## **Phyloinformatics**

#### Assignment-3

Perform a global pairwise sequence alignment based on the following parameter.

Sequence 1: PHWSSILPY Sequence 2: PHWALSY

Scoring parameters: BLOSUM62 Gap penalty: Linear gap penalty of 6

Q1- The report consist the equation, calculation, filled matrix, backtracking and optimal alignment

## 1. Filled Matrix:

The first and crucial step of the algorithm is matrix filling starting from the upper left hand corner of the matrix. To find the maximum score of each cell, it is required to know the neighbouring scores (diagonal, left and right) of the current position. From the assumed values, add the match or mismatch (assumed) score to the diagonal value. Similarly add the gap score to the other neighbouring values. Thus, we can obtain three different values, from that take the maximum among them and fill the i th and j th position with the score obtained.

		P	Н	W	S	S	I	L	P	Y
	0	-6	-12	-18	-24	-30	-36	-42	-48	-54
Р	-6	7	1	-5	-11	-17	-23	-29	-35	-41
Н	-12	1	15	9	3	-3	-9	-15	-21	-27
W	-18	-5	9	26	20	14	8	2	-4	-10
Α	-24	-11	3	20	27	21	15	9	3	-3
L	-30	-17	-3	14	21	25	23	19	13	7
S	-36	-23	-9	8	18	25	23	21	18	12
Υ	-42	-29	-15	2	12	19	24	22	18	25

## 2. Backtracking:

The second step in the algorithm is the trace back for the best alignment. In the above mentioned matrix, one can see the bottom right hand corner score as 25. The important point to be noted here is that there may be two or more alignments possible between the two sequences.

The current cell with value 25 has immediate predecessor, where the maximum score obtained is diagonally located and its value is 18. If there are two or more values which points back, suggests that there can be two or more possible alignments.

By continuing the trace back step by the above defined method, one would reach to the 0th row, 0th column. Following the above described steps, alignment of two sample sequences can be found. The best alignment among the alignments can be identified by using the maximum alignment score which may be user defined.

		P	Н	W	S	S	I	L	P	Y
	0	-6	-12	-18	-24	-30	-36	-42	-48	-54
		<b>—</b>	<b>—</b>	<b>—</b>	<b>—</b>	<b>—</b>	<b>←</b>	<b>←</b>	<b>←</b>	<b>←</b>
P	-6	7	1	-5	-11	-17	-23	-29	35	-41
						<b>—</b>	<b>—</b>		<b>—</b>	<b>—</b>
Н	-12	1	15	9	3	-3	-9	-15	-21	-27
W	-18	-5	9	26	20	14	8	2	-4	-10
VV	1 1			20	<b>4</b>	<b>—</b>	<b>—</b>	<b>—</b>	<b>←</b>	<del>-</del> 10
A	-24	-11	3	20	27	21	15	9	3	-3
L	-30	-17	-3	14	21	25	23	19	13	7
S	-36	-23	-9 <b>1</b>	8	18	25	23	21	18	12
Y	-42	-29	-15	2	12	19	24	22	18	25

# 3. Equations:

a)Linear gap penalty score:

$$\gamma(g) = -gd$$

b)Boundary conditions

$$F(i, 0) = -i d$$

$$F(j, 0) = -j d$$

$$F(i, j) = F(i-1, j-1) + s(xi, yj)$$

$$F(i, j) = F(i-1, j) - d$$

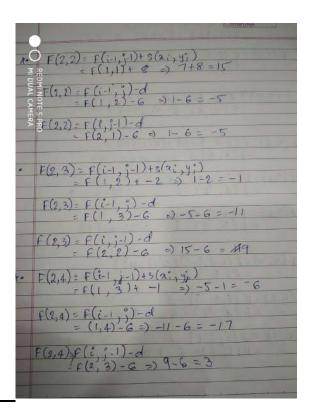
$$F(i, j) = F(i, j-1) - d$$

# 4.Calculation:

	BID - 1900 6.
O REDA	
II NOTE 5 PI JAL CAMER <i>i</i>	= -6-2 -) -8
RO	f(1,2) = f(i-1,j) = $f(0,2) - d$ = $-12 - 6 = -13$
	F(1,2) = f(i,j-1) = $F(1,1) - d$ =) $7-6 = 1$
•. #	F(1,3) = F(i-1, j-1) + s(xi, y;) = $F(0, 2) + -4$
ō	= -12-4 = -16 = -12-4 = -16 = -16 = -16 = -16 = -16 = -16 = -16 = -16 = -16 = -16
	= -18-6=-24
	= f(i, j-1)-d = $f(i, 2)-6=1-6$ = $i-5$

