Human adenovirus

**Abstract**

Adenoviruses are double-stranded DNA viruses that are non-enveloped and may infect a range of human organs. DNA homology among human adenovirus subgroups ranges from 48 percent to 99 percent. Human adenovirus subgroups, on the other hand, have a DNA homology of less than 20 percent. Adenovirus infections can produce a wide range of symptoms, including sore throats, runny noses, pink eye, and diarrhea, causing symptoms that are very similar to the common cold. The National Center for Biotechnology Information (NCBI), was the primary source for this study. Alignment of multiple sequences, protein coding, categorization, and analysis are all performed. We got a total of five Human Adenovirus (HAdV) strains as a consequence of our research. The oldest strain was used as a reference, in this case human adenovirus E4 strain RDU2954/New Jersey USA/1966. Similarity, mutations, transition/transversion ratios (TT ratios), gaps, insertions, and deletions, for the overall sequence as well as for coding and non-coding sections of the sequence were compared to the other four strains.

**Introduction**

Adenoviruses were first described over 60 years ago, when the medium from cultures of young children's adenoids in Washington, D.C. that degenerated within 1–4 weeks after isolation was discovered to contain a filterable agent (i.e. a virus) that transmitted the cytopathic effect to established human cell lines, such as HeLa cells. The same winter, an agent eliciting very comparable cytopathic reactions was identified having atypical pneumonia. (Flint, J. & Nemerow, G., 2017)

Adenoviruses are double-stranded DNA viruses that are non-enveloped and may infect a range of human organs. They have a diameter of 65 to 80 nanometers. The virion has 252-capsomere protein capsids and a nucleoprotein core containing the viral genome (26–46 kbp length with 23–46 protein-coding genes) and internal proteins. The capsid form is icosohedral, with 240 hexon components per virus particle and 12 pentons. (Jawetz, E. et al, 1955; Lennette, E. H. 1985). Each penton has a fiber-filled base plate. The length of the fibers differs depending on the serotype. (De Jong, J. C. et al, 1999).

Within the human adenovirus subgroups, DNA homology ranges from 48 percent to 99 percent. When compared to other human adenovirus subgroups, human adenovirus subgroup C serotypes demonstrated the greatest DNA homology (up to 99 percent). The DNA homology between human adenovirus subgroups, on the other hand, is less than 20%. (Walls, T. et al, 2003; Ghebremedhin B., 2014)

Viruses enter the body through the respiratory tract the most frequently. All viruses that infect the host through the respiratory tract do so by binding to epithelial cell receptors. (Burrell, C. J., et al 2017) Adenovirus infections are prevalent in late winter, early spring, and early summer, when flu season overlaps. Though these viral respiratory illnesses can be readily confused with the flu, there are several key distinctions to be aware of. Adenovirus infections can produce a wide range of symptoms, including sore throats, runny noses, pink eye, and diarrhea. More than 50 distinct strains of adenovirus have been identified throughout the world. (Duffy, S. et al, 2008)

Human adenoviruses belong to the Adenoviridae family, specifically the Mastadenovirus genus, which has seven identified species ranging from A to G. As a result of the hemagglutination and serum neutralization reactions, genomic data has lately uncovered numerous novel forms of adenovirus, including a number of novel and recombinant viruses. (Ghebremedhin B., 2014)

**Materials and Methods**

The National Center for Biotechnology Information (NCBI) advances science and health was used for virus and starin selection. The genome of a human adenovirus was assembled and annotated. A total of five strains were chosen for further investigation and comparison. For several strains, FASTA sequences were obtained. Strain 1 (reference strain) was collected in 1966 (GenBank: KX384948.1), strain 2 in 1971 (GenBank: KX384950.1), strain 3 in 1995 (GenBank: KX384951.1), strain 4 in 2002 (GenBank: KX384945.1), and strain 5 in 2004 (GenBank: AY594253.1).

Multiple Sequences Alignment

For multiple sequence alignment, the Clustal Omega tool was employed. Obtained FASTA sequences were uploaded for DNA sequence alignment. For additional research, a multiple sequence alignment file for chosen strains was saved.

