

ANALYSIS AND COMPARISON OF DIFFERENT HUMAN ORTHOPNEUMOVIRUS STRAINS

Abas Sezer

International University of Sarajevo

BIO402 Molecular Evolution

Prof. Dr. Muhamed Adilović

January, 2022

**ABSTRACT**

There are many virus species known today. As a human beings we are regularly exposed to variety of viruses. They can be harmful for human organism. One of them is Human orthopneumovirus. It’s also known as a Human respiratory syncytial virus or shortly RSV. RSV is capable of causing respiratory tract illnesses, especially in children. In this project different strains of RSV were compared and analyzed to certain reference points such as: similarity, mutations, transition/transversion ratio (TT ratios), gaps, insertions and deletions for overall sequences as well as for coding and non-coding parts of the sequences.

Keywords: viruses, Human orthopneumovirus, coding sequence, non-coding sequence

**TABLE OF CONTENTS**

[**LIST OF FIGURES** 1](#_Toc92559136)

[**LIST OF TABLES** 1](#_Toc92559137)

[**INTRODUCTION** 2](#_Toc92559138)

[**MATERIALS AND METHODS** 3](#_Toc92559139)

[Selection 3](#_Toc92559140)

[Multiple Sequence Alignment 3](#_Toc92559141)

[Protein Coding Gene Search 3](#_Toc92559142)

[Analysis and Classification 3](#_Toc92559143)

[**RESULTS** 4](#_Toc92559144)

[**DISCUSSION** 11](#_Toc92559145)

[**CONCLUSION** 13](#_Toc92559146)

[**REFERENCES** 14](#_Toc92559147)

# LIST OF FIGURES

[Figure 1: Graphical representation of similarity values. 5](#_Toc92558795)

[Figure 2: Graphical representation of mutations. 6](#_Toc92558796)

[Figure 3: Graphical representation of mutation rate 7](#_Toc92558797)

[Figure 4: Graphical representation of TT ratios. 8](#_Toc92558798)

[Figure 5: Graphical representation of overall gaps, insertions and deletions 9](#_Toc92558799)

[Figure 6: Graphical representation of coding sequence (CDS) gaps, insertions and deletions 9](#_Toc92558800)

[Figure 7: Graphical representation non-coding sequence (nonCDS) gaps, insertions and deletions 10](#_Toc92558801)

# LIST OF TABLES

[Table 1: Similarity values 4](#_Toc92558963)

[Table 2: Number of mutations accumulated 5](#_Toc92558964)

[Table 3: Mutation rate values 6](#_Toc92558965)

[Table 4: Transition/transversion ratios (TT ratios). 7](#_Toc92558966)

[Table 5: Number of gaps (Gap), insertions (Ins) and deletions (Del) 8](#_Toc92558967)

# INTRODUCTION

We are regularly exposed to variety of viruses. Some of them are capable to infect us and cause infection and even epidemic or pandemic. All viruses have unique structure. They differ in terms of the disease they cause and organs they attack (1). They also have unique mode of replication and they are capable of producing thousands of virus particles from a single virus (1). Viruses are known as a parasites, since they cannot replicate without host (1). Viral genome contains either sense (+)/antisense (-) single/double stranded DNA or RNA (2). Viruses are classified based on morphology. There are more than 30,000 different virus isolates known today, and they are grouped into more than 3,600 species, 164 genera and 71 families (2). Here, focus will be on analysis and comparison of certain strains of Human orthopneumovirus which has 71 strains in total.

Human respiratory syncytial virus (RSV) is a single-stranded antisense virus with genome length of approximately 15kb (3). RSV is a *Pneumovirinae* family member and member of *Pneumovirus* genus (3). It was firstly discovered and isolated in 1956 (3). Reports suggest that RSV is one of the most common respiratory viruses in children (4). Approximately 70% of hospitalized children for respiratory tract illness are affected by RSV (5). Up to date, there is no effective treatment nor vaccine available for RSV (3). Although there are few RSV vaccine candidates, but they are in clinical stage of development (3).

# MATERIALS AND METHODS

## Selection

Genome database of National Center for Biotechnology Information (NCBI) was used for virus and strain selection. Human orthopneumovirus genome assembly and annotation report was analyzed. Five strains were selected for further analysis and comparison. FASTA sequences were obtained for selected strains. The oldest one (reference strain), in further text strain 1, was released in December 1991 (GenBank: M74568.1), strain 2 was released in November 1997 (GenBank: M74568.1), strain 3 (GenBank: KY684758.1), strain 4 (GenBank: KY674983.1) and strain 5 (GenBank: KY674984.1) were released in March 2017.

