**Name and Surname: Adna Hrapovic BIO402 Molecular Evolution Project**

**Human Metapneumovirus (HMPV)**

**Abstract**

Human metapneumovirus has been discovered in 2001, but it existed even years before. It is a cold weather virus that causes upper and lower respiratory ilness in humans. HMPV mostly affects children and older adults. This RNA single negative-stranded virus has an envelope and its incubation lasts 3 to 5 days. The sequence of the RNA is 13411 bases long and it represents itself in five known human strains. In this project, each strain is compared to the first strain that is discovered. The strains are furhter explained with graphs that show mutation ratios, transition and transversion ratio, similarities and indels, gaps, deletion. In the future it is expected that each strain is going to be pathogenetically explained in details, along with the immune response of the human organism to the virus and proper medical treatment, for example vaccines. Until then, people need to be responsible and cautious – wash their hands, cover coughs and sneezes in order to stop this virus from spreading and causing consequences.

**Introduction**

The human metapneuovirus causes upper and lower respiratory illness in humans and it has been discovered in 2001. The cases of HMPV are found in people of all ages, but it mostly manifests during the winter time in young children and older adults. This cold-weather virus causes fever, cough, shortness of breath and nasal congestion. This infection can turn into bronchitis or pneumonia if the patient’s immune system is very weak. Some rare cases are showing symptoms like vomiting or diarrhea. HMPV is spread through contact of infected person by droplets like caugh or sneeze or by touching surfaces that have viruses on them – then eventually spreading in on the face,mouth, nose, eyes… After exposure to HMPV, infection develops 3 to 5 days and the symptoms lasts up to 2 weeks. In order to prevent this infection, we should avoid touching out face, wash our hands often, disinfect areas (especially childten’s toys) and cover mouth and nose while sneezing or coughing (*Human Metapneumovirus Clinical Features - NREVSS | CDC*, 2021; *Metapneumovirus | Intermountain Healthcare*, n.d.). The structure of human Metapneumovirus is defined as a single negative-stranded RNA virus with an envelope, classified in *Pneumovirinae* subfamily of *Paramyxovridae* family. HMPV is spherical and has diameter from 150 nm. RNA of the virus is about 13kb in size and codes for eight proteins. It is a second most common cause of lower respiratory tract infection in children (*Human Metapneumovirus*, 2021; *Metapneumovirus ~ ViralZone*, n.d.). The human Metapneumovirus has five known strands that are used in this project, the first one is isolated from children in the Netherlands. It was most related (homologus) to the avian metapneumovirus that infects chicken, ducks and turkeys. (Shafagati & Williams, 2018) The overall sequence is 13411 bases, with coding sequence length of 12463 and noncoding sequence length of 948 bases.  
Mutations are permanent changes in the DNA material, most of them occur because of the DNA copying mistakes or due to exposure of different mutagens. The most common mutations are transitions and transversions – “point mutations”. Transition occurs if the purine is substituted with another purine (A to G/G to A). Transversions occur when the pyrimidine is substituted with another pyrimidine (T to C or vice versa). Those mutations can be silent, nonsense or missense. The second type of mutation is deletion – it results as loss of one or more base pairs of DNA. It affects the final protein synthesis. Insertion is addition of base pairs in the DNA. (*DNA Mutation and Repair*, n.d.)

**Materials and Methods**

For the method of multiple alignment sequencing we used the NCBI and searched genome of the virus of our choice. The human metapneumovirus had five sequences available. By clicking on the “Replicons” we managed to read/copy the available sequence one by one. For the MSA, Clustal Omega is used. It is an online tool that compares the given sequences from txt file. The file is saved in txt format and later analyzed by a program. As an output we got an excel document that listed the information we needed about the sequences. The mutations that are mentioned in this project are insertions, deletions, transitions and transversions. Besides that, the output document shows similarities between the strains, including coding, noncoding and overall sequences and mutation frequency in them. All that data is converted into graphs that are explained later in the discussion part. The graphs are created using Excel.   
Mutation frequencies are calculated as division number of mutations of the strain and number of coding base pairs.

**Results**

Strain\_1 (marked as”1” in graphs) is the first known strain of this virus, so it is used as a “test sample” in order to compare other strains to it. Therefore, Strain\_1 parameters are either 0 or 100, depending on the information that the graphs are showing.

Strain\_2 is marked as “2”. Strain\_3 is marked as “3”. Strain\_4 is marked as “4”. Strain\_5 is marked as “5”.

Transition/transversion ratio graph:

From #1 - Strain\_1 shows no transition/transversion mutation.   
#2 – Strain\_2 has 5,489 TT ratio.  
#3 – Strain\_3 has 2.45 TT ratio.  
#4 – Strain\_4 has 2.54 TT ratio.  
#5 – Strain\_5 has 5.58 TT ratio.

Gaps/Indels graph shows numbers of indels and gaps, when compared to Strain\_1.

Shown in colors are: blue – number of gaps, red – number insertions, green – number of deletions.   
Strain\_2 has 88 gaps, 51 insertions and 37 deletions. Strain\_3 has 165 gaps, 111 insertions and 54 deletions. Strain\_4 has 148 gaps, 108 insertions and 40 deletions. Strain\_5 has 85 gaps, 50 insertions and 35 deletions.

%Similarity graph shows similarities coding sequences, noncoding sequences and overall sequences of one strain when compared to Strain\_1.

