

# CS 175

## Problem Set 2

Due: 22 October 2018, 11:59pm

### General Instructions

- The answer sheet for this problem set should be submitted as a PDF file. You may use any word processing software to create the answer sheet. The name of the PDF file to be submitted should follow the following format: [CS 175] < *Last Name, First Name* > – Problem Set 2.pdf. For example: [CS 175] De la Cruz, Juan - Problem Set 2.pdf. Include the answer sheet in the .zip file detailed in the 3rd bullet point.
- If you have consulted references (books, journal articles, online materials, other people), cite them as footnotes to the specific item where you used the resource/s as reference.
- This problem set will require you to submit a .zip file containing several files (including the answer sheet) as detailed below. The name of the .zip file to be submitted should follow the following format: [CS 175] < *Last Name, First Name* > – Problem Set 2.zip. For example: [CS 175] De la Cruz, Juan - Problem Set 2.zip
- Submission of the problem set answers should be done via e-mail. Attach the .zip file, and write as the subject header of the e-mail: [CS 175] < *Last Name, First Name* > – Problem Set 2. For example: [CS 175] De la Cruz, Juan - Problem Set 2. Send your answers to [jcyap@dcs.upd.edu.ph](mailto:jcyap@dcs.upd.edu.ph).
- **You should receive a confirmation e-mail from me stating receipt of your deliverable within 24 hours upon your submission of the problem set.** If you have not received any, forward your previous submission using the same subject header once more.
- If you have any questions regarding an item (EXCEPT the answer and solution) in the problem set, do not hesitate to e-mail me to ask them. However, **questions regarding this problem set forwarded/received on or after 12:01am of 19 October 2018 will NOT be entertained.**

### Questions

For this activity, you are given eight sequences derived from the DNA of eight meat samples bought from an undisclosed shop. These meats were labeled as ‘baboy ramo’ or the Philippine pig (*Sus philippinensis*) (samples B1-5) and ‘usa’ or Philippine deer (*Rusa marianna*) (Samples U1-5). Each sequence consists of approximately 460 nucleotides amplified through PCR from the cytochrome oxidase subunit I gene.

Sample B1

acacgaaaaaatcatcgttgtacttcaactacaagaaccttaatgaccaacatccgaaaa  
tcacaccactaataaaaattatcaacaacgcattcattgacctccagccccctcaaac  
atctcatcatgatgaaacttcggttcctcttaggcattctgcctaattcttgcaaatccta  
acaggcctgttcttagcaatacattacacatcagacacaacaacagctttctcatcagtt  
acacacatttgtcgagacgtaaattacggatgagttattcgctatctacatgcaaacgga  
gcatccatattctttatttgctattcatccacgtaggcccagggtctatactacggatcc  
tataatctcctagaaacatgaaacattggagtagtcctactattttaccggtatagcaaca  
gccttcataggctacgtcctgcatgaggacaaatTTTTTTTctgag

Sample B2

atgtgcctcatggcaggacgtagcctatgaaggctgttgctataacggtaaatagtaga  
actactccaatgtttcatgtttctaggaatatataggatccgtagtataggcctcggcct  
acgtggatgaataggcaataaagaacatggatgctccgtttgcatgtaggtagcgaata  
actcatccgtaatttacgtctcgacagatgtgtgtaactgatgagaaagctgttgttg  
tctgatgtgtaatgtattgctaagaacaggcctgttaggatttgcaagattaggcagatg  
cctaagagggaaccgaagtttcatcatgatgagatgtttgagggggctgggagggtcaatg  
aatgcgttgttgataatTTTTTattagtggggtgtgattttcggatgttggtcattaaggtt  
ctttagttgaagtacaacgatgatntttcgtgtc

Sample B3

cctcagaaaagtatttgcctcatggcaggacgtagcctatgaaggctgttgctataacg  
gtaaatagtagaactactccaatgtttcatgtttctaggaatatataggatccgtagtat  
aggcctcggcctacgtggatgaataggcaataaagaacatggatgctccgtttgcatgt  
aggtagcgaataactcatccgtaatttacgtctcgacagatgtgtgtaactgatgagaaa  
gctgttgttgtctgatgtgtaatgtattgctaagaacaggcctgttaggatttgcaag  
attaggcagatgcctaagagggaaccgaagtttcatcatgatgagatgtttgagggggct  
gggagggtcaatgaatgcgttgttgataatTTTTTattagtggggtgtgattttcggatgttg  
gtcattaaggttctttagttgaagtacaacgatgattTTTTTcgtgtc

Sample B5

cctcagaaaantatttgcctcatggcaggacgtagcctatgaaggctgttgctataacg  
gtaaatagtagaactactccaatgtttcatgtttctaggaatatataggatccgtagtat  
aggcctcggcctacgtggatgaataggcaataaagaacatggatgctccgtttgcatgt  
aggtagcgaataactcatccgtaatttacgtctcgacagatgtgtgtaactgatgagaaa  
gctgttgttgtctgatgtgtaatgtattgctaagaacaggcctgttaggatttgcaag  
attaggcagatgcctaagagggaaccgaagtttcatcatgatgagatgtttgagggggct  
gggagggtcaatgaatgcgttgttgataatTTTTTattagtggggtgtgattttcggatgttg  
gtcattaaggttctttagttgaagtacaacgatgattTTTTTcgtgtc

