TABLE OF CONTENTS

in Virology (https://www.frontiersin.org/journals/2000)
Emerging and Reemerging Viruses (https://www.frontiersin.org/journals/2000/sections/2012)

THIS ARTICLE IS PART OF THE RESEARCH TOPIC
Emerging and Re-emerging Viral Diseases View all 22 Articles › (https://www.frontiersin.org/research-topics/23391#articles)

PERSPECTIVE article
Front.Virol., 21 February 2022 | https://doi.org/10.3389/fviro.2022.834808 (https://doi.org/10.3389/fviro.2022.834808)

MSH3 Homology and Potential Recombination Link to SARS-CoV-2 Furin Cleavage Site

Balamurali K. Ambati[1], Akhil Varshney (https://www.frontiersin.org/people/u/1460400)[2], Kenneth Lundstrom (https://www.frontiersin.org/people/u/1020281)[3*], Giorgio Palú (https://www.frontiersin.org/people/u/236782)[4], Bruce D. Uhal (https://www.frontiersin.org/people/u/1659356)[5], Vladimir N. Uversky (https://www.frontiersin.org/people/u/1326652)[6] and Adam M. Brufsky[7]

[1]Knight's Campus for Accelerating Scientific Impact, University of Oregon, Eugene, OR, United States
[2]Dr. Shroff's Charity Eye Hospital, New Delhi, India
[3]PanTherapeutics, Lutry, Switzerland
[4]Department of Molecular Medicine, University of Padova, Padova, Italy
[5]Department of Physiology, Michigan State University, East Lansing, MI, United States
[6]Department of Molecular Medicine, Morsani College of Medicine, University of South Florida (USF) Health Byrd Alzheimer's Institute, University of South Florida, Tampa, FL, United States
[7]Division of Hematology/Oncology, Department of Medicine, University of Pittsburgh Medical Center (UPMC) Hillman Cancer Center, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

Among numerous point mutation differences between the SARS-CoV-2 and the bat RaTG13 coronavirus, only the 12-nucleotide furin cleavage site (FCS) exceeds 3 nucleotides. A BLAST search revealed that a 19 nucleotide portion of the SARS.Cov2 genome encompassing the furing cleavage site is a 100% complementary match to a codon-optimized proprietary sequence that is the reverse complement of the human mutS homolog (MSH3). The reverse complement sequence present in SARS-CoV-2 may occur randomly but other possibilities must be considered. Recombination in an intermediate host is an unlikely explanation. Single stranded RNA viruses such as SARS-CoV-2 utilize negative strand RNA templates in infected cells, which might lead through copy choice recombination with a negative sense SARS-CoV-2 RNA to the integration of the MSH3 negative strand, including the FCS, into the viral genome. In any case, the presence of the 19-nucleotide long RNA sequence including the FCS with 100% identity to the reverse complement of the MSH3 mRNA is highly unusual and requires further investigations.

Introduction

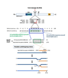
Based on a recent publication describing insertion variants of SARS-CoV-2 (1) we would like to bring the attention to our recent findings related to the sequence of the furin cleavage site (FCS) in SARS-CoV-2 Spike (S) protein. The SARS-CoV-2 causing the COVID-19 pandemic (2) has 82.3% amino acid identity to bat coronavirus SL-CoVZC45, 77.2% amino acid identity to SARS-CoV, and 96.2% genome sequence identity to bat coronavirus RaTG13. While numerous point mutation differences exist between SARS-CoV-2 and RaTG13, only one insertion and dissimilarity exceeding 3 nucleotides (nt): a 12-nucleotide insertion coding for four amino acids (aa 681-684, PRRA) in the SARS-CoV-2 S protein has been discovered. This polybasic FCS differentiates SARS-CoV-2 from other b-lineage betacoronaviruses or any other sarbecovirus (3). An FCS addition enhanced the infectivity of SARS Co-V-2 in 2019 (4). The absence of this FCS results in attenuated SARS-CoV-2 variants useful for animal vaccination, accentuating its relevance to human infection (5). This FCS is vital for human and ferret transmission (6), expands viral tropism to human cells (7), and is requisite for severe disease in two animal models of SARS-CoV-2 (8).

SARS-CoV-2 Spike Protein and MSH3

A peculiar feature of the nucleotide sequence encoding the PRRA furin cleavage site in the SARS-CoV-2 S protein is its two consecutive CGG codons. This arginine codon is rare in coronaviruses: relative synonymous codon usage (RSCU) of CGG in pangolin CoV is 0, in bat CoV 0.08, in SARS-CoV 0.19, in MERS-CoV 0.25, and in SARS-CoV-2 0.299 (9).

