**Rhinovirus**

Rhinoviruses were discovered in the 1950s during efforts to identify what cause commmon cold. Even nowadays scientists are searching for cure. HRV is common cause of respiratory tract infection. For the detection PCR based assays are commonly used and HRV is connected with other respiratory diseases like asthma and pneumonia. Rhinovirus is single stranded RNA virus(family of Enteroviruses, Picornaviridae), approximately 7200 bp long. Scientists identified more than 100 strains of rhinoviruses which are divided into two main groups, A and b. Rhinovirus is transmitted mosty directly( contact) and aerosol. Gathering the cultures of rhinoviruses is obtained by taking the sample from the nasopharynx. Rhinovirus is not destroying ephithelial cells like other respiratory viruses(influenza). Rhinovirus is geographically distributed all around the world and it follows „season“ pattern but it is always present and cirulates. Average incubation period is 2 days and symptom period is from 7 up to 14 days. Common symptoms include rhinorrhea, nasal congestion, sore throat, cough, headache, subjective fevers, and malaise. Production of vaccines for rhinovirus is big challenge for scientists because of the presence more than 100 strains. Our understanding of HRV etiology based on in vitro data and in vivo studies of experimental infections, primarily in healthy adults, means both direct viral effects and tissue damage due to the host`s immune response. In addition, whole-genome sequencing can provide insight into the differences observed in clinical symptoms and outcomes depending on the HRV strain. A better understanding of the mechanisms leading to the symptoms of HRV infection and the role of the host's immune response is needed to guide future efforts to prevent and treat HRV.

**Clustal Omega**

**MATERIALS AND METHODS**

**Selection**

Genome database of National Center for Biotechnology Information (NCBI) was used for virus and strain selection. Rhinovirus genome assembly and annotation report was analyzed. Five strains were selected for further analysis and comparison. FASTA sequences were obtained for selected strains.

**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**

Five strains of rhinovirus were compared. They are not compared between each other, the oldest strain was reference point. Length of strains were between 7,03 and 7,13 kb. 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.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Similarity |  |  |
| NO | OVERALL | CDS | nonCDS |
| STRAIN 2 | 72% | 73% | 66% |
| STRAIN 3 | 73% | 74% | 67% |
| STRAIN 4 | 74% | 75% | 72% |
| STRAIN 5 | 74% | 74% | 68% |

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).

**Mutations**

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 | 1966 | 1705 | 261 |
| STRAIN 3 | 1913 | 1663 | 250 |
| STRAIN 4 | 1806 | 1598 | 208 |
| STRAIN 5 | 1864 | 1624 | 240 |

Table 3: 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 RATIO |  |  |
| NO. | OVERALL | CDS | nonCDS |
| STRAIN 2 | 0.936975 | 0.940707 | 0.893333 |
| STRAIN 3 | 1.04878 | 1.018405 | 1.586957 |
| STRAIN 4 | 1.044957 | 1.026718 | 1.432432 |
| STRAIN 5 | 0.994246 | 0.98401 | 1.142857 |

Table 4: 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 |
| STRAIN2 | 122 | 9 | 113 | 3 | 3 | 0 | 119 | 6 | 113 |
| STRAIN3 | 149 | 24 | 125 | 18 | 18 | 0 | 131 | 6 | 125 |
| STRAIN4 | 123 | 7 | 116 | 5 | 5 | 0 | 118 | 2 | 116 |
| STRAIN5 | 131 | 12 | 119 | 11 | 11 | 0 | 120 | 1 | 119 |

**Discussion:**

What we intend to do is to characterize and understand the relationship between strains, to identify a potent strain that is highly and ideally similar to other strains, and to analyze all the proteins of all strains to identify the highly conserved amino acid sequences. This can be done by Clustal Omega W software. The findings suggest that these four strains are highly and ideally similar among themselves, as the similarity column indicates in the table below. This means that retrieved scores satisfied the minimum threshold, which is in this case above 70%. Everything above 70% is consiedered a sufficing similarity score, which is said to be the minimum pairwise similarity. Furthermore, over 1800 single mutations have been detected in four strains indicating the increase of infectivity. Regarding the Gap column values, we can say that its very much correlated with insertions/deletions due to the fact that algorithms try to optimize an objective function by minimizing mismatches and gaps. As a matter of fact, gaps can be inserted during the alignment process if deletion or insertion sites are detected by the algorithm, so that sites with identical nucleotides align together. Usually, fewer gap insertions give better alignment accuracy, which we tend to observe. Variations can be observed in final alignments obtained from Clustal W. The numbers vary from 122 to 149 gaps per sequence, which can further lead to differences in the following steps of constructing the phylogenetic tree. While most protein sequences are shorter than 13,000 residues, large numbers of difficult to align sequences may incorporate many gaps, thereby inflating the total alignment size to well beyond the length of the longest unaligned sequence. It is necessary to allow pre-aligned sequences (sequence profiles) to be used as input. To recognize that the input sequences are already aligned, Clustal Omega requires that all sequences have the same length and that at least one sequence contains one gap. However, this failed to recognize valid alignments that did not contain any gaps.

**Conclusion**

All organisms are changing and evolving over time, some of them in small scale some of them in large scale. Viruses are „masters“ of evolution. They are capable to survive and change in heavy conditions. In the future, beacuse of this abilities of viruses, we can expect new pandemics which can cause large problems for the humans.

**REFERENCES:**

1. Jacobs, S. E., Lamson, D. M., St. George, K., & Walsh, T. J. (2013). Human Rhinoviruses. Clinical Microbiology Reviews, 26(1), 135–162. <https://doi.org/10.1128/cmr.00077-12>
2. Sievers, F., & Higgins, D. G. (2013). Clustal Omega, Accurate Alignment of Very Large Numbers of Sequences. Methods in Molecular Biology, 105–116. <https://doi.org/10.1007/978-1-62703-646-7_6>
3. Zia, A., & Moses, A. M. (2011). Ranking insertion, deletion and nonsense mutations based on their effect on genetic information. BMC Bioinformatics, 12(1). https://doi.org/10.1186/1471-2105-12-299