Aiša Trebo  
IUS/GBE  
BIO 402/Molecular evolution

**Molecular Evolution Project**

**Human orthopneumovirus**

**Abstract**

In 1956, the human respiratory syncytial virus (hRSV, often known as RSV) was discovered in chimpanzees and later in newborns with severe lower respiratory tract illness. It is a non-segmented single-stranded negative-sense enveloped RNA virus belonging to the Paramyxoviridae family, genus Pneumovirus, subfamily Pneumovirinae. The RSV virus has a single serotype with two antigenic subgroups, A and B. Both subtypes of strains co-circulate often, however one of the subtypes predominates in most cases. Research was done on five different strains of virus using three different platforms (NCBI, Clustal Omega, and Microsoft Excel). Research showed similarity percentage, ratio of transition and trasnversion, gap/indel number and mutation frequency among five strains.

**Introduction**

Human respiratory syncytial virus (hRSV, here RSV) was first isolated from chimpanzees in 1956 and was subsequently recovered from infants with severe lower respiratory tract disease. [[1]](#footnote-1)

It is a non-segmented negative-sense single-stranded enveloped RNA virus that belongs to the family of Paramyxoviridae, genus Pneumovirus, subfamily Pneumovirinae. Its 10 genes encode 11 proteins since two overlapping open reading frames in the M2 mRNA yield two distinct matrix proteins, M2-1 and M2-2. The viral envelope contains three proteins, the G glycoprotein, the fusion (F) glycoprotein, and the small hydrophobic (SH) protein. The G protein functions in host cell attachment and the F protein is responsible for fusion and cell entry, whereas the SH protein is not required in either of these processes. The RSV virus comprises five other structural proteins, the large (L) protein, nucleocapsid (N), phosphoprotein (P), matrix (M), and M2-1, and two non-structural proteins (NS1 and NS2). There is a single RSV serotype with two major antigenic subgroups, A and B. Strains of both subtypes often co-circulate, but generally one of the subtypes predominates. [[2]](#footnote-2)

In temperate regions, RSV infections show a distinct seasonality with onset in late fall or early winter, a peak between mid-December and early February, and season offset in late spring. Some areas, in particular in northern Europe, report yearly alternations between an early large outbreak and a late small outbreak. In tropical regions, the patterns are less predictable and can include two yearly peaks in spring and fall or fairly constant infection rates throughout the year.[[3]](#footnote-3)

Primary infection with RSV is believed to be almost always symptomatic, although there are data suggesting that this may not actually be the case. The clinical manifestations range from mild upper respiratory tract illness (URTI) or otitis media to severe and potentially life-threatening lower respiratory tract involvement (LRTI).[[4]](#footnote-4)

**Materials and Methods**

* NCBI (National Center for Biotechnology Information)[[5]](#footnote-5)
* Clustal Omega (multiple sequence alignment program)[[6]](#footnote-6)
* Microsoft Excel
* Human orthopneumovirus [[7]](#footnote-7)

*Genomic Sequence Search -* go to “NCBI” website, select “Genome” database and enter your query. After entering the query, select “Genome Assembly and Annotation report” and then select “replicons”. To get your sequence, you can choose either to click on “FASTA” and copy the sequence, or click on “Send to – Format FASTA – Create File” to download the sequence.

*Multiple Sequence Alignment –* go to “Clustal Omega - [www.ebi.ac.uk](http://www.ebi.ac.uk)”, select “DNA” as the type of the sequence and copy your sequences that were previously ordered hierarchically. As an Output Format click “ClustalW”, click “More options – order – input”, click “Submit” and “Download Alignment file” (if not downloaded directly, you can download it by right click – save as).

*Microsoft Excel -* interpretation of the results using the “Microsoft Excel”.

**Results**

During this research, five different strains of Human orthopneumovirus were analyzed. Strains include:

1. Human respiratory syncytial virus B - MF001045.1 (Human respiratory syncytial virus B strain 6C3)[[8]](#footnote-8)
2. Human respiratory syncytial virus B - MF973155.1 (Human respiratory syncytial virus B strain 5T9, complete genome)[[9]](#footnote-9)
3. Human respiratory syncytial virus B - MG642059.1 (Human respiratory syncytial virus B strain RSVB/Homo sapiens/USA/MCRSV\_267/1983, complete genome)[[10]](#footnote-10)
4. Human respiratory syncytial virus B - KY924878.1 (strain RSVB/BCH-Y/2016, complete genome)[[11]](#footnote-11)
5. Human respiratory syncytial virus B isolate RSVB/England198/2012 - KY249672.1 (Human respiratory syncytial virus B strain 6A8, complete genome)[[12]](#footnote-12)

Figure 1 – Similarity between strains

After obtaining whole genomes sequences of each strain of virus from NCBI using format FASTA, multiple sequence alignment (using Clustal Omega) was done in order to generate alignments between sequences.

The first figure gives us an explanation about percentage of similarity between five strains. According to the chart, strain 2 is 98% similar to strain 1, strain 3 is 98% similar to strain 1, strain 4 is 98%, and strain 5 is 95% similar to strain 1.

