

**COMPARING 5 STRAINS OF MIDDLE EAST RESPIRATORY SYNDROME**

Submitted to:

Prof. Mohamed Adilovic

By:

Nour Elbechnak

12th JAN 2022

Table of Contents

[**Abstract** 2](#_Toc92905336)

[**1.** **Introduction** 2](#_Toc92905337)

[**2.** **Method** 2](#_Toc92905338)

[**3.** **Result and Discussion** 2](#_Toc92905339)

[**3.1. Mutation** 2](#_Toc92905340)

[**3.2 Similarities** 4](#_Toc92905341)

[**3.3 TT-Ratio** 4](#_Toc92905342)

[**3.4 Gaps** 4](#_Toc92905343)

[**Conclusion** 5](#_Toc92905344)

[**Reference** 5](#_Toc92905345)

[**APPENDIX** 5](#_Toc92905346)

# **Abstract**

For this study, KJ813439, KP209306, KT156560, NC\_009019 and KC545386 strains were collected from NCBI and they were analyzed in terms of similarities, mutations, deletions, insertions, gaps and TT-Ratios. The study observed very close relationship between strains, even between the strain collected from humans to the bat. This provides insight about the slow mutation rate of the virus, even if they were collected from 8 years apart.

# **Introduction**

Middle East respiratory syndrome-related coronavirus (MERS-CoV) is a species of ssRNA (+) virus in the family Coronaviridae which contains 30.1 kb genome and 10 coding genes (NCBI). The virus uses dipeptidyl peptidase 4 (DPP4) to enter host cells (Petrosillo, N., Viceconte, G., Ergonul, O., Ippolito, G., & Petersen, E. 2020). It was first emerged in human in 2012 and number of cases were around 2000 by 2020, whereas numbers of cases from SARS-CoV-2 by 2020 were exceeding millions. We can definitely say that MERS is not as infectible as SARS-Cov-2.

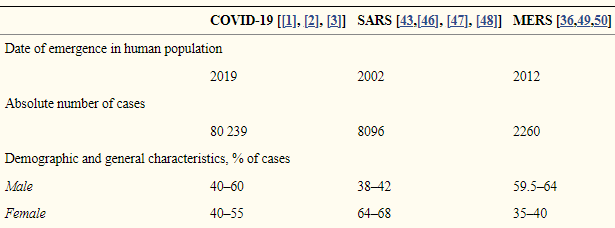


Figure 1- Clinical characteristics of COVID-19, SARS and MERS, source: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7176926/>

# **Method**

I took 5 strains of Middle Eastern Respiratory Syndrome Virus in FASTA format from NCBI website. 3 of them are from human host and 2 from bat host. I then ran a MSA on the Clustal Omega website. The aligned sequence is then analyzed by a program written on Python.

# **Result and Discussion**

For this study I have taken 5 strains of MERS virus and ran a MSA on the website Clustal Omega. Then a python-based program was used to analyze the data. The result is given below.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | source | Similarity | Mutations | TT ratio | Gaps | Insertions | Deletions |
| 1 | KJ813439.1 | Human | 0.68 | 9840 | 0.81255 | 770 | 411 | 359 |
| 2 | KP209306.1 | Human | 0.68 | 9839 | 0.813437 | 770 | 411 | 359 |
| 3 | KT156560.1 | Human | 0.68 | 9839 | 0.813075 | 770 | 411 | 359 |
| 4 | NC\_009019.1 | Bat | 0.66 | 10517 | 0.725867 | 809 | 349 | 460 |
| 5 | KC545386.1 | Bat | 1 | 0 | 0 | 0 | 0 | 0 |

My reference strain was KC545386, which was collected on 17-OCT-2013 from a bat. NC\_009019 was also collected from a bat. The first 3 stains were collected from human sample. But interestingly, we can see that the strains did not diverge a lot. The similarity between strain 4 and 5 are 66% and similarity between strains 1-3 and strain 5 is 68%. Even if there are many mutations with significant number of insertions and deletions, we do not see much divergence in the actual strains.

