BIO402 – Molecular Evolution

IUS – FENS – GBE

Hamza Omerović ID:180301001

**Molecular Evolution Project**

**Human Enterovirus D**

**Abstract**

This paper will show Human Enterovirus D with its strains on molecular level. Human Enterovirus D is small virus, specie of Enteroviruses, and also is part of the of Picornavirus.

It is primary respiratory virus which is most of the time infecting childer and teenagers. This type of virus is transmitted direct way, by sneezing, coughing and also by getting in touch with infected surfaces or samples.

Human Enterovirus D brings out five strains, and each of them is observed and analyzed in this paper. Throughout this paper we will presenent informations about similarity, mutations and mutation rate, transition/transversion ratio, Gaps and also informations about insertions and deletions between sequences.

**Introduction**

**Virus species:** Enterovirus D is species of *Enterovirus* which is member of *Picornavirus*

**Virus strains:**

* >NC\_001430.1 Human enterovirus D
* >AY355268.1 Human rhinovirus 87 VP1
* >NC\_038308.1 Human enterovirus 68
* >DQ916376.1 Human enterovirus 94
* >KF040080.1 Enterovirus D120

Enteroviruses are considered to be present and evolved with humanity for thousands of years. One of the Egyptian carving thought to illustrate a priest with weakened and small limb, which represents a basic feature of past polio infection. This is considered to be a one of the oldest records of enterovirus. It was the first eneterovirus discovered, and it was also the most widespread and virus which cause huge morbidity and mortality between all of the enterovirus genotypes. Poliovirus as it is, can result in variety of symptoms. One of the most sirious one is AFP (Accute Flaccid Paralysis), which can lead to lifelong disability and eventually results in death. Throughout the twentieth century, people became more aware of polio. This is largely credited to President Franklin D. Roosevelt, who was paralyzed by polio and was instrumental in establishing the National Foundation for Infantile Paralysis, which launched mass global vaccination campaigns. Later on, enterovirus D and its strains occured as ''polio-like'' illnes with similar symptoms. [1]

Enterovirus D was first time isolated from respiratory samples of pediatric patients in 1962, in California - United States. One of the main contributors to this cognition and establishment of enterovirus D is Albert Sabin. He isolated many of the enterovirus types and established them like main agents of human diseases. [1] Later on as the technology for working with viruses developed, they were futher more explained. Now they are recognized as the constituting one of the genera of  *picornavirus* family. This picornavirus stands for pico – small viruses which have RNA genome. The enterovirus genus includes the human polioviruses, coxsackieviruses, and echoviruses, as well as a number of lower animal enteroviruses such as monkeys, mice etc. [2]

Enterovirus is type of virus that enters the body through the gastrointestinal tract and from there often moves on and aims to attack even the nervous system. They are small viruses which are made out of RNA and protein. Enteroviruses are found in the respiratory secretions (saliva, sputum) and stool of infected patient. People become infected by direct contact with infected persons, secretions from infected person, or even by touching contaminated surfaces or objects. Infections caused by this virus are most likely to occur during the summer and fall. Infected people develop mild upper respiratory symptoms, a flu-like illness with fever, and illness with rash. Sometimes, entering of this virus inside the organism can cause meningitis, it can attack the heart (myocarditis) or the brain (encephalitis). During 2014, in early fall, an outbreak of infection by enterovirus strain D68 occurs in many children across United States. It had a dramatic increase between children that many of them needed hospital-intesive care. Later on by investigatin this enterovirus D and all its strains, infection by them can be a reason for development of type 1 diabetes. Also, by investigation they are considered to be the most dangerous for newborns, who can ends up by death. [3]

**Materials and Methods**

Programs used for this research:

* NCBI (National Center for Biotechnology Information)
* Clustal Omega (multiple sequence alignment program)
* Microsoft Excel

Material which is used is genome of Enterovirus, following strains:

* >NC\_001430.1 Human enterovirus D
* >AY355268.1 Human rhinovirus 87 VP1
* >NC\_038308.1 Human enterovirus 68
* >DQ916376.1 Human enterovirus 94
* >KF040080.1 Enterovirus D120

**Methods:**

* Genomic Sequence Search

1. Go to NCBI, select GENOME database and type your virus of interest;
2. Click on the ''Genome Assembly and Annotation Report'' and then in order to get the sequence click on the replicons (of each virus version);
3. Two options to save the sequence: click FASTA and then copy the sequence, or go to ''Send to'', format FASTA, and create file to download the sequence.

