C	L	C	R	P	M
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13
A	-1	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
J	-2	0	1	0	-1	-1	-2	-3	-4	-5	-6	-7	-8	-9
C	-3	-1	0	2	1	0	-1	-2	-2	-3	-4	-5	-6	-7
J	-4	-2	-1	1	2	2	1	0	-1	-2	-3	-4	-5	-6
N	-5	-3	-2	0	2	2	2	1	0	-1	-2	-3	-4	-5
R	-6	-4	-3	-1	1	2	3	2	1	0	-1	-1	-2	-3
C	-7	-5	-4	-2	0	1	2	3	3	2	1	0	-1	-2
K	-8	-6	-5	-3	-1	0	1	2	3	3	2	1	0	-1
C	-9	-7	-6	-4	-2	-1	0	1	3	3	4	3	2	1
R	-10	-8	-7	-5	-3	-2	0	0	2	3	3	5	3	2
B	-11	-9	-7	-6	-4	-3	-1	0	1	2	3	4	4	3
P	-12	-10	-8	-7	-5	-4	-2	-1	0	1	2	3	5	

Needleman & Wunsch Global Pairwise Alignment Traceback

Rules:


Start at bottom right corner, proceed through highest scoring path at each cell, (follow the arrows)

		A	B	C	N	J	R	O	C	L	C	R	P	M
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13
A	-1	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
J	-2	0	1	0	-1	-1	-2	-3	-4	-5	-6	-7	-8	-9
C	-3	-1	0	2	1	0	-1	-2	-2	-3	-4	-5	-6	-7
J	-4	-2	-1	1	2	2	1	0	-1	-2	-3	-4	-5	-6
N	-5	-3	-2	0	2	2	2	1	0	-1	-2	-3	-4	-5
R	-6	-4	-3	-1	1	2	3	2	1	0	-1	-1	-2	-3
C	-7	-5	-4	-2	0	1	2	3	3	2	1	0	-1	-2
K	-8	-6	-5	-3	-1	0	1	2	3	3	2	1	0	-1
C	-9	-7	-6	-4	-2	-1	0	1	3	3	4	3	2	1
R	-10	-8	-7	-5	-3	-2	0	0	2	3	3	5	3	2
B	-11	-9	-7	-6	-4	-3	-1	0	1	2	3	4	4	3
P	-12	-10	-8	-7	-5	-4	-2	-1	0	1	2	3	5	●

Needleman & Wunsch Global Pairwise Alignment Traceback

Rules:

Start at bottom right corner, proceed through highest scoring path at each cell, (follow the arrows)

		A	B	C	N	J	R	O	C	L	C	R	P	M
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13
A	-1	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
J	-2	0	1	0	-1	-1	-2	-3	-4	-5	-6	-7	-8	-9
C	-3	-1	0	2	1	0	-1	-2	-2	-3	-4	-5	-6	-7
J	-4	-2	-1	1	2	2	1	0	-1	-2	-3	-4	-5	-6
N	-5	-3	-2	0	2	2	2	1	0	-1	-2	-3	-4	-5
R	-6	-4	-3	-1	1	2	3	2	1	0	-1	-1	-2	-3
C	-7	-5	-4	-2	0	1	2	3	3	2	1	0	-1	-2
K	-8	-6	-5	-3	-1	0	1	2	3	3	2	1	0	-1
C	-9	-7	-6	-4	-2	-1	0	1	3	3	4	3	2	1
R	-10	-8	-7	-5	-3	-2	0	0	2	3	3	5	3	2
B	-11	-9	-7	-6	-4	-3	-1	0	1	2	3	4	4	3
P	-12	-10	-8	-7	-5	-4	-2	-1	0	1	2	3	5	

Needleman & Wunsch Global Pairwise Alignment Traceback

Rules:

Start at bottom right corner, proceed through highest scoring path at each cell, (follow the arrows)

