Gotoh-Algorithm

Motivation

- Obervation: In nature are a few long gaps much more likelier as lot of short gabs

- Needleman-Wunsch is ignoring this

- New scoring system: gap penalty g(x) have to subadditive:

$$g(x+y) \leq g(x) + g(y)$$

- “Cheaper to insert a existing gap as to insert a new one”

- g(x) have to be affine:

$$g(x) = \alpha + \beta * x$$

- Algorithm of Smith-Waterman-Beyer is exact but needs a runtime of $O(n^3)$
- Use a better approach instead —> Gotoh's algorithm

Gotoh's algorithm

- Works similar to Needleman-Wunsch
- Uses three matrices:
  - D: The cost for the alignment like in N-W
  - P: The cost if the alignment ends with an gap in b
  - Q: The cost if the alignment ends with an gap in a

Gotoh's algorithm

- Recursion:
- First matrix D:

$$D_{i,j} = \min \begin{cases} D_{i-1,j-1} + w(a_i, b_j) \\ P_{i,j} \\ Q_{i,j} \end{cases}$$

- Second matrix P:

$$P_{i,j} = \min \begin{cases} D_{i-1,j} + g(1) \\ P_{i-1,j} + \beta \end{cases}$$

- Third matrix Q:

$$Q_{i,j} = \min \begin{cases} D_{i,j-1} + g(1) \\ Q_{i,j-1} + \beta \end{cases}$$

Formulas from [1]

Gotoh's algorithm

$D_{0,0} = 0$

$D_{0,j} = g(j)$

$D_{i,0} = g(i)$

$P_{i,0}$ = not used

$P_{0,j}$ = infinite

$Q_{i,0}$ = ininite

$Q_{0,j}$ = not used

for i = 1 to n

for j = 1 to n

$P_{i,j}$

$Q_{i,j}$

$D_{i,j}$

end

end

Example

D

	ϵ	**A**	**C**
ϵ	0	3	4
A	3	0	3
G	4	3	1
C	5	4	3

P

	ϵ	**A**	**C**
ϵ	-	inf	inf
A	-	6	7
G	-	3	6
C	-	4	4

Q

	ϵ	**A**	**C**
ϵ	-	-	-
A	inf	6	3
G	inf	7	6
C	inf	8	7

$$w(a,b) = \begin{cases} 0, & a = b \\ 1, & a \neq b \end{cases}$$

g(k) = 2 + 1k

$$D_{i,j} = \min \begin{cases} D_{i-1,j-1} + w(a_i, b_j) \\ P_{i,j} \\ Q_{i,j} \end{cases}$$

$$Q_{i,j} = \min \begin{cases} D_{i,j-1} + g(1) \\ Q_{i,j-1} + \beta \end{cases}$$

$$P_{i,j} = \min \begin{cases} D_{i-1,j} + g(1) \\ P_{i-1,j} + \beta \end{cases}$$

Formulas from [1]

Example - Traceback

D

	ϵ	**A**	**C**
ϵ	0	3	4
A	3	0	3
G	4	3	1
C	5	4	3

P

	ϵ	**A**	**C**
ϵ	-	inf	inf
A	-	6	7
G	-	3	6
C	-	4	4

Q

	ϵ	**A**	**C**
ϵ	-	-	-
A	inf	6	3
G	inf	7	6
C	inf	8	7

$$D_{i,j} = D_{i-1,j-1} + w(a_i, b_j)$$
$$D_{i,j} = Q_{i,j},$$
$$D_{i,j} = P_{i,j};$$

$D_{3,2} = 3 == D_{3\text{-}1,2\text{-}1} + w(C, C) =$

$3+0 = 3$

$D_{3,2} = 3 != P_{3,2} = 4$

$D_{3,2} = 3 != Q_{3,2} = 7$

Formulas from [1]

Example - Traceback

D

	ϵ	**A**	**C**
ϵ	0	3	4
A	3	0	3
G	4	3	1
C	5	4	3

P

	ϵ	**A**	**C**
ϵ	-	inf	inf
A	-	6	7
G	-	3	6
C	-	4	4

Q

	ϵ	**A**	**C**
ϵ	-	-	-
A	inf	6	3
G	inf	7	6
C	inf	8	7

$$D_{i,j} = D_{i-1,j-1} + w(a_i, b_j)$$
$$D_{i,j} = Q_{i,j},$$
$$D_{i,j} = P_{i,j};$$

