

Experiment-5

Aim- To perform pairwise sequence alignment between Theropithecus gelada HUS1 checkpoint clamp component (HUS1) sequence and other orthologous genes of your teammates using

a) EMBOSS DOTMATCHER b) EMBOSS NEEDLE (Global) and Water (Local)

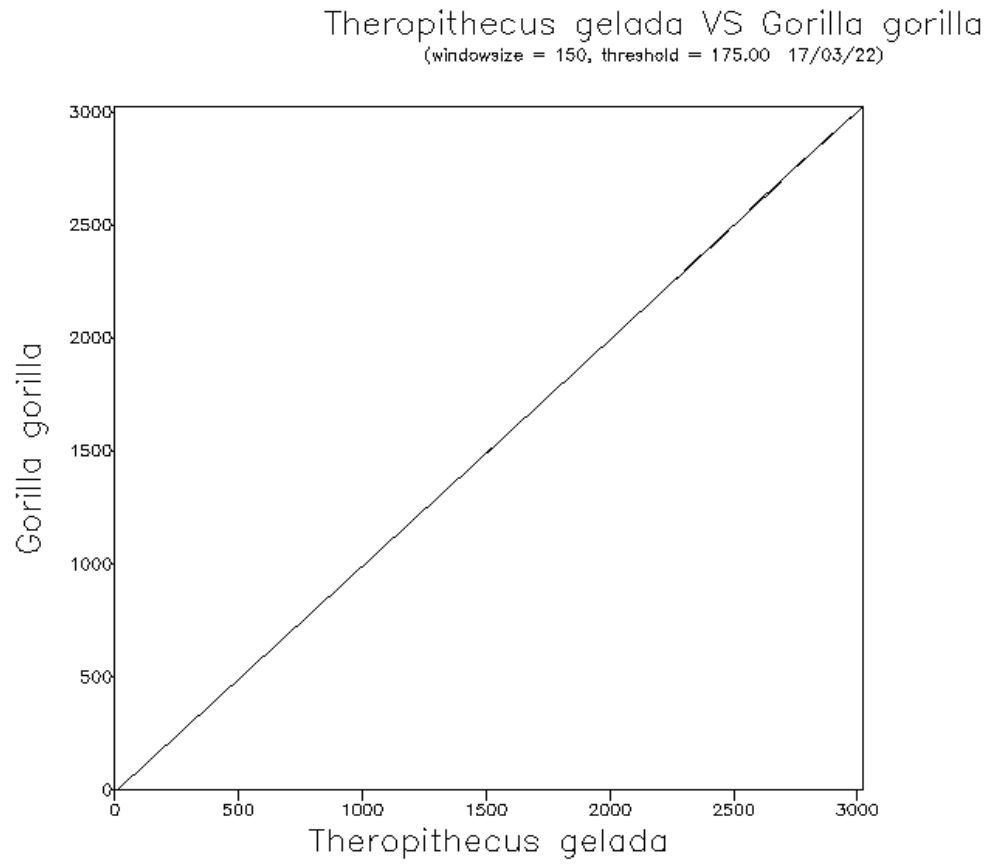
Materials Required- ORF Finder, gene sequence of Theropithecus gelada HUS1 checkpoint clamp component (HUS1)

Procedure:

- 1) Collect orthologous genes and retrieve their nucleotide sequence in FASTA format from NCBI and make a Multiple Sequence Format (MSF) file out of it.
- 2) Open the home page of DOTMATCHER
https://www.ebi.ac.uk/Tools/seqstats/emboss_dotmatcher/ or
(<https://www.bioinformatics.nl/cgi-bin/emboss/dotmatcher>), EMBOSS NEEDLE
(https://www.ebi.ac.uk/Tools/psa/emboss_needle/) and EMBOSS WATER
(https://www.ebi.ac.uk/Tools/psa/emboss_water/) respectively in separate tabs for performing Pairwise Sequence Alignment (PSA).
- 3) Paste the appropriate combination of sequences in a given space or upload your respective sequence for which you want to perform PSA.
- 4) Select appropriate parameters like word size, threshold value, output format, Gap open, and extension penalty in respective tools.
- 5) Take relevant screenshots of observations and make a comparative table for the dynamic programming (Needle and Water) algorithm.
- 6) Analyze the results.

