Ivan Capistran @01282924

Homework 2

1)

1) F(m,n) = f(m - 1, n) + f(m, n-1) + f(m-1, n-1)

With initial conditions  
 f(0,m) = f(n,0) n,m

2)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 | 21 |
| 1 | 5 | 13 | 25 | 41 | 61 | 85 | 113 | 145 | 181 | 221 |
| 1 | 7 | 25 | 63 | 129 | 231 | 377 | 575 | 833 | 1159 | 1561 |
| 1 | 9 | 41 | 129 | 321 | 681 | 1289 | 2241 | 3649 | 5641 | 8361 |
| 1 | 11 | 61 | 231 | 681 | 1683 | 3653 | 7183 | 13073 | 22363 | 36365 |
| 1 | 13 | 85 | 377 | 1289 | 3653 | 8989 | 19825 | 40081 | 75517 | 134245 |
| 1 | 15 | 113 | 575 | 2241 | 7183 | 19825 | 48639 | 108545 | 224143 | 433905 |
| 1 | 17 | 145 | 833 | 3649 | 13073 | 40081 | 108545 | 265729 | 598417 | 1256465 |
| 1 | 19 | 181 | 1159 | 5641 | 22363 | 75517 | 224143 | 598417 | 1462563 | 3317445 |
| 1 | 21 | 221 | 15651 | 8361 | 36365 | 154245 | 433905 | 1256465 | 3317445 | 8097453 |

3)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| -ab-  x-yz | -a-b  x-yz | a-b-  xy-z | -ab-  xy-z | --ab  xy-z |
| a--b-  -xy-z | a---b  -xyz- | -ab--  x--yz | -a-b-  x-y-z | -a--b  x-yz- |
| a--b  -xyz | --ab  xyz- | a--b  xyz- | -a-b  xyz- | a-b--  -x-yz |
| --ab  xy--z | --a-b  xy-z- | ab---  --xyz | ---ab  xyz-- | ab--  x-yz |
| ab- xyz | a-b  xyz | -ab  xyz | ab--  xyz | a-b-  -xyz |

3.)

Achieving linear space for the Smith-Waterman algorithm can be attained using Hirschberg's insight. Hirschberg’s alignment algorithm gives optimal alignment using linear space. The way the algorithm works is that it splits checking the score pass into smaller subproblems. First the algorithm makes a forward score-only pass stopping at the middle row or half of the space. Then there is a backward score-only pass that stops at the middle or half way point. Thus, at each point along the middle row, there is now optimal score from the starting point (0,0) to the middle points and the optimal score from the middle points to the end (M,N). Adding these numbers gives the optimal score over all paths from the starting point (0,0) to the end point (M,N). Checking the sums of the middle section determines a point (mid,j) where an optimal path crosses the middle row. This reduces the problem to finding an optimal path from (0,0) to (mid,j) and an optimal path from (mid,j) to (M,N), which is done recursively. The corresponding alignment can be found by maximizing the forward score pass and the backward score pass. The path can be traced from the middle to find the best alignment.

4)  
 a) Did not get to complete the question and did not write the program.  
  
 b) I started on the program but have it incomplete. Was trying to configure the gap penalty and could not finish it in time. I have included the source code in the zip folder as biohw2 which is the folder to the java project in eclipse. The program is written in Java and it outputs the original matrix and the sequence alignment. Does not output the information required by the problem. The outputs are located in the source code folder. Example the output for seq\_25.txt is called seq\_25\_output.txt. All the outputs are in the biohw2 folder.

5)  
 a)  
 Accession number - NP\_000607

458 Amino acids

MNRGV first amino acids

b)   
 Gorilla Gorilla Gorilla (Western Gorilla)  
 Pan Paniscus (Bondoo Pygmy Chimpanzee)

Nomascus Leucogenys (Northern white cheeked gibbon)  
 Rhinopithecus Roxellana (Golden snub nosed monkey)  
 Macaca Mulatta (Rhesus Macaque)

c)

S’ = ( (.203)x274 - ln(0.041) )/ln2   
 S’ = 84.85

E = 458 x 16,719,817,031 x 2^ 84.85   
 E = 2.19635x10^-13

d)

Accession - NM\_000616

Exons 1 … 3116

e)

Word size 28

Math/Mismatch- 1,-2

Gap costs - Linear

122 hits found

Last significant hit : sequence accession XM\_004800856.1

Score 113

E-Value 4e^-21

Alignment Length 2442

Percent of identities - 131/165 79%

Percent of gaps - 3/165 1%

f)

Word size : 11

Math/Mismatch : 2,-3

Gap cost - Existence: 5 Extension: 2

155 blast hits

Yes, still in the search

Score 737

E-Value 0.0

Length 2442

Percent Identities : 1105/1571 70%

Gaps 75/1571 4%

Bonus - Spent about 3 days on the assignment. Discussed with Matthew and Marcela about number 5. 8-10 difficulty. No comments.