## Multiple Sequence Alignment

Clustal Omega tool was used for multiple sequence alignment. Obtained FASTA sequences were uploaded for DNA sequence alignment. Multiple sequence alignment file for selected strains was saved for further analysis.

## Protein Coding Gene Search

Selected strains were analyzed from genome assembly and annotation report. Protein coding genes were found through replication information. Then script was used on tables of protein coding genes and obtained results were analyzed.

## Analysis and Classification

Results were classified in excel, with all the information obtained from script. Tables and graphs were made based on excel data.

# RESULTS

In total five strains of the Human orthopneumovirus were compared. It is important to mention that strains were not compared between each other. The reference strain was the oldest one. Therefore, four strains were compared to the oldest one. Length of multiple sequence alignment was 15308 bp, from which 13609 bp was coding sequence and 1699 bp was non coding sequence. Four strains were compared to certain reference points such as: similarity, mutations, transition/transversion ratio (TT ratios), gaps, insertions and deletions for overall sequences as well as for coding and non-coding parts of the sequences.

Strain 2 was 80.73% similar to strain 1, strain 3 was 79.75% similar to strain 1, strain 4 was 80.18% similar to strain 1 and strain 5 was 79.34% similar to strain 1. Regarding the comparison of coding sequence similarity, strain 2 was 83.16% similar to strain 1, strain 3 was 82.93% similar to strain 1, strain 4 was 83.15% similar to strain 1 and strain 5 was 82.90% similar to strain 1. Regarding non-coding sequence similarity, strain 2 was 61.27% similar to strain 1, strain 3 was 54.32% similar to strain 1, strain 4 was 56.32% similar to strain 1 and strain 5 was 50.85% similar to strain 1.

Table 1: Similarity values for strain 2, 3, 4 and 5 compared to strain 1. They were collected based on overall similarity, coding sequence similarity (CDS) and non-coding sequence similarity (non-CDS).

|  |  |  |  |
| --- | --- | --- | --- |
| Similarity | | | |
| No. | Overall | CDS | nonCDS |
| Strain 2 | 80.73% | 83.16% | 61.27% |
| Strain 3 | 79.75% | 82.93% | 54.32% |
| Strain 4 | 80.18% | 83.15% | 56.32% |
| Strain 5 | 79.34% | 82.90% | 50.85% |

Figure 1: Graphical representation of similarity values showed in Table 1.

When it comes to the mutations, strain 2 accumulated 2949 mutations, strain 3 accumulated 3099 mutations, strain 4 accumulated 3034 mutations and strain 5 accumulated 3162 mutations compared to strain 1. Regarding the comparison of coding sequence mutations, strain 2 accumulated 2291 mutations, strain 3 accumulated 2323 mutations, strain 4 accumulated 2292 mutations and strain 5 accumulated 2327 mutations compared to strain 1. Regarding the comparison of non-coding sequence mutations, strain 2 accumulated 658 mutations, strain 3 accumulated 776 mutations, strain 4 accumulated 742 mutations and strain 5 accumulated 835 mutations compared to strain 1. Overall mutation rate for strain 2 was 0.192 mutations per nucleotide, coding sequence mutation rate for strain 2 was 0.168 mutations per nucleotide and non-coding sequence mutation rate for strain 2 was 0.387 mutations per nucleotide. Overall mutation rate for strain 3 was 0.202 mutations per nucleotide, coding sequence mutation rate for strain 3 was 0.170 mutations per nucleotide and non-coding sequence mutation rate for strain 3 was 0.456 mutations per nucleotide. Overall mutation rate for strain 4 was 0.198 mutations per nucleotide, coding sequence mutation rate for strain 4 was 0.168 mutations per nucleotide and non-coding sequence mutation rate for strain 4 was 0.436 mutations per nucleotide. Overall mutation rate for strain 5 was 0.206 mutations per nucleotide, coding sequence mutation rate for strain 5 was 0.170 mutations per nucleotide and non-coding sequence mutation rate for strain 5 was 0.491 mutations per nucleotide.

Table 2: Number of mutations accumulated for strain 2, 3, 4 and 5 compared to strain 1 for whole sequence, coding sequence (CDS) and non-coding sequence (non-CDS).

|  |  |  |  |
| --- | --- | --- | --- |
| Mutations | | | |
| No. | Overall | CDS | nonCDS |
| Strain 2 | 2949 | 2291 | 658 |
| Strain 3 | 3099 | 2323 | 776 |
| Strain 4 | 3034 | 2292 | 742 |
| Strain 5 | 3162 | 2327 | 835 |

Figure 2: Graphical representation of mutations showed in Table 2.

