

Sequence comparison: Local alignment

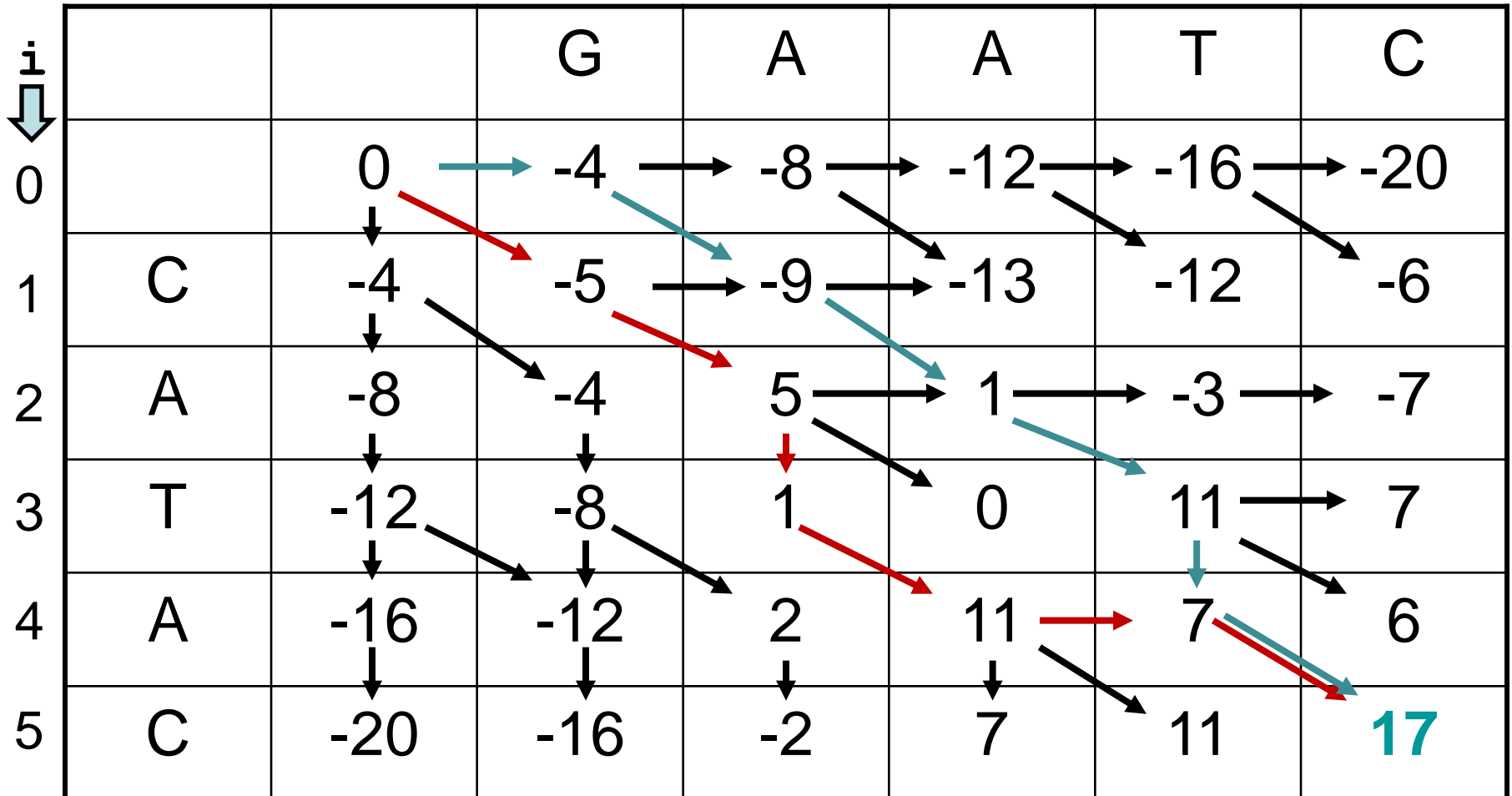
Genome 559: Introduction to Statistical
and Computational Genomics
Prof. James H. Thomas