	Corrière
C MI BUA	$F(1,+) = F(i-1,3-1) + 3/ac, y; i)$ $= F(0,3) - 1.$ $\Rightarrow -18 + -1 \Rightarrow -19$
AL CAMERA	F(14) = F(t, 3) - d - f(t, 3) - 6 11
	F(1,4) = F(1-1,1) -d = F(0,1) - 6 => -24 - 6 = -30
· ±	$f(i, s) = f(i-1, j-1) + f(\alpha i, \gamma j)$ = $f(0, 4) + -1 = -2 s$
	F(1,5) = F(1-1,1) - d = $F(0,5) - 6 = 1 - 36$
	F(1,5)=F(i,j-1)-d=(1,4)-6=-17 F(1,6)=F(i-1,j-1)+s(xi,yj) =F(0,5)+5=)-3-30=-33
	F(1,6) = F(1-1,0) -d = (0,6) -6 = -42
1	-(1,6) = F (1/51) -d= (1,5)-6 -17-6 = -23
-	

```
(F(1,7) = F(i-1,j-1) + s(xi, yj)
 = f(0,6)+s=) -36-3=3-39
*F(1,7)=) F(i-1,3) -d
 DD = F(0,7)-6=)-42-6=-48
 E(37) = f(i, j-1) -d
 R PRO = (1,6)-6 =) -23-6 =)-29.
+ F(1,8) = F(i-1, j-1)+s(ai,yj)
       = F(0,7)+7=)-42+7--45
  F(1,8) = F(0,8)-6 =)-48-6 =-54
 F(1,8) = F(1,7)-6=)-29-6=-35
.. F(1,9) = f(i-1,1)+s(ai,y1)
=)f(0,8)+-3=-48-3=)-s1
F(1,9)=F(i-1), 1)-d
     = $10,9-6=)-54-6=-60
F(1,9)=F(c, j-1)-d
    =f(1,8)-6= -35-6
         =) -41.
```



```
F(2,5) = F(i-1, i-1)+s(2i, yi)

= F(1, 4) + -1 = -11-1 = -12

F(2,5) = F(i-1)i) -d

= (1,5) - G = -17-G = -23

F(2,5) = F(i,i-1) - G

= F(2,6) = F(i-1,i-1)+s(2i,yi)

= F(1,6) + -3 = -17-3 = -20

F(2,6) = F(i-1,i-1)+s(2i,yi)

= F(1,6) + -3 = -17-3 = -20

F(2,6) = F(i-1,i-1)+s(2i,yi)

= F(1,6) - G = -23-G = -29

F(2,6) = F(i-1,i-1)+s(2i,yi)

= F(1,6) + -3 = -23-3 = -26

F(2,7) = F(i-1,i-1)+s(2i,yi)

= F(2,7) - G = -3-G = -9

F(2,7) = F(i-1,i-1)+s(2i,yi)

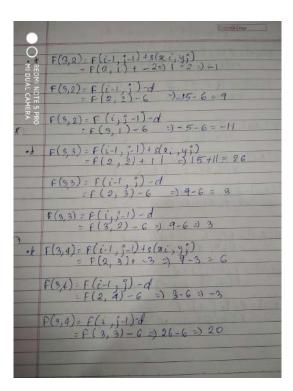
= F(2,7) - G = -3-G = -9

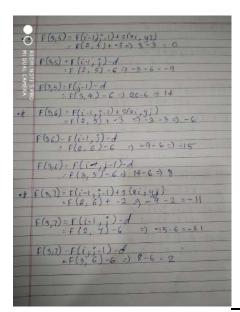
F(2,7) = F(i-1,i-1)+s(2i,yi)

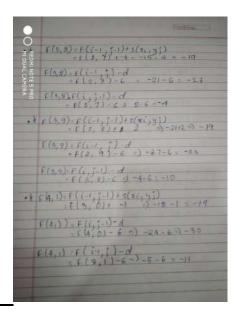
= F(2,7) - G = -3-G = -9

F(2,7) = F(i-1,i-1)-d = -29-G = -35

= F(2,7) = F(i-1,i-1)-d = -29-G = -35
```







```
F(a,8)=F(i-1,j-1)+s(2i,9j)
= F(3,7)+-1=0 2-1=1

F(4,8)=F(i-1,j)-d
= F(3,8)-G=0-4-G=-10

F(4,8)=F(i,j-1)-d
= F(4,9)=F(i-1,j-1)+s(2i,9j)
= F(3,8)+-2=0-4-2=-6

F(4,9)=F(i-1,j)-d
= F(3,9)-G=0-10-G=-16

F(4,9)=F(i-1,j)-d
= F(4,8)-G=0-2

F(6,1)=F(i-1,j)-d
= F(4,0)+-3=0-2

F(5,1)=F(i-1,j)-d
= F(4,0)-d
```

```
F(5,2) = F(i-1) \frac{1}{3} - 1 + 3(2i) \frac{1}{3} \frac{1}{3}
= F(4,1) + -3 = -14
F(5,2) = F(i-1) \frac{1}{3} - d
= F(4,2) - 6 = 1 \cdot 3 - 6 = -3
F(5,2) = F(i-1) \frac{1}{3} - 1 + d
= F(5,2) - F(i-1) - 6 = -17 - 6 = 1 - 23
F(5,3) = F(i-1) - 6 = -17 - 6 = 1 - 23
F(5,3) = F(i-1) - 6 = 1 - 17 - 6 = 1 - 23
F(5,3) = F(i-1) - 1 + 5(2i) \frac{1}{3} - 6 = -14
F(5,3) = F(i-1) - 1 - 6 = 16 - 9
F(5,4) = F(i-1) - 1 + 5(2i) \frac{1}{3} - 6 = -16
F(5,4) = F(i-1) - 1 + 3(2i) \frac{1}{3} - 6 = 18
F(5,4) = F(i-1) - 1 - 6 = 16 - 16 - 16 = 16
F(5,4) = F(i-1) - 1 - 6 = 16 - 16 = 16
F(5,4) = F(i-1) - 1 - 6 = 16 - 16 = 16
F(5,4) = F(i-1) - 1 - 6 = 16 - 16 = 16
= F(5,3) - 6 = 16 - 16 = 16
```

```
• F(5,5)=F(i-1,j-1)+s(xi)y_j^2

• F(4,4)+-2=)27-2=25

• F(4,5)-6=)21-6=15

• F(5,6)=F(i-1,j-1)+s(xi)y_j^2

• F(5,7)=F(i-1,j-1)+s(xi)y_j^2

• F(5,7)=F(i-1,j-1)+s(xi)y_j^2
```

```
F(5,8) = f(i-1,3-i) + s(x+1,3)
F(4,x) + -3 + 3 + 6 = 6
F(6,8) = F(i-1,3) - d
F(5,8) = F(i-1,3-i) - d
F(5,9) = F(i-1,3-i) - d
F(5,9) = F(i-1,3-i) + s(x+1,3-i) = 2
F(5,9) = F(i-1,3-i) - d
F(5,9) = F(i-1,3-i) + s(x+1,3-i)
F(5,9) = F(i-1,3-i) - d
F(6,9) = F(6,9-i) - d
```

