

ASSIGNMENT

NAMA : Erio Yoshino

NIM : 2501994986

SOAL 1.

Gene 1:

GGAGTTAGATCCAACCCCAGAGCGGCTTTTGTGGGCACTGATTGCTCCAGCTTCTGCGTC
CTGCGTGAGGGAAGAGGATCCAGGCGTTAGACATGTATAGACACAAAAACAGCTGGAGAT
TG
GGCCTTAAATACCCACCAAGCTCCAAAGAAGAGACCCAAGTCCCCAAAACATTGATTTGAG
GGCTGCCAGCAGGGTGGGAGAGAAGCTCCAGTCAGCCCACAAGATGCCCTTGCCCCCGG
GCCTCCTGCTGCTGCTGCTCTCCGGGGCCACGGCCACCGCTGCCCTGCCCTGGAGGGC
GGCCCCACCGGCCAGACAGCGAGCATATGCAGGAAGCGGCAGGAATAAGGAAAAGCAG
CCTCCGTGACTTTCTCCGCTTGGTGGTTTGAGTGGACCTCCAGGTCCAGTGCTGGGCC
CCTTCATAGGGAGAGGAAGCCCCGGGAGGTGGCCAGGCGGCAGGAAGGCGCACCCCCC
CGGCAATTCCGCGCGCCGGGACAGAATGCCCTGCAGGAACCTTTCTTCTGGGAAGACCTTT
CTCCTCCTGCCATAAAATAAACTTTACCCCTAATGATTCACTCATGCAACGTGTAATGACAGA
CCTGAATAAAATGTATTAAGCAGCAGGTGATCTTCTCTCTCCTTCCCAAGTCATTTGAAA
AGTGTTTGTATTTACAGTTCTAGTAATGCCCAATACTGACGGTGTGTCTTGAGTAATTTGGA
ACCCAAAATGAAGATCTTTGATAAAGATTTTTTGTGTTTGTGTTTACTGGCCTGTGCTGAG
TGCGGGTCACTGGGCTTTTGCTTCTGATGTTTATTATGGTGCTGGGAAAGCTCTGTCTTTGA
TTTAAACTAAAGATAGCTAAAGAGCTACATAATTAAGATGTTTCAAGTAACATCTTATTTAGT
TTATGAATTGATATGAATTGTCTAATTTAAAAGATATTTCCCTCACATTAAAAGCAAATTTTAA
CATCA

Gene 2:

ATTAATTAATTCTGAGAGCTGCTGAGTTGTGTTTACTGAGAGATTGTGTATCTGCGAGAGAAG
TCTGTAGCAAGTAGCTAGACTGTGCTTGACCTAGGAACATATACAGTAGATTGCTAAATGTC
TCACTTGGGGAATTTTAGACTAAACAGTAGAGCATGTATAAAATACTCTAGTCAAGTGCTGC
TTTTGAAACAAATGATAAAACCACACTCCCATAGATGAGTGTCATGATTTTCATGGAGGAAGT
TAATATTCATCCTCTAAGTATACCCAGACTAGGGCCATTCTGATATAAAACATTAGGACTTAAG
AAAGATTAATAGACTGGAGTAAAGGAAATGGACCTCTGTCTCTCTCGCTGTCTCTTTTTTGTG
GACTTTGTGGTCAGTGGGGCTGGAATAAAAGTAGAATAGACCTGCACCTGCTGTGGCATCC
ATTCACAGAGTAGAAGCAAGCTCACAATAGTGAAGATGTCAGTAAGCTTGAATAGTTTTTCAG
GAACTTTGAATGCTGATTTAGATTTGAACTGAGGCTCTGACCATAACCAAATTTGCACTATT
TATTGCTTCTTGAACTTATTTGCCTGGTATGCCTGGGCTTTTGATGGTCTTAGTATAGCTTGC
AGCCTTGTCCCTGCAGGGTATTATGGGTAATAGAAAGAAAAGTCTGCGTTACACTCTAGTCA
TACTAAGTAACTACCATTGGAAAAGCAACCCCTGCCTTGAAGCCAGGATGATGGTATCTGCA
GCAGTTGCCAACACAAGAGAAGGATCCATAGTTCATCATTTAAAAAAGAAAACAAAATAGAAA
AAGGAAAACATTTCTGAGCATAAGAAGTTGTAGGGTAAGTCTTTAAGAAGGTGACAATTTCT
GCCAATCAGGATTTCAAAGCTCTTGCTTTGACAATTTTGGTCTTTCAGAATACTATAAATATAA
CCTATATTATAATTTATAAAGTCTGTGCATTTTCTTTGACCCAGGATATTTGCAAAAGACATAT
TCAAACCTCCGCAGAACACTTTATTTACATATACATGCCTCTTATATCAGGGATGTGAAACAG

