Project report

1 Linear gap penalty

1.1 Data testing

The algorithm of linear gap penalty is contained in linear_gap _penalty.py under the folder of codes. To testing this algorithm, I choose input1.txt, input3.txt, parameter1.txt and parameter3.txt as the testing data because their score for initiating a gap are 0, so we just need to fill in a V table. The test results are as follows:

Input1:

input1.txt/parameter1.txt			
seq1=tgacaatccc			
seq2=tgagcatggt			
	score	entries	running time/s
DP	10	121	0.0013
banded-DP	10	65	0.0013
X-drop	10	43	0.0024

Input3:

Input3.txt/parameter3.txt			
seq1=gcggcgcaagatggcccaggagaaccccaagatgcacaactcggagatcagcaagcgcctgggcgccgagtggaaacttttgtcggagacggagaagcggccgttcatcgacgaggctaa			
seq2=gcggcgcaagatggcccaagagaaccccaagatgcacaactcggagatcagcaagcgcctgggcgccgagtggaaacttttgtcggagacggagaagcggccgttcatcgacgaggccaagcggctgcga			
	score	entries	running time/s
DP	234	15851	0.3374
banded-DP	234	841	0.1931
X-drop	234	602	0.6111

1.2 Table check

The V table we've filled in are as follows:

	0	1	2	3	4	5	6	7	8	9	10
0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1	0.00000	2.00000	1.00000	0.00000	0.00000	0.00000	0.00000	2.00000	1.00000	0.00000	2.00000
2	0.00000	1.00000	4.00000	3.00000	2.00000	1.00000	0.00000	1.00000	4.00000	3.00000	2.00000
3	0.00000	0.00000	3.00000	6.00000	5.00000	4.00000	3.00000	2.00000	3.00000	3.00000	2.00000
4	0.00000	0.00000	2.00000	5.00000	5.00000	7.00000	6.00000	5.00000	4.00000	3.00000	2.00000
5	0.00000	0.00000	1.00000	4.00000	4.00000	6.00000	9.00000	8.00000	7.00000	6.00000	5.00000
6	0.00000	0.00000	0.00000	3.00000	3.00000	5.00000	8.00000	8.00000	7.00000	6.00000	5.00000
7	0.00000	2.00000	1.00000	2.00000	2.00000	4.00000	7.00000	10.00000	9.00000	8.00000	8.00000
8	0.00000	1.00000	1.00000	1.00000	1.00000	4.00000	6.00000	9.00000	9.00000	8.00000	7.00000
9	0.00000	0.00000	0.00000	0.00000	0.00000	3.00000	5.00000	8.00000	8.00000	8.00000	7.00000
10	0.00000	0.00000	0.00000	0.00000	0.00000	2.00000	4.00000	7.00000	7.00000	7.00000	7.00000

DP(input1)

	0	1	2	3	4	5	6	7	8	9	10
0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1	0.00000	2.00000	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2	0.00000	1.00000	4.00000	3.00000	2.00000	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000
3	0.00000	0.00000	3.00000	6.00000	5.00000	4.00000	3.00000	0.00000	0.00000	0.00000	0.00000
4	0.00000	0.00000	2.00000	5.00000	5.00000	7.00000	6.00000	5.00000	0.00000	0.00000	0.00000
5	0.00000	0.00000	1.00000	4.00000	4.00000	6.00000	9.00000	8.00000	7.00000	0.00000	0.00000
6	0.00000	0.00000	0.00000	3.00000	3.00000	5.00000	8.00000	8.00000	7.00000	6.00000	0.00000
7	0.00000	0.00000	0.00000	0.00000	2.00000	4.00000	7.00000	10.00000	9.00000	8.00000	8.00000
8	0.00000	0.00000	0.00000	0.00000	0.00000	4.00000	6.00000	9.00000	9.00000	8.00000	7.00000
9	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	5.00000	8.00000	8.00000	8.00000	7.00000
10	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	7.00000	7.00000	7.00000	7.00000

banded-DP(input1)