Beetiles.
O F (6,2) = F(1-1,5-1)+5(x1,3/1) = F(5,1)+0-1-3-17-1-3-18
$= \begin{cases} F(\xi, 2) = F(i-1, 1) - d \\ F(i-1, 2) - G(i-1, 2) $
F(6,2)= F(i, 0-1)-d = F(6,1)-6 => -23-6=>-29
+ F(6,3) = F(1-1, 1-1)+3(ac, 45) • F(5,2)+ -3 = -3-3=-6
F(6,3) = F(1-1,1) -d = F(5,3) -6 => 14-6> 9
F(6,3)=F(6,3)-6 3-9-63-15
+ F(6,4)= F(6-1,1-1)+ s(xc,y;) = F(6,3) + 4 =) 14+4=18
F(6,4)=F(1-1°)-d =F(5,4)-6 =) 21-6=) 15
F(6,4) = F(1,7-1)-d = F(6,3)-6 =>) 8 ~6 = 2

```
F(6,5) = F(i-1,j-1) + S(x_i,y_j)
= F(5,d) + d \Rightarrow 51 + 9 = 25
F(6,5) = F(i-1,j-1) + d \Rightarrow 25 - c = 19
F(6,5) = F(i,j-1) - d \Rightarrow 18 - 6 \Rightarrow 12
F(6,6) = F(i-1,j-1) + S(x_i,y_j)
= F(5,j-1) - d \Rightarrow 23 - 6 = 19
F(6,6) = F(i-1,j-1) - d \Rightarrow 23 - 6 = 19
F(6,6) = F(i,j-1) - d \Rightarrow 23 - 6 = 19
F(6,6) = F(i,j-1) - d \Rightarrow 23 - 6 = 19
F(6,7) = F(i-1,j-1) + S(x_i,y_j)
= F(5,j-1) - d \Rightarrow 23 - 6 = 19
F(6,7) = F(i-1,j-1) + S(x_i,y_j)
= F(5,j-1) - d \Rightarrow 23 - 6 = 19
F(6,7) = F(i-1,j-1) + S(x_i,y_j)
= F(5,j-1) - d \Rightarrow 23 - 6 = 19
F(6,7) = F(i-1,j-1) - d \Rightarrow 23 - 6 = 19
F(6,7) = F(i,j-1) - d \Rightarrow 23 - 6 = 19
```

```
F(6,3) = F(i-1,i-1) + S(2i,4i)
F(6,3) = F(i-1,i-1) + S(2i,4i)
F(6,3) = F(i-1,i-1) - d
F(6,7) - G
F(7,1) = F(i-1,j-1) + S(2i,4j)
F(7,1) = F(i-1,j-1) - G
```

```
. + F(7,2) = F(i-1,j-1)+s(ai,yj)
= F(e,1) + 2 => -2
 -) -9-6=-15
 (7,2) = f(1,j-1) - d
      =) F(7, 1) -6 =) -29-6 =)-35
· + F(7,3) =) F(i-1);-1)+5(xi,yi)
         =) F(6,2) + 2 =) -9+2=-7
 f(7,3) = f(i-1) 9) -d
       = f(6,3)-6 => 8-6=) 2
  F(78) = F(i, j-1) -d
       = f(7,2)-6 => -15-6= -21
* + F(7.4) = F(i-1,j-1)+s(ai,y;)
      = [(6,3)+ -2 =)8-2=6
 F(7,4) = F(i-1, 1) -d
      = f(6,4)-6 =) 18-6=) 12
F(7,4)=F(i, j-1)-d
      = F(7, 3)-6 =) 2-6=-4
```