GGTCTTGAAAACGTGTCTAAATCTAAAACAATGCTAATGCAGGTTTAAATTTAATAAAAATAAAATC
CAAAATCTAACAGCCAAGTCAAATCTGTATGTTTTAACATTTAAAATATTTTAAAGACGTCTTTT
CCCAGGATTCAACATGTGAAATCTTTTCTCAGGGATACACGTGTGCCTAGATCCTCATTGCTT
TAGTTTTTTTACAGAGGAATGAATATAAAAAGAAAATACTTAAATTTTATCCCTCTTACCTCTATAA
TCATACATAGGCATAATTTTTTAACTAGGCTCCAGATAGCCATAGAAGAACCAAACACTTTTCT
GCGTGTGTGAGAATAATCAGAGTGAGATTTTTTTCACAAGTACCTGATGAGGGTTGAGACAGG
TAGAAAAAGTGAGAGATCTCTATTTATTTAGCAATAATAGAGAAAGCATTTAAGAGAATAAAGC
AATGGAAATAAGAAATTTGTAAATTTCTTCTGATAACTAGAAATAGAGGATCCAGTTTCTTTT
GGTTAACCTAAATTTTATTTTCAATTTTATTGTTTTATTTTATTTTATTTTATTTTGTGTAATCGTAGT
TTCAGAGTGTTAGAGCTGAAAGGAAGAAGTAGGAGAAACATGCAAAGTAAAAGTATAACACT
TTCCTTACTAAACCGACATGGGTTTTCCAGGTAGGGGCAGGATTCAGGATGACTGACAGGGC
CCTTAGGGAACACTGAGACCCTACGCTGACCTCATAAATGCTTGCTACCTTTGCTGTTTTAAT
TACATCTTTTAAATAGCAGGAAGCAGAACTCTGCACTTCAAAAGTTTTTCTCACCTGAGGAGT
TAATTTAGTACAAGGGGAAAAAGTACAGGGGGATGGGAGAAAGGCGATCACGTTGGGAAGC
TATAGAGAAAGAAGAGTAAATTTTAGTAAAGGAGGTTTAAACAAACAAAATATAAAGAGAAATA
GGAACCTGAATCAAGGAAATGATTTTAAACGCGAGTATTCTTAGTGGACTAGAGGAAAAAAT
AATCTGAGCCAAGTAGAAGACCTTTTCCCCTCCTACCCCTACTTTCTAAGTCACAGAGGCTT
TTTGTTCCCCCAGACACTCTTGCGAGATTAGTCCAGGCAGAAACAGTTAGATGTCCCCAGTTA
ACCTCCTATTTGACACCACTGATTACCCCATTTGATAGTCACACTTTGGGTTGTAAGTGACTTT
TTATTTATTTGTATTTTTGACTGCATTAAGAGGTCTCTAGTTTTTTACCTCTTGTTTCCCAAAC
CTAATAAGTAACTAATGCACAGAGCACATTGATTTGTATTTATTCTATTTTTAGACATAATTTATT
AGCATGCATGAGCAAATTAAGAAAAACAACAACAAATGAATGCATATATATGTATATGTATGTG
TGTATATATACATATATATATATATTTTTCTTACCAGAAGGTTTTAATCCAAATCAGGAGAAGATAT
GCTTAGAACTGAGGTAGAGTTTTTCATCCATTCTGTCCTGTAAGTATTTTGCATATTCTGGAGA
CGCAGGAAGAGATCCATCTACATATCCCAAAGCTGAATTATGGTAGACAAAACCTTCCACTT
TTAGTGCATCAATTTCTTATTTGTGTAATAAGAAAATTGGGAAAACGATCTTCAATATGCTTACC
AAGCTGTGATTCCAAATATTACGTAAATACACTTGCAAAGGAGGATGTTTTTAGTAGCAATTTG
TACTGATGGTATGGGGCCAAGAGATATATCTTAGAGGGAGGGCTGAGGGTTTGAAGTCCAA
CTCCTAAGCCAGTGCCAGAAGAGCCAAGGACAGGTACGGCTGTCATCACTTAGACCTCACC
CTGTGGAGCCACACCCTAGGGTTGGCCAATCTACTCCCAGGAGCAGGGAGGGCAGGAGC
CAGGGCTGGGCATAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGACACAAC
TGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCACCTGACTCCTGAGGAGAAGTCTG
CCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCA
GGTTGGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAAACTGGGCATGTGGAGA
CAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCTCTGCCTATTGGTCTATTTTCCC
ACCCTTAGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTTTGGGGATC
TGTCCTACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAGTGCT
CGGTGCCTTTAGTGATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACT
GAGTGAGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGGTGAGTCTATG
GGACCCTTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTTCATGTCATAGGAAGGGGAT
AAGTAACAGGGTACAGTTTAGAATGGGAAACAGACGAATGATTGCATCAGTGTGGAAGTCTC
AGGATCGTTTTAGTTTCTTTTATTTGCTGTTTATAACAATTGTTTTCTTTTGTTTAATTCCTGCT
TTCTTTTTTTTTCTTCTCCGCAATTTTTACTATTATACTTAATGCCTTAACATTGTGTATAACAAA
AGGAAATATCTCTGAGATACATTAAGTAACTTAAAAAAAACCTTACACAGTCTGCCTAGTACA