Table 3: Mutation rate values for strain 2, 3, 4 and 5 compared to strain 1 for whole sequence, coding sequence (CDS) and non-coding sequence (nonCDS).

|  |  |  |  |
| --- | --- | --- | --- |
| Mutation rate | | | |
| No. | Overall | CDS | nonCDS |
| strain 2 | 0.192 | 0.168 | 0.387 |
| strain 3 | 0.202 | 0.17 | 0.456 |
| strain 4 | 0.198 | 0.168 | 0.436 |
| strain 5 | 0.206 | 0.17 | 0.491 |

Figure 3: Graphical representation of mutation rate showed in Table 3.

Overall TT ratio for strain 2 was 1.92, for strain 3 was 1.90, for strain 4 was 1.94 and for strain 5 was 1.91 compared to the strain 1. TT ratio for coding sequence for strain 2 was 2.22, for strain 3 was 2.16, for strain 4 was 2.22 and for strain 5 was 2.19 compared to strain 1. TT ratio for non-coding sequence for strain 2 was 1.16, for strain 3 was 1.18, for strain 4 was 1.21 and for strain 5 was 1.12.

Table 4: Transition/transversion ratios (TT ratios) for strain 2, 3, 4 and 5 compared to strain 1 for whole sequence, coding sequence (CDS) and non-coding sequence (nonCDS).

|  |  |  |  |
| --- | --- | --- | --- |
| TT ratios | | | |
| No. | Overall | CDS | nonCDS |
| Strain 2 | 1.9269777 | 2.2263083 | 1.1684588 |
| Strain 3 | 1.9039039 | 2.169863 | 1.1821561 |
| Strain 4 | 1.9420142 | 2.2231638 | 1.2181818 |
| Strain 5 | 1.9163265 | 2.1928375 | 1.1259843 |

Figure 4: Graphical representation of TT ratios showed in Table 4.

When it comes to the gaps, insertions and deletions in whole sequences, strain 2 accumulated 63 gaps, 30 insertions and 33 deletions, strain 3 accumulated 198 gaps, 112 insertions and 86 deletions, strain 4 accumulated 142 gaps, 109 insertions and 33 deletions and strain 5 accumulated 304 gaps, 226 insertions and 78 deletions compared to strain 1. Regarding the comparison of gaps, insertions and deletions in coding sequences, strain 2 accumulated 10 gaps, 10 insertions, strain 3 accumulated 9 gaps, 9 insertions, strain 4 accumulated 10 gaps, 10 insertions and strain 5 accumulated 9 gaps and 9 insertions compared to strain 1. No deletions were observed in coding sequence. Regarding the comparison of gaps, insertions and deletions in non-coding sequences, strain 2 accumulated 53 gaps, 20 insertions and 33 deletions, strain 3 accumulated 189 gaps, 103 insertions and 86 deletions, strain 4 accumulated 132 gaps, 99 insertions and 33 deletions and strain 5 accumulated 295 gaps, 217 insertions and 78 deletions compared to strain 1.

Table 5: Number of gaps (Gap), insertions (Ins) and deletions (Del) in strain 2, 3, 4 and 5 compared to strain 1 for overall sequence, coding sequence (CDS) and non-coding sequence (nonCDS).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Overall | | | CDS | | | nonCDS | | |
| No. | Gap | Ins | Del | Gap | Ins | Del | Gap | Ins | Del |
| Strain 2 | 63 | 30 | 33 | 10 | 10 | 0 | 53 | 20 | 33 |
| Strain 3 | 198 | 112 | 86 | 9 | 9 | 0 | 189 | 103 | 86 |
| Strain 4 | 142 | 109 | 33 | 10 | 10 | 0 | 132 | 99 | 33 |
| Strain 5 | 304 | 226 | 78 | 9 | 9 | 0 | 295 | 217 | 78 |

Figure 5: Graphical representation of overall gaps, insertions and deletions showed in Table 5.

Figure 6: Graphical representation of coding sequence (CDS) gaps, insertions and deletions showed in Table 5.

Figure 7: Graphical representation non-coding sequence (nonCDS) gaps, insertions and deletions showed in Table 5.

# DISCUSSION

Based on the obtained results coding sequence similarity was highest compared to overall and non-coding sequence similarity for all selected strains (strain 2: 83.16%; strain 3: 82.93%; strain 4: 83.15%; strain 5: 82.90%). On the other hand, non-coding sequence similarity was lowest similarity compared to overall and coding sequence similarity (strain 2: 61.27%; strain 3: 54.32%; strain 4: 56.32%; strain 5: 50.85%). However, this is expected since we know that coding sequences in general are more resistant to certain changes over time and non-coding sequences are more prone to changes. Moreover, strains differ much more in non-coding sequences. The reason for that is because non-coding sequence is much shorter than coding sequence and certain change will not affect the percentage on the same way.