Review - global alignment

GAAT-C
-CATAC

GA-ATC
CATA-C

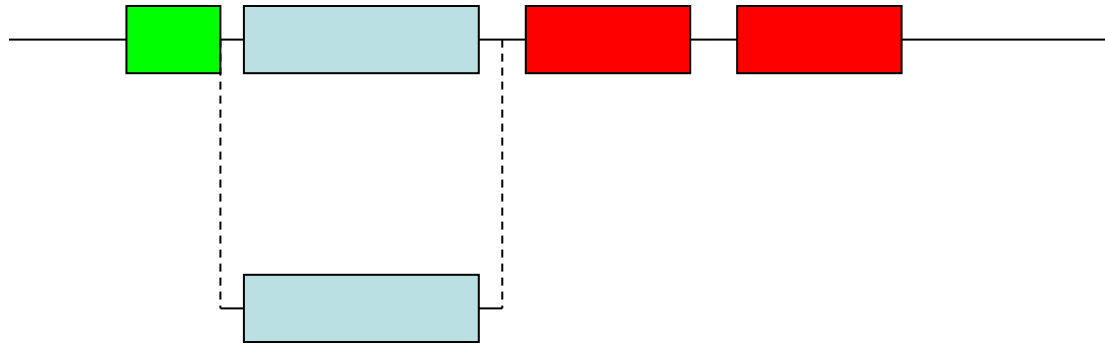
	A	C	G	T
A	10	-5	0	-5
C	-5	10	-5	0
G	0	-5	10	-5
T	-5	0	-5	10



Review - three legal moves

- A diagonal move aligns a character from each sequence.
- A vertical move aligns a gap in the sequence along the top edge.
- A horizontal move aligns a gap in the sequence along the left edge.
- The move you **keep** is the best scoring of the three.

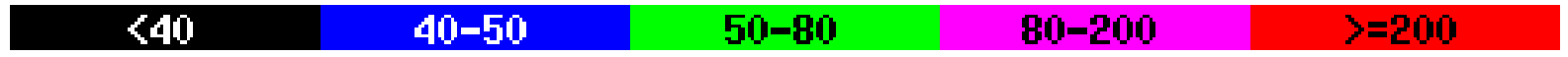
Local alignment



- A single-domain protein may be similar only to one region within a multi-domain protein.
- A DNA query may align to a small part of a genome.
- An alignment that spans the complete length of both sequences may be undesirable.

BLAST allows local alignments

Color Key for Alignment Scores



Review - global alignment DP

- Align sequence x and y .
- F is the DP matrix; s is the substitution matrix; d is the linear gap penalty.

$$F(0,0) = 0$$


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

Local alignment DP

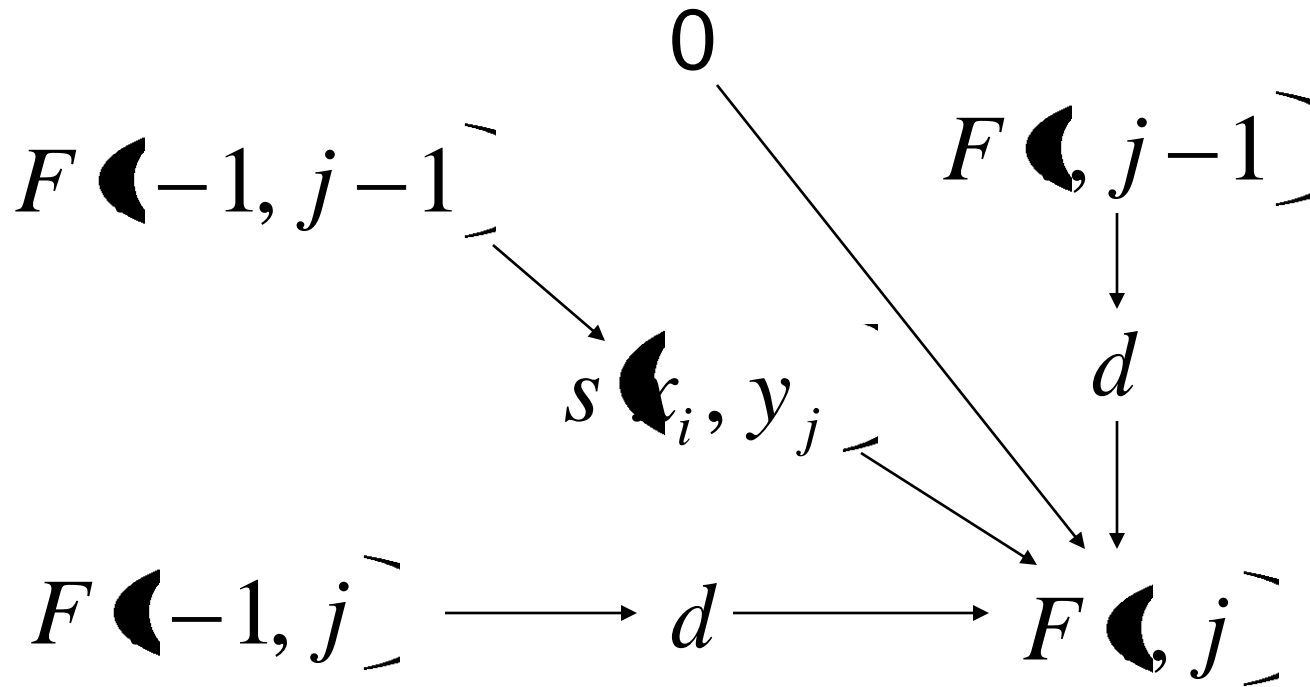
- Align sequence x and y .
- F is the DP matrix; s is the substitution matrix; d is the linear gap penalty.

