

Biological Science

KEYWORDS:

IS THERE A DIRECT RELATIONSHIP IN SPIKE GLYCOPROTEIN PERCENT IDENTITY, PEPTIDE COUNT, AND MORTALITY RATE BETWEEN DAUGHTER HUMAN CORONAVIRUS STRAINS AND THE "PARENT" (HCOV-229E)?



Volume - 7, Issue - 9, September- 2022

ISSN (O): 2618-0774 | ISSN (P): 2618-0766

Ebrahim Waleed**Ebrahim Ahmed Al**

Applied Technology High School Umm Al Quwain Campus, UAE

Mansoori**Hamdan Mohamed****Khalfan Khairi Al**

Applied Technology High School Umm Al Quwain Campus, UAE*Corresponding Author

Noobi Al Hammadi***Mohammed Abdalla****Rashid Abdalla**

Applied Technology High School Umm Al Quwain Campus, UAE

RashidINTERNATIONAL JOURNAL
OF PURE MEDICAL RESEARCH**Abstract**

In this investigation, we sought to identify the following relationships between the Human Coronavirus HCoV-229E strain ("Parent") and subsequent daughter strains (HCoV-NL63, HCoV-HKU1, HCoV-OC43, MERS-CoV, SARS-CoV, and SARS-CoV2) in their Spike Glycoproteins: Percent Identity, Peptide Differences, and Mortality Rate. We also aimed to propose a possible reason as to why a given strain of Human Coronavirus is more fatal than others. To compare percent identity between each daughter strain's Spike glycoprotein sequence to that of the "Parent", we used an online amino acid sequence alignment database called Protein BLAST. To identify peptide-count differences between Parent and daughter Human coronavirus strains' Spike Glycoprotein sequences, we used an online protein classification database called InterPro. To compare mortality rates between each Human coronavirus strain, we used several online science websites and journals, including the World Health Organization, to obtain reliable data. During our analysis, we determined that no direct relationship exists between low or high percent Spike Glycoprotein sequence identity and mortality between daughters and parent. That is, if a daughter's sequence was closer to that of the "Parent", it did not mean that its mortality rate would be similar. We also learned that there is no direct relationship between a high or low number of additional peptides found on the Spike Glycoprotein of daughter strains and their mortality rate, as compared to the "Parent". Meaning, strains having similar peptides to that of the "Parent" were not similarly fatal. We learned that Daughter Four (MERS-CoV) has the highest mortality rate among all the other human Coronavirus strains, possibly due to the presence of three unique peptides.

Introduction

For the last three years, Coronavirus has been a relevant topic of discussion for everyone. Even though there is a voluminous amount of information online about genomes, spike Glycoproteins, protein segments/peptides, and mortality rates of the various Human Coronavirus strains, there were no articles that answered the following question: Is There A Direct Relationship In Spike Glycoprotein Percent Identity, Peptide Count, and Mortality Rate Between Daughter Human Coronavirus Strains And The "Parent" (HCoV-229E)? There were also no tabulated data charts online that compared the percent identify and protein segments/peptides

between Spike Glycoprotein sequences of each Human Coronavirus strain. We were also not readily able to find a tabulated list of mortality rate between the various strains of Human Coronaviruses. We knew that this gap in literature and knowledge had to be filled and we sought to fill it. Materials and Methodology

All Human Coronavirus Spike glycoprotein sequences for the following strains were obtained from the National Center for Biotechnology Information (NCBI) database: HCoV-229E, HCoV-NL63, HCoV-HKU1, HCoV-OC43, MERS-CoV, SARS-CoV, and SARS-CoV2. NCBI, a genome database has not only mapped that genome of each Human Coronavirus strain, but it has also classified where each sequence is found on the virus. As a result, we were able to easily find the Spike Glycoprotein sequence of each virus by searching the genome pages for "Spike" or "Glycoprotein". We determined percent identity between the Spike Glycoprotein of each 'daughter' Human Covid strain (HCoV-NL63, HCoV-HKU1, HCoV-OC43, MERS-CoV, SARS-CoV, and SARS-CoV2) and the 'parent' (HCoV-229E) using the U.S. National Library of Medicine (NLM) protein alignment service called Protein BLAST. Protein BLAST only allows for two sequences to be aligned at once. We did not change any of the algorithm parameters. All parameters were left as 'Default'. To determine the protein segments/peptides found on each strain's Spike Glycoprotein, we used the online database provided by InterPro called Classification of Protein Families. We did not change any of the algorithm parameters. All parameters were left as 'Default'. With the results from InterPro, we manually color-coded all protein segment/peptides that were identical between 'daughters' and 'parent'. As a result, we were also able to identify protein segment/peptide differences between them. To determine the mortality rate of each strain of Human Coronavirus, we used the following research articles:

1. Comparison of the clinical characteristics and mortality of adults infected with human coronaviruses 229E and OC43,
2. Increased Incidence, Morbidity, and Mortality in Human Coronavirus NL63 Associated with ACE Inhibitor Therapy and Implication in SARS-CoV-2 (COVID-19),
3. Emergence of Bat-Related Betacoronaviruses: Hazard and Risks,
4. Comparison of the clinical characteristics and mortality of adults infected with human coronaviruses 229E and OC43,
5. Middle East respiratory syndrome coronavirus (MERS-CoV),
6. Estimates of SARS death rates revised upward, and
7. Infection fatality rate of SARS-CoV2 in a super-spreading event in Germany.

Data Analysis

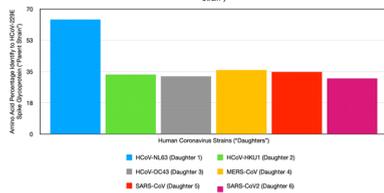
Existing research indicates that Mers-Covid high mortality rate is due to the presence of a Fusion Peptide (Xia et al., 2014). Our research found this to be true. But what makes our research finding and data synthesis different, is that currently there are no easily available online data tables/graphs that show comparison between the different human Coronavirus strains' Spike Glycoprotein in terms of their percent identity, 'peptide' similarity, amino acid base count comparison, and mortality rate. Our investigation has filled this gap.

We acknowledge that we loosely used the term 'peptide' in our research. The term 'peptide' in this investigation means protein segment or segments found on the Spike Glycoprotein sequences of each Human coronavirus strain that are protein related. Irrespective of the term used, we are confident that we accurately identified, aligned, protein-classified, amino acid base counted, and compared the mortality rates of the Spike Glycoprotein sequences of each Human Coronavirus strain.

In our investigation, we sought to determine whether there was a direct relationship between mortality rate of each Human Coronavirus strain and their Spike Glycoprotein sequence. We hypothesized that 'daughter' strains that are closely related to the 'parent' strain (HCoV-229E) would have similar mortality rates. Our finding showed that our hypothesis was not correct.

Rather, there was no direct relationship between being genetically close in identity to the 'parent' (HCoV-229E) and being similarly fatal. What we discovered was that the presence of special 'peptides' found within/on the Spike glycoprotein sequence of each Human Coronavirus strain determined their mortality rate.

Figure 1: Human Coronavirus Strain Amino Acid Percent Identity to HCoV-229E Spike Glycoprotein ("Parent Strain")



Human Coronavirus Strain	Amino Acid Percentage Identity to HCoV-229E Spike Glycoprotein ("Parent Strain")
HCoV-229E (Parent)	100
HCoV-NL63 (Daughter 1)	64
HCoV-HKU1 (Daughter 2)	36
HCoV-OC43 (Daughter 3)	33.29
MERS-CoV (Daughter 4)	35
SARS-CoV (Daughter 5)	14.5
SARS-CoV2 (Daughter 6)	2.2

Figure 1: Data shows daughter Spike Glycoprotein Human Coronavirus Strains' Amino Acid Percent Identity to HCoV-229E ("Parent Strain"). The graph shows that HCoV-NL63 (Daughter 1) has the closest present Spike glycoprotein identity to that of the "Parent", at a rate of 64%.

MERS-CoV (Daughter 4), SARS-CoV (Daughter 5), HCoV-HKU1 (Daughter 2), and HCoV-OC43 (Daughter 3) have a 36%, 35%, 33.29%, and 32.3% respectively.

Table 1

Human Coronavirus Strain ("Daughters")	Spike Glycoprotein Amino Acid base Difference to HCoV-229E ("Parent")
HCoV-NL63 (Daughter 1)	236
HCoV-HKU1 (Daughter 2)	2,609
HCoV-OC43 (Daughter 3)	1,424
MERS-CoV (Daughter 4)	2,802
SARS-CoV (Daughter 5)	2,434
SARS-CoV2 (Daughter 6)	2,586

Figure 2: Spike Glycoprotein Amino Acid Difference to HCoV-229E ("Parent")

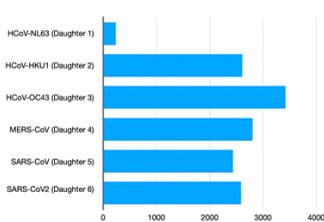


Figure 2: Data shows additional Spike Glycoprotein Amino Acids that each daughter Human Coronavirus has compared to HCoV-229E ("Parent"). In this graph, HCoV-229E ("Parent") Spike Glycoprotein amino acid count is zero. The graph shows that the Human coronavirus strain with the greatest additional amino acids

compared to that of the Parent is HCoV-OC43 (Daughter 3). Daughter 3 has 3,424 additional amino acids. Daughters 4, 2, 6, and 5 have 2,802, 2,609, 2,586, and 2,434 respectively.