Protein Coding Gene Search

From the genome assembly and annotation publication, we looked at a few strains. Replication information was used to find protein-coding genes. The script was then run on tables of protein coding genes, and the results examined.

Classification and Analysis

All information from the script was used to classify the results in Microsoft Excel. Excel data were used to create tables and graphs.

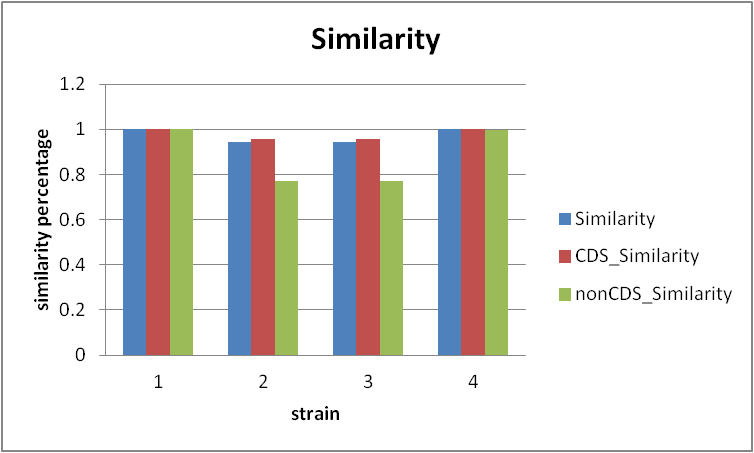
**Results**

A total of five Human Adenovirus (HAdV) strains were compared. The oldest strain was used as a reference, in this case human adenovirus E4 strain RDU2954/New Jersey USA/1966. As a result, four strains were compared to the first one, the oldest. The total length of sequences is 35991 bp, with 33615 bp of coding sequence and 2663 bp of non-coding sequences. Similarity, mutations, transition/transversion ratios (TT ratios), gaps, insertions, and deletions, for the overall sequence as well as for coding and non-coding sections of the sequence were compared to the other four strains.

Strain 2 is 99.96% similar to strain 1, strain 3 is 94.37% similar to strain 1, strain 4 is 94.36% similar to strain 1 and strain 5 is 99.94% similar to strain 1. Regarding the comparison of coding sequence similarity, strain 2 is 99.97% similar to strain 1, strain 3 is 95.72% similar, strain 4 is 95.71% similar and strain 5 is 99.97% similar to strain 1. Regarding the non-coding sequence similarity, strain 2 is 99.92% similar to strain 1, strain 3 is 77.24% similar to strain 1, strain 4 is 77.28% similar to strain 1 and strain 5 is 99.51% similar to strain 1 (calculated from table 1).

*Table 1: Sequence similarity and comparison of coding and non-coding sequences*

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Similarity** | **CDS\_Similarity** | **nonCDS\_Similarity** |
| **Human\_adenovirus\_1966** | 1.0 | 1.0 | 1.0 |
| **Human\_adenovirus\_1971** | 1.0 | 1.0 | 1.0 |
| **Human\_adenovirus\_1995** | 0.94 | 0.96 | 0.77 |
| **Human\_adenovirus\_2002** | 0.94 | 0.96 | 0.77 |
| **Human\_adenovirus\_2004** | 1.0 | 1.0 | 1,0 |

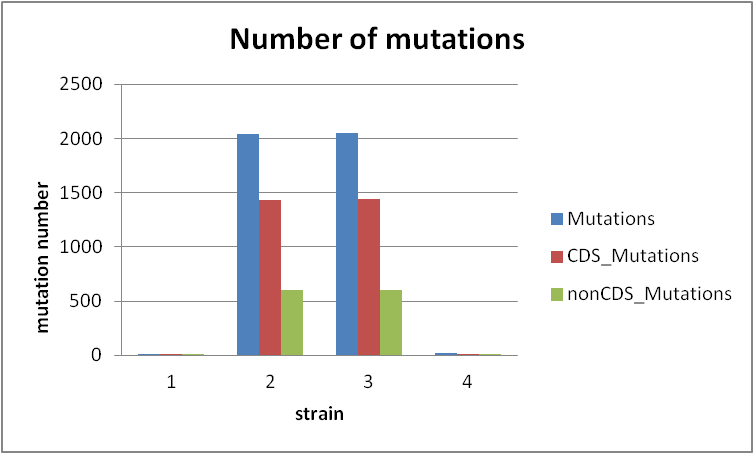
****

*Figure 1: representation of sequence similarity and comparison of coding and non-coding sequences*