$$F(0,0) = 0$$

$$F(i,j) = \max \left\{ \begin{array}{l} F(i-1, j-1) + s(x_i, y_j) \\ F(i-1, j) + d \\ F(i, j-1) + d \\ 0 \end{array} \right.$$


(corresponds to NO alignment)

Local DP in equation form

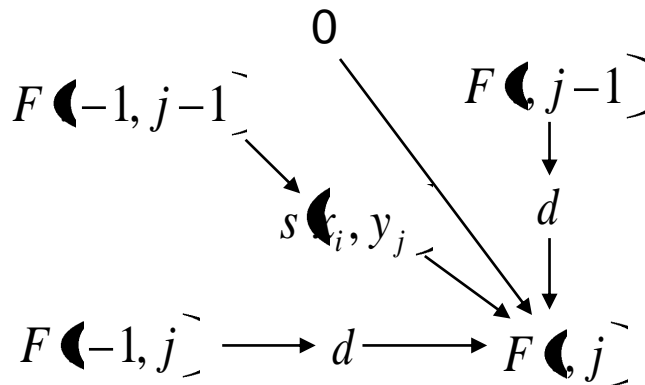


keep max of these
four values

A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

$d = -5$



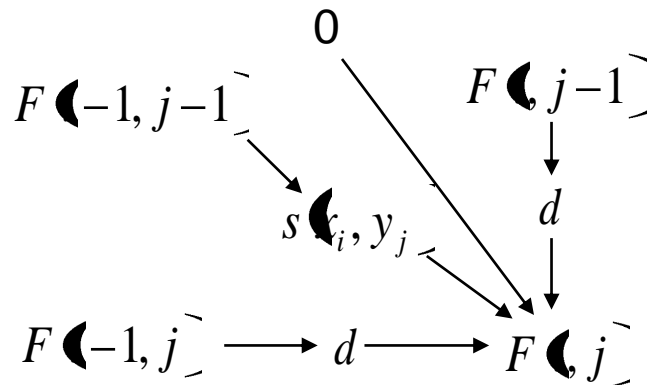
initialize the same way
as for global alignment

		A	A	G
	0			
A				
G				
C				

A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

$d = -5$

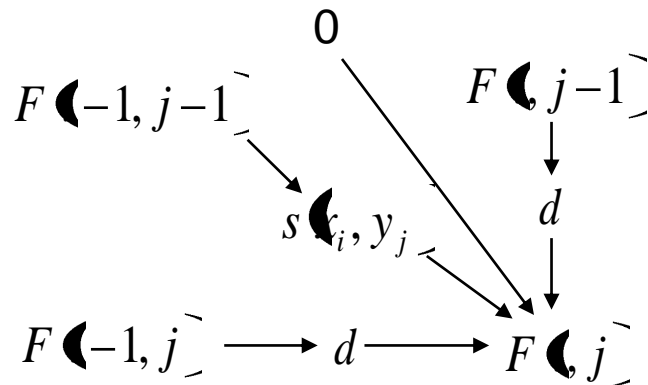


		A	A	G
	0	?	?	?
A	?			
G	?			
C	?			