Human Coronavirus Strain	HCoV-229E (Parent)	HCoV-NL63 (Daughter 1)	HCoV-HKU1 (Daughter 2)	HCoV-OC43 (Daughter 3)	MERS-CoV (Daughter 4)	SARS-CoV (Daughter 5)	SARS-CoV2 (Daughter 6)
Peptide Similarity To HCoV-229E Spike Glycoprotein ("Parent")	100	64	36	33.29	35	14.5	2.2
Peptide Difference To HCoV-229E Spike Glycoprotein ("Parent")	0	236	2,609	1,424	2,802	2,434	2,586
Percent Mortality Rate	25	12.5	0.5	9.1	35	14.5	2.2

Figure 3: Peptide Count Similarity And Difference Between Human Coronavirus Strains

Figure 3: The table shows peptide similarities and differences between Human Coronavirus strains. Out of the 24 peptides found on the Parent strain (HCoV-229E), 100% of them are found in Daughter 1 (HCoV-NL63 (Daughter 1) with one additional non-parent peptide. 75% are found in Daughter 2 (HCoV-HKU1) with 10 additional non-parent peptides. 83% are found in Daughter 3 (HCoV-OC43) with 12 additional non-parent peptides. Nearly 71% are found in Daughter 4 (MERS-CoV) with 11 additional non-parent peptides. Nearly 67% are found in Daughter 5 (SARS-CoV) with 12 additional non-parent peptides, and nearly 67% are found in Daughter 6 (SARS-CoV2) with 14 additional non-parent peptides.

Human Coronavirus Strain	Percent Mortality Rate
HCoV-229E (Parent)	25
HCoV-NL63 (Daughter 1)	12.5
HCoV-HKU1 (Daughter 2)	0.5
HCoV-OC43 (Daughter 3)	9.1
MERS-CoV (Daughter 4)	35
SARS-CoV (Daughter 5)	14.5
SARS-CoV2 (Daughter 6)	2.2

Figure 4: Percent Mortality Rate for Human Coronavirus Strains

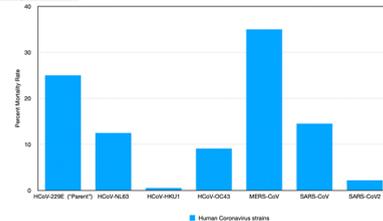


Figure 4: Data shows percent mortality rate for Human Coronavirus Strains. The "Parent" (HCoV-229E) has a mortality rate of 25%. Daughter 1 (HCoV-NL63 (Daughter 1) has a rate of 12.5%, Daughter 2 (HCoV-HKU1) 0.5%, Daughter 3 (HCoV-OC43) 9.1%, Daughter 4 (MERS-CoV) 35%, Daughter 5 (SARS-CoV) 14.5%, and Daughter 6 (SARS-CoV2) has a mortality rate of 2.2%.

Figure 5: Looking For Relationships between Mortality And Other Variables

Human Coronavirus Strains	Spike Glycoprotein Amino Acid Percent Identity To HCoV-229E ("Parent")	Spike Glycoprotein Amino Acid Difference To HCoV-229E ("Parent")	Spike Glycoprotein Peptide Difference To HCoV-229E ("Parent")	Percent Mortality Rate
HCoV-229E ("Parent")	100	0	0	25%
HCoV-NL63 (Daughter 1)	64	236	1	12.5%
HCoV-HKU1 (Daughter 2)	33.29	2,609	10	0.5%
HCoV-OC43 (Daughter 3)	32.3	3,424	12	9.1%
MERS-CoV (Daughter 4)	36	2,802	11	35%
SARS-CoV (Daughter 5)	35	2,434	12	14.5%
SARS-CoV2 (Daughter 6)	31	2,586	14	2.2%

Figure 5: Data shows that there is no direct relationship between Human Coronavirus strains' mortality rate and their Spike Glycoprotein peptide count, amino acid count, and percent identity to the "Parent" strain.