* Protein Coding Gene Search

1. Click on the organism name;
2. Click on the number below ''Protein'' column in ''Replicon info'' table;
3. Download the table of protein genes in .csv file.

* Multiple Sequence Alignment (MSA)

1. Go to Clustal Omega page, select ''DNA'' as a sequence type;
2. Upload your sequences (they have to be ordered hierarchically);
3. Select ''ClustalW'' as output;
4. By clicking on ''More options'' you will set in ''Order'' to be ''Input'' and like that submit it;
5. Then download alignment file (which will be in clustal or txt file).

* Microsoft Excel – used for presenting results.

**Results**

* Coding sequence lenght = 6585
* Non-coding sequence length = 817
* Whole sequence length: = 7402

**Similarity:**

In this figure we can see and observe similarity between five strains, in percentage. Based on the chart above and results given in excel we can conclude following:

* Similarity between strain 2 and strain 1 is 8%; between strain 3 and strain 1 is 72%; between strain 4 and strain 1 is 76%; between strain 5 and strain 1 is 13%.
* CDS\_Similarity between strain 2 and strain 1 is 9%; between strain 3 and strain 1 is 72%; between strain 4 and strain 1 is 75%; between strain 5 and strain 1 is 15%.
* nonCDS\_Similarity between strain 2 and strain 1 is 1%; between strain 3 and strain 1 is 69%; between strain 4 and strain 1 is 80%; between strain 5 and strain 1 is 1%.

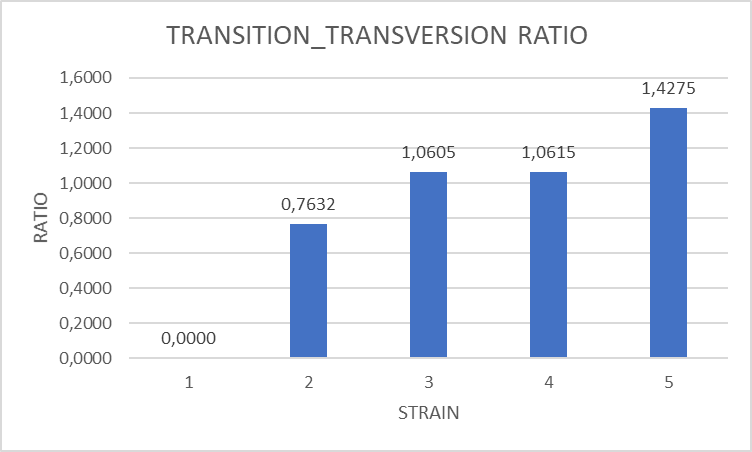
**Mutation:**

* Mutation number between strain 2 and strain 1 is 6789; between strain 3 and strain 1 is 2087, between strain 4 and strain 1 is 1780; between strain 5 and strain 1 is 6432.
* CDS\_Mutation number between strain 2 and strain 1 is 5984, between strain 3 and strain 1 is 1833; between strain 4 and strain 1 is 1617; between strain 5 and strain 1 is 5627.
* nonCDS\_Mutation number between strain 2 and strain 1 is 805; between strain 3 and strain 1 is 254; between strain 4 and strain 1 is 163; between strain 5 and strain 1 is 805.

Based on the length of coding, non-coding and whole sequence length we can get overall mutation frequency.