A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

$d = -5$

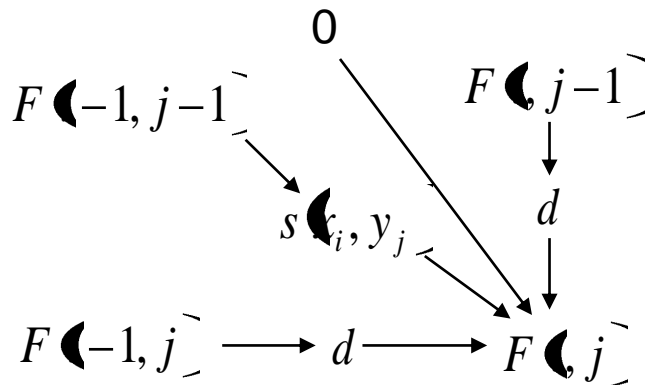


		A	A	G
	0	0	0	0
A	0	?		
G	0			
C	0			

A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

$d = -5$

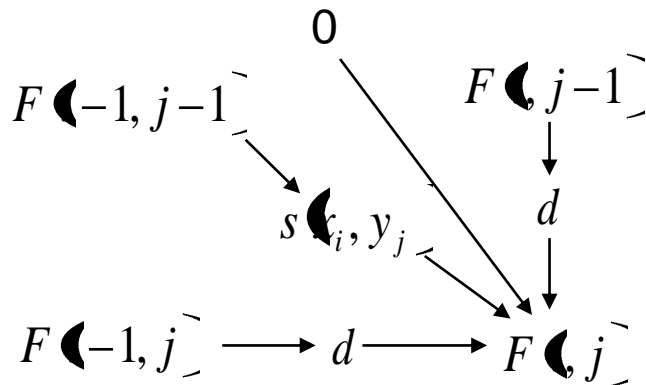


		A	A	G
	0	0	0	0
A	0	2	-5	
G	0	-5	0	
C	0			

A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

$d = -5$

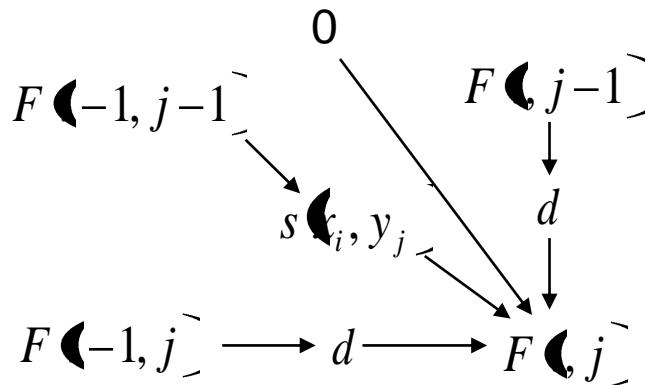


		A	A	G
	0	0	0	0
A	0	2		
G	0			
C	0			

A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

$d = -5$

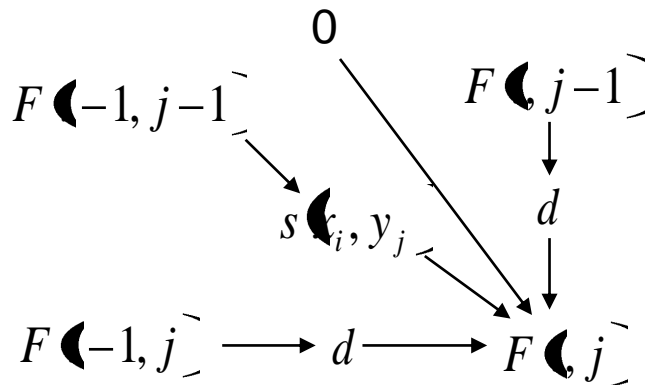


		A	A	G
	0	0	0	0
A	0	2		
G	0	?		
C	0	?		

