

## TASK 1

NC\_012920.1

## TASK 2

NC\_001643

## TASK 3



## TASK 4

Przez wykres przeciąga się jedna ukośna linia od 0,0 do 400,400. Im większa wartość "Window size over which to test threshold" tym mniej szczegółów/kropek jest widocznych, podobnie z "treshold".

## TASK 5

Ograniczeniem jest na przykład ilość danych, które można wyświetlić. Najlepiej sprawdzają się dla małych ilości danych. Są ciężkie do rozczytania w przypadku większej ilości danych.

## TASK 6

Jako, że wartości zanikają i jest to porównanie dwóch sekwencji, opisałbym treshold (pol. próg), jako próg podobieństwa, przy którym jest "stawiana" kropka.

"Window size over which to test threshold" spróbowałbym określić jako *przybliżenie* brane pod uwagę podczas rysowania wykresu.

## TASK 7

rs333

## TASK 8

```
>3 dna:chromosome chromosome:GRCh38:3:46373053:46373887:1
GTCATCCTCATCCTGATAAACTGCAAAAGGCTGAAGAGCATGACTGACATCTACCTGCTC
AACCTGGCCATCTCTGACCTGTTTTTCCTTCTTACTGTCCCCTTCTGGGCTCACTATGCT
GCCGCCAGTGGGACTTTGGAAATACAATGTGTCAACTCTTGACAGGGCTCTATTTTATA
GGCTTCTTCTCTGGAATCTTCTTCATCATCCTCCTGACAATCGATAGGTACCTGGCTGTC
GTCCATGCTGTGTTTGCTTTAAAGCCAGGACGGTCACCTTTGGGGTGGTGACAAGTGTG
ATCACTTGGGTGGTGGCTGTGTTTGCGTCTCTCCCAGGAATCATCTTTACCAGATCTCAA
AAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGG
AAGAATTTCCAGACATTAAAGATAGTCATCTTGGGGCTGGTCCTGCCGCTGCTTGTCATG
GTCATCTGCTACTCGGGAATCCTAAAACTCTGCTTCGGTGTGAAATGAGAAGAAGAGG
CACAGGGCTGTGAGGCTTATCTTCAACCATCATGATTGTTTATTTTCTTCTGCGCTCCC
TACAACATTGTCCTTCTCCTGAACACCTTCCAGGAATTCTTTGGCCTGAATAATTGCAGT
AGCTCTAACAGGTTGGACCAAGCTATGCAGGTGACAGAGACTCTTGGGATGACGCACTGC
```

```
TGCATCAACCCCATCATCTATGCCTTTGTCTGGGGAGAAGTTCAGAACTACCTCTTAGTC
TTCTTCCAAAAGCACATTGCCAAACGCTTCTGCAAATGCTGTTCTATTTTCCAGC
```

## TASK 9

KU382465.1

## TASK 10

```
>KU382465.1 Chlorocebus pygerythrus CCR5 gene, complete cds
ATGGATTATCAAGTGTCAAGTCCAACCTATGACATCAATTATTATACATCGGAGCCCTGCCAAAAAATCA
ACGTGAAGCAAATTGCAGCCCGCCTCCTGCCTCCGCTCTACTACTGGTGTTCATCTTTGGTTTTGTGGG
CAACATACTGGTCGTCCTCATCCTGATAAACTGCAAAAGGCTGAAAAGCATGACTGACATCTACCTGCTC
AACCTGGCCATCTCTGACCTGCTTTTCCTTCTTACTGTCCCCTTCTGGGCTCACTATGCTGCTGCCCAGT
GGGACTTTGGAAATACAATGTGTCAACTCTTGACAGGGCTCTATTTTATAGGCTTCTTCTCTGGAATCTT
CTTCATCATCCTCCTGACAATCGATAGGTACCTGGCTATCGTCCATGCTGTGTTTGCTTTAAAGCCAGG
ACAGTCACCTTTGGGGTGGTGACAAGTGTGATCACTTGGGTGGTGGCTGTGTTTGCCTCTCTCCCAAGAA
TCATCTTTACCAGATCTCAGAGAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTA
TCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATCTTGGGGCTGGTCCTGCCGCTGCTTGTCTATG
GTCATCTGCTACTCGGGAATCCTGAAAACCTCTGCTTCGGTGTGAAACGAGAAGAAGAGGCACAGGGCTG
TGAGGCTCATCTTACCATCATGATTGTTTATTTTCTTCTTGGGCTCCCTACAACATTGTCCTTCTCCT
GAACACCTTCCAGGAATCTTTGGCCTGAATAATTGCAGTAGCTCTAACAGGTTGGACCAAGCCATGCAG
GTGACAGAGACTCTTGGGATGACACACTGCTGCATCAACCCCATCATCTATGCCTTCGTCGGGGAGAAGT
TCAGAAACTACCTCTTAGTCTTCTTCCAAAAGCACATTGCCAAACGCTTCTGCAAATGCTGTTCCATTTT
CCAGCAAGAGGCTCCCGAGCGAGCAAGTTCAGTTTACACCCGATCCACTGGGGAGCAGGAAACATCTGTG
GGCTTGTGA
```