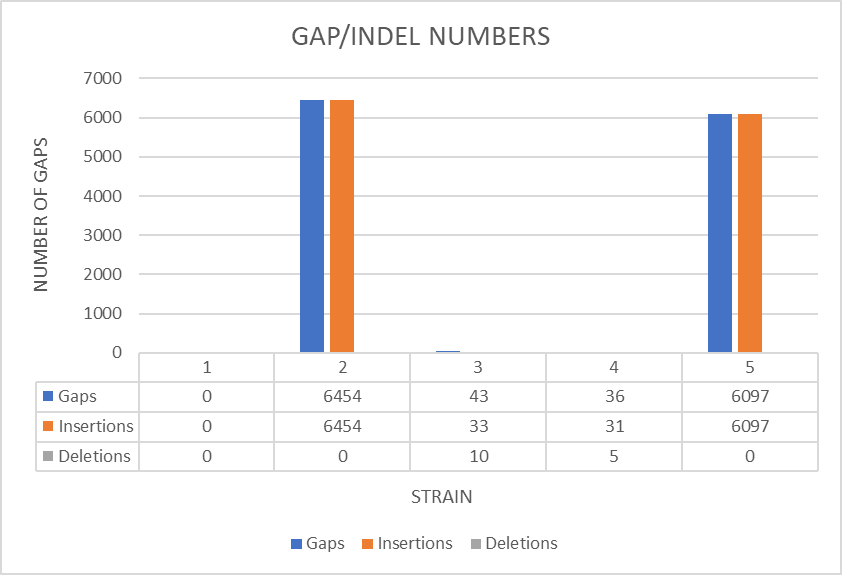
* Mutation frequency between strain 2 and strain 1 (6789/7402) is **0.917**; between strain 3 and strain 1 (2087/7402) is **0.281**; between strain 4 and strain 1 (1780/7402) is **0.240**; between strain 5 and strain 1 (6432/7402) is **0.868**.
* CDS\_Mutation frequency between strain 2 and strain 1 (5984/6585) is **0.908**; between strain 3 and strain 1 (1833/6585) is **0.278**; between strain 4 and strain 1 (1617/6585) is **0.245**; between strain 5 and strain 1 (5627/6585) is **0.854**.
* nonCDS\_Mutation frequency between strain 2 and strain 1 (805/817) is **0.985**; between strain 3 and strain 1 (254/817) is **0.310**; between strain 4 and strain 1 (163/817) is **0.199**; between strain 5 and strain 1 (805/817) is **0.985**.

**T/T ratio:**



Transition\_Transversion Ratio is observed in this graph. T/T ratio between strain 2 and strain 1 is 0,76; T/T ratio between strain 3 and strain 1 is 1,06; T/T ratio between strain 4 and strain 1 is 1,06; T/T ratio between strain 5 and strain 1 is 1,42.

**GAP/INDEL number:**



In this graph we have presented number of gaps, insertions and deletions between five strains. Gap number between strain 2 and strain 1 is 6454, number of insertion is 6454, with 0 deletions. Gap number between strain 3 and strain 1 is 43, number of insertion is 33 with number of deletions of 10. Gap number between strain 4 and strain 1 is 36, number of insertion is 31, with 5 deletions. Gap number between strain 5 and strain 1 is 6097, number of insertion is 6097 with 0 deletions.

**Discussion**

* Similarity

Strain of the virus represents genetically different virus lineage, which can be recognised by mutation, which are possible occuring, between different strains. In order to see how strains biologically differ between each other, human immune system will definitely have certain response to one which tends to make bigger disorder. Two strains would differ between each other if they, after infecting certain cell, starts to make copies inside the cell, because that will certainly undergo many mutations.[4] In this case, the biggest similarity to strain 1 has strain 4 of non-coding sequence which is 80%. Strain 3 shows similarity of 72%, so as the strain 3 in coding sequnence. Strain 4 show 76% of similarity to strain 1, which is close to similarity of strain 4 (75%) of coding sequence. Lowest percent of similarity shows strain 5 of non-coding sequence which is 1% similar to strain 1. Sequence similarity is derived out of empirical relationship between the sequences. Score of similarity, like in this case, serves as a ''tool'' which will give close picture of evolutionary distance occuring between a pair of nucleotide or protein sequences. As common objective for finding sequence similarity is in establishing possible chance that observed sequences comes and evolved from common ancestor. [5]

* Mutation

The highest mutation frequency is found in non-coding sequence of strains 2 and 5 compared to strain 1. Similar mutation rate is between strain 2 and strain 1, and also between coding strain 2 comparing it to non-coding strain 2. Mutations in regions of non-coding sequences of DNA can also cause many diseases. There are regions in non-coding sequence which has certain role in controling gene activity. Also, there are parts of non-coding regions which are together with RNA moleculer plays a role in assembly of proteins.[6] When there is present higher mutation rate, either in conding or non-coding sequences, that will lead new virulent forms to have impact on disease as it is.[7]

* T/T ratio

Estimating this ratio is actually very important. For collection of aligned nucleotide sequences it can be of crucial meaning, because making this ratio can present model of molecular evolution process for those sequences.[8] Strain 3 and 4 have almost same ratio towards strain 1, while strain 5 show the biggest ratio which is 1,42.

* GAP/INDEL numbers

GAP, if occurs in one of the sequences or strains, present deletion of one or more aminoacid residues from certain sequence, which is directly proportional to insertion of it to another sequence. That is the reason why it is calculated together.[9] The biggest gap is in strain 2, 6454 insertions in sequence, and also second biggest is strain 5, with 6097 insertions, both of them with 0 deletions. Strain 3 and 4 are much smaller, with gaps of 43 and 36, insertions 33 and 31, and deletions 10 and 5.