$$F(7,8) = F(i-1,j-1)+S(xi,j-1)$$

$$= F(6,7) + -3 = 21 - 3 = 18$$

$$O(7,8) = F(i-1,j-1) - d$$

$$= F(6,8) - G = 18 - 6 = 12$$

$$= F(7,7) - G = 22 - 6 = 16$$

$$-x(7,9) = F(i-1)j-1)+s(xi,yj)$$

$$= F(6,8) + 7 = 18 + 7 = 25$$

$$F(7,9) = F(i-1,j) - d$$

$$= F(6,9) - G = 12 - G = G$$

$$F(7,9) = F(i,j-1) - d$$

$$= F(7,8) - G$$

$$= 18 - G = 12$$

#### **5.Optimal alignment:**

P H W S S I L P Y
P H W A L S Y

**BLOSUM**: BLOcks SUbstitution Matrix, developed by Henikoff and Henikoff in 1992, used conserved regions. These matrices are actual percentage identity values. Simply to say, they depend on similarity. **Blosum 62** means there is 62 % similarity.

Gap score or gap penalty: Dynamic programming algorithms use gap penalties to maximize the biological meaning. Gap penalty is subtracted for each gap that has been introduced. There are different gap penalties such as gap open and gap extension. The gap score defines a penalty given to alignment when we have insertion or deletion. During the evolution, there may be a case where we can see continuous gaps all along the sequence, so the linear gap penalty would not be appropriate for the alignment. Thus gap open and gap extension has been introduced when there are continuous gaps (five or more). The open penalty is always applied at the start of the gap, and then the other gaps following it is given with a gap extension penalty which will be less compared to the open penalty. Typical values are -12 for gap opening, and -4 for gap extension.

Date:18/1/21	
Roll No: BID 19006	11