Strain 2 accumulated 12 mutations, strain 3 accumulated 2042 mutations, strain 4 accumulated 2046 mutations, and strain 5 gained 21 mutations when compared to strain 1. When coding sequence mutations were compared to strain 1, strain 2 accumulated 10 mutations, strain 3 - 1436 mutations, strain 4 - 1441 mutations, and strain 5 - 8 mutations. In comparison to strain 1, strain 2 gained 2 non-coding mutations, strain 3 - 606 mutations, strain 4 - 605 mutations, and strain 5 - 13 mutations. The total mutation rate for sequence number 2 is 0.00033 mutations per nucleotide, for the third sequence it is 0.056 mutations, for the sequence number 4 it is also 0.056, and for the sequence number 5 – 0.00057 mutations per nucleotide. The total mutation rate for coding sequences for the second sequence is 0.00029, third sequence it is 0.042, fourth – also 0.042, and for the fifth sequence - 0.0002 mutations per nucleotide. Non-coding sequences seeing overall small number of mutations per nucleotide for all of the sequences. (visible and calculated in table 2)

*Table 2: Number of mutations (overall, in CDS and nonCDS)*

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Mutations** | **CDS\_Mutations** | **nonCDS\_Mutations** |
| **Human\_adenovirus\_1966** | 0 | 0 | 0 |
| **Human\_adenovirus\_1971** | 12 | 10 | 2 |
| **Human\_adenovirus\_1995** | 2042 | 1436 | 606 |
| **Human\_adenovirus\_2002** | 2046 | 1441 | 605 |
| **Human\_adenovirus\_2004** | 21 | 8 | 13 |

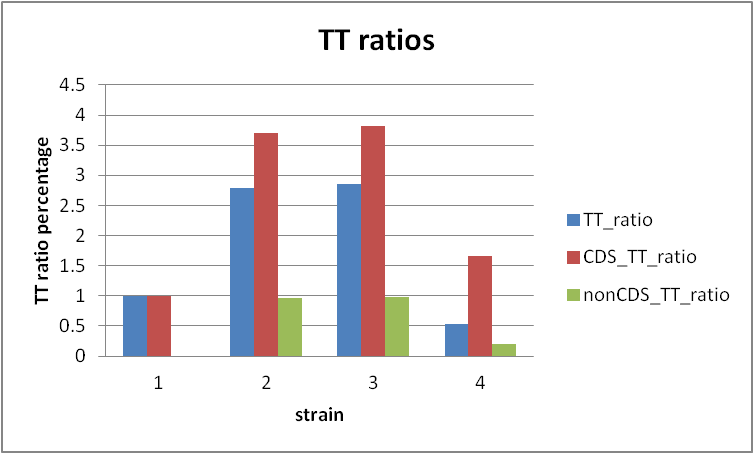
****

*Figure 2: Representation of number of mutations (overall, in CDS and nonCDS)*

In comparison to strain 1, the overall TT ratio for strain 2 is 1, strain 3 is 2.79, strain 4 is 2.85, and strain 5 is 0.53. In comparison to strain 1, the TT ratio for the coding sequence for strain 2 is 1, for strain 3 it is 3.7, for strain 4 it is 3.82, and for strain 5 it is 1.66. Strain 2 has a TT ratio of 0, strain 3 has a TT ratio of 0.96, strain 4 has a TT ratio of 0.97, and strain 5 has a TT ratio of 0.2.

