

This is a file to explain the Needleman-Wunsch Algorithm used for global sequence alignment implemented in this program.

How the program works:

- The program reads from the file two sequence alignments
- The algorithm supports different scoring systems that the user provides in the input files.
- The output of the program is the grid calculated to get the final score and the optimal alignment of the two sequences.

How the program was implemented (function description): • void

Parsingfile(): This function starts by checking if the file is empty or not then it parses through the input file line by line and starts reading the sequence after each space of the line according to the format of the file and then puts each sequence in the substring and after that, it reads the values of the scoring system of match, mismatch, and gap.

Sequence1: seq_1

Sequence2: seq_2

Match: match_value

Mismatch: mismatch_value

Gap: gap_value

- **void initialize()**: This function initializes an empty grid except for the first row and the first column that is initialized in the grid by the gap value.
- **void evaluategrid()**: The program after initializing the grid fills in it according to the scoring system and by comparing the current grade with the grid in the previous column or in the previous row or the both previous row and column and chooses the greatest value possible to be updated in the current position in the grid.
-
- **void alignment()**: the function starts by checking if it is a match or not and by comparing if the [current position -i] [current position -j] is greater than the [current position-i][j] and

the current position [i][j-1], if it is a match, it adds the scoring value of the match to the [current position -i][current position -j], if it is not a match it compares between [the current position-i] and [the current position-j] and based on the greater number, it adds the gap to the sequence.

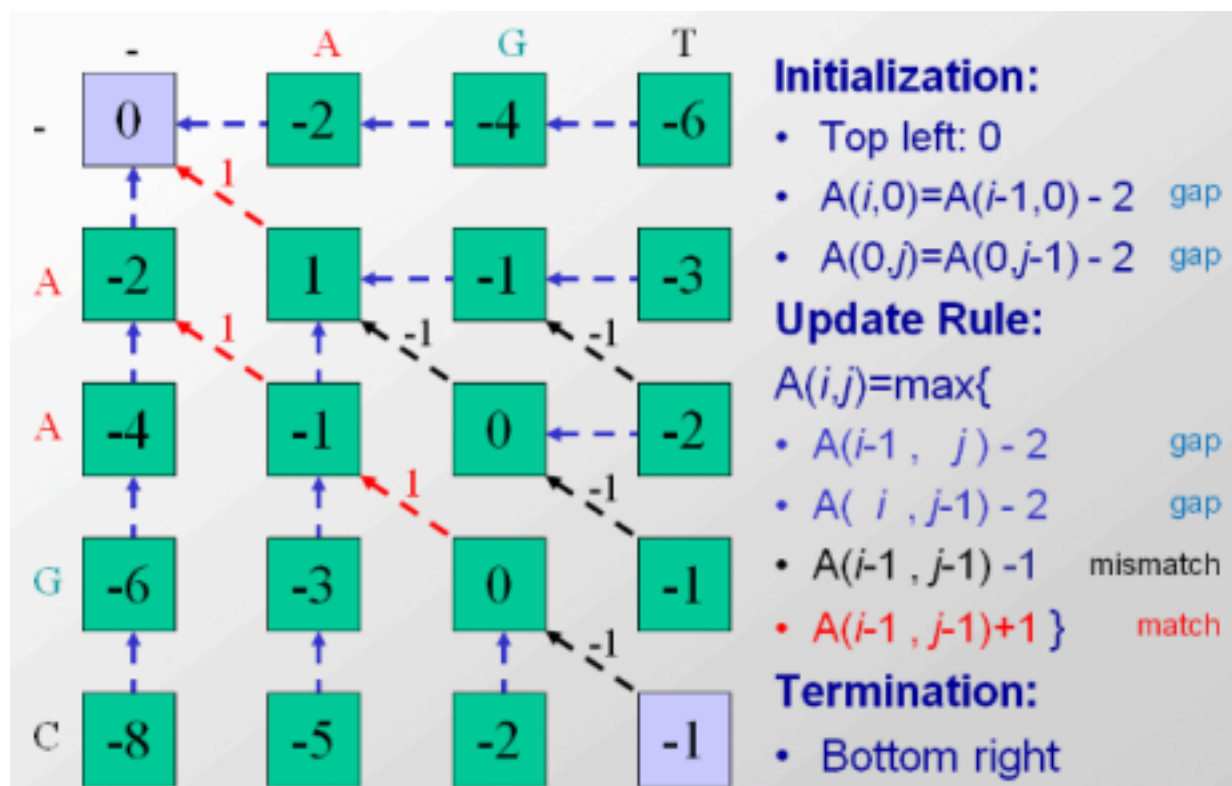
- **void reverse()**: This function is called in the alignment function to reverse the resulting sequences after the alignment because the alignment starts from the end of the grid.
- **int max()**: This function takes three parameters and returns the greatest value of them.
- In the main function, the user is required to enter the name of the file with the system and the scoring system in the file, and then the main calls the function and prints the output out.