A BLAST search for the 12-nucleotide insertion led us to a 100% reverse match in a proprietary sequence (SEQ ID11652, nt 2751-2733) found in the US patent 9,587,003 filed on Feb. 4, 2016 (10) (Figure 1). Examination of SEQ ID11652 revealed that the match extends beyond the 12-nucleotide insertion to a 19-nucleotide sequence: 5′-CTACGTGCCCGCCGAGGAG-3′ (nt 2733-2751 of SEQ ID11652), such that the resulting mRNA would have 3′- GAUGCACGGGCGGCUCCUC-5′, or equivalently 5′- CU **CCU CGG CGG GCA** CGU AG-3′ (nucleotides 23547-23565 in the SARS-CoV-2 genome, in which the four bold codons yield PRRA, amino acids 681–684 of its spike protein). This is very rare in the NCBI BLAST database.

FIGURE 1



(https://www.frontiersin.org/files/Articles/834808/fviro-02-834808-HTML/image_m/fviro-02-834808-g001.jpg)
Figure 1. The origin of the furin sequence in SARS-CoV-2. Comparison of the protein sequences at the S1/S2 junction in SARS-CoV, RaTG13, and SARS-CoV-2 demonstrating the presence of the furin cleavage site (FCS) PRRA only in SARS-CoV-2. Based on a BLAST search of the 12-nucleotide stretch coding for the FCS PRRA, a 19-nucleotide long identical sequence was identified in the patented (US 958 7003) sequence Seq ID11652. SEQ ID11652 is transcribed to a MSH3 mRNA that appears to be codon optimized for humans. This 19-nucleotide sequence including 12 nucleotides coding for the FCS PRRA, present in the human MSH3 gene might have been introduced into the SARS-CoV-2 genome by the illustrated copy choice recombination mechanism in SARS-CoV-2 infected human cells overexpressing the MSH3 gene.

The correlation between this SARS-CoV-2 sequence and the reverse complement of a proprietary mRNA sequence is of uncertain origin. Conventional biostatistical analysis indicates that the probability of this sequence randomly being present in a 30,000-nucleotide viral genome is 3.21×10^{-11} (Figure 2).

FIGURE 2

(https://www.frontiersin.org/files/Articles/834808/fviro-02-834808-HTML/image_m/fviro-02-834808-g002.jpg)
Figure 2. Calculations of the probability of natural occurrence of the 19nt sequence under study. The SARS-CoV-2 genome is ~30,000 nucleotides long (P1). The patented sequence is ~3,300 nucleotides long (P2). The patented library encompasses 24'712 sequences of varying lengths with median length being in the range of 3,300 nucleotides. Conventional probability calculations are given of the probability of the presence of a 19-nucleotide sequence in the human genome and in one of the patented library sequences.

The proprietary sequence SEQ ID11652, read in the forward direction, encodes a 100% amino acid match to the human mut S homolog 3 (MSH3) (9). MSH3 is a DNA mismatch repair protein (part of the MutS beta complex) (11). SEQ ID11652 is transcribed to a MSH3 mRNA that appears to be codon optimized for humans (12). We did not find the 19-nucleotide sequence CTCCTCGGCGGGCACGTAG in any eukaryotic or viral genomes except SARS-CoV-2 with 100% coverage and identity in the BLAST database (Supplementary Tables 1–3).

Discussion

MSH3 replacement with a codon-optimized mRNA sequence for human expression likely has applications in cancers with mismatch repair deficiencies. While a portion of a reverse complement sequence being present in SARS-CoV-2 could be a random coincidence, other possibilities merit consideration.

Overexpression of MSH3 is known to interfere with mismatch repair (MSH2 sequestration from the MutS alpha complex comprising MSH2 and MSH6 results in MSH6 degradation and MutS alpha depletion) (13), which holds virologic importance. Induction of DNA mismatch repair deficiency results in permissiveness of influenza A virus (IAV) infection of human respiratory cells and increased pathogenicity (14). Mismatch repair deficiency may extend shedding of SARS-CoV-2 (15, 16).

