# CSE 4065 - Introduction to Computational Genomics

## Project 1 - Report

### 150114051 - Mehmet MUM
### 150115067 - Tayfun YURDAER

## WUHAN - Hamming Distance 0

```
WUHAN:          Hamming Distance:0
---------------------------------
5-mer
-----------------
k-mer:    aacaa
rev comp: ttgtt
freq:     201

6-mer
-----------------
k-mer:    taacaa
rev comp: ttgtta
freq:     70

7-mer
-----------------
k-mer:    ttgttaa
rev comp: ttaacaa
freq:     28

8-mer
-----------------
k-mer:    aaattgtt
rev comp: aacaattt
freq:     12

9-mer
-----------------
k-mer:    taaacgaac
rev comp: gttcgttta
freq:     7
```

```
10-mer
-----------------
k-mer:    ctaaacgaac
rev comp: gttcgtttag
freq:     5

11-mer
-----------------
k-mer:    tctaaacgaac
rev comp: gttcgtttaga
freq:     3

12-mer
-----------------
k-mer:    gttgatggtgtt
rev comp: aacaccatcaac
freq:     3

13-mer
-----------------
k-mer:    ggctactaacaat
rev comp: attgttagtagcc
freq:     2

14-mer
-----------------
k-mer:    caacctgaagaaga
rev comp: tcttcttcaggttg
freq:     2

15-mer
-----------------
k-mer:    atcagacaactacta
rev comp: tagtagttgtctgat
freq:     2
```

```
16-mer
-----------------
k-mer:    atcagacaactactat
rev comp: atagtagttgtctgat
freq:     2

17-mer
-----------------
k-mer:    atcagacaactactatt
rev comp: aatagtagttgtctgat
freq:     2

18-mer
-----------------
k-mer:    acactggtaattaccagt
rev comp: actggtaattaccagtgt
freq:     2

19-mer
-----------------
k-mer:    acactggtaattaccagtg
rev comp: cactggtaattaccagtgt
freq:     2

20-mer
-----------------
k-mer:    acactggtaattaccagtgt
rev comp: acactggtaattaccagtgt
freq:     2
```

## WUHAN - Hamming Distance 1

```
WUHAN:          Hamming Distance:1
---------------------------------
5-mer
-----------------
k-mer:    aaaaa
rev comp: ttttt
freq:     2315

6-mer
-----------------
k-mer:    aaaaaa
rev comp: tttttt
freq:     761

7-mer
-----------------
k-mer:    tgttgtt
rev comp: aacaaca
freq:     290

8-mer
-----------------
k-mer:    tgttgtta
rev comp: taacaaca
freq:     113

9-mer
-----------------
k-mer:    tgttgttaa
rev comp: ttaacaaca
freq:     47

10-mer
-----------------
k-mer:    tgatgttgtt
rev comp: aacaacatca
freq:     25
```

```
11-mer
-----------------
k-mer:    aatttaggtga
rev comp: tcacctaaatt
freq:     9

12-mer
-----------------
k-mer:    tatgaagatttt
rev comp: aaaatcttcata
freq:     5

13-mer
-----------------
k-mer:    ttatgaagatttt
rev comp: aaaatcttcataa
freq:     4

14-mer
-----------------
k-mer:    ttgtacagaaagtg
rev comp: cactttctgtacaa
freq:     3

15-mer
-----------------
k-mer:    gtaacaaaccaacca
rev comp: tggttggtttgttac
freq:     2

16-mer
-----------------
k-mer:    ctactagtgaagctgt
rev comp: acagcttcactagtag
freq:     2
```

```
17-mer
-----------------
k-mer:    ctactagtgaagctgtt
rev comp: aacagcttcactagtag
freq:     2

18-mer
-----------------
k-mer:    ctactagtgaagctgttg
rev comp: caacagcttcactagtag
freq:     2

19-mer
-----------------
k-mer:    tacactggtaattaccagt
rev comp: actggtaattaccagtgta
freq:     2

20-mer
-----------------
k-mer:    tacactggtaattaccagtg
rev comp: cactggtaattaccagtgta
freq:     2

21-mer
-----------------
k-mer:    tacactggtaattaccagtgt
rev comp: acactggtaattaccagtgta
freq:     2
```