A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

$d = -5$



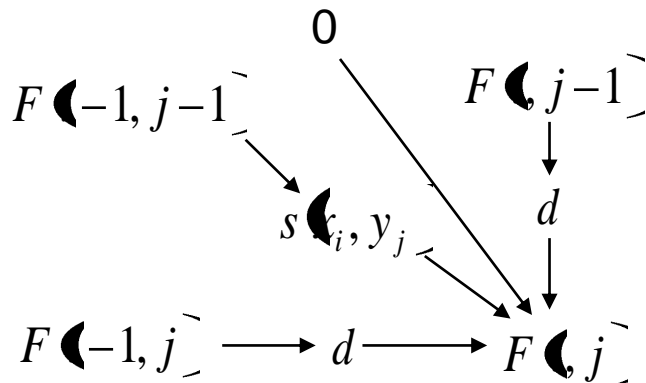
		A	A	G
	0	0	0	0
A	0	2		
G	0	0		
C	0	0		

(you can signify no preceding alignment with no arrow)

A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

$d = -5$



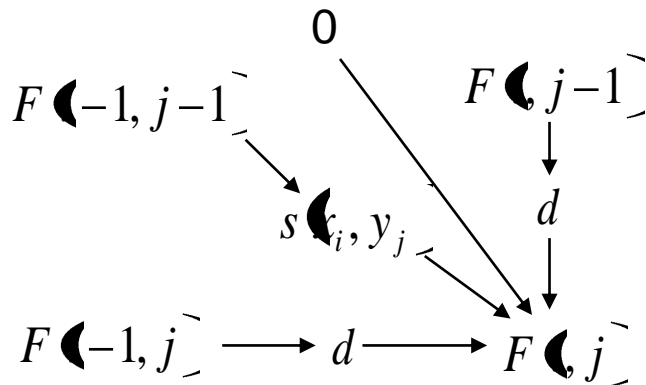
		A	A	G
	0	0	0	0
A	0	2	?	
G	0	0	?	
C	0	0	?	

(you can signify no preceding alignment with no arrow)

A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

$d = -5$



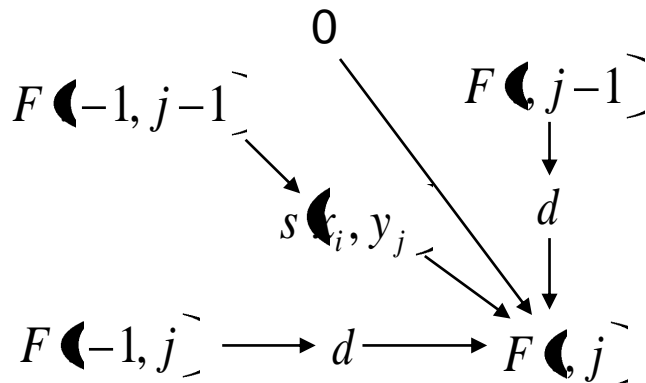
		A	A	G
	0	0	0	0
A	0	2	2	
G	0	0	0	
C	0	0	0	

(you can signify no preceding alignment with no arrow)

A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

$d = -5$



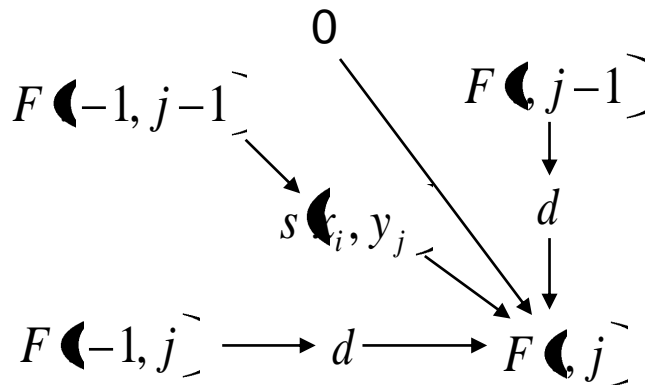
		A	A	G
	0	0	0	0
A	0	2	2	?
G	0	0	0	?
C	0	0	0	?