## TASK 11

```
#####
# Program: needle
# Rundate: Sun 21 May 2023 23:58:23
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20230521-235805-0033-24972353-p1m.asequence
#   -bsequence emboss_needle-I20230521-235805-0033-24972353-p1m.bsequence
#   -datafile EDNAFULL
#   -gapopen 10.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -aformat3 pair
#   -snucleotide1
#   -snucleotide2
# Align_format: pair
# Report_file: stdout
#####
```

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1059
# Identity:      818/1059 (77.2%)
# Similarity:    818/1059 (77.2%)
# Gaps:          224/1059 (21.2%)
# Score: 4022.0
#
#=====

EMBOSS_001      1  ----- 0
EMBOSS_001      1 ATGGATTATCAAGTGTCAAGTCCAACCTATGACATCAATTATTATACATC 50
EMBOSS_001      1  ----- 0
EMBOSS_001     51 GGAGCCCTGCCAAAAAATCAACGTGAAGCAAATTGCAGCCCGCCTCCTGC 100
EMBOSS_001      1  ----- 0
EMBOSS_001    101 CTCCGCTCTACTCACTGGTGTTTCATCTTTGGTTTTGTGGGCAACATACTG 150
EMBOSS_001      1 GTCATCCTCATCCTGATAAACTGCAAAGGCTGAAGAGCATGACTGACAT 50
    |||.|||||||||||||||||||||||||||||.|||||||||||
EMBOSS_001    151 GTCGTCCTCATCCTGATAAACTGCAAAGGCTGAAAAGCATGACTGACAT 200
EMBOSS_001     51 CTACCTGCTCAACCTGGCCATCTCTGACCTGTTTTTCCTTCTTACTGTCC 100
    |||||||||||||||||||||||||||||.|||||||||||||||
EMBOSS_001    201 CTACCTGCTCAACCTGGCCATCTCTGACCTGCTTTTCCTTCTTACTGTCC 250
EMBOSS_001    101 CCTTCTGGGCTCACTATGCTGCCGCCAGTGGGACTTTGGAAATACAATG 150
    |||||||||||||||||||||.|||||||||||||||||||||
EMBOSS_001    251 CCTTCTGGGCTCACTATGCTGCTGCCAGTGGGACTTTGGAAATACAATG 300
EMBOSS_001    151 TGTCAACTCTTGACAGGGCTCTATTTTATAGGCTTCTTCTCTGGAATCTT 200
    |||||||||||||||||||||||||||||||||||||||||
EMBOSS_001    301 TGTCAACTCTTGACAGGGCTCTATTTTATAGGCTTCTTCTCTGGAATCTT 350
EMBOSS_001    201 CTTTCATCATCCTCCTGACAATCGATAGGTACCTGGCTGTGCGTCCATGCTG 250
    |||||||||||||||||||||||||||||.|||||||||||
EMBOSS_001    351 CTTTCATCATCCTCCTGACAATCGATAGGTACCTGGCTATCGTCCATGCTG 400
EMBOSS_001    251 TGTTTGCTTTAAAAGCCAGGACGGTCACCTTTGGGGTGGTGACAAGTGTG 300
    |||||||||||||||||||||.|||||||||||||||||||||
EMBOSS_001    401 TGTTTGCTTTAAAAGCCAGGACAGTCACCTTTGGGGTGGTGACAAGTGTG 450
```

EMBOSS_001	301	ATCACTTGGGTGGTGGCTGTGTTTGCGTCTCTCCCAGGAATCATCTTTAC	350
EMBOSS_001	451	ATCACTTGGGTGGTGGCTGTGTTTGCCTCTCTCCAAGAATCATCTTTAC	500
EMBOSS_001	351	CAGATCTCAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCAT	400
EMBOSS_001	501	CAGATCTCAGAGAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCAT	550
EMBOSS_001	401	ACAGTCAGTATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATC	450
EMBOSS_001	551	ACAGTCAGTATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATC	600
EMBOSS_001	451	TTGGGGCTGGTCCTGCCGCTGCTTGTCATGGTCATCTGCTACTCGGGAAT	500
EMBOSS_001	601	TTGGGGCTGGTCCTGCCGCTGCTTGTCATGGTCATCTGCTACTCGGGAAT	650
EMBOSS_001	501	CCTAAAACTCTGCTTCGGTGTCGAAATGAGAAGAAGAGGCACAGGGCTG	550
		.	
EMBOSS_001	651	CCTGAAAACTCTGCTTCGGTGTCGAAACGAGAAGAAGAGGCACAGGGCTG	700
EMBOSS_001	551	TGAGGCTTATCTTCACCATCATGATTGTTTATTTTCTTCTTGGGCTCCC	600
		.	
EMBOSS_001	701	TGAGGCTCATCTTCACCATCATGATTGTTTATTTTCTTCTTGGGCTCCC	750
EMBOSS_001	601	TACAACATTGTCCTTCTCCTGAACACCTTCCAGGAATTCTTTGGCCTGAA	650
EMBOSS_001	751	TACAACATTGTCCTTCTCCTGAACACCTTCCAGGAATTCTTTGGCCTGAA	800
EMBOSS_001	651	TAATTGCAGTAGCTCTAACAGGTTGGACCAAGCTATGCAGGTGACAGAGA	700
		.	
EMBOSS_001	801	TAATTGCAGTAGCTCTAACAGGTTGGACCAAGCCATGCAGGTGACAGAGA	850
EMBOSS_001	701	CTCTTGGGATGACGCACTGCTGCATCAACCCCATCATCTATGCCTTTGTC	750
		.	
EMBOSS_001	851	CTCTTGGGATGACACACTGCTGCATCAACCCCATCATCTATGCCTTCGTC	900
EMBOSS_001	751	GGGGAGAAGTTCAGAACTACCTCTTAGTCTTCTTCCAAAAGCACATTGC	800
EMBOSS_001	901	GGGGAGAAGTTCAGAACTACCTCTTAGTCTTCTTCCAAAAGCACATTGC	950
EMBOSS_001	801	CAAACGCTTCTGCAAATGCTGTTCTATTTTCCAGC-----	835
		.	
EMBOSS_001	951	CAAACGCTTCTGCAAATGCTGTTCCATTTTCCAGCAAGAGGCTCCCGAGC	1000
EMBOSS_001	836	-----	835
EMBOSS_001	1001	GAGCAAGTTCAGTTTACACCCGATCCACTGGGGAGCAGGAAACATCTGTG	1050
EMBOSS_001	836	-----	835
EMBOSS_001	1051	GGCTTGTGA	1059