*Table 3: Transition/Transversion Ratio*

|  |  |  |  |
| --- | --- | --- | --- |
|  | **TT\_ratio** | **CDS\_TT\_ratio** | **nonCDS\_TT\_ratio** |
| **Human\_adenovirus\_1966** | 0.00 | 0.00 | 0.00 |
| **Human\_adenovirus\_1971** | 1.00 | 1.00 | 0.00 |
| **Human\_adenovirus\_1995** | 2.79 | 3.71 | 0.97 |
| **Human\_adenovirus\_2002** | 2.86 | 3.82 | 0.98 |
| **Human\_adenovirus\_2004** | 0.54 | 1.67 | 0.20 |

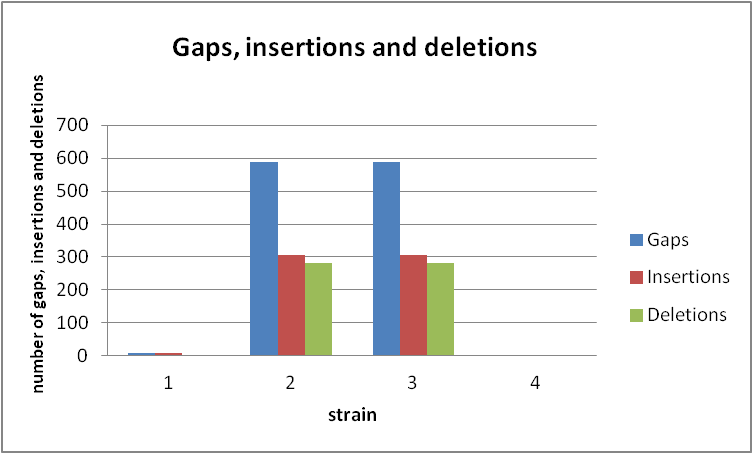
****

*Figure 3: representation of TT ratios*

In contrast to strain 1, strain 2 had 8 gaps, 8 insertions, and 0 deletions, strain 3 had 590 gaps, 307 insertions, and 283 deletions, strain 4 had 588 gaps, 307 insertions, and 281 deletions, and strain 5 had 1 gap, 1 insertion, and 0 deletions.

*Table 4: Gap, insertion and deletion represenation*

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Gaps** | **Insertions** | **Deletions** |
| **Human\_adenovirus\_1966** | 0 | 0 | 0 |
| **Human\_adenovirus\_1971** | 8 | 8 | 0 |
| **Human\_adenovirus\_1995** | 590 | 307 | 283 |
| **Human\_adenovirus\_2002** | 588 | 307 | 281 |
| **Human\_adenovirus\_2004** | 1 | 1 | 0 |

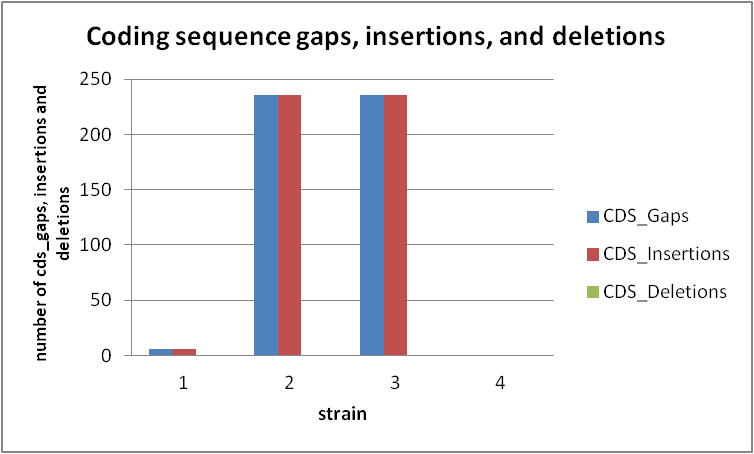


*Figure 4: representation of gap, insertions and deletions*

In terms of coding sequence gaps, insertions, and deletions, strain 2 accrued 6 gaps and 6 insertions. Strain 3 had 236 gaps and 236 insertions, whereas strain 4 had 236 gaps and 236 insertions as well. In comparison to strain 1, strain 5 accumulated 0 gaps and 0 insertions. There were no deletions in the coding sequence.