	0	1	2	3	4	5	6	7	8	9	10
0	0.00000	-1.00000	-2.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000
1	-1.00000	2.00000	1.00000	0.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000
2	-2.00000	1.00000	4.00000	3.00000	2.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000
3	-999.00000	0.00000	3.00000	6.00000	5.00000	4.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000
4	-999.00000	-999.00000	2.00000	5.00000	5.00000	7.00000	6.00000	5.00000	-999.00000	-999.00000	-999.00000
5	-999.00000	-999.00000	-999.00000	4.00000	4.00000	6.00000	9.00000	8.00000	7.00000	-999.00000	-999.00000
6	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	5.00000	8.00000	8.00000	7.00000	-999.00000	-999.00000
7	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	7.00000	10.00000	9.00000	8.00000	-999.00000
8	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	9.00000	9.00000	8.00000	-999.00000
9	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	8.00000	8.00000	8.00000	-999.00000
10	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	0.00000

X-drop(input1)

2 Affine gap penalty

2.1 Data testing

The algorithm of affine gap penalty is contained in affine_gap _penalty.py under the folder of codes. To testing this algorithm, I choose input2.txt and parameter2.txt as the testing data because their score for initiating a gap are not 0, so we need to fill in V, E and F three tables. The test results are as follows:

Input2:

Input2.txt/parameter2.txt			
seq1=AWGHEE			
seq2=AWHEAEHEA			
	score	entries	running time/s
DP	24	70	0.0011
banded-DP	24	43	0.0013
X-drop	24	21	0.0035

2.2 Table check

The V, E and F table we've filled in are as follows:

DP:

	0	1	2	3	4	5	6	7	8	9
0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1	0.00000	5.00000	0.00000	0.00000	0.00000	5.00000	0.00000	0.00000	0.00000	5.00000
2	0.00000	0.00000	20.00000	8.00000	6.00000	4.00000	2.00000	0.00000	0.00000	0.00000
3	0.00000	0.00000	8.00000	18.00000	6.00000	6.00000	2.00000	0.00000	0.00000	0.00000
4	0.00000	0.00000	6.00000	18.00000	18.00000	6.00000	6.00000	12.00000	0.00000	0.00000
5	0.00000	0.00000	4.00000	6.00000	24.00000	17.00000	12.00000	8.00000	18.00000	6.00000
6	0.00000	0.00000	2.00000	4.00000	12.00000	23.00000	23.00000	12.00000	14.00000	17.00000

DP(input2, V table)

	0	1	2	3	4	5	6	7	8	9
0	-999.00000	-12.00000	-14.00000	-16.00000	-18.00000	-20.00000	-22.00000	-24.00000	-26.00000	-28.00000
1	-999.00000	-12.00000	-7.00000	-9.00000	-11.00000	-12.00000	-7.00000	-9.00000	-11.00000	-12.00000
2	-999.00000	-12.00000	-12.00000	8.00000	6.00000	4.00000	2.00000	0.00000	-2.00000	-4.00000
3	-999.00000	-12.00000	-12.00000	-4.00000	6.00000	4.00000	2.00000	0.00000	-2.00000	-4.00000
4	-999.00000	-12.00000	-12.00000	-6.00000	6.00000	6.00000	4.00000	2.00000	0.00000	-2.00000
5	-999.00000	-12.00000	-12.00000	-8.00000	-6.00000	12.00000	10.00000	8.00000	6.00000	6.00000
6	-999.00000	-12.00000	-12.00000	-10.00000	-8.00000	0.00000	11.00000	11.00000	9.00000	7.00000

DP(input2, E table)