# WUHAN - Hamming Distance 2

```
WUHAN:            Hamming Distance:2
------------------------------------
5-mer
------------------
k-mer:     aaaaa
rev comp: ttttt
freq:      11057

6-mer
------------------
k-mer:     aaaaaa
rev comp: tttttt
freq:      4903

7-mer
------------------
k-mer:     taaaaaa
rev comp: tttttta
freq:      1973

8-mer
------------------
k-mer:     tttaaaat
rev comp: attttaaa
freq:      777

9-mer
------------------
k-mer:     aattttaaa
rev comp: tttaaaatt
freq:      312

10-mer
------------------
k-mer:     aactttaaaa
rev comp: ttttaaagtt
freq:      126
```

```
11-mer
------------------
k-mer:     taatggtgtta
rev comp: taacaccatta
freq:      50

12-mer
------------------
k-mer:     ttaatggtgtta
rev comp: taacaccattaa
freq:      27

13-mer
------------------
k-mer:     tgttaatggtgtt
rev comp: aacaccattaaca
freq:      15

14-mer
------------------
k-mer:     ttgttaatggtgtt
rev comp: aacaccattaacaa
freq:      9

15-mer
------------------
k-mer:     tggtgtttattctgt
rev comp: acagaataaacacca
freq:      6

16-mer
------------------
k-mer:     aaggcattttgatgaa
rev comp: ttcatcaaaatgcctt
freq:      4
```

```
17-mer
------------------
k-mer:     tgttggtgattattttg
rev comp: caaaataatcaccaaca
freq:      3

18-mer
------------------
k-mer:     aaagtcaacatcaatatt
rev comp: aatattgatgttgacttt
freq:      2

19-mer
------------------
k-mer:     aaagtcaacatcaatattg
rev comp: caatattgatgttgacttt
freq:      2

20-mer
------------------
k-mer:     gctcttcaacctgaagaaga
rev comp: tcttcttcaggttgaagagc
freq:      2

21-mer
------------------
k-mer:     tagtgagtacactggtaatta
rev comp: taattaccagtgtactcacta
freq:      2

22-mer
------------------
k-mer:     tagtgagtacactggtaattac
rev comp: gtaattaccagtgtactcacta
freq:      2
```

```
23-mer
------------------
k-mer:     tagtgagtacactggtaattacc
rev comp: ggtaattaccagtgtactcacta
freq:      2

24-mer
------------------
k-mer:     tagtgagtacactggtaattacca
rev comp: tggtaattaccagtgtactcacta
freq:      2

25-mer
------------------
k-mer:     tagtgagtacactggtaattaccag
rev comp: ctggtaattaccagtgtactcacta
freq:      2

26-mer
------------------
k-mer:     tagtgagtacactggtaattaccagt
rev comp: actggtaattaccagtgtactcacta
freq:      2

27-mer
------------------
k-mer:     tagtgagtacactggtaattaccagtg
rev comp: cactggtaattaccagtgtactcacta
freq:      2

28-mer
------------------
k-mer:     tagtgagtacactggtaattaccagtgt
rev comp: acactggtaattaccagtgtactcacta
freq:      2
```

## NEPAL - Hamming Distance 0

```
NEPAL:              Hamming Distance:0
------------------------------------------
5-mer
------------------
k-mer:      aacaa
rev comp: ttgtt
freq:      201

6-mer
------------------
k-mer:      taacaa
rev comp: ttgtta
freq:      70

7-mer
------------------
k-mer:      ttgttaa
rev comp: ttaacaa
freq:      28

8-mer
------------------
k-mer:      aaattgtt
rev comp: aacaattt
freq:      12

9-mer
------------------
k-mer:      taaacgaac
rev comp: gttcgttta
freq:      7

10-mer
------------------
k-mer:      ctaaacgaac
rev comp: gttcgtttag
freq:      5
```

```
11-mer
------------------
k-mer:      tctaaacgaac
rev comp: gttcgtttaga
freq:      3

12-mer
------------------
k-mer:      gttgatggtgtt
rev comp: aacaccatcaac
freq:      3

13-mer
------------------
k-mer:      ggctactaacaat
rev comp: attgttagtagcc
freq:      2

14-mer
------------------
k-mer:      caacctgaagaaga
rev comp: tcttcttcaggttg
freq:      2

15-mer
------------------
k-mer:      atcagacaactacta
rev comp: tagtagttgtctgat
freq:      2

16-mer
------------------
k-mer:      atcagacaactactat
rev comp: atagtagttgtctgat
freq:      2
```

```
17-mer
------------------
k-mer:      atcagacaactactatt
rev comp: aatagtagttgtctgat
freq:      2

18-mer
------------------
k-mer:      acactggtaattaccagt
rev comp: actggtaattaccagtgt
freq:      2

19-mer
------------------
k-mer:      acactggtaattaccagtg
rev comp: cactggtaattaccagtgt
freq:      2

20-mer
------------------
k-mer:      acactggtaattaccagtgt
rev comp: acactggtaattaccagtgt
freq:      2
```