```
#-----  
#-----
```

TASK 12

```
#####  
# Program: needle  
# Rundate: Mon 22 May 2023 00:05:18  
# Commandline: needle  
#   -auto  
#   -stdout  
#   -asequence emboss_needle-I20230522-000515-0142-1745463-p2m.asequence  
#   -bsequence emboss_needle-I20230522-000515-0142-1745463-p2m.bsequence  
#   -gapopen 10.0  
#   -gapextend 0.5  
#   -endopen 10.0  
#   -endextend 0.5  
#   -aformat3 pair  
#   -sprotein1  
#   -sprotein2  
# Align_format: pair  
# Report_file: stdout  
#####  
  
#=====
```

#		
# Aligned_sequences: 2		
# 1: EMBOSS_001		
# 2: EMBOSS_001		
# Matrix: EBLOSUM62		
# Gap_penalty: 10.0		
# Extend_penalty: 0.5		
#		
# Length: 1059		
# Identity: 845/1059 (79.8%)		
# Similarity: 845/1059 (79.8%)		
# Gaps: 214/1059 (20.2%)		
# Score: 5097.0		
#		
#		
#=====		

EMBOSS_001	1	-----	0
EMBOSS_001	1	ATGGATTATCAAGTGTCAAGTCCAACCTATGACATCAATTATTATACATC	50
EMBOSS_001	1	-----	0
EMBOSS_001	51	GGAGCCCTGCCAAAAAATCAACGTGAAGCAAATTGCAGCCCGCCTCCTGC	100
EMBOSS_001	1	-----CAACATACTG	10

EMBOSS_001	101	CTCCGCTCTACTCACTGGTGTTTCATCTTTGGTTTTGTGGGCAACATACTG	150
EMBOSS_001	11	GTCGTCCTCATCCTGATAAACTGCAAAGGCTGAAAAGCATGACTGACAT	60
EMBOSS_001	151	GTCGTCCTCATCCTGATAAACTGCAAAGGCTGAAAAGCATGACTGACAT	200
EMBOSS_001	61	CTACCTGCTCAACCTGGCCATCTCTGACCTGCTTTTCCTTCTTACTGTCC	110
EMBOSS_001	201	CTACCTGCTCAACCTGGCCATCTCTGACCTGCTTTTCCTTCTTACTGTCC	250
EMBOSS_001	111	CCTTCTGGGCTCACTATGCTGCTGCCAGTGGGACTTTGGAAATACAATG	160
EMBOSS_001	251	CCTTCTGGGCTCACTATGCTGCTGCCAGTGGGACTTTGGAAATACAATG	300
EMBOSS_001	161	TGTCAACTCTTGACAGGGCTCTATTTTATAGGCTTCTTCTCTGGAATCTT	210
EMBOSS_001	301	TGTCAACTCTTGACAGGGCTCTATTTTATAGGCTTCTTCTCTGGAATCTT	350
EMBOSS_001	211	CTTCATCATCCTCCTGACAATCGATAGGTACCTGGCTATCGTCCATGCTG	260
EMBOSS_001	351	CTTCATCATCCTCCTGACAATCGATAGGTACCTGGCTATCGTCCATGCTG	400
EMBOSS_001	261	TGTTTGCTTTAAAAGCCAGGACAGTCACCTTTGGGGTGGTGACAAGTGTG	310
EMBOSS_001	401	TGTTTGCTTTAAAAGCCAGGACAGTCACCTTTGGGGTGGTGACAAGTGTG	450
EMBOSS_001	311	ATCACTTGGGTGGTGGCTGTGTTTGCCTCTCTCCCAAGAATCATCTTTAC	360
EMBOSS_001	451	ATCACTTGGGTGGTGGCTGTGTTTGCCTCTCTCCCAAGAATCATCTTTAC	500
EMBOSS_001	361	CAGATCTCAGAGAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCAT	410