	0	1	2	3	4	5	6	7	8	9
0	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000
1	-12.00000	-12.00000	-12.00000	-12.00000	-12.00000	-12.00000	-12.00000	-12.00000	-12.00000	-12.00000
2	-14.00000	-7.00000	-12.00000	-12.00000	-12.00000	-7.00000	-12.00000	-12.00000	-12.00000	-7.00000
3	-16.00000	-9.00000	8.00000	-4.00000	-6.00000	-8.00000	-10.00000	-12.00000	-12.00000	-9.00000
4	-18.00000	-11.00000	6.00000	6.00000	-6.00000	-6.00000	-10.00000	-12.00000	-12.00000	-11.00000
5	-20.00000	-12.00000	4.00000	6.00000	6.00000	-6.00000	-6.00000	0.00000	-12.00000	-12.00000
6	-22.00000	-12.00000	2.00000	4.00000	12.00000	5.00000	0.00000	-2.00000	6.00000	-6.00000

DP(input2, F table)

banded-DP:

	0	1	2	3	4	5	6	7	8	9
0	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1	0.00000	5.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2	0.00000	0.00000	20.00000	8.00000	6.00000	4.00000	0.00000	0.00000	0.00000	0.00000
3	0.00000	0.00000	8.00000	18.00000	6.00000	6.00000	2.00000	0.00000	0.00000	0.00000
4	0.00000	0.00000	6.00000	18.00000	18.00000	6.00000	6.00000	12.00000	0.00000	0.00000
5	0.00000	0.00000	4.00000	6.00000	24.00000	17.00000	12.00000	8.00000	18.00000	0.00000
6	0.00000	0.00000	0.00000	4.00000	12.00000	23.00000	23.00000	12.00000	14.00000	17.00000

banded-DP(input2, V table)

	0	1	2	3	4	5	6	7	8	9
0	-999.00000	-12.00000	-14.00000	-16.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1	-999.00000	-12.00000	-7.00000	-9.00000	-11.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2	-999.00000	-12.00000	-12.00000	8.00000	6.00000	4.00000	0.00000	0.00000	0.00000	0.00000
3	-999.00000	-12.00000	-12.00000	-4.00000	6.00000	4.00000	2.00000	0.00000	0.00000	0.00000
4	0.00000	-2.00000	-4.00000	-6.00000	6.00000	6.00000	4.00000	2.00000	0.00000	0.00000
5	0.00000	0.00000	-2.00000	-4.00000	-6.00000	12.00000	10.00000	8.00000	6.00000	0.00000
6	0.00000	0.00000	0.00000	-2.00000	-4.00000	0.00000	11.00000	11.00000	9.00000	7.00000

banded-DP(input2, E table)

	0	1	2	3	4	5	6	7	8	9
0	-999.00000	-999.00000	-999.00000	-999.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1	-12.00000	-12.00000	-12.00000	-12.00000	-2.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2	-14.00000	-7.00000	-12.00000	-12.00000	-4.00000	-2.00000	0.00000	0.00000	0.00000	0.00000
3	-16.00000	-9.00000	8.00000	-4.00000	-6.00000	-4.00000	-2.00000	0.00000	0.00000	0.00000
4	0.00000	-11.00000	6.00000	6.00000	-6.00000	-6.00000	-4.00000	-2.00000	0.00000	0.00000
5	0.00000	0.00000	4.00000	6.00000	6.00000	-6.00000	-6.00000	0.00000	-2.00000	0.00000
6	0.00000	0.00000	0.00000	4.00000	12.00000	5.00000	0.00000	-2.00000	6.00000	-2.00000

banded-DP(input2, F table)

X-drop:

	0	1	2	3	4	5	6	7	8	9
0	0.00000	-12.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000
1	-12.00000	5.00000	-7.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000
2	-999.00000	-7.00000	20.00000	8.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000
3	-999.00000	-999.00000	8.00000	18.00000	6.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000
4	-999.00000	-999.00000	-999.00000	18.00000	18.00000	6.00000	-999.00000	-999.00000	-999.00000	-999.00000
5	-999.00000	-999.00000	-999.00000	6.00000	24.00000	17.00000	-999.00000	-999.00000	-999.00000	-999.00000
6	-999.00000	-999.00000	-999.00000	-999.00000	12.00000	23.00000	23.00000	11.00000	-999.00000	0.00000