TTACTATTTGGAATATATGTGTGCTTATTTGCATATTCATAATCTCCCTACTTTATTTTCTTTTATT
TTTAATTGATACATAATCATTATACATATTTATGGGTAAAGTGTAATGTTTTAATATGTGTACACA
TATTGACCAAATCAGGGTAATTTTGCATTTGTAATTTTAAAAAATGCTTTCTTCTTTTAATATACT
TTTTTGTTTATCTTATTTCTAATACTTTCCCTAATCTCTTTCTTTCAGGGCAATAATGATACAATG
TATCATGCCTCTTTGCACCATTCTAAAGAATAACAGTGATAATTTCTGGGTAAAGGCAATAGCA
ATATTTCTGCATATAAATATTTCTGCATATAAATTGTAAGTATGTAAGAGGTTTCATATTGCTAA
TAGCAGCTACAATCCAGCTACCATTCTGCTTTTATTTTATGGTTGGGATAAGGCTGGATTATTC
TGAGTCCAAGCTAGGCCCTTTTGTAAATCATGTTTCATACCTCTTATCTTCCTCCCACAGCTCC
TGGGCAACGTGCTGGTCTGTGTGCTGGCCCATCACTTTGGCAAAGAATTCACCCCACCAGT
GCAGGCTGCCTATCAGAAAGTGGTGGCTGGTGTGGCTAATGCCCTGGCCCACAAGTATCAC
TAAGCTCGCTTTCTTGCTGTCCAATTTCTATTAAAGGTTCTTTGTTCCCTAAGTCCAATACTACT
AAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATT
TTCATTGCAATGATGTATTTAAATTATTTCTGAATATTTTACTAAAAAGGGAATGTGGGAGGTCA
GTGCATTTAAACATAAAGAAATGAAGAGCTAGTTCAAACCTTGGGAAAATACACTATATCTTA
AACTCCATGAAAGAAGGTGAGGCTGCAAACAGCTAATGCACATTGGCAACAGCCCCTGATG
CCTATGCCTTATTCATCCCTCAGAAAAGGATTCAAGTAGAGGCTTGATTTGGAGGTTAAAGTT
TTGCTATGCTGTATTTTACATTACTTATTGTTTTAGCTGTCCTCATGAATGTCTTTTCACTACCC
ATTTGCTTATCCTGCATCTCTCAGCCTTGACTCCACTCAGTTCTCTTGCTTAGAGATACCACC
TTTCCCCTGAAGTGTTCTTCCATGTTTTACGGCGAGATGGTTTCTCCTCGCCTGGCCACTC
AGCCTTAGTTGTCTCTGTTGTCTTATAGAGGTCTACTTGAAGAAGGAAAAACAGGGGTCATG
GTTTGACTGTCCTGTGAGC

Prosedur:

1. a. Gen apakah yang dikode oleh kedua urutan di atas?