Summary and Conclusion

"Daughter Four" shows uniqueness in mortality rate. The W.H.O has reported that MERS-CoV ("Daughter Four") has nearly a 35% mortality rate (Middle East respiratory syndrome coronavirus

(MERS-COV)), the highest rate among the other Human Coronavirus strains, compared to the 25% rate of the "Parent" (HCoV-229E). To identify a possible reason for its enhanced fatality, we compared its peptides to the other Human Coronavirus strains. Out of the 28 peptides found on the Spike Glycoprotein of MERS-CoV (Daughter Four), nearly 61% are found on the Parent's Spike Glycoprotein. Out of the remaining peptides, nearly 29% are found among the other daughter strains. But no other strain is as fatal as MERS-CoV (Daughter Four). Therefore, peptides that are unique to Daughter Four could provide a possible explanation for its high mortality rate.

We set out to answer the following question: Is There A Direct Relationship In Spike Glycoprotein Percent Identity, Peptide Count, and Mortality Rate Between Daughter Human Coronavirus Strains And The "Parent" (HCoV-229E)? We discovered that no direct relationship exists between low or high percent Spike Glycoprotein sequence identity and mortality rate between daughters and parent. That is, if a daughter's sequence was close to that of the "Parent", it did not mean that its mortality rate would be similar. We also learned that there is no direct relationship between a high or low number of additional peptides found on the Spike Glycoprotein of daughter strains and their mortality rate, as compared to the "Parent". That is, strains having similar peptide count to that of the "Parent" were not similarly fatal. We learned that Daughter Four (MERS-CoV) has the highest mortality rate among all the other human Coronavirus strains, possibly due to the presence of three unique peptides. The peptides that are unique to Daughter Four are: cd21626, G3DSA:2.20.210.30, and cd21486. Peptides cd21626 and cd21486, both viral receptor binding domains, are associated with a genus of virus called the merbecovirus. These viruses are associated with pigs, cows, bats, and human. The peptide G3DSA:2.20.210.30, which also functions as a receptor binding domain and a viral fusion-cell entry mediator, is associated with viruses in bats, hedgehogs, and rodents.

MERS-CoV (Daughter Four)'s trans-species viral receptors, fusion and entry mediation abilities may have contributed to its hyper-mortality. Recommendations This investigation relied entirely on online data sources. Every Human Coronavirus Spike Glycoprotein sequence and alignment was obtained/conducted through one online source (National Center for Biotechnology Information). Every protein segment/'peptide' identified on each sequence of Human Coronavirus strain was obtained through one online protein- sequence classification database (InterPro). Every data relating to the mortality rate of each Human CoVid strain was sourced from various online academic database/website/journals. The parameter that we selected for each Spike Glycoprotein sequence alignment and protein classification was 'default'. That means, if any error existed in those primary sources, then the results of our investigation and data produced would reflect such error.

We recommend, for future investigations, that multiple genome database be referenced to determine if there exists any variation in the Spike Glycoprotein sequences that are available online. Our second recommendation is to use multiple sequence alignment and protein classification databases to determine if there exist variations in their results.

Among the many things for which we are proud, during the course of this research, is the data that we produced/tabulated. Currently, there are no readily available tabulated data online that compares the percent identity, peptide similarity and difference, amino acid base count, and mortality rate between the Spike Glycoprotein of each Human Coronavirus strain. The products of our investigation have filled that gap. Acknowledgment

We first would like to thank Allah. Next, we would like to thank the government of the U.A.E for providing us with the opportunity to have an excellent education. Then, we would like to thank our Principal Mr. Karim for creating a wonderful and encouraging

learning environment. Lastly, we would like to thank Mr. Ted for guiding us through the steps of our investigation and research.