X-drop(input2, V table)

	0	1	2	3	4	5	6	7	8	9
0	-999.00000	-12.00000	-14.00000	-16.00000	-18.00000	-20.00000	-22.00000	-24.00000	-26.00000	-28.00000
1	-999.00000	-24.00000	-7.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2	-999.00000	-1001.00000	-19.00000	8.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
3	-999.00000	0.00000	-2.00000	-4.00000	6.00000	0.00000	0.00000	0.00000	0.00000	0.00000
4	-999.00000	0.00000	0.00000	-2.00000	6.00000	6.00000	0.00000	0.00000	0.00000	0.00000
5	-999.00000	0.00000	0.00000	-2.00000	-4.00000	12.00000	0.00000	0.00000	0.00000	0.00000
6	-999.00000	0.00000	0.00000	0.00000	-2.00000	0.00000	11.00000	11.00000	0.00000	0.00000

X-drop(input2, E table)

	0	1	2	3	4	5	6	7	8	9
0	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000	-999.00000
1	-12.00000	-24.00000	-1001.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2	-14.00000	-7.00000	-19.00000	-2.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
3	-16.00000	0.00000	8.00000	-4.00000	-2.00000	0.00000	0.00000	0.00000	0.00000	0.00000
4	-18.00000	0.00000	0.00000	6.00000	-4.00000	-2.00000	0.00000	0.00000	0.00000	0.00000
5	-20.00000	0.00000	0.00000	6.00000	6.00000	-4.00000	0.00000	0.00000	0.00000	0.00000
6	-22.00000	0.00000	0.00000	0.00000	12.00000	5.00000	-2.00000	-2.00000	0.00000	0.00000

X-drop(input2, F table)

2.3 Other sequences' analysises

For more analysises, I download four pairs of homology protein sequences from http://eggnogdb.embl.de/#/app/downloads. The sequence file are under the folder of analytical_input. The test results are as follows:

analysis_input1.txt:

analysis_input1.txt/parameter.txt			
seq1=MGLSTKTIHVHRHQQHLKWSRDIPPVLTVNSGETVTFDALDGSNGMIARDSGDAILKFDVARADPGFGPIYVNEAEPGDVLKVEILDLQLGDWAWTAIVPNFGLLADEFPEARLKIWDIDPSLPYVKFKEGVHIRKAPFLGIMAVAPGVEGEFSTIPPLETGGNMDCRYLTAGSTLYLPIRTKGALFSCGDGHTAQGDGEVCGTAIETTLTASLRLTVVKNQPWVTSPQFQTPPLKQMLSAADIEADKGEYATMGIDCDLLEATRKATRNMIEWLTRTKDLTREEAYMLTSVAGCLKMAEVVDMPNYAIVMTIPLNIFVAPDWPNGTTT			
seq2=MGLSTKTIHVHRHQQHLKWSRDIPPVLTIDSGETVTFDALDGSNGMITRESGDEAILNFDVERADPGFGPIYVNDAEPGDVLKVDILDLQLGDWAWTAIVPGFGLLADDFLEARLKIWDIDSNLPYVKFKEGVHIRKAPFLGIMAVAPGVEGEFSTIPPLETGGNMDCRYLTTGSTLYLPIRTKGALFSCGDGHTAQGDGEVCGTAIETTLTASLRLTVCKNQPWVTSPQFQTPPLKQTLSASDIESDKGEYATMGIDSDLLEATRKATRSMIEWLTCTKDLTREEAYMLTSVAGCLKMAEVVDMPNYAIVMTMPLSIFVASDRPNGTTT			
	score	entries	running time/s
DP	2032	109230	6.1001
banded-DP	2032	2301	4.7067
X-drop	2032	991	11.005