Coding region similarity between strain 2 and strain 1 is 99%, between strain 3 and strain 1 is 98%, strain 4 and strain 1 99%, and for strain 5 and strain 1 results in 96%.

Regarding the non-coding region similarity, strain 2 is 93% similar to strain 1, strain 3 is 97% similar to strain 1, 93% of similarity is between strain 4 and strain 1, and 89% of similarity is found between strain 5 and strain 1.

Figure 2 – Transition/Transversion Ratio

Transition/Transversion ratio (T/T ratio) between sequence 2 and sequence 1 is 8,22. Sequence 3 and sequence 1 have T/T ratio of 4,90. Sequence 4 and sequence 1 have T/T ratio of 8 22, while sequence 5 and sequence 1 result in 4,79 T/T ratio (Fig.2).

Figure 3 – Gap / Indel numbers

Gap number between strain 2 and strain 1 is 82, insertion is 84 and deletions is 1. Between strain 3 and strain 1 gap number is 7, insertion number is 6 and deletion number is 1; strain 4 and 1 have gap number 85, insertion number 84 and deletion number 1. Strain 5 and strain 1 result in 128 number of gaps, 111 number of insertions and 17 number of deletions. (Fig.3)

Figure 4 – Mutation frequency

The fourth figure represents frequency of mutations per nucleotide. Mutation frequency between strain 2 and strain 1 is 0,016. Between strain 3 and strain 1 is 0,020. Strain 4 and strain 1 have 0,016. And between strain 5 and strain 1 mutation frequency is 0,04.

Coding region mutation frequency between strain 2 and 1 equals 0,009, between strain 3 and strain 1 equals 0,018. Strain 4 and strain 1 have CDS mutation frequency of 0,009. CDS mutation frequency of 0,038 is found between strain 5 and strain 1.

Regarding the non-coding region mutation frequency between strain 2 is 0,070 to strain 1, strain 3 has nonCDS mutation frequency of 0,032 comparing to strain 1, 0,070 frequency is between strain 4 and strain 1, and 0,109 frequency is found between strain 5 and strain 1.

**Discussion**

The similarity of 98% is found between strains 2,3 and 4, comparing to strain 1, while strain 5 shares 95% of similarity to strain 1. In coding High percentage of similarity shows the relatedness of sequences under the study. In coding regions, strain 2 and 4 show 99% of similarity to strain 1, strain 3 shows 98% and strain 5 shows 96% of similarity. In non-coding region high similarity is found between strain 3 resulting in 97%, while strain 2 and 4 share same percentage of similarity to strain 1, and 89% is found between strain 5 and 1.

If the two sequences share significant similarity, it is extremely unlikely that the extensive similarity between the two sequences has been acquired randomly, meaning that the two sequences must have derived from a common evolutionary origin. [[13]](#footnote-13) High similarity between sequences can be a proof that sequences are connected through evolutionary changes arising from same ancestry.

T/T ratio is same in strains 2 and 4 comparing to strain 1, while strain 3 and 5 have approximately same ratio to strain 1. Transitions (substitutions between purines and purines or between pyrimidines and pyrimidines) occur more frequently than transversions (substitutions between purines and pyrimidines). [[14]](#footnote-14) Evaluating the transitions/transversions ratio of aligned sequences is significant for better understanding of process of molecular evolution and can applied to further description the evolutionary process. [[15]](#footnote-15)

Gap/indel number between strain 2 and 4, comparing to strain 2 is same. Strain 3 has the lowest number of gap/indel number compared to strain 1, while strain 5 has the highest gap/indel number. A gap in one of the sequences simply means that one or more amino acid residues have been deleted from the sequence, or we could also say that there is an insertion in the second sequence. [[16]](#footnote-16)

Strain 5 has the highest mutation frequency compared to strain 1 (including CDS and nonCDS mutation frequency). Strain 2 and 4 share approximately same mutation frequency with strain 1 (including CDS and nonCDS mutation frequency as well). Higher nonCDS mutation frequency is observed than CDS mutation frequency.

**Conclusion**

Human respiratory syncytial virus (hRSV, here RSV) was first isolated from chimpanzees in 1956 and was subsequently recovered from infants with severe lower respiratory tract disease. It is a non-segmented negative-sense single-stranded enveloped RNA virus that belongs to the family of Paramyxoviridae, genus Pneumovirus, subfamily Pneumovirinae. In temperate regions, RSV infections show a distinct seasonality with onset in late fall or early winter, a peak between mid-December and early February, and season offset in late spring. [[17]](#footnote-17)

Study was done among five strains of Human orthopneumovirus using NCBI, Clustal Omega and Microsoft Excel.