$D2_{,2} = 3 == D_{2\text{-}1,2\text{-}1} + w(G, A)$

$= 3+1 = 4$

$D_{2,2} = 3 == P_{2,2} = 3$

$D_{2,2} = 3 \text{ != } Q_{2,2} = 7$

Formulas from [1]

Example - Traceback

D

	ϵ	**A**	**C**
ϵ	0	3	4
A	3	0	3
G	4	3	1
C	5	4	3

P

	ϵ	**A**	**C**
ϵ	-	inf	inf
A	-	6	7
G	-	3	6
C	-	4	4

Q

	ϵ	**A**	**C**
ϵ	-	-	-
A	inf	6	3
G	inf	7	6
C	inf	8	7

$$P_{i,j} = P_{i-1,j} + \beta,$$
$$P_{i,j} = D_{i-1,j} + g(1)$$

$P_{2,2} = 3 \; != P_{2\text{-}1,2} + 1 = 6+1 = 7$

$P_{2,2} = 3 == D_{2\text{-}1,2} + g(1) = 0+3$

Formulas from [1]

Example - Traceback

D

	ϵ	**A**	**C**
ϵ	0	3	4
A	3	0	3
G	4	3	1
C	5	4	3

P

	ϵ	**A**	**C**
ϵ	-	inf	inf
A	-	6	7
G	-	3	6
C	-	4	4

Q

	ϵ	**A**	**C**
ϵ	-	-	-
A	inf	6	3
G	inf	7	6
C	inf	8	7

$$D_{i,j} = D_{i-1,j-1} + w(a_i, b_j)$$
$$D_{i,j} = Q_{i,j},$$
$$D_{i,j} = P_{i,j};$$

$D_{1,1} = 0 == D_{1\text{-}1,1\text{-}1} + w(A, A) =$

$0+0 = 0$

$D_{1,1} = 0 != P_{1,1} = 6$

$D_{1,1} = 0 != Q_{1,1} = 6$

Formulas from [1]

Example - Traceback

D

	ϵ	**A**	**C**
ϵ	0	3	4
A	3	0	3
G	4	3	1
C	5	4	3

P

	ϵ	**A**	**C**
ϵ	-	inf	inf
A	-	6	7
G	-	3	6
C	-	4	4

Q

	ϵ	**A**	**C**
ϵ	-	-	-
A	inf	6	3
G	inf	7	6
C	inf	8	7

$D\nwarrow \quad P\bullet \quad D\uparrow \quad D\nwarrow$

$$
\begin{aligned}
D\nwarrow \in \mathrm{tr}^D_{i,j} &\Leftrightarrow D_{i,j} = D_{i-1,j-1} + w(a_i, b_j),\\
Q\bullet \in \mathrm{tr}^D_{i,j} &\Leftrightarrow D_{i,j} = Q_{i,j},\\
P\bullet \in \mathrm{tr}^D_{i,j} &\Leftrightarrow D_{i,j} = P_{i,j};
\end{aligned}
$$

$$
\begin{aligned}
P\uparrow \in \mathrm{tr}^P_{i,j} &\Leftrightarrow P_{i,j} = P_{i-1,j} + \beta,\\
D\uparrow \in \mathrm{tr}^P_{i,j} &\Leftrightarrow P_{i,j} = D_{i-1,j} + g(1);
\end{aligned}
$$

$$
\begin{aligned}
Q\leftarrow \in \mathrm{tr}^Q_{i,j} &\Leftrightarrow Q_{i,j} = Q_{i,j-1} + \beta,\\
D\leftarrow \in \mathrm{tr}^Q_{i,j} &\Leftrightarrow Q_{i,j} = D_{i,j-1} + g(1)
\end{aligned}
$$

Formulas from [1]

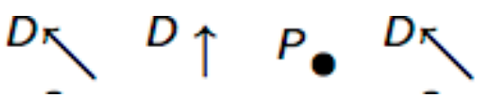
Example - Alignment

If you want insert gaps in:

- a: Arrow to the left: gap
- b: Arrow to the up: gap

Else: use character from a respectively b

- Point: Just change the matrix.
- Iterate over the list in reverse order



A	G	C
A	-	C

Formulas from [1]

List of references

- [1] Lecture “Sequence Alignment - Gap Penalties, Gotoh’s Algorithm and Smith/Waterman’s Local Alignment” Bioinformatics I, Prof. Backofen

  URL: http://www.bioinf.uni-freiburg.de//Lehre/Courses/2014_SS/V_Bioinformatik_1/gap-penalty-gotoh.pdf, visited: 13/11/2014