analysis_input2.txt:

analysis_input2.txt/parameter.txt			
seq1=MNKKRFILSLLFIVFVIASLLIGVSSELTLKSLLSGNKEAYNVLFRSRMPRTLAIILSASSLSIAGLIMQSLSRNKFISPQTAGTNDAAAFGVLISFIFLGSIPSYYRLIFSFVFALISSLLFIFILNKIKFKNIIYVPLIGLMYGGLLSAVTRLIAYETNTLQLLSSLNLGTFSHINVLNGSMLLILVPVLLIAIIFAEGFNVASLGEEFATNLGVNYKKVLFLGLIVTSLIAASTYLVVGPLPFLGLIIPNIVSMYYGDNLKKNIIDIALLGSIFVLANDIISRLVIFPFELSVGFTMGITGAIIFILLIFKQVKNND			
seq2=MNKKRFLIIALFFILLIISLLIGVSNSVTIGALLKGDKNAWFIFLESRVPRTLVIVLTASSLSIAGLIMQAISRNKFISPSTAGTTNAAVLGVLIGFLFFGTQSLYIRFIFAFVFALLSSILFITVLNKIKVKNVIYVPLIGMMYGTLISSITTLIAQRFNALQTVSMLNLGSFSHITRLNSTLLIVLIPSLVLAVIFASQFNIVSLGEDFSKNLGVNYKRIMFIGLIVISVISASTYIVVGPLPFLGLIIPNIVTVYYGDNLKRNLFDVAIFGATFVLLNDIISRLLIFPYEISVGFTMGISGAIIFLILIFRKVKQK			
	score	entries	running time/s
DP	1389	102720	5.3258
banded-DP	1389	2231	4.0357
X-drop	410	329	5.8234

Comparing the two table above we can find that the banded-DP is the fast, because it just need to calculate the entries within the bandwith, the DP is the second one and the X-drop is the slowest one because if the max value is the last entry of the V table, this algorithm not just need to calculate the whole table, but also need to judge if there is a consistent entry. But for the number of entries, the X-drop always be the fewest, because the number of consistent entry is few, and for DP and banded-DP, both of them need to fill in a specified number of entries.

But we can also find that in the second result table, the X-drop get a small score, is that wrong? No, because the X-drop is a extentional algorithm, and the extention depends on the threshold value X, once it judge there is no more consistent entry then it will stop, so the choose of threshold's value X is very important.

analysis_input3.txt:

<table>
<tr><td colspan="4">analysis_input3.txt/parameter.txt</td></tr>
<tr><td colspan="4">seq1=MAGSHPVAENPPAVQVASPFGFVRALEWAAAERLPQMATYWEIILYIQENLAMANGAVRPLEALKNWGDDFDPFAR
RAYSWLLVHHESNSSESAIAGISHTRPRLRQQIMDSLASHERINRNTNVKGAVDQTFPDSIKSEVDFWSQEAITKSINNRTPAA
SAIADIPQTRQQQAVYIREMFDALKNTQDIIEKASNHKVGTVKSTSGRVFQDIAWVLYREARNLQQGKPGVKPWCTSFKYK
QYPTMRARWNDMVEFFQTSKAAVANLIVAHSEKRFAGNPTKEKQRKTVNDKGNKKKATKMKDNEEKAARLDKVVKDA
KGKAAAAKSAPDQGTNEEWDAEASEDLGGDGEAEETDYEAEEFDDSAREREDNDHESDDEATHLLQNAQQDVDRYDDIP
EDQDSDDLKFQFGRAMGDSAASTVPLAGTSQPHAPAYPPLPGSISHRRQVRAIFPDEGHPPSIMPSNDHESYHPCYSGGLVSP
NVQGQCIDDARSNDLGSSPLSGQPMDYLEHTLHRYISPSSPELTSGIGFMGQAQPQEGRGSDTAVGTSRIAHMPLNTGTSGV
GGGAQNNSDPQDSTNSSRSRARRNRHSSEDGSKTRAEAPPRRRPRI</td></tr>
<tr><td colspan="4">seq2=MAGNHPIAGSPPVVQVASPFGFVRALECAAAEMLPQMSTYWEIILYIQENLAMANGAVRPLEALKNWGDDFDPFARR
AYSWLLVHHESNSSESAIAGISHTRPRLRQQIMDSLASHERINRNTHVKGAVDQTFPDSIKSEVDFWSQEAITKSINNRTPAAS
AIADIPRTRQEQAVYIREMFDALKNTQDIIEKASNHKVGTVKSTSGRVLQDLAWVLYREARNLQQGKPGVNPWCTSFKYKQ
YPTLRARWNDMVEFFHTSKAAVANLIVTHSEKRFAGNPTKEKQRKTVNDKCNKKKATRMKDNEEKAATLDKVVKDAKG
KVATAESTPNKGIDEERDAEDSVDLDDEAEEFDDSEGEPKDTDNESDDEATHLLQNAQQDVGRYDDIKEEQESDDLNPQSG
QAMGDSAASTTAFAASSHPHAPAYPPSPGLIGRRRQVRAISPDEGHHSSIMPSIEHESYGPFHAGGVVSQNDHDRYVDDANG
NDRGSLQLPSQPVNYPEQPLHQHTSPSPPELPSGIGFVKEEEIPQSVVTTSHIAPMPLNTGTSGVGGGAQDNTDLRDSTNSSRS
RARRNRHSSEDSSDSHAQARARRRPRI</td></tr>
<tr><td></td><td>score</td><td>entries</td><td>running time/s</td></tr>
<tr><td>DP</td><td>3117</td><td>362355</td><td>72.9901</td></tr>
<tr><td>banded-DP</td><td>2115</td><td>4159</td><td>73.5448</td></tr>
<tr><td>X-drop</td><td>2102</td><td>1141</td><td>97.8880</td></tr>
</table>

analysis_input4.txt:

analysis_input4.txt/parameter.txt

seq1=MLKFIKNNKWWVAIISVFAVFLSSFGIFAKSFVDSNKQKVVNKVENYIKASSYAVQSRILKETENLNEDYLNQKIGKKT
LLNEFNNNFIWQPNNNKTTSLDTISDLWKTYFGSTDNVLRKDLQIQYQNNNKLENIKSSKGEITPKNIDFLLTISKSLEKFLNG
FAPSLASLGISFLQSTVLQHRDDPNFEKYKSGITSVANAIEDNRETFNYLGKVLTPANLDENYYKNLTVKQAITKNINKLASVI
TNNKDFASQTDVDKLPEALDKLLVDLELDSITNIFSDIDFKNFKAEKISEIFPKIKNLFNNGTFTKLKQKSLEILNKITPHLATY
LYSELFFGLYYVANSDLKNPEDLLKQKVDDSVFIAFTKNKLDLNVLVDGLVKVLKNKKDFERLYDFIFKRFDQNKIFNNINT
LGTDTGTDSLLYDLINWLEKKLYNVSNTLSTIIRFAELALNDANIQKTIKEKLVEFIKSKLPKISVGSWIVNFNKDELEVSLKIY
LGFIPLRTPLQLKADFFGKTGLINNLINILKSFNSSINYAFEEFFKYIKNTLYLNTNKKFTLSPFVELVNNLNAFLKNDKNVYIS
LAKGLIGDLEIKTIYDFITLPYNKEFLNSLVQKYAGKDLQPVLDKIKAFLESLKTYGFIKEPEKLKEQFPQYLENLSKHLATYE
NNDLLNFNLLDSLYEGNIISDFISKWINFLTKDITKEDNPILPILRAINKSDKLEKLEQIKNQWTSKVSDLAKKIDNYSNISKIRN
IKLQLPKELVEQFGLQSLDGLNITELLEGLSNYIKDYLKANPNKVIGFNISSIGMMLYALTVKVGVEFKKELSKNNFLYNKNT
KQDKSKTVLKALADGFDSHDNSSDVGRDSTLNRKDQSYYNWDKIYFYINGFDKPYILDRTNLKEEFSYSPLHMLIGINPDKT
TYFKGSIGYAIGSLFGGLNTTDPNYNLSIENKNDATGILNVFNYVLDQKDKELKKHEDQIATQYYDKDAWETKVINSSEDEI
NYELIRLKSSKTQESKQLGSRFKVKLAKKKHSSYWEITQIIAVDYKAA

