

Aim: In this assignment, given n DNA sequences of different species in fasta file format, we ask to implement UPGMA to construct a phylogenetic relationship for given DNA sequences.

You will use Needleman-Wunsch algorithm to construct distance matrix, where match score = 4, mismatch score = -3, and gap penalty = -2. **You can reuse your code from HW3.**

You will use the aligned sequences to create the distance matrix for UPGMA.

Use the following formula: distance = number of mismatches + number of gaps

Input file format (sequences.fasta):

```
>species 1
GAACAGGATTAAATGAGAGT
>species 2
GAACGGCATTAAATTGG
>species 3
GCTCAATTAG
>species 4
GCAATTCAAACCC
```

Output file format (upgma.tree.txt):

```
[[species 1:3.5-species 2:3.5]:1.625-[species 3:4.0-species 4:4.0]:1.125]:0.0
```

In the output file, you will give the resulting tree in the given format. You will give the relationships between species ([species x - species y]) and the distance between nodes and their parent nodes (:distance_to_its_parent_node).

Notes:

- Your code must be written by yourself. Sufficient evidence of plagiarism will be treated the same as for plagiarism or cheating.
- **Non-compiling submissions will not be evaluated.**
- **Do not submit the program binary.**
- You must submit the following items:
 - **For C & C++:** All of the header and source files (.h & .cpp) or (.h & .c) and a script to compile the source code and produce the binary (Makefile).
 - For Java:** All of the source files (.java)
 - For Python:** All of the source files (.py)
- Submit your answers to Fatma Kahveci at 'fatma.balci@bilkent.edu.tr'.
- Use 'CS481 Assignment-4' in the subject line of your e-mail.
- Zip your files and send them **in only one zipped file**.
File name format='surname_name_hw4.zip'
- Name your source and header files **properly**.
- **C / C++, Python, or Java** will be used as programming language.
- **Do not add input and output files.** You can assume that input file is found in the same directory with your header and/or sources files.
- All submissions must be made by **23:59, December 21, 2017**.
- Assignments that are sent after the deadline will not be accepted **in any reason**.
- **Bonus** will be given for the fastest code. The fastest wins.