EMBOSS_001	501	CAGATCTCAGAGAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCAT	550
EMBOSS_001	411	ACAGTCAGTATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATC	460
EMBOSS_001	551	ACAGTCAGTATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATC	600
EMBOSS_001	461	TTGGGGCTGGTCCTGCCGCTGCTTGTCATGGTCATCTGCTACTCGGGAAT	510
EMBOSS_001	601	TTGGGGCTGGTCCTGCCGCTGCTTGTCATGGTCATCTGCTACTCGGGAAT	650
EMBOSS_001	511	CCTGAAAACCTCTGCTTCGGTGTCGAAACGAGAAGAAGAGGCACAGGGCTG	560
EMBOSS_001	651	CCTGAAAACCTCTGCTTCGGTGTCGAAACGAGAAGAAGAGGCACAGGGCTG	700
EMBOSS_001	561	TGAGGCTCATCTTCACCATCATGATTGTTTATTTTCTCTTCTGGGCTCCC	610
EMBOSS_001	701	TGAGGCTCATCTTCACCATCATGATTGTTTATTTTCTCTTCTGGGCTCCC	750
EMBOSS_001	611	TACAACATTGTCCTTCTCCTGAACACCTTCCAGGAATTCTTTGGCCTGAA	660
EMBOSS_001	751	TACAACATTGTCCTTCTCCTGAACACCTTCCAGGAATTCTTTGGCCTGAA	800

EMBOSS_001	661	TAATTGCAGTAGCTCTAACAGGTTGGACCAAGCCATGCAGGTGACAGAGA	710
EMBOSS_001	801	TAATTGCAGTAGCTCTAACAGGTTGGACCAAGCCATGCAGGTGACAGAGA	850
EMBOSS_001	711	CTCTTGGGATGACACACTGCTGCATCAACCCCATCATCTATGCCTTCGTC	760
EMBOSS_001	851	CTCTTGGGATGACACACTGCTGCATCAACCCCATCATCTATGCCTTCGTC	900
EMBOSS_001	761	GGGGAGAAGTTCAGAAACTACCTCTTAGTCTTCTTCCAAAAGCACATTGC	810
EMBOSS_001	901	GGGGAGAAGTTCAGAAACTACCTCTTAGTCTTCTTCCAAAAGCACATTGC	950
EMBOSS_001	811	CAAACGCTTCTGCAAATGCTGTTCCATTTTCCAGC-----	845
EMBOSS_001	951	CAAACGCTTCTGCAAATGCTGTTCCATTTTCCAGCAAGAGGCTCCCGAGC	1000
EMBOSS_001	846	-----	845
EMBOSS_001	1001	GAGCAAGTTCAGTTTACACCCGATCCACTGGGGAGCAGGAAACATCTGTG	1050
EMBOSS_001	846	-----	845
EMBOSS_001	1051	GGCTTGTGA	1059
#-----			
#-----			

TASK 13

Podobieństwo jest wyższe, ponieważ usunęliśmy fragmenty, które odpowiadały za różnice w obydwu sekwencjach.

Świadczy o tym, że bez przycinania wynik porównywania dwóch różnych ciągów, jest niższy.

TASK 14

Wybieram dwie sekwencje DNA:

- Sekwencja 1: ATTAGGGGCGCCTGACAGTAAACGTTGGATAAGACTCCCAATTAGCCGAG
- Sekwencja 2: GAGTACGTAATATAATAGCGTTTAGCTATTTCGGGTGCGGTCTTCAAGGAC

Wynik:

EMBOSS_001	1	ATTAGGGGCGCCTGACAGTAAACGTTGGATAAGACTCCCAATTAGC----	46
		.       ...     .. ..	
EMBOSS_001	1	-----GAGT--ACGTAATATAA---TAGCGTTTAGCTATT	30
EMBOSS_001	47	CGAG-----	50