Results of the study show similarity percentage between five strains, including CDS and nonCDS similarity. Highest similarity (98%) is between strains 2,3 and 4, comparing to strain 1. High similarity between sequences can be a proof that sequences are connected through evolutionary changes arising from same ancestry. Calculating the ratio between transition and transversion shows importance in understanding of process of molecular evolution and can be applied in further description. Transition/Transversion ratio is same in strains 2 and 4 comparing to strain 1, while strain 3 and 5 have approximately same ratio to strain 1. . A gap in one of the sequences means that one or more amino acid residues have been deleted from the sequence, or that there is an insertion in the second sequence. Strain 2 and 4 share same number compared to strain 1. Mutation frequency is highly observed in strain 5 compared to strain 1.

**References**

Borchers, A. T., Chang, C., Gershwin, M. E., & Gershwin, L. J. (2013). Respiratory syncytial virus—A comprehensive review. *Clinical Reviews in Allergy & Immunology*, *45*(3), 331–379. https://doi.org/10.1007/s12016-013-8368-9

*Clustal Omega < Multiple Sequence Alignment < EMBL-EBI*. (n.d.). Retrieved January 9, 2022, from https://www.ebi.ac.uk/Tools/msa/clustalo/

*Genome List—Genome—NCBI*. (n.d.). Retrieved January 9, 2022, from https://www.ncbi.nlm.nih.gov/genome/browse/?fbclid=IwAR2dLXWgb2LDYtozCEUBV3Jbukt-h5ImXORTP3r6\_UoyAT-P-42kMme8iDY#!/viruses/5144/

*Human respiratory syncytial virus B isolate RSVB/England198/2012, complete genome* (1133566796). (2017). [Data set]. NCBI Nucleotide Database. http://www.ncbi.nlm.nih.gov/nuccore/KY249672.1

*Human respiratory syncytial virus B strain 5T9, complete genome* (1276729273). (2018). [Data set]. NCBI Nucleotide Database. http://www.ncbi.nlm.nih.gov/nuccore/MF973155.1

*Human respiratory syncytial virus B strain 6C3, complete genome* (1196646692). (2018). [Data set]. NCBI Nucleotide Database. http://www.ncbi.nlm.nih.gov/nuccore/MF001045.1

*Human respiratory syncytial virus B strain RSVB/BCH-Y/2016, complete genome* (1214784559). (2017). [Data set]. NCBI Nucleotide Database. http://www.ncbi.nlm.nih.gov/nuccore/KY924878.1

*Human respiratory syncytial virus B strain RSVB/Homo sapiens/USA/MCRSV\_267/1983, complete genome* (1301645931). (2017). [Data set]. NCBI Nucleotide Database. http://www.ncbi.nlm.nih.gov/nuccore/MG642059.1

Information, N. C. for B., Pike, U. S. N. L. of M. 8600 R., MD, B., & Usa, 20894. (n.d.). *National Center for Biotechnology Information*. Retrieved January 9, 2022, from https://www.ncbi.nlm.nih.gov/

Kristina Strandberg, A. K., & Salter, L. A. (2004). A comparison of methods for estimating the transition:transversion ratio from DNA sequences. *Molecular Phylogenetics and Evolution*, *32*(2), 495–503. https://doi.org/10.1016/j.ympev.2004.01.013

*Sequence Alignment: Scores, Gaps and Gap Penalties*. (n.d.). Retrieved January 9, 2022, from http://proteinstructures.com/sequence//sequence/sequence-alignment/index.php

Xiong, J. (2006). *Essential Bioinformatics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511806087

1. (Borchers et al., 2013) [↑](#footnote-ref-1)
2. (Borchers et al., 2013) [↑](#footnote-ref-2)
3. (Borchers et al., 2013) [↑](#footnote-ref-3)
4. (Borchers et al., 2013) [↑](#footnote-ref-4)
5. (Information et al., n.d.) [↑](#footnote-ref-5)
6. (*Clustal Omega < Multiple Sequence Alignment < EMBL-EBI*, n.d.) [↑](#footnote-ref-6)
7. (*Genome List - Genome - NCBI*, n.d.) [↑](#footnote-ref-7)
8. (*Human Respiratory Syncytial Virus B Strain 6C3, Complete Genome*, 2018) [↑](#footnote-ref-8)
9. (*Human Respiratory Syncytial Virus B Strain 5T9, Complete Genome*, 2018, p. 9) [↑](#footnote-ref-9)
10. (*Human Respiratory Syncytial Virus B Strain RSVB/Homo Sapiens/USA/MCRSV\_267/1983, Complete Genome*, 2017) [↑](#footnote-ref-10)
11. (*Human Respiratory Syncytial Virus B Strain RSVB/BCH-Y/2016, Complete Genome*, 2017) [↑](#footnote-ref-11)
12. (*Human Respiratory Syncytial Virus B Isolate RSVB/England198/2012, Complete Genome*, 2017) [↑](#footnote-ref-12)
13. (Xiong, 2006) [↑](#footnote-ref-13)
14. (Xiong, 2006) [↑](#footnote-ref-14)
15. (Kristina Strandberg & Salter, 2004) [↑](#footnote-ref-15)
16. (*Sequence Alignment*, n.d.) [↑](#footnote-ref-16)
17. (Borchers et al., 2013) [↑](#footnote-ref-17)