		A	B	C	N	J	R	O	C	L	C	R	P	M
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13
A	-1	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
J	-2	0	1	0	-1	-1	-2	-3	-4	-5	-6	-7	-8	-9
C	-3	-1	0	2	1	0	-1	-2	-2	-3	-4	-5	-6	-7
J	-4	-2	-1	1	2	2	1	0	-1	-2	-3	-4	-5	-6
N	-5	-3	-2	0	2	2	2	1	0	-1	-2	-3	-4	-5
R	-6	-4	-3	-1	1	2	3	2	1	0	-1	-1	-2	-3
C	-7	-5	-4	-2	0	1	2	3	3	2	1	0	-1	-2
K	-8	-6	-5	-3	-1	0	1	2	3	3	2	1	0	-1
C	-9	-7	-6	-4	-2	-1	0	1	3	3	4	3	2	1
R	-10	-8	-7	-5	-3	-2	0	0	2	3	3	5	3	2
B	-11	-9	-7	-6	-4	-3	-1	0	1	2	3	4	4	3
P	-12	-10	-8	-7	-5	-4	-2	-1	0	1	2	3	5	●

Needleman & Wunsch Global Pairwise Alignment Traceback

Rules:

Start at bottom right corner, proceed through highest scoring path at each cell, (tie goes to the diagonal)

ABCNJRCLCRPM

AJC-JNRCKCRBP

		A	B	C	N	J	R	O	C	L	C	R	P	M
	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13
A	-1	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11
J	-2	0	1	0	-1	-1	-2	-3	-4	-5	-6	-7	-8	-9
C	-3	-1	0	2	1	0	-1	-2	-2	-3	-4	-5	-6	-7
J	-4	-2	-1	1	2	2	1	0	-1	-2	-3	-4	-5	-6
N	-5	-3	-2	0	2	2	2	1	0	-1	-2	-3	-4	-5
R	-6	-4	-3	-1	1	2	3	2	1	0	-1	-1	-2	-3
C	-7	-5	-4	-2	0	1	2	3	3	2	1	0	-1	-2
K	-8	-6	-5	-3	-1	0	1	2	3	3	2	1	0	-1
C	-9	-7	-6	-4	-2	-1	0	1	3	3	4	3	2	1
R	-10	-8	-7	-5	-3	-2	0	0	2	3	3	5	3	2
B	-11	-9	-7	-6	-4	-3	-1	0	1	2	3	4	4	3
P	-12	-10	-8	-7	-5	-4	-2	-1	0	1	2	3	5	5

Smith & Waterman Local Pairwise Alignment

- Construct scoring matrix (no negative numbers, lowest cell score possible is 0)
- Trace path through scoring matrix by selecting local maximum (start at global high score, trace toward origin until hitting a 0)

Smith & Waterman Local Pairwise Alignment Traceback

Rules:

Start at highest score, proceed through highest scoring path at each cell, (through the source from each cell), until you hit a 0

NJROCLCRPM

CJNRCKCRBP

or

NJROCLCR-P

CJNRCKCRBP

	A	B	C	N	J	R	O	C	L	C	R	P	M
A	1	0	0	0	0	0	0	0	0	0	0	0	0
J	0	1	0	0	1	0	0	0	0	0	0	0	0
C	0	0	2	1	0	1	0	1	0	1	0	0	0
J	0	0	1	2	2	1	1	0	1	0	1	0	0
N	0	0	0	2	2	2	1	1	0	1	0	0	0
R	0	0	0	1	2	3	2	1	1	0	2	1	0
C	0	0	1	0	1	2	3	3	2	2	1	2	1
K	0	0	0	1	0	1	2	3	3	2	2	1	2
C	0	0	1	0	1	0	1	3	3	4	3	2	1
R	0	0	0	1	0	2	1	2	3	3	5	4	3
B	0	1	0	0	1	1	2	1	2	3	4	5	4
P	0	0	1	0	0	1	1	2	1	2	3	4	5

Practice Run!

Go to GitHub page, download activity ppt

Next up: Multiple Sequence Alignment

Download MEGA

Homework #2 will be posted this afternoon

Please read: Needleman & Wunsch 1970, Smith & Waterman 1981