

Identifying Neisseria Species

UVIC

UNIVERSITAT DE VIC
UNIVERSITAT CENTRAL
DE CATALUNYA

○ ○ ○ ○

Alba Burillo
Final Task
Python

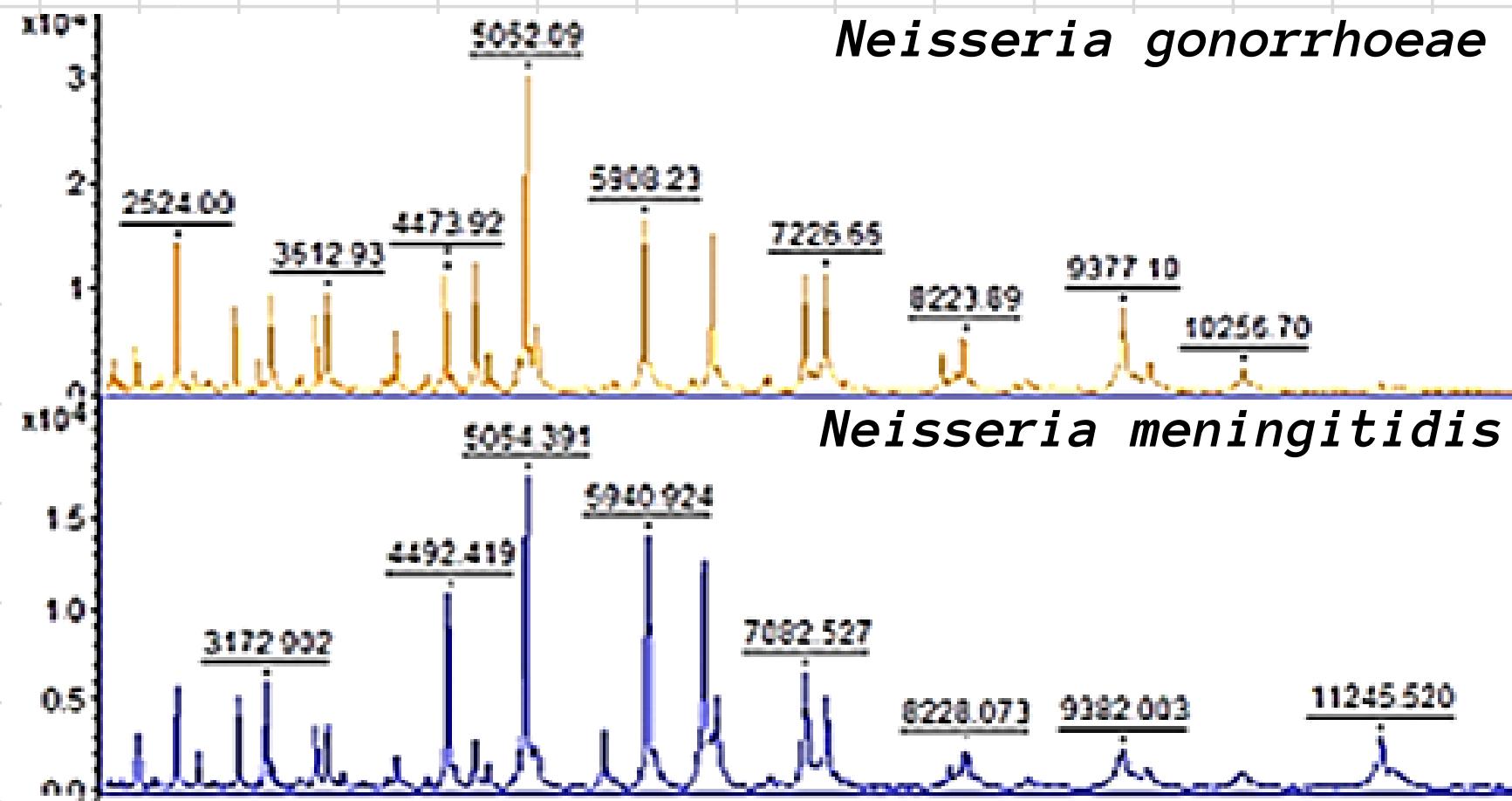
Index

○ ○ ○ ○

- 1 Biological problem
- 2 Program structure
- 3 Used modules
- 4 Difficulties

Biological problem

○ ○ ○ ○



- Bacterial genera isolated in the oral cavity are: *Streptococcus*, *Veillonella*, *Porphyromonas*, *Treponema*, *Neisseria*, *Haemophilus*, *Eubacteria*, *Lactobacterium*, *Capnophycum*, *Pseudomonas*...
- *Neisseria meningitidis* and *Neisseria gonorrhoeae*.
- MALDI-TOF MS → problems for the differentiation of these two pathogenic species.

