

Group 6:

Tiffany Cook

Drashti Mehta

Joshua Mikombo

BINF 6201 Lab Report 1

- **Part 1: Find and download the hexokinase sequences [3 pts total]**
 - 1A. Human HK 1: **NP_000179.2**
 - 1B. Dog HK 1: **XP_005619024.1**
 - 1C. Crocodile HK1: **XP_019394065.1**
 - 1D. Human HK 2: **NP_000180.2**
 - 1E. Dog HK 2: **XP_038309165.1**
 - 1F. Crocodile HK 2: **XP_019401522.1**
- **Part 2: Pairwise Alignments of HK1 with Different Parameters [7 pts total]**
 - 2A. Run the alignment using the default parameters. What scoring matrix is being used? What are the gap open and gap extend penalties? For the resulting alignment, what is the percent identity? What is the percent similarity? How many gaps are there? (1 pt)

```
# Aligned_sequences: 2
# 1: NP_000179.2
# 2: XP_005619024.1
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 943
# Identity:      859/943 (91.1%)
# Similarity:    886/943 (94.0%)
# Gaps:          34/943 ( 3.6%)
# Score: 4446.5
```
 - Human and Dog HK1
 - **It uses BLOSUM62 as the default parameter. Gap open penalties are 10 and Extension penalties are 0.5. The rest of the scoring is shown in the figure.**
 - 2B. Re-do the alignment with the PAM10 matrix. Did this change your results from part A? In what way did the results change? (1 pt)

```
# Aligned_sequences: 2
# 1: NP_000179.2
# 2: XP_005619024.1
# Matrix: EPAM10
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 955
# Identity:      857/955 (89.7%)
# Similarity:    857/955 (89.7%)
# Gaps:          58/955 ( 6.1%)
# Score: 6823.0
#
```

Human and Dog HK1 Pam 10

- In using PAM10, the results had minor changes. The gap open and gap extend stayed the same while the length increased by 38 nucleotides. The percent identity decreased to 89.7% and the percent similarity decreased to 89.7%. The percent similarity and percent identity are the same here. Which means the part of sequences which are similar are also identical. The gaps in the sequence are showing 6.1% whereas, in the BLOSUM62, it was 3.6% Lastly, the score did not change within 2 aligned sequences.**
- 2C. Re-do the alignment again, now with the PAM250 matrix. Did this seem to do better or worse than the alignment with the PAM10 matrix? How do these results compare to the results with the default parameters? (1 pt)

```
# Aligned_sequences: 2
# 1: NP_000179.2
# 2: XP_005619024.1
# Matrix: EPAM250
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 935
# Identity:      859/935 (91.9%)
# Similarity:    901/935 (96.4%)
# Gaps:          18/935 ( 1.9%)
# Score: 4297.0
#
```

Human and Dog HK1 Pam 250

- Likewise, in part B, the gap open and extend did not change. PAM 250 seemed to have done better because PAM250 has less divergence meaning higher similarity than using BLOSUM62. In the default test, BLOSUM62 showed 94% similarity and PAM250 showed 96.4% similarity. This test goes to show that PAM250 is a better test to use because it has a more relevant test in regards to the sequence vs BLOSUM62 showing substitutions and conserved positions in blocks representing most-alike common regions in related sequences.**

- 2D. Re-do the alignment one more time using the PAM250 matrix. Set the gap open and the gap extend penalties to the highest possible values. Now, what does your alignment look like in terms of identity, similarity, and number of gaps? (1 pt)

```
# Aligned_sequences: 2
# 1: NP_000179.2
# 2: XP_005619024.1
# Matrix: EPAM250
# Gap_penalty: 100.0
# Extend_penalty: 10.0
#
# Length: 935
# Identity:      856/935 (91.6%)
# Similarity:    894/935 (95.6%)
# Gaps:          18/935 ( 1.9%)
# Score: 4286.0
```

- Human and Dog HK1 PAM250 with other parameters
 - **The gap identity and similarity are the same at maximum. In the previous tests, the identity and similarity seemed to decrease in comparison to those at maximum. The gap is the same for both PAM250 tests, but it is the lowest in PAM250 when compared to BLOSUM62 and PAM10.**
- 2E. Why do you think lowering the number of the PAM matrix had the effect that it did on the alignments? (3 pts)
 - **The lower the PAM matrix is the lower the time between divergence thus only giving a short window in the analysis between the sequences. This leads to similarity to be lower when comparing PAM matrices of a lower number than those of a higher one.**

- Part 3: Use the Cluster to Align HK2 [1 pt total]

- **#!/bin/bash**
- **#SBATCH --partition=Centaurus**
- **#SBATCH --job-name=needle**
- **#SBATCH --nodes=1**
- **#SBATCH --ntasks-per-node=1**
- **#SBATCH --time=0:30:00**
- **module load emboss**
- **needle -auto **
- **-asequence Dog_HK2.fasta -sformat1 fasta -sprotein1 Y **
- **-bsequence Human_HK2.fasta -sformat2 fasta -sprotein2 Y **

- **-datafile EBLOSUM62 -gapopen 10 -gapextend 0.5 **
- **-outfile HK2_alignment**
- Question 3: What does the alignment of Human and Dog HK2 sequences look like? How does it compare to the HK1 alignment with the same parameters? (1 pt)

```
#####
# Program: needle
# Rundate: Mon 30 Jan 2023 12:57:07
# Commandline: needle
#   -auto
#   -asequence dog_HK2.fasta
#   -sformat1 fasta
#   -sprotein1
#   -bsequence human_HK2.fasta
#   -sformat2 fasta
#   -sprotein2
#   -datafile EBLOSUM62
#   -gapopen 10
#   -gapextend 0.5
#   -outfile HK2_alignment
# Align_format: srspair
# Report_file: HK2_alignment
#####

#=====
#
# Aligned_sequences: 2
# 1: XP_038309165.1
# 2: NP_000180.2
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 917
# Identity:      877/917 (95.6%)
# Similarity:    896/917 (97.7%)
# Gaps:          0/917 ( 0.0%)
# Score: 4567.0
#
#=====
```