Observations-

A) DOTMATCHER



B) EMBOSS NEEDLE

Theropithecus gelada VS Gorilla gorilla

```
#=====
#
# Aligned_sequences: 2
# 1: XM_025380797.1
# 2: XM_004045421.3
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 3042
# Identity:      2866/3042 (94.2%)
# Similarity:    2866/3042 (94.2%)
# Gaps:          33/3042 ( 1.1%)
# Score: 13631.5
#
#
#=====

XM_025380797.      1  -----AACCA-TGTGCGCGCCGCCGTGTTTCC|
                        ||||| |||.|||||||||||||||
XM_004045421.      1  GGGCGGAACCACTGTTTCGCGCCGCCGTGTTTCC|

XM_025380797.     44  GCGCGACACTCATCTGTTACCCACAGAG-GCCG|
                        |||||.|.|.|||||||||||||||.|||
XM_004045421.     51  GCGCGACGCTTGTCTGTTACCCACAGAGTCCCG|
```

Theropithecus gelada VS Pan paniscus

```
#=====
#
# Aligned_sequences: 2
# 1: XM_025380797.1
# 2: XM_003805830.5
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 3052
# Identity:      1101/3052 (36.1%)
# Similarity:    1101/3052 (36.1%)
# Gaps:          1919/3052 (62.9%)
# Score: 5357.0
#
#
#=====

XM_025380797.      1  -----AACCA-TGTGCGCG|
                        ||||| |||.||||
XM_003805830.      1  CGTAGCCCGAGCCGGGCGGAACCACTGTTTCGCG|

XM_025380797.     31  GGGGACACTCAGGGCGCGACACTCATCTGTTAC|
                        |||||||||||||.|||.|||||
XM_003805830.     51  GGGGACACTCAGGGCGCGACGCTTTTCTGTTAC|
```

Theropithecus gelada VS Homo Sapiens

```
#=====
#
# Aligned_sequences: 2
# 1: XM_025380797.1
# 2: NM_004507.4
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 3034
# Identity:   2867/3034 (94.5%)
# Similarity: 2867/3034 (94.5%)
# Gaps:       30/3034 ( 1.0%)
# Score: 16531.0
#
#=====

XM_025380797.      1 AACCATGTGCGCGCCGCCGTGTTTCCGGGCGGGC
                  |||.|||||.|||||.|||||.|||||.
NM_004507.4        1 -AC--TGTTCGCGCTGCCGTGTTTCCGGGCGGGC

XM_025380797.     51 ACTCATCTGTTACCCACAGAGG-CCGTCGCGGC
                  .|||.|||||.|||||.|||||.|||||.
NM_004507.4       48 GCTTTTCTGTTACCCACAGAGGCCCGCCGCGGC
```

Theropithecus gelada VS Macaca mulatta

```
#=====
#
# Aligned_sequences: 2
# 1: XM_025380797.1
# 2: XM_001088000.4
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 3035
# Identity:   2995/3035 (98.7%)
# Similarity: 2995/3035 (98.7%)
# Gaps:       8/3035 ( 0.3%)
# Score: 14816.0
#
#=====

XM_025380797.      1 -AACCA-TGTGCGCGCCGCCGTGTTTCCGGGC
                  |||||.|||||.|||||.|||||.|||||.
XM_001088000.      1 GAACCACTGTGCGCGCCGCCGTGTTTCCGGGC

XM_025380797.     49 AACTCATCTGTTACCCACAGAGGCCGTCGCGC
                  |||||.|||||.|||||.|||||.|||||.
XM_001088000.     51 AACTTATCTGTTACCCACAGAGGCCGTCGCGC
```