Jawab:

Gen 1 =Pongo abelii

Gen 2 = Homo sapiens

-
- b. Organisme apa saja yang terdapat dalam gen tersebut?

Jawab:

Gen 1 =Pongo abelii cortistatin (CORT), mRNA

Gen 2 = Homo sapiens isolate HbA-Dgn58 beta globin (HBB) gene, complete cds

-
-
- c. Berapa panjang gen tersebut?

Jawab:

Gen1: 990

Gen2: 5139

2. Perhatikan hasil BLAST yang anda peroleh. Pada hasil tersebut terlihat beberapa parameter hasil BLAST seperti di bawah ini. Berikan 3 contoh teratas dari deskripsi nilai dari parameter parameter berikut? dan jelaskan.

GEN1:

- a. Scientific name: *Pongo abelii*
 - a. Description: *Pongo abelii* cortistatin (CORT), mRNA
 - b. Max Score: 1823
 - c. Total Score: 1823
 - d. Query Score: 100%
 - e. E value: 0
 - f. Ident: 99.90%
 - g. Accession: 991

Penjelasan:

Dari gen 1 yang diberikan, *Pongo abelii* memiliki kemiripan paling baik, dengan skor alignment terbesar 1823, dan total score 1823. Query score 100% menandakan bahwa gen dari *Pongo abelii* memiliki kesamaan 100% dengan gen 1. E value = 0 menandakan bahwa alignment merupakan skor terbaik dari gen *Pongo abelii*, ident = 99.90% menandakan bahwa gen *Pongo abelii* memiliki 99.90% residu yang sama dengan gen. *Pongo abelii* memiliki identifier 991 pada database NCBI.

- b. Scientific name: *Pan troglodytes*
 - a. Description: *Pan troglodytes* cortistatin (CORT), mRNA
 - b. Max Score: 1362
 - c. Total Score: 1362
 - d. Query Score: 100%
 - e. E value: 0
 - f. Ident: 91.91%
 - g. Accession: 1232

Penjelasan:

Dari gen 1 yang diberikan, *Pan troglodytes* memiliki kemiripan cukup baik, dengan skor alignment terbesar 1362, dan total score 1362. Query score 100% menandakan bahwa gen dari *Pan troglodytes* memiliki kesamaan 100% dengan gen 1. E value = 0 menandakan bahwa alignment merupakan skor terbaik dari gen *Pan troglodytes*, ident = 91.91% menandakan bahwa gen *Pan troglodytes* memiliki 91.91% residu yang sama dengan gen. *Pongo abelii* memiliki identifier 1232 pada database NCBI.

- c. Scientific name: *Pongo pygmaeus*
 - a. Description: PREDICTED: *Pongo pygmaeus* cortistatin (CORT), mRNA
 - b. Max Score: 1338
 - c. Total Score: 1338
 - d. Query Score: 88%
 - e. E value: 0
 - f. Ident: 94.97%
 - g. Accession: 904

Penjelasan:

Dari gen 1 yang diberikan, Pan troglodytes memiliki kemiripan cukup baik, dengan skor alignment terbesar 1338, dan total score 1338. Query score 88% menandakan bahwa gen dari Pongo pygmaeus memiliki kesamaan 88% dengan gen 1. E value = 0 menandakan bahwa alignment merupakan skor terbaik dari gen Pongo pygmaeus, ident = 94.97% menandakan bahwa gen Pan troglodytes memiliki 94.97% residu yang sama dengan gen. Pongo pygmaeus memiliki identifier 904 pada database NCBI.

GEN2:

- a. Scientific name: Homo sapiens
 - a. Description: Homo sapiens isolate HbA-Dgn58 beta globin (HBB) gene, complete cds
 - b. Max Score: 9491
 - c. Total Score: 9491
 - d. Query Score: 100%
 - e. E value: 0
 - f. Ident: 100%
 - g. Accession: 5139

Penjelasan:

Dari gen 2 yang diberikan, Homo sapiens memiliki kemiripan paling baik, dengan skor alignment terbesar 9491, dan total score 9491. Query score 100% menandakan bahwa gen dari Homo sapiens memiliki kesamaan 100% dengan gen 2. E value = 0 menandakan bahwa alignment merupakan skor terbaik dari gen Homo sapiens, ident = 100% menandakan bahwa gen Pan troglodytes memiliki 100% residu yang sama dengan gen. Pongo pygmaeus memiliki identifier 5139 pada database NCBI.