Code



Programme structure

- 1 FASTA file with known DNA sequences of the rplF and rpmH genes, we obtain the RNA, and the proteins sequence.
- 2 We create a function to obtain the molecular weight of the proteins.
- 3 We create a function to identify at the gender level.
- 4 We create a function to identify at the species level.
- 5 We make a **dendrogram** to see the relationships between the sequences.

```
def mol_weigth(proteins):
    # We create an empty list called listrw to save the roundweights of all the proteins in the fasta file
    listrw = []
    for i in proteins:
        # The weight of each amino acid
        weights = {'A': 71.04, 'C': 103.01, 'D': 115.03, 'E': 129.04, 'F': 147.07,
                   'G': 57.02, 'H': 137.06, 'I': 113.08, 'K': 128.09, 'L': 113.08,
                   'M': 131.04, 'N': 114.04, 'P': 97.05, 'Q': 128.06, 'R': 156.1,
                   'S': 87.03, 'T': 102.05, 'V': 99.07, 'W': 186.08, 'Y': 163.06}
        weight = sum(weights[p] for p in i)
        # We add the molecular weight of one water molecule (H2O) = 18.01524 Da
        weight = (weight + 18.01524)
        # We round to two decimal places
        roundweight = round(weight, 2)
        listrw.append(roundweight)
        # Print the sum of the molecular weights
        # print("The molecular weight of this protein is:", roundweight, "Da")
    return listrw
```

```
def neisseria_genus(listrw):  
    for i in listrw:  
        if i == 5051.96:  
            print("The molecular weight of this protein is:", i, "Da")  
            print("Belongs to the genus Neisseria")  
        else:  
            print("The molecular weight of this protein is:", i, "Da")  
            print("Does NOT belong to the genus Neisseria")
```

The molecular weight of the 50S ribosomal protein L34 in the different species of the Neisseria genus is the same (5051.96 Da / ~ 5052 Da).

The molecular weight of the 50S ribosomal protein L6 is different in the species of the genus Neisseria.

```
def neisseria_species(listrw):
    for i in listrw:
        if i == 18879.89:
            print("The molecular weight of this protein is:", i, "Da")
            print("Neisseria flavescens")
        elif i == 18897.87:
            print("The molecular weight of this protein is:", i, "Da")
            print("Neisseria meningitidis")
        elif i == 18916.86:
            print("The molecular weight of this protein is:", i, "Da")
            print("Neisseria gonorrhoeae")
        elif i == 18907.91:
            print("The molecular weight of this protein is:", i, "Da")
            print("Neisseria mucosa")
        elif i == 18894.93:
            print("The molecular weight of this protein is:", i, "Da")
            print("Neisseria lactamica")
        elif i == 18936.95:
            print("The molecular weight of this protein is:", i, "Da")
            print("Neisseria sicca")
        elif i == 18893.9:
            print("The molecular weight of this protein is:", i, "Da")
            print("Neisseria elongata")
        else:
            print("The molecular weight of this protein is:", i, "Da")
            print("Unidentified")
```

Used modules



```
# We need to install different modules
# pip install biopython
from Bio import SeqIO
# pip install scipy
from scipy.cluster.hierarchy import dendrogram, linkage
from matplotlib import pyplot as plt
```

We used SeqIO
from Biopython

We used pyplot
from matplotlib

Dendrogram and linkage
from
scipy.cluster.hierarchy

Difficulties

◦ ◦ ◦ ◦

1

When obtaining the molecular weight of the protein, using only the sum of the amino acids was not the expected weight.

- The correct weight was obtained by adding the weight of a water molecule to the sum.

2

When applying the `mol_weight()` function to the L34 and L6 protein lists, it only returned the last value.

- We create an empty list inside the function called **listrw** to store the weights of all proteins in the fasta file.

Difficulties

3

To apply the `neisseria_genus()` and `neisseria_species()` functions we create an object where we first apply the `mol_weight()` function to the list of proteins.

4

To make the dendrograms we needed to create a list, so we directly used the lists of the molecular weights of the proteins (`L6_MW` and `L34_MW`).

1. **Serrano-Col** HA, Sánchez-Jiménez M, Cardona-Castro N. Conocimiento de la microbiota de la cavidad oral a través de la metagenómica. Rev. CES Odont 2015; 28(2): 112-118

2. Morel, F., Jacquier, H., Desroches, M., Fihman, V., Kumanski, S., Cambau, E., Decousser, J. y Berçot, B. 2018, "Use of Andromas and Bruker MALDI-TOF/MS MS in the identification of Neisseria", European Journal of Clinical Microbiology & Infectious Diseases, vol. 37, no. 12, pp. 2273-2277.

3. Bennett JS, Watkins ER, Jolley KA, Harrison OB, Maiden MCJ. Identifying Neisseria species by use of the 50S ribosomal protein L6 (rplF) gene. J Clin Microbiol. 2014;52(5):1375-81.

4. Malakhova, M.M., Maier, T., Kubanova, A.A., Govorun, V.M., Svistunova, T.S., Gazarian, A.O., Borovskaya, A.D., Kostrzewa, M., Ilina, E.N., Vereshchagin, V.A. y Kruglov, A.N. 2009, "Direct Bacterial Profiling by Matrix-Assisted Laser Desorption-Ionization Time-ofFlight Mass Spectrometry for Identification of Pathogenic Neisseria", Journal of Molecular Diagnostics, The, vol. 11, no. 1, pp.75-86