C) EMBOSS WATER

Theropithecus gelada VS Gorilla gorilla

```
#=====
#
# Aligned_sequences: 2
# 1: XM_025380797.1
# 2: XM_004045421.3
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 3034
# Identity:      2866/3034 (94.5%)
# Similarity:    2866/3034 (94.5%)
# Gaps:          25/3034 ( 0.8%)
# Score: 13631.5
#
#
#=====

XM_025380797.      1 AACCA-TGTGCGCGCCGCGTGTTCGG
                   ||||| |||.|||||||||||||||
XM_004045421.      7 AACCACTGTTCGCGCCGCGTGTTCGG

XM_025380797.     50 CACTCATCTGTTACCCACAGAG-GCCGTC
                   |.|||.|||||||||||||||.|||.|
XM_004045421.     57 CGCTTGTCTGTTACCCACAGAGTCCCGCC
```

Theropithecus gelada VS Pan paniscus

```
#=====
#
# Aligned_sequences: 2
# 1: XM_025380797.1
# 2: XM_003805830.5
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1135
# Identity:      1101/1135 (97.0%)
# Similarity:    1101/1135 (97.0%)
# Gaps:          2/1135 ( 0.2%)
# Score: 5357.0
#
#
#=====

XM_025380797.      1 AACCA-TGTGCGCGCCGCGTGTTCGGGCGGG
                      ||||| |||.|||||||||||||||||||
XM_003805830.     20 AACCACTGTTGCGCGCCGCGTGTTCGGGCGGG

XM_025380797.     50 CACTCATCTGTTACCCACAGAGG-CCGTCGCGGC
                      |.|||.||||||||||||||||||| |||.|||||
XM_003805830.     70 CGCTTTTCTGTTACCCACAGAGGCCGCGCGGC
```

Theropithecus gelada VS Homo Sapiens

```
#=====
#
# Aligned_sequences: 2
# 1: XM_025380797.1
# 2: NM_004507.4
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 3029
# Identity:      2865/3029 (94.6%)
# Similarity:    2865/3029 (94.6%)
# Gaps:          27/3029 ( 0.9%)
# Score: 13649.5
#
#
#=====

XM_025380797.      6 TGTGCGCGCCGCGTGTTCGGGCGGGGACAC
                      |||.|||||.|||||||||||||||||||
NM_004507.4        3 TGTTCGCGCTGCCGTGTTCGGGCGGGGACAC

XM_025380797.     56 TCTGTTACCCACAGAGG-CCGTCGCGGCTGCGG
                      |||||||||||||||| |||.|||||||.
NM_004507.4       53 TCTGTTACCCACAGAGGCCGCGCGGCTGCGC
```

Theropithecus gelada VS Macaca mulatta

```
#=====
#
# Aligned_sequences: 2
# 1: XM_025380797.1
# 2: XM_001088000.4
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 3032
# Identity:      2995/3032 (98.8%)
# Similarity:    2995/3032 (98.8%)
# Gaps:          5/3032 ( 0.2%)
# Score: 14816.0
#
#
#=====

XM_025380797.      1 AACCA-TGTGCGCGCCGCCGTGTTTCCGGGCGGC
                  |||||
XM_001088000.      2 AACCCTGTGCGCGCCGCCGTGTTTCCGGGCGGC

XM_025380797.      50 CACTCATCTGTTACCCACAGAGGCCGTCGCGGC
                  |||.|||||
XM_001088000.      52 CACTTATCTGTTACCCACAGAGGCCGTCGCGGC
```

Interpretation of Results-

A) The Pairwise Sequence Alignment (PSA) using DOTMATCHER of **Theropithecus gelada** HUS1 checkpoint clamp component (HUS1) with **Gorilla gorilla** HUS1 checkpoint clamp component (HUS1) shows maximum alignment by selecting **window size 150** and **threshold value 175**, respectively.

B) The Pairwise Sequence Alignment (PSA) using EMBOSS NEEDLE between orthologous genes shows maximum similarity with *Macaca mulatta* with 98.7% similarity as shown:

With Gorilla gorilla : 94.2%
With Pan paniscus : 36.1%
With Homo sapiens : 94.5%
With Macaca mulatta : 98.7%

C) The Pairwise Sequence Alignment (PSA) using EMBOSS WATER between orthologous genes shows maximum similarity with *Macaca mulatta* with 98.7% similarity as shown:

With Gorilla gorilla : 94.2%

With Pan paniscus : 36.1%

With Homo sapiens : 94.5%

With Macaca mulatta : 98.7%