DNA Alignment – Worksheet 1

*CDS 230*

*Spring 2018*

# Problem

What is the purpose of alignment algorithms?

The purpose of alignment algorithms is to present an accurate, but shortened version of a DNA strand in about 300 base pairs instead of the 3000 that are actually in the nucleotide sequence. Using this method to align DNA in multiple sequence alignment also allows us to compare similar genomes between organisms.

# Problem

Load and run the code presented in slide 10 of the video (Code 25.1 in the text).

# simplealign.py

import numpy as np

def SimpleScore ( s1, s2):

a1 = np.array(list(s1))

a2 = np.array(list(s2))

score = (a1==a2).astype(int).sum()

score = score -(a1!=a2 ).astype(int).sum()

ngaps = s1.count( '-' ) + s2.count('-')

score = score - ngaps

print(score)

return score

# # CDS\_HW12.py

import simplealign as sal

sal.SimpleScore( 'AGTCGATCGATT', 'AGTCGATCGATT')

sal.SimpleScore( 'AGTCGATCGATT', 'AGTCGATCGAAT')

sal.SimpleScore( 'AGTCGATCGATT', 'AGTCGATCGA-T')

*12*

*10*

*9*

# Problem

Load and run the code presented in slides 16 and 18 of the video (Code 25.2 and 25.4 in the text).

import blosum

print(blosum.BLOSUM50 [:4 ,:4])

# array ([[5, -2, -1, -2],

# [-2, 7, -1, -2],

# [-1, -1, 7, 2],

# [-2, -2, 2, 8]])

print(blosum. PBET )

# 'ARNDCQEGHILKMFPSTWYV'

print(blosum.PBET.index('D'))

print(blosum.PBET.index('Q'))

print(blosum.BLOSUM50[3,5])

.

*[[ 5 -2 -1 -2]*

*[-2 7 -1 -2]*

*[-1 -1 7 2]*

*[-2 -2 2 8]]*

*ARNDCQEGHILKMFPSTWYV*

*3*

*5*

*0*