*Table 5: coding sequence gaps, insertions, and deletions*

|  |  |  |  |
| --- | --- | --- | --- |
|  | **CDS\_Gaps** | **CDS\_Insertions** | **CDS\_Deletions** |
| **Human\_adenovirus\_1966** | 0 | 0 | 0 |
| **Human\_adenovirus\_1971** | 6 | 6 | 0 |
| **Human\_adenovirus\_1995** | 236 | 236 | 0 |
| **Human\_adenovirus\_2002** | 236 | 236 | 0 |
| **Human\_adenovirus\_2004** | 0 | 0 | 0 |

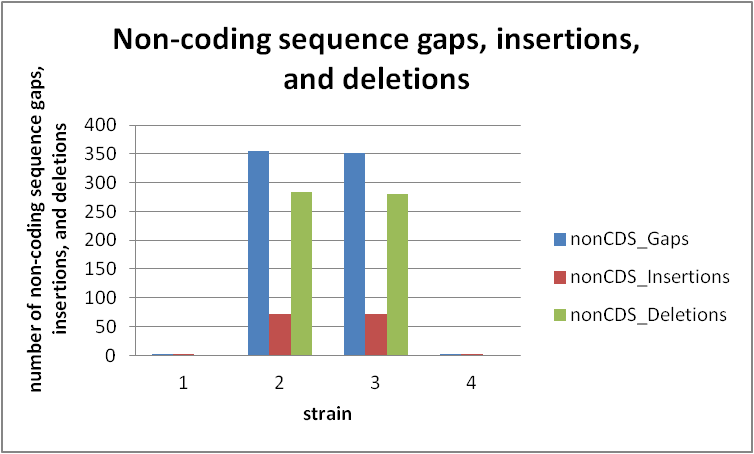
****

*Figure 5: representation of coding sequence gaps, insertions, and deletions*

In relation to strain 1, strain 2 accumulated 2 non-coding gaps, 2 non-coding insertions, and 0 non-coding deletions, strain 3 collected 354 non-coding gaps, 71 non-coding insertions, and 283 non-coding deletions, strain 4 accumulated 352 non-coding gaps, 71 non-coding insertions, and 281 non-coding deletions, and strain 5 accumulated 1 non-coding gap, 1 non-coding insertion, and 0 non-coding deletions.

*Table 6: non-coding sequence gaps, insertions, and deletions*

|  |  |  |  |
| --- | --- | --- | --- |
|  | **nonCDS\_Gaps** | **nonCDS\_Insertions** | **nonCDS\_Deletions** |
| **Human\_adenovirus\_1966** | 0 | 0 | 0 |
| **Human\_adenovirus\_1971** | 2 | 2 | 0 |
| **Human\_adenovirus\_1995** | 354 | 71 | 283 |
| **Human\_adenovirus\_2002** | 352 | 71 | 281 |
| **Human\_adenovirus\_2004** | 1 | 1 | 0 |

****

*Figure 6: representation of non-coding sequence gaps, insertions, and deletions*

**Discussion and conclusion**

Human adenoviruses are members of the Adenoviridae family, especially the Mastadenovirus genus, which includes seven species ranging in size from A to G. More than 50 serotypes have been discovered as a consequence of hemagglutination and serum neutralization responses; however, genomic data has recently revealed several additional forms of adenovirus, including a number of novel and recombinant viruses. (Ghebremedhin B., 2014)

In this study we have aimed to analyze the percentage of similarity of 5 human adenoviruses, focusing on similarities in deletions, insertions and gaps. As seen in table 1 and figure 1, the sequence similarity of coding and non-coding sequences is minimum 0.94 (94%) and maximum 1 (100%).The most similarity is observed between first reference strain Human\_adenovirus\_1966, second strain Human\_adenovirus\_1971 and fifth strain Human\_adenovirus\_2004, as shown 1 or 100%. Same similarity is visible between Human\_adenovirus\_1995 and Human\_adenovirus\_2002 0.94 (94%).

In table 2 and figure 2 we present the number of mutations in coding and non coding sequences. There we saw that most mutations are seen in Human\_adenovirus\_1995 and Human\_adenovirus\_2002 (2041 and 2046) and the least number of mutations are observed in Human\_adenovirus\_1971, with 12 total mutations. These results are in line with the current literature, claiming that certain DNA viruses have the same high rate of mutations as RNA viruses, implying that viral evolutionary rates are influenced by a variety of factors other than polymerase fidelity, such as genomic architecture and replication speed. (Duffy, S., et al, 2008).