Human and Dog HK2

- **The HK2 alignment shows greater identity and similarity percentages between the human and dog HK2 gene, compared to the HK1 alignment.**

-
- Part 4: Alignment with a More Distant Ortholog [2 pts total]

```
#=====
#
# Aligned_sequences: 2
# 1: NP_000179.2
# 2: XP_019394065.1
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 917
# Identity:      805/917 (87.8%)
# Similarity:    870/917 (94.9%)
# Gaps:          0/917 ( 0.0%)
# Score: 4251.0
```

- Human and Croc HK1

```
# Aligned_sequences: 2
# 1: NP_000180.2
# 2: XP_019401522.1
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 917
# Identity:      812/917 (88.5%)
# Similarity:    876/917 (95.5%)
# Gaps:          0/917 ( 0.0%)
# Score: 4300.0
```

- Human and Croc HK2

- Question 4: How do the identity, similarity, and gap scores compare to the Human/Dog alignments?

- 4A. For HK1? (1 pt)

```
# Aligned_sequences: 2
# 1: NP_000179.2
# 2: XP_005619024.1
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 943
# Identity:      859/943 (91.1%)
# Similarity:    886/943 (94.0%)
# Gaps:          34/943 ( 3.6%)
# Score: 4446.5
```

- Human and Dog HK1

```

=====
#
# Aligned_sequences: 2
# 1: NP_000179.2
# 2: XP_019394065.1
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 917
# Identity:      805/917 (87.8%)
# Similarity:    870/917 (94.9%)
# Gaps:          0/917 ( 0.0%)
# Score: 4251.0

```

- Human and Croc HK1

- **The Dog HK1 ortholog shows more identity to the human HK1 when compared to the Crocodile HK1. But here, the similarity of humans HK1 to crocodile is a little more than that of the dog.**

- 4B. For HK2? (1 pt)

```

# Aligned_sequences: 2
# 1: NP_000180.2
# 2: XP_019401522.1
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 917
# Identity:      812/917 (88.5%)
# Similarity:    876/917 (95.5%)
# Gaps:          0/917 ( 0.0%)
# Score: 4300.0

```

- Human and Croc HK2

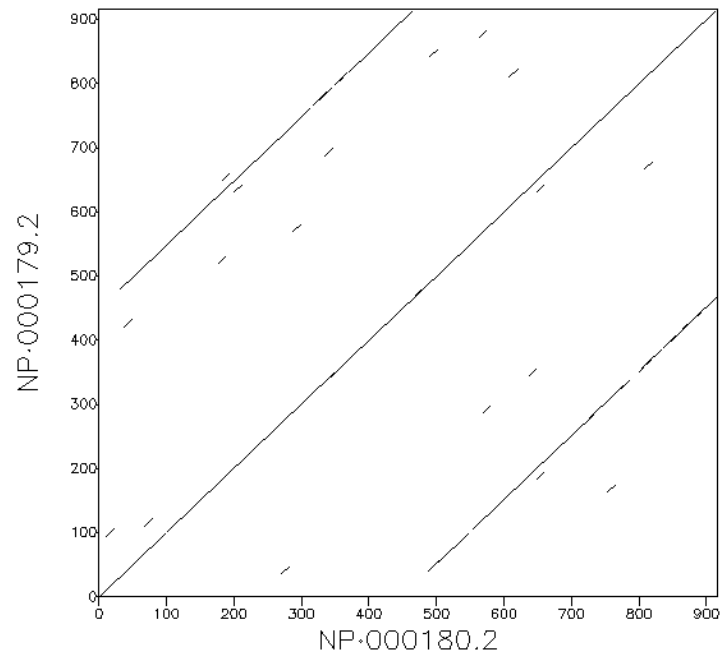
```
#####
# Program: needle
# Rundate: Mon 30 Jan 2023 12:57:07
# Commandline: needle
#   -auto
#   -asequence dog_HK2.fasta
#   -sformat1 fasta
#   -sprotein1
#   -bsequence human_HK2.fasta
#   -sformat2 fasta
#   -sprotein2
#   -datafile EBLOSUM62
#   -gapopen 10
#   -gapextend 0.5
#   -outfile HK2_alignment
# Align_format: srspair
# Report_file: HK2_alignment
#####

#=====
#
# Aligned_sequences: 2
# 1: XP_038309165.1
# 2: NP_000180.2
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 917
# Identity:      877/917 (95.6%)
# Similarity:    896/917 (97.7%)
# Gaps:          0/917 ( 0.0%)
# Score: 4567.0
#
#=====
```

Human and Dog HK2

- The Dog HK2 ortholog shows more similarity and identical to the human HK2 when compared to the Crocodile HK2. Comparing it to the HK1 results, the similarity of Dog ortholog is higher which was not the case for HK1. Also, % difference is also higher in the case of HK2.
- Part 5: Dot-plot Comparison of HK1 and HK2 [2 pts total]
 - 5A. Show your resulting dot-plot. (1 pt)

Dotmatcher: fasta::emboss-dotmatcher-I20230130-182102-04...
(windowsize = 10, threshold = 23.00 30/01/23)



-
- 5B. What kind of macro-scale genomic pattern or mutation is indicated by your dot-plot? (1 pt)
- The dot-plot indicates “repeats” in the alignment of the human HK1 and HK2 protein sequences. But they are also inverted.