## **3.1. Mutation**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Mutations | CDS\_Mutations | nonCDS\_Mutations |
| 1 | KJ813439.1 | 9840 | 9277 | 563 |
| 2 | KP209306.1 | 9839 | 9276 | 563 |
| 3 | KT156560.1 | 9839 | 9276 | 563 |
| 4 | NC\_009019.1 | 10517 | 9787 | 730 |
| 5 | KC545386.1 | 0 | 0 | 0 |

As we can see from strain number 5 KC545386 which was collected from a bat comparing to strain number 4 NC\_009019 which is also from a bat there is outstanding difference, but on the other hand the first three strains which were collected from humans there isn’t that much of a difference between them but comparing strain number 1 KJ813439 to strain number 4 NC\_009019 there is a quiet noticeable difference. Strain number 4 has 10517 mutations with 9787 in the CDS region, however strain 1-3 have also around ~9000 mutations, but they were collected from humans, interestingly most of them are from CDS regions. From this observation we can say that MERS may have low mutation frequency and that was the reason why it did not spread as much as SARS-CoV

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Mutations | Insertions | Deletions |
| 1 | KJ813439.1 | 9840 | 411 | 359 |
| 2 | KP209306.1 | 9839 | 411 | 359 |
| 3 | KT156560.1 | 9839 | 411 | 359 |
| 4 | NC\_009019.1 | 10517 | 349 | 460 |
| 5 | KC545386.1 | 0 | 0 | 0 |
|  |  |  |  |  |

identical but in the bat one the fourth strain NC\_009019 the indels are a lot more different comparing them to strain number 1. The reason that happened might be either by point mutations or proofreading error, and the point mutation is large number of mutation that was occurred due to a single nucleotide change in the sequence, meaning A might have swapped with C,G, and T or one nucleotide might have been deleted in the process. Proofreading error is when polymerase add incorrect nucleotide to the strand.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | CDS\_Mutations | CDS\_Insertions | CDS\_Deletions |
| 1 | KJ813439.1 | 9277 | 392 | 0 |
| 2 | KP209306.1 | 9276 | 392 | 0 |
| 3 | KT156560.1 | 9276 | 392 | 0 |
| 4 | NC\_009019.1 | 9787 | 294 | 0 |
| 5 | KC545386.1 | 0 | 0 | 0 |

The same thing we can observe in the coding region. We know that coding region codes for the protein that is crucial for mutation. we can see that around 9000 mutations occurring in the CDS region whereas only around 300 insertions and no deletions.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | nonCDS\_Mutations | nonCDS\_Insertions | nonCDS\_Deletions | |
| 1 | KJ813439.1 | 563 | 19 | 359 | |
| 2 | KP209306.1 | 563 | 19 | 359 | |
| 3 | KT156560.1 | 563 | 19 | 359 | |
| 4 | NC\_009019.1 | 730 | 55 | 460 | |
| 5 | KC545386.1 | 0 | 0 | 0 | |
|  |  |  |  |  |  | |

Here in this table, we can see nonCDS region of the strains the nonCDS deletion in the first three are about 300 base pair and that could be to natural selection or due to evolutionary reasons. In the “The Role of Mutation Bias in Adaptive Evolution” article, Erik I. Svensson and David Berger wrote, “natural selection can shape the phenotypic effects of mutations, giving the false impression that direct mutations are driving adaptive evolution.”

## **3.2 Similarities**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Similarity | CDS\_Similarity | nonCDS\_Similarity |
| 1 | KJ813439.1 | 0.68 | 0.684015123 | 0.584501845 |
| 2 | KP209306.1 | 0.68 | 0.684049184 | 0.584501845 |
| 3 | KT156560.1 | 0.68 | 0.684049184 | 0.584501845 |
| 4 | NC\_009019.1 | 0.66 | 0.666643959 | 0.461254613 |
| 5 | KC545386.1 | 1 | 1 | 1 |

It is very interesting to see that the first three strains which were from humans are the almost the same percentage, which is 68%. the first one was taken in 12-MAY-2014, second taken in 26-NOV-2014 and the last one was from 12-JUNE-2015, but the fourth one that was taken from a bat was one of the earliest in 20-FEB-2007 differ only with 2% to the first three and what that tell us is MERS mutation rate in slow.