Sample U1

cctcagaaaagatatttgtcctcatggcaggacgtagcctatgaaggctgttgctataac  
ggtaaatagtagaactactccaatgtttcatgtttctaggaatatataggatccgtagta  
taggcctcggcctacgtggatgaataggcaaataaagaacatggatgctccgtttgcag  
taggtagcgaataactcatccgtaatttacgtctcgacagatgtgtgtaactgatgagaa  
agctgttgttgtctgatgtgtaatgtattgctaagaacaggcctgttaggatttgcaa  
gattaggcagatgcctaagagggaaccgaagtttcatcatgatgagatgtttgagggggc  
tgggaggtcaatgaatgcgttgttgataatTTTTATTAGTGGGTGTGATTTTCGGATGTT  
GGTCATTAAGGTTCTTGTAGTTGAAGTACAACGATGATTTTTTCGTGTC

Sample U2

cctcagaaaagatatttgtcctcatggcaggacgtagcctatgaaggctgttgctataac  
ggtaaatagtagaactactccaatgtttcatgtttctaggaatatataggatccgtagta  
taggcctcggcctacgtggatgaataggcaaataaagaacatggatgctccgtttgcag  
taggtagcgaataactcatccgtaatttacgtctcgacagatgtgtgtaactgatgagaa  
agctgttgttgtctgatgtgtaatgtattgctaagaacaggcctgttaggatttgcaa  
gattaggcagatgcctaagagggaaccgaagtttcatcatgatgagatgtttgagggggc  
tgggaggtcaatgaatgcgttgttgataatTTTTATTAGTGGGTGTGATTTTCGGATGTT  
GGTCATTAAGGTTCTTGTAGTTGAAGTACAACGATGATTTTTTCGTGTC

Sample U3

gacacgaaaaatcatcgttgtacttcaactacaagaaccttaatgaccaacatccgaaaa  
tcacaccactaataaaaattatcaacaacgcattcattgacctccagccccctcaaac  
atctcatcatgatgaaacttcggttcctcttaggcactctgcctaactcttgcaaatccta  
acaggcctgttcttagcaatacattacacatcagacacaacaacagctttctcatcagtt  
acacacatctgtcgagacgtaaattacggatgagttattcgctacctacatgcaaacgga  
gcatccatgttctttatttgcctattcatccacgtaggcaggcctatactacggatcc  
tatatattcctagaaacatgaaacattggagtagttctactatttaccgttatagcaaca  
gccttcataggctacgtcctgcatgaggacaaatactTTTTCTGAGG

Sample U5

tcagaaaagatatttgtcctcatggcaggacgtagcctatgaaggctgttgctataacgg  
taaataagtagaactactccaatgtttcatgtttctaggaatatataggatccgtagtata  
ggcctcggcctacgtggatgaataggcaaataaagaacatggatgctccgtttgcagta  
ggtagcgaataactcatccgtaatttacgtctcgacagatgtgtgtaactgatgagaaag  
ctgttgttgtctgatgtgtaatgtattgctaagaacaggcctgttaggatttgcaaga  
ttaggcagatgcctaagagggaaccgaagtttcatcatgatgagatgtttgagggggctg  
ggaggtcaatgaatgcgttgttgataatTTTTATTAGTGGGTGTGATTTTCGGATGTTGG  
TCATTAAGGTTCTTGTAGTTGAAGTACAACGATGATTTTTTCGTGTC

Next, perform a sequence database search on each of the sequences above using Nucleotide BLAST in the NCBI BLAST homepage (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). For the Choose Search Set menu, select nucleotide collection (nr/nt) along Database. For the Program Selection, select Highly similar sequences (megablast). Identify the closest match for each of the samples above based on percentage coverage, bit score and percentage identity. Take note of the accession number as well.

1. Create a table as indicated below, filling it up with the relevant information obtained using the BLAST search.

Sequence	GenBank Taxon with Closest Taxon	Accession No. of Closest Match	% Identity	E-value
B1				
B2				
B3				
B5				
U1				
U2				
U3				
U5				

2. Based on the obtained results, were the meats properly labelled?

Now, let's check if phylogenetic analysis will somehow concur with the findings of the BLAST search. Create two FASTA formatted files: one containing samples B1, B2, B3, and B5; and another containing samples U1, U2, U3, and U5. Reconstruct phylogenetic trees for each of the FASTA formatted files using MAFFT for the alignment, then RaxML for the phylogenetic tree reconstruction (use GTRCAT as model parameter and 12345 as seed) . Generate a visualization of the trees using FigTree. Include the FASTA-formatted files you created as input, the output files of RaxML, and an image of the tree visualization using FigTree in the .zip file mentioned in the General Instructions.

3. For this item, place the screenshots of the trees generated, grouping them according to the file used to generate the tree. Is there a concurrence with the trees generated by all of the techniques used for samples B1, B2, B3, and B5? How about for the trees generated using samples U1, U2, U3, and U5?
4. Is there a concurrence between the findings of the BLAST search (i.e. if the meats were properly labelled) and the resulting phylogenetic trees (i.e. were outlier sequences detected in the generated phylogenetic tree)? Expound on your answer.