**Conclusion**

This research paper is based on molecular analysis of Enterovirus D and its strains. This paper presented similarity, mutation rates, transition/transversion rates, and also insertion and deletions of sequences. Number of strains which are covered in this paper is 5, and those are following strains: enterovirus D, rhinovirus, enterovirus D68, enterovirus D94 and enterovirus D120. As most common strains, which mostly cause diseases, are enterovirus D68 and rhinovirus. In this paper results which are derived will point out the importance of molecular evolution of certain virus.

Similarity of all strains appears to be the biggest in non-coding sequence, strain 5, and it is 80%. Number of mutation is the smallest in non-coding sequence of strain 4 and number is 163, while the biggest in coding sequence is 5984. The highest mutation frequency occurs, comparing to strain 1, in strain 2 and 5 of non-coding sequence and frequency is 0.985. Gap number of strain 2 which was 6454, was also the biggest gap out of all 5 strains analyzed.

As a small reminder and also mentioning how much important is to know what causes everyday illnes - Enterovirus D68 is a respiratory virus, which can occur as asymptomatic illnes or it can occur as very hard disease. It is type of enterovirus which causes people pain in neck, arms, and back, difficulty in swallowing and difficulty with breathing. It is also spread direct way, or by air. Children and teenagers are most likely to get infected with this kind of enterovirus.[10] Rhinovirus is infection which cause most common typical cold. Also, it can cause ear infections, sore throath or infection of sinuses. It is also type of virus which is transferred direct way.[11]

**References**

[1] H. Cassidy, R. Poelman, M. Knoester, C. C. Van Leer-Buter, and H. G. M. Niesters, “Enterovirus D68 – The New Polio?,” *Front. Microbiol.*, vol. 9, p. 2677, Nov. 2018, doi: 10.3389/fmicb.2018.02677.

[2] J. L. Melnick, “The discovery of the enteroviruses and the classification of poliovirus among them,” *Biol. J. Int. Assoc. Biol. Stand.*, vol. 21, no. 4, pp. 305–309, Dec. 1993, doi: 10.1006/biol.1993.1088.

[3] “Enterovirus Definition: Understand EV-D68 Infection,” *MedicineNet*. https://www.medicinenet.com/enterovirus/definition.htm (accessed Jan. 13, 2022).

[4] C. the science explained-UKRI, “Are there different strains of the SARS-CoV-2 virus circulating?” https://coronavirusexplained.ukri.org/en/article/cad0013/ (accessed Jan. 14, 2022).

[5] J. Heringa, “Sequence Similarity,” 2008. doi: 10.1002/9780470015902.a0005317.pub2.

[6] “Can changes in noncoding DNA affect health and development?: MedlinePlus Genetics.” https://medlineplus.gov/genetics/understanding/mutationsanddisorders/noncondingdnahealth/ (accessed Jan. 14, 2022).

[7] E. K. Wabalo, A. D. Dubiwak, M. W. Senbetu, and T. S. Gizaw, “<p>Effect of Genomic and Amino Acid Sequence Mutation on Virulence and Therapeutic Target of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS COV-2)</p>,” *Infect. Drug Resist.*, vol. 14, pp. 2187–2192, Jun. 2021, doi: 10.2147/IDR.S307374.

[8] A. K. Kristina Strandberg and L. A. Salter, “A comparison of methods for estimating the transition:transversion ratio from DNA sequences,” *Mol. Phylogenet. Evol.*, vol. 32, no. 2, pp. 495–503, Aug. 2004, doi: 10.1016/j.ympev.2004.01.013.

[9] “Sequence Alignment: Scores, Gaps and Gap Penalties.” http://proteinstructures.com/sequence//sequence/sequence-alignment/index.php (accessed Jan. 14, 2022).

[10] “Enterovirus D68 (EV-D68) | CDC,” Aug. 30, 2021. https://www.cdc.gov/non-polio-enterovirus/about/ev-d68.html (accessed Jan. 14, 2022).

[11] “Rhinovirus Infections,” *HealthyChildren.org*. https://www.healthychildren.org/English/health-issues/conditions/ear-nose-throat/Pages/Rhinovirus-Infections.aspx (accessed Jan. 14, 2022).