seq2=MLKFIKNSKWWVIIISIFAIFFSSFGIFAKSYVDSNKQKVVNKIQNYVQASSYAVQSRILKETENLNEDYLNEKIGKKSLL
NEFNSDFIWRPNNSKTTSSDTISDLWKTYFGSSNNVLNKNLQIQYKNNNEYKNIESSKGEITPKNIDFLLSISKSAEKFLNGFAP
SLASLGLSFLQNKVLNNRDDSKLQSYKDGINKFADVIENNKDLFSYLGKLFTPEPLEKDYYKDLTVQQALVKNINQIAAVIA
DNKEFAKETEVDKIPEALDKILAELGLDSLGEIIGELISSKNGFQNLGKIFSKIKNIFTSQNFQKLKTKGLELINKITPYLVTYLY
SEIFFGLYYVTHEEFKNPSDLTKQKVNVSDFSALVNNKIDLNVLINGFLKVLKDKKSFDRFYNFIFKRFNENEIFNSKNNIGSN
NGVGNLIFDLINWVEKKLHNFSNLLETLIKFVDFAMSDEKIKKTFEENIKNFITKKLNELGTPLGKWHVEIKNGILNISASSLW
LGSLTAKIEVFSKDGIIPKIIDVLKKIHGIINDSSETIIKHIKEIFYLAKDNILDFSITSKNISEIIMTFKELLINKKILNVNVKALFIK
LLDISSIYDILELPYSNSALKWILERFAKDKIEPHLKKIKSISETLQKNKFIQNADKIKEQLPKYLDLFSSHFKKYEKSSDLKFNL
NISLYKGNIIEDLILKWINFLLQDAENKENPLLPIIRLITKKNSLKTLDDIKNEWVSKITNISKKIESFQNISKIRDKKIDIPKELLK
YFGLESINNQTILRLLEILGKYFDDYLSKNPNKVVGVNISSLGKVLTALTIKVGVEYRPENKDKNFLYSKDIKENKTKTILKAL
VYGFDTHDNSSDVGSDSIKNRKPESYYNWDKIHFYINGSSQAITLDRTKLKDDSSYSPLHMLLGINPDKSSYLKDSLGYVFGT
LFGGLPASDSNYQLSIENKTDVTSILNVFNYVLDKKDKELKKQEEQIATKYYDKNAWSTKILSSNENEIIYQLIRLKSSDTKES
KRLGTRFEVRLLKNKNNPYWSINRVIALGYKAV

	score	entries	running time/s
DP	3954	1103550	767.9343
banded-DP	3954	7341	732.0790
X-drop	1386	862	903.6858

Comparing the last two table above we can find that with the extention of the sequence, all of the algorithm becoming very slow, but there is also reasonable, because the matrix is become bigger and bigger, which leads to the algorithm's running time increasing in an exponential way.