(you can signify no preceding alignment with no arrow)

A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

$d = -5$



		A	A	G
	0	0	0	0
A	0	2	2	0
G	0	0	0	4
C	0	0	0	0

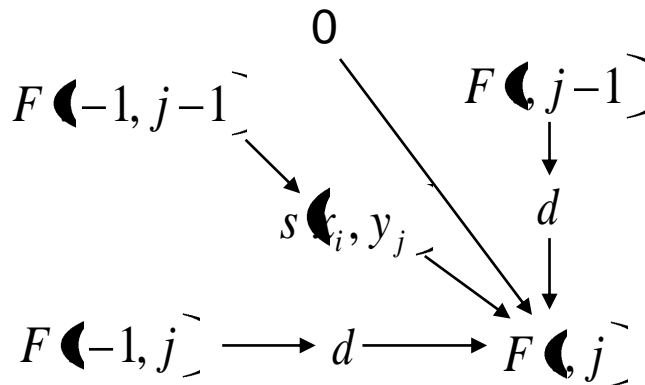
(you can signify no preceding alignment with no arrow)

Traceback alignment AG

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

$d = -5$

		A	A	G
	0	0	0	0
A	0	2	2	0
G	0	0	0	4
C	0	0	0	0



Start at highest score
anywhere in matrix, follow
 arrows back until you reach 0

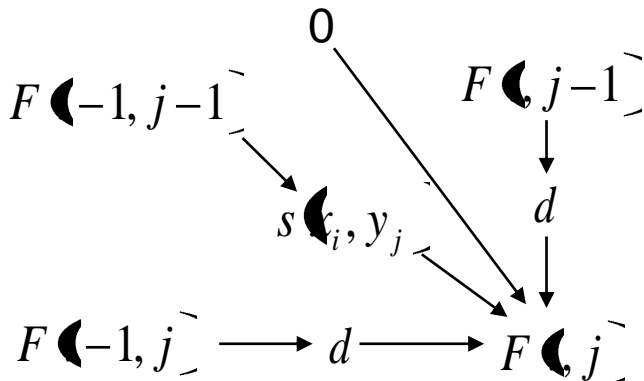
Local alignment

- Two differences from global alignment:
 - If a score is negative, replace with 0.
 - Traceback from the highest score in the matrix and continue until you reach 0.
- Global alignment algorithm: *Needleman-Wunsch*.
- Local alignment algorithm: *Smith-Waterman*.

Another example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

Find the optimal local alignment of *AAG* and *GAAGGC*.
Use a gap penalty of $d = -5$.



		A	A	G
	0	0	0	0
G	0	0	0	2
A	0	2	2	0
A	0	2	4	0
G	0	0	0	6
G	0	0	0	2
C	0	0	0	0

Traceback

		A	A	G
	0	0	0	0
G	0	0	0	2
A	0	2	2	0
A	0	2	4	0
G	0	0	0	6
G	0	0	0	2
C	0	0	0	0

AAG

AAG

DP matrix

		A	A	G
G	0	0	0	0
A	0	0	0	2
A	0	2	2	0
G	0	2	4	0
G	0	0	0	6
G	0	0	0	2
C	0	0	0	0

Traceback matrix

You don't actually need first row and column

(-10)	(-10)	(-10)	(-10)
(-10)	-10	-10	0
(-10)	0	0	-10
(-10)	0	0	-10
(-10)	-10	-10	0
(-10)	-10	-10	0
(-10)	-10	-10	-10

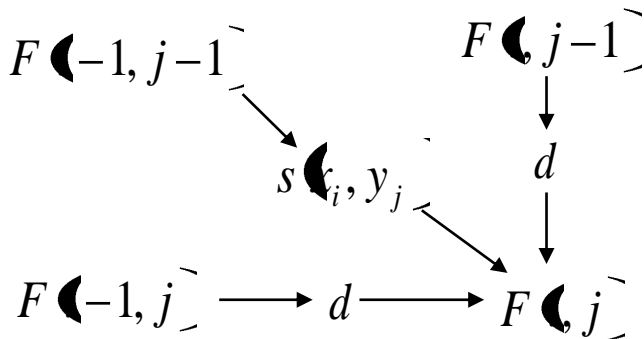
0 = diagonal, -1 = gap left, +1 = gap top, -10 = no alignment

Problem - find the best GLOBAL alignment

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

Find the optimal global alignment of AAG and GAAGGC.
Use a gap penalty of $d = -5$.

		A	A	G
	0	-5	-10	-15
G	-5			
A	-10			
A	-15			
G	-20			
G	-25			
C	-30			



(contrast with the best local alignment)