Colors show type of sequences that are compared. Blue shows overall similarity of strains when compared to Strain\_1. Red shows similarities in coding sequences, while green shows similarities in noncoding sequences.

Strain\_2 has 92.23% similarities in overall sequence when compared to Strain\_1, CDS of Strain\_2 is 93.03% similar to CDS of Strain\_1 and nonCDS is 80.91% similar.   
Strain\_3 is 80.26% similar to Strain\_1, with 82.24% similarity in coding sequences and 54.22% similarity in noncoding sequences.   
Strain\_4 is 80.25% similar to Strain\_1. Coding sequences are 82.14% similar, while noncoding sequences are 55.38% similar.

Strain\_5 is 92.35% similar to Strain\_1, with 93.18% similarities in coding sequences and 81.43% similarities in noncoding sequences.

The mutation frequency graph shows the frequencies of overall mutations (blue), mutations in the coding regions of the strains (red) and mutations in the noncoding regions of the strains (green).

As graph shows: mutation frequencies of Strain\_2 are 0.08 for overall strain, 0.07 for coding sequence mutations and 0.19 for noncoding sequence mutations. Strain\_3 has overall mutation ratio of 0.19, where ratio of 0.18 is coding sequence mutation and 0.46 is noncoding sequence mutation ratio. Strain\_4 has a mutation ratio of 0.19, in coding sequences the ratio is 0.18 and in noncoding sequences the ratio is 0.45. Strain\_5 has mutation ratio of 0.08, mutation ratio of coding sequences is 0.07 and for noncoding sequences it is 0.19.

**Discussion**

As seen from the charts, all strains show different values in the given parameters.   
Strain\_5 showed the most similarities with Strain\_1 in all sequences. It has the lowest number of gaps, insertions and deletions - however the mutation frequencies matched with Strain\_2. The most of mutations happened in the noncoding region of the strain and the least happened in the coding region. Strain\_5 also showed the highest rate of TT ratio – from this we can conclude that the TT mutations did not have much effect on the coding region of the RNA of the virus, since the similarity to the Strain\_1 is the highest or that the outcome protein of the virus is very similar to the outcome proteins of the Strain\_1 – possibly showing the most similar symptoms. Strain\_3 seems to be the least similar to Strain\_1 – which means that it had the most mutations. As graph for mutation frequencies show Strain\_3 has the most mutations in the noncoding region, while in the coding regions mutations are significantly higher than the rest of the strains. However, the TT ratio of the Strain\_3 is the lowest. Strain\_3 and Strain\_4 are the most similar to each other. Both show the same mutation frequency in the coding sequences and mutations in the whole sequence and both are very similar in the TT ratio, gaps, indels and deletions. We can conclude that those strains have the most similar alterations of the protein and possibly show similar (if not the same) symptoms. It is important to mention that Strain\_3 and Strain\_4 have slightly different symptoms that Strain\_1 and Strain\_2 and Strain\_5.

**Conclusion**

Viruses mutate fast and it is happening all the time. That is because the virus needs to copy itself in order to survive and sometimes a mistake can occur, which turns out to be advantageous for the virus. If mutations occur many times, it can result as a new strain of the virus. In some cases, the virus can adapt to certain conditions – for example, warmer weather, humid places (geographically different locations) or it can become more resistant to some medical treatments. When the virus adapts it becomes more contagious or even more dangerous (*How Do Viruses Mutate?*, n.d.).

Even though HMPV is relatively new (it has been discovered in 2001) scientists understand it’s mechanism of replication and expression. However, evolutionary studies showed that the human metapneumovirus has been circulating for years before it has been discovered. For now, they established the animal models that will help them to create a possible vaccine and antibodies for the virus. Even so, they still can’t explain the pathogenesis of the different strains and our immunity response to them – that is something that needs to be discovered in the future. (Shafagati & Williams, 2018)

Human metapneumovirus causes the flu mostly among children and elderly. HMPV can sometimes cause asthma flare up and pneumonia, which can be very serious. This is a sign that the humanity needs to be more cautious about spreading this infection in order to save kids and elderly from experiencing uncomfortable symptoms that are caused by any strain of this virus.

**References**

*DNA Mutation and Repair*. (n.d.). Retrieved January 7, 2022, from http://www2.csudh.edu/nsturm/CHEMXL153/DNAMutationRepair.htm

*How do viruses mutate? | COVID Vaccine Facts*. (n.d.). Retrieved January 8, 2022, from https://www.covidvaccinefacts.org/questions/how-do-viruses-mutate

*Human Metapneumovirus Clinical Features—NREVSS | CDC*. (2021, April 14). https://www.cdc.gov/surveillance/nrevss/hmpv/clinical.html

*Human Metapneumovirus: Practice Essentials, Background, Pathophysiology*. (2021). https://emedicine.medscape.com/article/237691-overview

*Metapneumovirus | Intermountain Healthcare*. (n.d.). Intermountainhealthcare.Org. Retrieved January 6, 2022, from https://intermountainhealthcare.org/health-information/germwatch/germ-school/metapneumovirus/

*Metapneumovirus ~ ViralZone*. (n.d.). Retrieved January 6, 2022, from https://viralzone.expasy.org/89?outline=all\_by\_species

Shafagati, N., & Williams, J. (2018). Human metapneumovirus—What we know now. *F1000Research*, *7*. https://doi.org/10.12688/f1000research.12625.1