Regarding the third table and figure, we observed transition/transversion ratio. A transition is a mutation (or substitution) that involves the replacement of one pyrimidine or purine with another. A purine is converted to a pyrimidine or vice versa during a transversion. (Bromham, L., 2016) In molecular evolution, transitions are preferred over transversions. (Lyons, D. M., & Lauring, A. S. 2017) This is true in terms of observable mutations, as well as the fact that transitions are more likely to result in synonymous mutations that are neutral, rather than nonsynonymous alterations that are subject to negative selection. (Deng, Z. L., 2021) The transition/transversion ratio is highest in Human\_adenovirus\_2002, and what is interesting to see is that lowest TT ratio is visible in Human\_adenovirus\_2004 with ratio of 0.54. For example in Human\_adenovirus\_1971 we can observe that there is equal number of transitions and transversions because the ratio is 1. In Human\_adenovirus\_2002 thare is highest TT ratio of 2.86, and after that high ratio is visible also in Human\_adenovirus\_1995, which is 2.79.

Next we observed insertions and deletions in five strains of human adenovirus. Gap events, or base pair deletions and insertions in DNA, are one of the most common drivers of evolutionary change at the molecular level. Nature’s article we found proposed that deletions in proteins are expressed more frequently than insertions based on very little evidence. (de Jong, W. W., & Rydén, L. 1981) We analyzed the currently known human adenovirus strains and found out that deletions are less frequent in all strains of the virus. The number of insertions is not large but in any case higher. For example Human\_adenovirus\_1971 has 8 insertions, and 0 deletions, Human\_adenovirus\_1995 307 insertions and 283 deletions, Human\_adenovirus\_2002 has also 307 insertions and 281 deletions. The last one, Human\_adenovirus\_2004 has 1 insertion and 0 deletions. The second sequence has very few mutations, but most of them are insertions (8 out of 12) which is unusual. The third and fourth sequences are logical - out of a total of ~ 2000 mutations, approximately one quarter (590 and 588, respectively) are indels, which means less than point mutations (transitions or transverions) which has logic. The latter is the same case, only one indel versus 20 point mutations.

These insertions and deletions might be thought of as hitches in the genome copying and maintenance process. Any gaps in a functional sequence that do not retain the codon structure, such as single or double base insertions or deletions are unlikely to occur. (Bromham, L., 2016)

As visible in table 5 and figure 5, only cds\_insertions are seen in Human\_adenovirus\_1971 which has 6, Human\_adenovirus\_1995 has 236, as well as Human\_adenovirus\_2002, also 236. Deletions for all of the strains are 0. The reason for no cds\_deletions can be because this is double-stranded DNA virus, which is more stable, meaning there is less mutations, also less or no deletions. Unlike in cds\_insertions and cds\_deletions, nonCDS\_insertions and nonCDS\_deletions are shown in slightly larger number. Similar number of nonCDS\_gaps can be seen in Human\_adenovirus\_1995 – 354 and Human\_adenovirus\_2002 – 352, and for the Human\_adenovirus\_1971 are 2. The last one Human\_adenovirus\_2004 has 1nonCDS\_gap.

The main aim of this project was to find relationship between strains, and their relation. Through this project and research we can better understand the virus itself, the structure and the way it works. As we can see while observing the sequences we worked on, the root of all viruses is from the territory of the USA in the period from 1966 to 2004. One of the reasons they have such a high degree of similarity can be the place of origin. In each of the tables shown, excluding the reference sequence, the second and fourth sequences (Human\_adenovirus\_1971 and Human\_adenovirus\_2004) have the most similarities, as do the third and fourth sequences (Human\_adenovirus\_1995 and Human\_adenovirus\_2002). It is interesting to note that instead of, for example, the number of mutations growing from year to year, the fifth sequence has a reduced number of mutations.

Many studies rely on mutation rate assumptions such a constant rate of mutation across time, across the genome, or between people and species. Because DNA repair mechanisms are so effective at what they do, the chances of any nucleotide having a mutation at any given time are vanishingly small. (Bromham, L., 2016) RNA viruses mutate more quickly than DNA viruses, and genome size appears to be negatively related to mutation rate. (Sanjuán, R., Domingo-Calap, P., 2016) As previously stated, HAdV is a double-stranded DNA virus, which might explain, among other things, why it has such a low frequency of mutations (that double stranded stability).