- b. Scientific name: Homo sapiens
 - a. Description: Homo sapiens isolate HbC-Dgn99 beta globin (HBB) gene, complete cds
 - b. Max Score: 9474
 - c. Total Score: 9474
 - d. Query Score: 100%
 - e. E value: 0
 - f. Ident: 99.94%
 - g. Accession: 5139

Penjelasan:

Gen dengan kemiripan baik kedua merupakan gen yang sama dengan yang pertama, yaitu homo sapiens. Namun gen kedua memiliki Homo sapiens isolate HbC-Dgn99 beta globin (HBB) gene, complete cds. Sementara gen pertama memiliki Homo sapiens isolate HbA-Dgn58 beta globin (HBB) gene, complete cds.

- c. Scientific name: Homo sapiens
 - a. Description: Homo sapiens isolate HbA-Cmn087 beta globin (HBB) gene, complete cds

- b. Max Score: 9474
- c. Total Score: 9474
- d. Query Score: 100%
- e. E value: 0
- f. Ident: 99.94%
- g. Accession: 5139

Penjelasan:

Gen ketiga juga sama dengan gen pertama & kedua, namun memiliki organisme Homo sapiens isolate HbA-Cmn087 beta globin (HBB) gene, complete cds

SOAL 2

Hitunglah score sequence alignment dari set DNA berikut menggunakan algoritma Dynamic Programming

Query : G A C T C A T T A

Subject : G T G T C A C T T C A

Algoritma1:

diff = 2 if match, -1, if mismatch

value (x, y) = max {

Value (x-1, y-1)+ diff,

Value (x, y-1) + diff,

Value (x-1, y)+ diff,

}

A	-22	-5	1	4	7	10	13	12	11	13
C	-20	-4	-1	5	8	11	10	9	11	11
T	-18	-3	0	3	9	8	7	10	12	11
T	-16	-2	1	4	7	6	5	8	10	9
C	-14	-1	2	5	4	6	6	6	5	6
A	-12	0	3	3	3	4	7	6	5	7
C	-10	1	1	4	3	5	4	3	4	4
T	-8	2	2	1	3	2	1	3	5	4
G	-6	3	2	1	1	1	0	1	3	3
T	-4	1	1	0	2	1	0	2	4	3
G	-2	2	1	0	-1	-2	-3	-4	-5	-6
gap	0	-2	-4	-6	-8	-10	-12	-14	-16	-18
	gap	G	A	C	T	C	A	T	T	A

G A C T C A T T \bar{T} \bar{C} A
 G T G T C A C T \bar{T} \bar{C} A

Mismatch = 3

Gap = 2

Match = 6

Value = $6(2) + 3(-1) + 2(-2) = 5$

Menurut saya terdapat algoritma yang lebih baik untuk kasus ini, yaitu Needleman-Wunsch, dimana untuk mendapat nilai (x, y), nilai dari (x, y-1) dan (x-1, y) akan dikenakan gap penalty (-2).

Algoritma2: Needleman-Wunsch

diff = 2 if match, -1, if mismatch

gap_penalty = -2

value (x, y) = max {

Value (x-1, y-1)+ diff,

Value (x, y-1) + diff + gap_penalty,

Value (x-1, y)+ diff + gap_penalty,

}

A	-22	-15	-8	-4	0	4	8	6	4	8
C	-20	-13	-9	-2	2	6	4	5	6	7
T	-18	-11	-7	-3	4	2	0	7	8	6
T	-16	-9	-5	-1	3	1	2	6	7	5
C	-14	-7	-3	1	-1	3	4	5	3	1
A	-12	-5	-1	0	1	2	6	4	2	3
C	-10	-3	-2	2	0	4	2	0	-2	-1
T	-8	-1	0	-2	2	0	-2	-1	0	-2
G	-6	1	-1	0	-2	-1	-3	-5	-4	-3
T	-4	0	1	-1	0	-2	-4	-3	-2	-4
G	-2	2	0	-2	-4	-6	-8	-10	-12	-14
gap	0	-2	-4	-6	-8	-10	-12	-14	-16	-18
	gap	G	A	C	T	C	A	T	T	A

G A C T C A T T A
 G T G T C A C T T C A

Mismatch = 2

Gap = 2

Match = 7

Value = $7(2) + 2(-1) + 2(-2) = 8$