The absence of CTCCTCGGCGGGCACGTAG from any eukaryotic or viral genome in the BLAST database makes recombination in an intermediate host an unlikely explanation for its presence in SARS-CoV-2. A human-codon-optimized mRNA encoding a protein 100% homologous to human MSH3 could, during the course of viral research, inadvertently or intentionally induce mismatch repair deficiency in a human cell line, which would increase susceptibility to SARS-like viral infection. Infection of SEQ ID11652-MSH3-transduced human cells by a SARS-like virus could enable copy choice recombination (15). Replication of SARS-CoV-2 and other single stranded RNA viruses with an RNA genome of positive polarity is initiated by the synthesis of negative strand RNA in the cytoplasm of infected cells (17) (Figure 1). The negative strand RNA is a template for synthesis of positive stranded RNA utilized for translation of non-structural proteins, the replication and transcription complex, or new virion capsids. Coronaviruses generate double stranded RNA at an early stage of infection through genomic replication and mRNA transcription (18).

Acquisition of the reverse complement FCS sequence from an overexpressed positive sense MSH3 mRNA could occur through copy choice recombination with a negative sense SARS-CoV-2 RNA intermediate (15), involving jumping from one template to another (19) (Figure 1). The homology between SARS-CoV-2 and other known coronaviruses is discontinued and most SARS-CoV-2 sequences derive from a relatively recent common ancestor with bat RaTG13. Moreover, similarity plots (SimPlots) have identified sudden changes in sequence identity between SARS-CoV-2 and RaTG13, signaling potential recombination events, which could explain the capability of SARS-CoV-2 binding to ACE2 through its RBD, which is not the case for the RaTG13 RBD (15).

A criticism of this hypothesis is that the identified sequence is on the opposite strand of the open reading frame in SEQ ID11652. However, cells transfected with MSH3, which induce mismatch repair deficiency could have targeted double-stranded cDNA encoding SEQ ID11652. Such cells co-transfected with a SARS-like virus expressing RdRp could attach to this 19-nucleotide sequence (15) and permit integration of a fragment from

(https://www.frontiersin.org/search?tab=top-results&origin=https://www.frontiersin.org/articles/10.3389/fviro.2022.834808/full)

the negative strand into the viral genome, including the FCS, despite being on the opposite strand of the open reading frame. Mismatch repair mechanisms have enabled integration of short fragments from antisense strands in experimental models (20, 21). Microhomology can direct recombination between the MSH3 and a SARS-like virus, which could take place at the 19-nucleotide sequence of interest.

The presence in SARS-CoV-2 of a 19-nucleotide RNA sequence encoding an FCS at amino acid 681 of its spike protein with 100% identity to the reverse complement of a proprietary MSH3 mRNA sequence is highly unusual. Potential explanations for this correlation should be further investigated.

Data Availability Statement

Publicly available datasets were analyzed in this study. This data can be found here: SEQ ID11652.

Author Contributions

The investigation and original draft were initiated by BA, AV, and AB. The visualization was performed by BA, AV, AB, and GP. The writing and editing were done by BA, AV, KL, GP, BU, VU, and AB. All authors contributed to the article and approved the submitted version.

Conflict of Interest

KL was employed by PanTherapeutics.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Acknowledgments

We are thankful for the contribution of Dr. Jian Ying, University of Utah, Eureka, USA to the probability analysis.

Supplementary Material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fviro.2022.834808/full#supplementary-material (https://www.frontiersin.org/articles/10.3389/fviro.2022.834808/full#supplementary-material)

References

1. Garushyants SK, Rogozin IB, Koonin EV. Template switching and duplications in SARS-CoV-2 genomes give rise to insertion variants that merit monitoring. *Commun Biol.* (2021) 4:1343. doi: 10.1038/s42003-021-02858-9

   PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/34848826) | CrossRef Full Text (https://doi.org/10.1038/s42003-021-02858-9) | Google Scholar (http://scholar.google.com/scholar_lookup?author=SK+Garushyants&author=IB+Rogozin&author=EV+Koonin+&publication_year=2021&title=Template+switching+and+duplications+in+SARS-CoV-2+genomes+give+rise+to+insertion+variants+that+merit+monitoring&journal=Commun+Biol.&volume=4&pages=1343)

2. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature.* (2020) 579:265–9. doi: 10.1038/s41586-020-2008-3

   PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/32296181) | CrossRef Full Text (https://doi.org/10.1038/s41586-020-2008-3) | Google Scholar (http://scholar.google.com/scholar_lookup?author=F+Wu&author=S+Zhao&author=B+Yu&author=YM+Chen&author=W+Wang&author=ZG+Song+&publication_year=2020&title=A+new+coronavirus+associated+with+human+respiratory+disease+in+China&journal=Nature&volume=579&pages=265-9)

3. Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res.* (2020) 176:104742. doi: 10.1016/j.antiviral.2020.104742

   PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/32057769) | CrossRef Full Text (https://doi.org/10.1016/j.antiviral.2020.104742) | Google Scholar (http://scholar.google.com/scholar_lookup?author=B+Coutard&author=C+Valle&author=X+de+Lamballerie&author=B+Canard&author=NG+Seidah&author=E+Decroly+&publication_year=2020&title=The+spike+glycoprotein+of+the+new+coronavirus+2019-nCoV+contains+a+furin-like+cleavage+site+absent+in+CoV+of+the+same+clade&journal=Antiviral+Res.&volume=176&pages=104742)

4. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell.* (2020) 181:281–92.e286. doi: 10.1016/j.cell.2020.02.058

   PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/33306958) | CrossRef Full Text (https://doi.org/10.1016/j.cell.2020.02.058) | Google Scholar (http://scholar.google.com/scholar_lookup?author=AC+Walls&author=YJ+Park&author=MA+Tortorici&author=A+Wall&author=AT+McGuire&author=D+Veesler+&publication_year=2020&title=Structure,+Function,+and+Antigenicity+of+the+SARS-CoV-2+Spike+Glycoprotein&journal=Cell&volume=181&pages=281-92)

5. Lau SY, Wang P, Mok BWY, Zhang AJ, Chu H, Lee ACY, et al. Attenuated SARS-CoV-2 variants with deletions at the S1/S2 junction. *Emerg Microbes Infect.* (2020) 9:837–42. doi: 10.1080/22221751.2020.1756700

   PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/32301390) | CrossRef Full Text (https://doi.org/10.1080/22221751.2020.1756700) | Google Scholar (http://scholar.google.com/scholar_lookup?author=SY+Lau&author=P+Wang&author=BWY+Mok&author=AJ+Zhang&author=H+Chu&author=ACY+Lee+&publication_year=2020&title=Attenuated+SARS-CoV-2+variants+with+deletions+at+the+S1%2FS2+junction&journal=Emerg+Microbes+Infect&volume=9&pages=837-42)

6. Peacock TP, Goldhill DH, Zhou J, Baillon L, Frise R, Swann OC, et al. The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat Microbiol.* (2020) 6:899–909. doi: 10.1038/s41564-021-00908-w

   PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/33907312) | CrossRef Full Text (https://doi.org/10.1038/s41564-021-00908-w) | Google Scholar (http://scholar.google.com/scholar_lookup?author=TP+Peacock&author=DH+Goldhill&author=J+Zhou&author=L+Baillon&author=R+Frise&author=OC+Swann+&publication_year=2020&title=The+furin+cleavage+site+in+the+SARS-CoV-2+spike+protein+is+required+for+transmission+in+ferrets&journal=Nat+Microbiol&volume=6&pages=899-909)

7. Xia S, Lan Q, Su S, Wang X, Xu W, Liu Z, et al. The role of furin cleavage site in SARS-CoV-2 spike protein-mediated membrane fusion in the presence or absence of trypsin. *Signal Transd. Targeted Ther.* (2020) 5:92. doi: 10.1038/s41392-020-0184-0

   PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/32532959) | CrossRef Full Text (https://doi.org/10.1038/s41392-020-0184-0) | Google Scholar (http://scholar.google.com/scholar_lookup?author=S+Xia&author=Q+Lan&author=S+Su&author=X+Wang&author=W+Xu&author=Z+Liu+&publication_year=2020&title=The+role+of+furin+cleavage+site+in+SARS-CoV-2+spike+protein-mediated+membrane+fusion+in+the+presence+or+absence+of+trypsin&journal=Signal+Transd.+Targeted+Ther&volume=5&pages=92)

8. Hou W. Characterization of codon usage pattern in SARS-CoV-2. *Virology J.* (2020) 17:138. doi: 10.1186/s12985-020-01395-x

   PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/32928234) | CrossRef Full Text (https://doi.org/10.1186/s12985-020-01395-x) | Google Scholar (http://scholar.google.com/scholar_lookup?author=W+Hou+&publication_year=2020&title=Characterization+of+codon+usage+pattern+in+SARS-CoV-2&journal=Virology+J.&volume=17&pages=138)

9. Kandeel M, Ibrahim A, Fayez M, Al-Nazawi M. From SARS and MERS CoVs to SARS-CoV-2: Moving toward more biased codon usage in viral structural and nonstructural genes. *J Med Virol.* (2020) 92:660–6. doi: 10.1002/jmv.25754

   PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/32159237) | CrossRef Full Text (https://doi.org/10.1002/jmv.25754) | Google Scholar (http://scholar.google.com/scholar_lookup?author=M+Kandeel&author=A+Ibrahim&author=M+Fayez&author=M+Al-Nazawi+&publication_year=2020&title=From+SARS+and+MERS+CoVs+to+SARS-CoV-2%3A+Moving+toward+more+biased+codon+usage+in+viral+structural+and+nonstructural+genes&journal=J+Med+Virol&volume=92&pages=660-6)

10. Bancel S, Chakraborty T, De Fougerolles A, Elbashir SM, John M, Roy A, et al. *Modified Polynucleotides for the Production of Oncology-Related Proteins and Peptides*. Cambridge, MA: United States Patent. (2016).