Because the "mutation rate genome", the component of a genome containing DNA replication and repair processes, is sensitive to genetic change, the mutation rate can develop, in larger or smaller number. (Bromham, L., 2016) That might explain why the fifth sequence had such a large mutation difference. The frequency with which mutations are detected in a particular viral population should not be confused with the mutation rate. The latter is a measure of genetic variety that is influenced by natural selection, random genetic drift, recombination, and other factors. (Sanjuán, R., Domingo-Calap, P., 2016) Rates of mutation should not be confused with rates of molecular evolution. The neutral theory of molecular evolution proposes a linear connection between these two rates, although molecular evolution refers to the fixation of novel alleles in populations. Mutation on the other hand, is a biochemical/genetic process. (Sanjuán, R., 2012; Duffy, S. et al, 2008)

1. Flint, J., & Nemerow, G. (2017). *Human adenoviruses*. Hackensack, N.J.: World Scientific.
2. Jawetz, E., Kimura, S., Nicholas, A. N., Thygeson, P., & Hanna, L. (1955). New type of APC virus from epidemic keratoconjunctivitis. *Science*, *122*(3181), 1190-1191.
3. Lennette, E. H. (Ed.). (1985). *Manual of clinical microbiology*. ASM Press.
4. De Jong, J. C., Wermenbol, A. G., Verweij-Uijterwaal, M. W., Slaterus, K. W., Wertheim-Van Dillen, P., Van Doornum, G. J., ... & Hierholzer, J. C. (1999). Adenoviruses from human immunodeficiency virus-infected individuals, including two strains that represent new candidate serotypes Ad50 and Ad51 of species B1 and D, respectively. *Journal of clinical microbiology*, *37*(12), 3940-3945.
5. Walls, T., Shankar, A. G., & Shingadia, D. (2003). Adenovirus: an increasingly important pathogen in paediatric bone marrow transplant patients. *The Lancet infectious diseases*, *3*(2), 79-86.
6. Ghebremedhin B. (2014). Human adenovirus: Viral pathogen with increasing importance. *European journal of microbiology & immunology*, *4*(1), 26–33. <https://doi.org/10.1556/EuJMI.4.2014.1.2>
7. Bamford, D., & Zuckerman, M. *Encyclopedia of virology*.
8. Burrell, C. J., Howard, C. R., & Murphy, F. A. (2017). Pathogenesis of Virus Infections. *Fenner and White's Medical Virology*, 77–104. <https://doi.org/10.1016/B978-0-12-375156-0.00007-2>
9. Duffy, S., Shackelton, L. A., & Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nature Reviews Genetics*, *9*(4), 267-276.
10. Bromham, L. (2016). An Introduction to Molecular Evolution and Phylogenetics (2nd UK ed.). Oxford University Press, USA.
11. Lyons, D. M., & Lauring, A. S. (2017). Evidence for the selective basis of transition-to-transversion substitution bias in two RNA viruses. *Molecular biology and evolution*, *34*(12), 3205-3215.
12. Deng, Z. L., Dhingra, A., Fritz, A., Götting, J., Münch, P. C., Steinbrück, L., ... & McHardy, A. C. (2021). Evaluating assembly and variant calling software for strain-resolved analysis of large DNA viruses. *Briefings in bioinformatics*, *22*(3), bbaa123.
13. de Jong, W. W., & Rydén, L. (1981). Causes of more frequent deletions than insertions in mutations and protein evolution. *Nature*, *290*(5802), 157-159.
14. Sanjuán, R., & Domingo-Calap, P. (2016). Mechanisms of viral mutation. *Cellular and molecular life sciences : CMLS*, *73*(23), 4433–4448. <https://doi.org/10.1007/s00018-016-2299-6>
15. Sanjuán, R. (2012). From molecular genetics to phylodynamics: evolutionary relevance of mutation rates across viruses. *PLoS pathogens*, *8*(5), e1002685.
16. Duffy, S., Shackelton, L. A., & Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nature Reviews Genetics*, *9*(4), 267-276.