REFERENCES

1. U.S. National Library of Medicine. (n.d.). Human Coronavirus 229E, complete genome - nucleotide - NCBI. National Center for Biotechnology Information. Retrieved January 29, 2022, from https://www.ncbi.nlm.nih.gov/nuccore/NC_002645
2. U.S. National Library of Medicine. (n.d.). Human Coronavirus NL63, complete genome - nucleotide - NCBI. National Center for Biotechnology Information. Retrieved January 29, 2022, from https://www.ncbi.nlm.nih.gov/nuccore/NC_005831
3. U.S. National Library of Medicine. (n.d.). Human coronavirus HKU1, complete genome - nucleotide - NCBI. National Center for Biotechnology Information. Retrieved January 29, 2022, from https://www.ncbi.nlm.nih.gov/nuccore/NC_006577
4. U.S. National Library of Medicine. (n.d.). Human coronavirus OC43 strain ATCCVR-759, complete genome - nucleotide - NCBI. National Center for Biotechnology Information. Retrieved January 29, 2022, from https://www.ncbi.nlm.nih.gov/nuccore/NC_006213
5. U.S. National Library of Medicine. (n.d.). Middle East respiratory syndrome-related coronavirus isolate hcov-EMC - nucleotide - NCBI. National Center for Biotechnology Information. Retrieved January 29, 2022, from https://www.ncbi.nlm.nih.gov/nuccore/NC_019843
6. U.S. National Library of Medicine. (n.d.). SARS coronavirus tor2, complete genome - nucleotide - NCBI. National Center for Biotechnology Information. Retrieved January 29, 2022, from https://www.ncbi.nlm.nih.gov/nuccore/NC_004718
7. U.S. National Library of Medicine. (n.d.). Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-hu-1, co - nucleotide - NCBI. National Center for Biotechnology Information. Retrieved January 29, 2022, from https://www.ncbi.nlm.nih.gov/nuccore/NC_045512
8. U.S. National Library of Medicine. (n.d.). Protein blast: Align two or more sequences using blast. National Center for Biotechnology Information. Retrieved January 29, 2022, from https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins&PROGRAM=blastp&BLAST_P
9. PROGRAMS=blastp&PAGE_TYPE=blastSearch&BLAST_SPEC=blast2seq&DATABASE=E%2F%2F&QUERY=&SUBJECTS=
10. Blum M, Chang H, Chuguransky S, Grego T, Kandasamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, Richardson L, Salazar GA, Williams L, Bork P, Bridge A, Gough J, Haft DH, Letunic I, Marchler-Bauer A, Mi H, Natale DA, Nuccicci M, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sililtoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A and Finn RD The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, Nov 2020, (doi: 10.1093/nar/gkaa977)
11. Choi, W.-I., Kim, I. B., Park, S. J., Ha, E.-H., & Lee, C. W. (2021, February 24). Comparison of the clinical characteristics and mortality of adults infected with human coronaviruses 229E and OC43. *Nature News*. Retrieved January 29, 2022, from <https://www.nature.com/articles/s41598-021-83987-3>
12. Krvavac, A., Patel, T. P., Karle, E. M., Epstein, N. B., Reznikov, E. A., Gates, L. G., & Holliday, Z. M. (2020). Increased incidence, morbidity, and mortality in human coronavirus NL63 associated with ACE inhibitor therapy and implication in SARS-COV-2 (COVID-19). *Missouri medicine*. Retrieved January 29, 2022, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7431067/>
13. Frutos R; Serra-Cobo J; Pinault L; Lopez Roig M; Devaux CA; (n.d.). Emergence of bat-related betacoronaviruses: Hazard and risks. *Frontiers in microbiology*. Retrieved January 29, 2022, from <https://pubmed.ncbi.nlm.nih.gov/33790874/>
14. Choi W; Kim IB; Park SJ; Ha EH; Lee CW; (n.d.). Comparison of the clinical characteristics and mortality of adults infected with human coronaviruses 229E and OC43. *Scientific reports*. Retrieved January 29, 2022, from <https://pubmed.ncbi.nlm.nih.gov/33627764/>
15. World Health Organization. (n.d.). Middle East respiratory syndrome coronavirus (MERS-COV). World Health Organization. Retrieved January 29, 2022, from https://www.who.int/health-topics/middle-east-respiratory-syndrome-coronavirus-mers#tab=tab_1
16. Roos , R. (2003, May 7). Estimates of SARS death rates revised upward. *CIDRAP*. Retrieved January 29, 2022, from <https://www.cidrap.umn.edu/news-perspective/2003/05/estimates-sars-death-rates-revised-upward>
17. Streeck, H., Schulte, B., Kümmerer, B. M., Richter, E., Höller, T., Fuhrmann, C., Bartok, E., Dolscheid-Pommerich, R., Berger, M., Wessendorf, L., Eschbach-Bludau, M., Kellings, A., Schwaiger, A., Coenen, M., Hoffmann, P., Stoffel-Wagner, B., Nöthen, M. M., Eis-Hübing, A. M., Exner, M., ... Hartmann, G. (2020, November 17). Infection fatality rate of SARS-cov2 in a super-spreading event in Germany. *Nature News*. Retrieved January 29, 2022, from <https://www.nature.com/articles/s41467-020-19509-y>
18. Xia, S., Liu, Q., Wang, Q., Sun, Z., Su, S., Du, L., Ying, T., Lu, L., & Jiang, S. (2014, December 19). Middle East respiratory syndrome coronavirus (MERS-COV) entry inhibitors targeting Spike protein. *Virus research*. Retrieved February 11, 2022, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7114414/>