## **3.3 TT-Ratio**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | TT\_ratio | CDS\_TT\_ratio | nonCDS\_TT\_ratio |
| 1 | KJ813439.1 | 0.81254996 | 0.813265306 | 0.778846154 |
| 2 | KP209306.1 | 0.813437313 | 0.814171942 | 0.778846154 |
|  | KT156560.1 | 0.81307477 | 0.813801552 | 0.778846154 |
| 4 | NC\_009019.1 | 0.725866667 | 0.725686239 | 0.733870968 |
| 5 | KC545386.1 | 0 | 0 | 0 |

We know various types of mutations happen at different rates. In genetics, TT ratios referred to transversion and transition ratio. We know that in evolutionary pathway, mutations have transition biases over transversion. As we can see the TT ratio are all around one and what that means the viral strains follow the molecular evolutionary pathway and the mutation bias was inclining towered transition over transversion (Stoltzfus, A., & Norris, R. W.., 2016)..).

## 

## **3.4 Gaps**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Gaps | CDS\_Gaps | nonCDS\_Gaps |
| 1 | KJ813439.1 | 770 | 392 | 378 |
| 2 | KP209306.1 | 770 | 392 | 378 |
| 3 | KT156560.1 | 770 | 392 | 378 |
| 4 | NC\_009019.1 | 809 | 294 | 515 |
| 5 | KC545386.1 | 0 | 0 | 0 |

From this table we can see that Gaps and ratio are related to each other. according to this study, “Gap genes are large areas of the normal cuticular pattern are deleted in individuals with mutant phenotypes” (Marjorie A.Hoy, 2019). We can see that the gap patterns correlate with the mutation pattern. From the CDS deletions, we didn’t see any deletions, whereas the entire genomes have around 300-460 bp deleted. From that, we can say the deleted sequence probably have some evolutionary causations, which is not reflected in the actual mutation.

# **Conclusion**

In conclusion, we can say that MERS is not a highly mutated virus, even if we saw a large number of mutations in the coding region. I have made a mistake in collecting the stains, my intention was to collect and compare strains from human only, I however ended up collecting 2 strains from bat host. Surprisingly the bat viruses did not differ much from the human viruses. They both have shown slow mutation rate. The earliest strain was collected in 2007 and the newest one in 2015. The 2007 was 66% similar to the strain that was collected in 2013 and the latest one, which was collected in 2015 was 68% similar to the first strain.

# **Reference**

1. Stoltzfus, A., & Norris, R. W. (2016). On the Causes of Evolutionary Transition:Transversion Bias. Molecular biology and evolution, 33(3), 595–602. <https://doi.org/10.1093/molbev/msv274>
2. DOI: <https://doi.org/10.1016/j.tree.2019.01.015>
3. <https://doi.org/10.1016/B978-0-12-815230-0.00004-2>
4. <https://www.ncbi.nlm.nih.gov/search/all/?term=MERS>
5. Petrosillo, N., Viceconte, G., Ergonul, O., Ippolito, G., & Petersen, E. (2020). COVID-19, SARS and MERS: are they closely related?. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, *26*(6), 729–734. <https://doi.org/10.1016/j.cmi.2020.03.026>

# **APPENDIX**

1. KJ813439, 12-MAY-2014, [Middle East respiratory syndrome coronavirus strain Indiana/USA-1\_Saud - Nucleotide - NCBI (nih.gov)](https://www.ncbi.nlm.nih.gov/nuccore/KJ813439.1)
2. KP209306, 26-NOV-2014, [Middle East respiratory syndrome coronavirus strain Abu Dhabi\_UAE\_8\_20 - Nucleotide - NCBI (nih.gov)](https://www.ncbi.nlm.nih.gov/nuccore/KP209306.1)
3. KT156560, 12-JUN-2015, [Middle East respiratory syndrome coronavirus strain Hu/Oman\_2285\_2013, - Nucleotide - NCBI (nih.gov)](https://www.ncbi.nlm.nih.gov/nuccore/KT156560)
4. NC\_009019, 20-FEB-2007, [Tylonycteris bat coronavirus HKU4, complete genome - Nucleotide - NCBI (nih.gov)](https://www.ncbi.nlm.nih.gov/nuccore/NC_009019) (BAT)
5. KC545386, 17-OCT-2013, [Betacoronavirus Erinaceus/VMC/DEU/2012 isolate ErinaceusCoV/2012-216/G - Nucleotide - NCBI (nih.gov)](https://www.ncbi.nlm.nih.gov/nuccore/KC545386) (BAT)