Another interesting thing we can find is that the score of banded-DP also become smaller like the X-drop we've talked above, is that wrong? Of course not. As we all know, the standard Smith Waterman dynamic programing need to fill in the whole table, so whatever the dimension of the matrix is, it can always get the max entry. But for banded-DP algorithm, it just need to calculate the entries within the bandwith section, so what if the two sequence are not the same length and the max entry is the last one of the matrix, it means the max entry is not in the section of bandwith, so it can't find the correct max entry. Here is an example below:

	587	588	589	590	591	592	593	594
589	.00000	3005.00000	3003.00000	3001.00000	2999.00000	2997.00000	2995.00000	2993.00000
590	.00000	3012.00000	3010.00000	3008.00000	3006.00000	3004.00000	3002.00000	3000.00000
591	.00000	3020.00000	3018.00000	3016.00000	3014.00000	3012.00000	3010.00000	3008.00000
592	.00000	3030.00000	3028.00000	3026.00000	3024.00000	3022.00000	3020.00000	3018.00000
593	.00000	3032.00000	3030.00000	3028.00000	3026.00000	3024.00000	3022.00000	3020.00000
594	.00000	3039.00000	3037.00000	3035.00000	3033.00000	3031.00000	3029.00000	3027.00000
595	.00000	3040.00000	3042.00000	3040.00000	3038.00000	3032.00000	3034.00000	3028.00000
596	.00000	3046.00000	3042.00000	3041.00000	3039.00000	3037.00000	3034.00000	3033.00000
597	.00000	3046.00000	3053.00000	3049.00000	3048.00000	3038.00000	3044.00000	3034.00000
598	.00000	3061.00000	3051.00000	3051.00000	3047.00000	3047.00000	3043.00000	3043.00000
599	.00000	3057.00000	3061.00000	3053.00000	3051.00000	3049.00000	3047.00000	3045.00000
600	.00000	3064.00000	3062.00000	3060.00000	3058.00000	3056.00000	3054.00000	3052.00000
601	.00000	3065.00000	3061.00000	3059.00000	3057.00000	3068.00000	3056.00000	3054.00000
602	.00000	3074.00000	3062.00000	3060.00000	3058.00000	3067.00000	3065.00000	3053.00000
603	00000	3062.00000	3081.00000	3069.00000	3067.00000	3065.00000	3074.00000	3062.00000
604	.00000	3069.00000	3069.00000	3088.00000	3076.00000	3074.00000	3072.00000	3070.00000
605	.00000	3067.00000	3076.00000	3076.00000	3095.00000	3083.00000	3081.00000	3079.00000
606	.00000	3066.00000	3065.00000	3074.00000	3083.00000	3105.00000	3093.00000	3091.00000
607	.00000	3054.00000	3073.00000	3072.00000	3081.00000	3093.00000	3112.00000	3100.00000
608	.00000	3062.00000	3061.00000	3070.00000	3079.00000	3091.00000	3100.00000	3117.00000

The real max entry is the last one of this matrix

	587	588	589	590	591	592	593	594
589	4.00000	2057.00000	2055.00000	2064.00000	2067.00000	2077.00000	0.00000	0.00000
590	2.00000	2055.00000	2056.00000	2062.00000	2064.00000	2066.00000	2076.00000	0.00000
591	000	2051.00000	2055.00000	2060.00000	2062.00000	2063.00000	2066.00000	2072.00000
592	000	0.00000	2049.00000	2058.00000	2060.00000	2061.00000	2062.00000	2062.00000
593	000	0.00000	0.00000	2056.00000	2058.00000	2059.00000	2060.00000	2058.00000
594	000	0.00000	0.00000	0.00000	2056.00000	2057.00000	2058.00000	2057.00000
595	000	0.00000	0.00000	0.00000	0.00000	2055.00000	2060.00000	2055.00000
596	000	0.00000	0.00000	0.00000	0.00000	0.00000	2054.00000	2059.00000
597	000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	2050.00000
598	000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
599	000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
600	000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
601	000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
602	000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
603	000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
604	000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
605	000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
606	000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
607	000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
608	000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

The bandwith section dosen't contains the real max entry

So we can see, the choose of the bandwith value is important for banded-DP algorithm, if the two sequence are not the same length, you might need to set a bigger value of bandwith to make sure you will not miss the max entry.