## NEPAL - Hamming Distance 1

```
NEPAL:              Hamming Distance:1
------------------------------------------
5-mer
------------------
k-mer:      aaaaa
rev comp: ttttt
freq:      2315

6-mer
------------------
k-mer:      aaaaaa
rev comp: tttttt
freq:      761

7-mer
------------------
k-mer:      tgttgtt
rev comp: aacaaca
freq:      290

8-mer
------------------
k-mer:      tgttgtta
rev comp: taacaaca
freq:      113

9-mer
------------------
k-mer:      tgttgttaa
rev comp: ttaacaaca
freq:      47

10-mer
------------------
k-mer:      tgatgttgtt
rev comp: aacaacatca
freq:      25
```

```
11-mer
------------------
k-mer:      aatttaggtga
rev comp: tcacctaaatt
freq:      9

12-mer
------------------
k-mer:      tatgaagatttt
rev comp: aaaatcttcata
freq:      5

13-mer
------------------
k-mer:      ttatgaagatttt
rev comp: aaaatcttcataa
freq:      4

14-mer
------------------
k-mer:      ttgtacagaaagtg
rev comp: cactttctgtacaa
freq:      3

15-mer
------------------
k-mer:      gtaacaaaccaacca
rev comp: tggttggtttgttac
freq:      2

16-mer
------------------
k-mer:      ctactagtgaagctgt
rev comp: acagcttcactagtag
freq:      2
```

```
17-mer
------------------
k-mer:      ctactagtgaagctgtt
rev comp: aacagcttcactagtag
freq:      2

18-mer
------------------
k-mer:      ctactagtgaagctgttg
rev comp: caacagcttcactagtag
freq:      2

19-mer
------------------
k-mer:      tacactggtaattaccagt
rev comp: actggtaattaccagtgta
freq:      2

20-mer
------------------
k-mer:      tacactggtaattaccagtg
rev comp: cactggtaattaccagtgta
freq:      2

21-mer
------------------
k-mer:      tacactggtaattaccagtgt
rev comp: acactggtaattaccagtgta
freq:      2
```

# NEPAL - Hamming Distance 2

```
NEPAL:            Hamming Distance:2
------------------------------------
5-mer
------------------
k-mer:     aaaaa
rev comp:  ttttt
freq:      11057

6-mer
------------------
k-mer:     aaaaaa
rev comp:  tttttt
freq:      4903

7-mer
------------------
k-mer:     taaaaaa
rev comp:  ttttttta
freq:      1973

8-mer
------------------
k-mer:     tttaaaat
rev comp:  attttaaa
freq:      777

9-mer
------------------
k-mer:     aattttaaa
rev comp:  tttaaaatt
freq:      312

10-mer
------------------
k-mer:     aactttaaaa
rev comp:  ttttaaagtt
freq:      126
```

```
11-mer
------------------
k-mer:     taatggtgtta
rev comp:  taacaccatta
freq:      50

12-mer
------------------
k-mer:     ttaatggtgtta
rev comp:  taacaccattaa
freq:      27

13-mer
------------------
k-mer:     tgttaatggtgtt
rev comp:  aacaccattaaca
freq:      15

14-mer
------------------
k-mer:     ttgttaatggtgtt
rev comp:  aacaccattaacaa
freq:      9

15-mer
------------------
k-mer:     tggtgtttattctgt
rev comp:  acagaataaacacca
freq:      6

16-mer
------------------
k-mer:     aaggcattttgatgaa
rev comp:  ttcatcaaaatgcctt
freq:      4
```

```
17-mer
------------------
k-mer:     tgttggtgattattttg
rev comp:  caaataatcaccaaca
freq:      3

18-mer
------------------
k-mer:     aaagtcaacatcaatatt
rev comp:  aatattgatgttgacttt
freq:      2

19-mer
------------------
k-mer:     aaagtcaacatcaatattg
rev comp:  caatattgatgttgacttt
freq:      2

20-mer
------------------
k-mer:     gctcttcaacctgaagaaga
rev comp:  tcttcttcaggttgaagagc
freq:      2

21-mer
------------------
k-mer:     tagtgagtacactggtaatta
rev comp:  taattaccagtgtactcacta
freq:      2

22-mer
------------------
k-mer:     tagtgagtacactggtaattac
rev comp:  gtaattaccagtgtactcacta
freq:      2
```

```
23-mer
------------------
k-mer:     tagtgagtacactggtaattacc
rev comp:  ggtaattaccagtgtactcacta
freq:      2

24-mer
------------------
k-mer:     tagtgagtacactggtaattacca
rev comp:  tggtaattaccagtgtactcacta
freq:      2

25-mer
------------------
k-mer:     tagtgagtacactggtaattaccag
rev comp:  ctggtaattaccagtgtactcacta
freq:      2

26-mer
------------------
k-mer:     tagtgagtacactggtaattaccagt
rev comp:  actggtaattaccagtgtactcacta
freq:      2

27-mer
------------------
k-mer:     tagtgagtacactggtaattaccagtg
rev comp:  cactggtaattaccagtgtactcacta
freq:      2

28-mer
------------------
k-mer:     tagtgagtacactggtaattaccagtgt
rev comp:  acactggtaattaccagtgtactcacta
freq:      2
```

- The algorithm finds the most frequent k-mer from k is 5 to no frequent k-mer. We run the algorithm for both Nepal and Wuhan with hamming distance 0, 1 and 2

    For Wuhan
    - It finds with hamming distance 0 until 20-mer
    - It finds with hamming distance 1 until 21-mer
    - It finds with hamming distance 2 until 28-mer

    For Nepal
    - It finds with hamming distance 0 until 20-mer
    - It finds with hamming distance 1 until 21-mer
    - It finds with hamming distance 2 until 28-mer

- The most frequent k-mer changes with hamming distance change. Some of them may not change but mostly it changes.

- Also, the number of occurance of the most frequent k-mer increases with hamming distance increases.

- Wuhan and Nepal, the algorithm find exactly same the most frequent k-mer and number of occurence for both of them with same hamming distance. So, we can say that there are not mutations in the virus.