```
EMBOSS_001      ||.|
                 31 CGGGTGCGGTCTTCAAGGAC      50
```

"|" pomiędzy oznacza match

" " pomiędzy oznacza mismatch

"." pomiędzy oznacza gap

Informacje bazowałem na [dokumentacji](#).

## TASK 15

Korzystam z narzędzia FinchTV.

Są to oznaczenia tych nukleotydów, które nie zostały zidentyfikowane.

- N jako nukleotydy, których nie udało się zidentyfikować jako:
  - adeniny (A),
  - cytozyny (C),
  - guaniny (G)
- M jako nukleotydy, których nie udało się zidentyfikować jako:
  - tyminy (T),
  - uracyl (U)

## TASK 16

Jako dwa pierwsze rekordy z % identyfikacją spekuluję, że jest to **Lysinibacillus sp**, czyli *Lysinibacillus sphaericus* - bakteria występująca w glebie.

## TASK 17

*Lysinibacillus sphaericus* zistał wyizolowany poraz pierwszy w 1982 roku.

Bakteria została odizolowana z ciała larwy *Culiseta incidens* - mchówka z rodziny komarowatych.

## TASK 18

BG2-10 16S ribosomal gene

## TASK 19

*Lysinibacillus* to rodzaj gram-dodatnich **bakterii**.

## TASK 20

```
TGC-
AGTCGAGCGAACAGAGAAGGAGCTTGCTCCTTCGACGTTAGCGGCGGACGGGTGAGTAACACGTGGGCAACCTACCTTATAG
TTTGGGATAACTCCGGGAAACCGGGGCTAATACC
```



# TASK 21

"." - dowolny nukleotyd, tożsamość/identity, zgodność z innymi

"-" - brak nukleotydu

Jeśli w miejscu kropki znajduje się nukleotyd, oznacza to, że jest to jednoliterowy skrót niezgodność pomiędzy badanymi sekwencjami.

# TASK 22

```
Query: None Query ID: 1c1|Query_106413 Length: 119

Query range 1: 1 to 120

Query      1      TGC-
AGTCGAGCGAACAGAGAAGGAGCTTGCTCCTTCGACGTTAGCGGCGGACGGGTGAGTAACACGTGGGCAACCTACCTTATAG
TTTGGGATAACTCCGGGAAACCGGGGCTAATACC  119
NR_112628.1  29
...A.....
..... 148
NR_042072.1  39
...A.....
..... 158
NR_112569.1  29
...A.....
..... 148
NR_134074.1  21
...-.....T.....C.
..... 139
NR_114207.1  29
...A.....T.....
..... 148
NR_118146.1  19
...-.....T.....C.
..... 137
NR_114920.1  29
...A.....C.
..... 148
```

# TASK 23

![[image-20230524020307273]](C:\Users\klips\Documents\PJATK\6 - Letni\PBIO\PBIO\Krzysztof\_Lipski\_lab\_6\task23)

Są to:

- A
- A
- A

- T, C
- A, T
- T, C
- A, C

## TASK 24

Jest to adenina (A). Wszystkie sekwencje się nim różnią od mojej, a w dwóch go brakuje (-).