Bonus implemented:

- Implement the ability to read sequences from a file.
- Provide a graphical representation of the alignment matrix.
- Providing different scoring matrices.

The reference test files are from the lecture slides, assignments and A

Computational Biology Genomes Networks and Evolution book by Kellis et al.

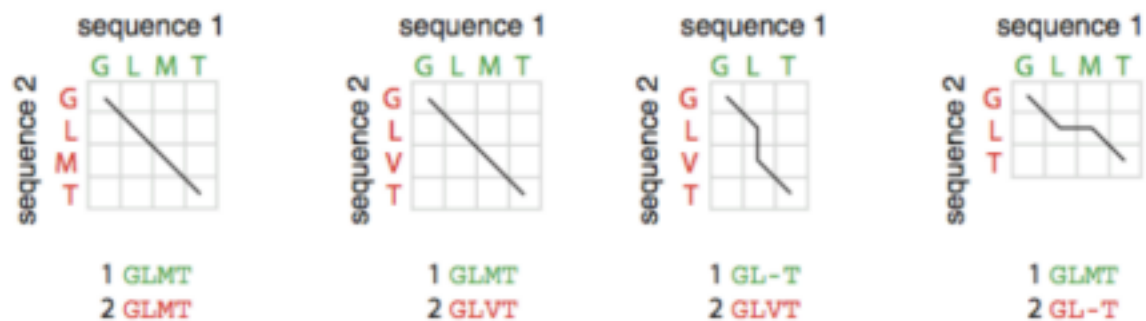


Global pairwise alignment using Needleman-Wunsch

(b)

		Sequence 2									
		F	M	D	T	P	L	N	E		
Sequence 1		0	-2	-4	-6	-8	-10	-12	-14	-16	
	F	-2	+1	-1	-3	-5	-7	-9	-11	-13	
	K	-4	-1	-1	-3	-5	-7	-9	-11	-13	
	H	-6	-3	-3	-5	-7	-9	-11	-13		
	M	-8	-5	-2	-4	-5	-7	-9	-11	-13	
	E	-10	-7	-4	-4	-6	-7	-9	-11	-10	
	D	-12	-9	-6	-3	-5	-7	-9	-11	-12	
	P	-14	-11	-8	-5	-5	-4	-6	-8	-10	
	L	-16	-13	-10	-7	-7	-6	-3	-5	-7	
	E	-18	-15	-12	-9	-9	-8	-5	-5	-4	

		+1	-1	-3	-2	-4	-3	-5	-4	-3	-5	-4
Sequence 1	F	K	H	M	E	D	-	P	L	-	E	
Sequence 2	F	-	-	M	-	D	T	P	L	N	E	



0	G -2	A -4	A -6	C -8			
C -2	-1	-3	-5	-5			
A -4	-3	0	-2	-4			
A -6	-5	-2	1	-1			
G -8	-5	-4	-1	0			
A -10	-7	-4	-3	-2			
C -12	-9	-6	-5	-2			