Based on the results, all strains accumulated mutations. Significant number of overall mutations was observed in all strains, but interestingly in strain 2 accumulated much more mutations over short period of time, since the difference in released date between strain 1 and 2 was 6 years. Strains 3, 4 and 5 accumulated slightly more mutations but over a period of 26 years. It’s known that RNA viruses usually show higher mutation rate than DNA viruses because of the error rate of enzymes responsible for RNA replication and it could be the reason for high number of mutations (2). Furthermore, coding sequence mutation number was higher than non-coding sequence mutation number but the rate of mutation was two to three times larger in non-coding sequences compared to mutation rate of coding ones. This result confirms previously mentioned statement about resistance of coding sequences and tendency of non-coding sequences to mutate.

When it comes to the transition/transversion ratio (TT ratio), obtained results showed that transitions are more common than transversions. Interestingly, TT ratio in coding sequences for all strains was above 2, and TT ratio in non-coding sequences for all strains was slightly higher than 1. It’s possible that sequence length plays an important role here, since transitions and transversions usually arise from point mutations. Shorter sequence means smaller chance for point mutation and vice versa.

Regarding the results of gaps, insertions and deletions in all strains, results showed that strains 3, 4 and 5 have much more gaps, insertions and deletions than strain 2 compared to strain 1. However, huge portion of all gaps, insertions and deletions are located in non-coding sequences, and very few in coding sequences. It’s important to mention that no deletions were observed in coding sequences, and the number of gaps and insertions match with each other. This is expected since gaps are equal to all insertions and deletions in sequence.

Viruses are genetically diverse and they are able to evolve into different strains and species (6). It is possible that viruses need evolution in order to cause infection and even epidemic/pandemic(6). If it’s true, most of them were harmless for humans at one point. But over a certain period of time they evolve. Basically they evolve and form new strains in order to survive in host without being spotted or destroyed. Moreover, with the evolution they are improving their “attacking” mechanism so by new strains they are becoming more dangerous. Regarding the evolution, there is a term called “pathogen pyramid” which is known as a framework for thinking about evolution of viruses (6). Based on pathogen pyramid there are four levels of interactions between pathogens and humans (6). First level is called exposure level which is the way of acquiring the pathogen, in this case virus (water, blood, food, salvia) (6). Second level is infection level, in which virus is capable to cause an infection and overcome the defense mechanism of body directly affecting both, human (host) and virus on molecular level (6). Third level is transmission, in which certain virus can be transmitted from one human to another (6). Fourth level is epidemic and it’s defined as a level in which virus can case major outbreaks (6).

# CONCLUSION

Organisms accumulate changes in their genome over time. The same goes for viruses. This analysis proved that viruses are capable of changing and evolving. They are trying to find most efficient way to survive and reproduce, but at the same time protecting most important parts of their genome, coding part. Their effort to mutate is mostly based on mutation in non-coding regions of their genome, since those parts are not at great importance, at least not in our understanding. As a result of evolution we can expect that in future more aggressive strains of viruses will emerge, and that they will be more harmful for humans as well as for other living organisms.

# REFERENCES

1. Taylor MW. What Is a Virus? In: Viruses and Man: A History of Interactions [Internet]. Cham: Springer International Publishing; 2014 [cited 2022 Jan 6]. p. 23–40. Available from: http://link.springer.com/10.1007/978-3-319-07758-1\_2

2. Gelderblom HR. Structure and Classification of Viruses. In: Baron S, editor. Medical Microbiology [Internet]. 4th ed. Galveston (TX): University of Texas Medical Branch at Galveston; 1996 [cited 2022 Jan 6]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK8174/

3. Jorquera PA, Anderson L, Tripp RA. Human Respiratory Syncytial Virus: An Introduction. In: Tripp RA, Jorquera PA, editors. Human Respiratory Syncytial Virus [Internet]. New York, NY: Springer New York; 2016 [cited 2022 Jan 6]. p. 1–12. (Methods in Molecular Biology; vol. 1442). Available from: http://link.springer.com/10.1007/978-1-4939-3687-8\_1

4. Øymar K, Skjerven HO, Mikalsen IB. Acute bronchiolitis in infants, a review. Scand J Trauma Resusc Emerg Med. 2014 Dec;22(1):23.

5. Oshansky CM, Zhang W, Moore E, Tripp RA. The host response and molecular pathogenesis associated with respiratory syncytial virus infection. Future Microbiol. 2009 Apr;4(3):279–97.

6. Woolhouse M, Scott F, Hudson Z, Howey R, Chase-Topping M. Human viruses: discovery and emergence. Philos Trans R Soc B Biol Sci. 2012 Oct 19;367(1604):2864–71.