-

Sequence Y: C A A G A C

Sequence X: G A A _ _ C

Results:

```
Microsoft Visual Studio Debu X
Enter the file name
input1.txt
AGT
AAGC
Grid Initialization
0 -2 -4 -6
-2 0 0 0
-4 0 0 0
-6 0 0 0
-8 0 0 0

After Evaluation
0 -2 -4 -6
-2 1 -1 -3
-4 -1 0 -2
-6 -3 0 -1
-8 -5 -2 -1

The score of this alignment = -1
Sequence Alignment
Sequence 1: A_GT
Sequence 2: AAGC

C:\Users\User\Desktop\BioinformaticsBonus\vs84\Debug\BioinformaticsBonus.exe (process 15312) exited with code 0.
To automatically close the console when debugging stops, enable Tools->Options->Debugging->Automatically close the console when debugging stops.
Press any key to close this window . . .
```

```
Microsoft Visual Studio Debu X
-4 0 0 0 0 0 0 0 0
-6 0 0 0 0 0 0 0 0
-8 0 0 0 0 0 0 0 0
-10 0 0 0 0 0 0 0 0
-12 0 0 0 0 0 0 0 0
-14 0 0 0 0 0 0 0 0
-16 0 0 0 0 0 0 0 0
-18 0 0 0 0 0 0 0 0

After Evaluation
0 -2 -4 -6 -8 -10 -12 -14 -16
-2 1 -1 -3 -5 -7 -9 -11 -13
-4 -1 -1 -3 -5 -7 -9 -11 -13
-6 -3 -3 -3 -5 -7 -9 -11 -13
-8 -5 -2 -4 -5 -7 -9 -11 -13
-10 -7 -4 -4 -6 -7 -9 -11 -10
-12 -9 -6 -3 -5 -7 -9 -11 -12
-14 -11 -8 -5 -5 -4 -6 -8 -10
-16 -13 -10 -7 -7 -6 -3 -5 -7
-18 -15 -12 -9 -9 -8 -5 -5 -4

The score of this alignment = -4
Sequence Alignment
Sequence 1: F__M_DTPLNE
Sequence 2: FKHMED_PL_E
```

```
Microsoft Visual Studio Debug
Enter the file name
input2.txt
GLT
GLVT
Grid Initialization
  0  -2  -4  -6
-2  0  0  0
-4  0  0  0
-6  0  0  0
-8  0  0  0

After Evaluation
  0  -2  -4  -6
-2  1  -1  -3
-4  -1  2  0
-6  -3  0  1
-8  -5  -2  1

The score of this alignment = 1
Sequence Alignment
Sequence 1:GL_T
Sequence 2:GLVT

C:\Users\User\Desktop\BioinformaticsBonus\vs4\Debug\BioinformaticsBonus.exe (process 19832) exited with code 0.
To automatically close the console when debugging stops, enable Tools->Options->Debugging->Automatically close the console when debugging stops.
Press any key to close this window . . .
```

```
Microsoft Visual Studio Debug
Enter the file name
input1.txt
GAAAC
CAAGAC
Grid Initialization
  0  -2  -4  -6  -8
-2  0  0  0  0
-4  0  0  0  0
-6  0  0  0  0
-8  0  0  0  0
-10 0  0  0  0
-12 0  0  0  0

After Evaluation
  0  -2  -4  -6  -8
-2  -1  -3  -5  -5
-4  -3  0  -2  -4
-6  -5  -2  1  -1
-8  -5  -4  -1  0
-10 -7  -4  -3  -2
-12 -9  -6  -5  -2

The score of this alignment = -2
Sequence Alignment
Sequence 1:GAA__C
Sequence 2:CAAGAC

C:\Users\User\Desktop\BioinformaticsBonus\vs4\Debug\BioinformaticsBonus.exe (process 20468) exited with code 0.
To automatically close the console when debugging stops, enable Tools->Options->Debugging->Automatically close the console when debugging stops.
Press any key to close this window . . .
```