    Google Scholar (http://scholar.google.com/scholar_lookup?author=S+Bancel&author=T+Chakraborty&author=A+De+Fougerolles&author=SM+Elbashir&author=M+John&author=A+Roy+&publication_year=2016&title=Modified+Polynucleotides+for+the+Production+of+Oncology-Related+Proteins+and+Peptides)

11. MacRae SL, McKnight Croken M, Calder RB, Aliper A, Milholland B, White RR, et al. DNA repair in species with extreme lifespan differences. *Aging.* (2015) 7:1171–84. doi: 10.18632/aging.100866

    PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/26729707) | CrossRef Full Text (https://doi.org/10.18632/aging.100866) | Google Scholar (http://scholar.google.com/scholar_lookup?author=SL+MacRae&author=M+McKnight+Croken&author=RB+Calder&author=A+Aliper&author=B+Milholland&author=RR+White+&publication_year=2015&title=DNA+repair+in+species+with+extreme+lifespan+differences&journal=Aging&volume=7&pages=1171-84)

12. Mauro VP, Chappell SA. A critical analysis of codon optimization in human therapeutics. *Trends Mol Med.* (2014) 20:604–13. doi: 10.1016/j.molmed.2014.09.003

    PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/25263172) | CrossRef Full Text (https://doi.org/10.1016/j.molmed.2014.09.003) | Google Scholar (http://scholar.google.com/scholar_lookup?author=VP+Mauro&author=SA+Chappell+&publication_year=2014&title=A+critical+analysis+of+codon+optimization+in+human+therapeutics&journal=Trends+Mol+Med.&volume=20&pages=604-13)

13. Marra G, Iaccarino I, Lettieri T, Roscilli G, Delmastro P, Jiricny J. Mismatch repair deficiency associated with overexpression of the MSH3 gene. *Proc Natl Acad Sci USA.* (1998) 95:8568. doi: 10.1073/pnas.95.15.8568

    PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/9671718) | CrossRef Full Text (https://doi.org/10.1073/pnas.95.15.8568) | Google Scholar (http://scholar.google.com/scholar_lookup?author=G+Marra&author=I+Iaccarino&author=T+Lettieri&author=G+Roscilli&author=P+Delmastro&author=J+Jiricny+&publication_year=1998&title=Mismatch+repair+deficiency+associated+with+overexpression+of+the+MSH3+gene&journal=Proc+Natl+Acad+Sci+USA&volume=95&pages=8568)

14. Chambers BS, Heaton BE, Rausch K, Dumm RE, Hamilton JR, Cherry S, et al. DNA mismatch repair is required for the host innate response and controls cellular fate after influenza virus infection. *Nat Microbiol.* (2019) 4:1964–77. doi: 10.1038/s41564-019-0509-3

    PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/31358986) | CrossRef Full Text (https://doi.org/10.1038/s41564-019-0509-3) | Google Scholar (http://scholar.google.com/scholar_lookup?author=BS+Chambers&author=BE+Heaton&author=K+Rausch&author=RE+Dumm&author=JR+Hamilton&author=S+Cherry+&publication_year=2019&title=DNA+mismatch+repair+is+required+for+the+host+innate+response+and+controls+cellular+fate+after+influenza+virus+infection&journal=Nat+Microbiol&volume=4&pages=1964-77)

15. Gallaher WR. A palindromic RNA sequence as a common breakpoint contributor to copy-choice recombination in SARS-COV-2. *Arch Virol.* (2020) 165:2341–8. doi: 10.1007/s00705-020-04750-z

    PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/32737584) | CrossRef Full Text (https://doi.org/10.1007/s00705-020-04750-z) | Google Scholar (http://scholar.google.com/scholar_lookup?author=WR+Gallaher+&publication_year=2020&title=A+palindromic+RNA+sequence+as+a+common+breakpoint+contributor+to+copy-choice+recombination+in+SARS-COV-2&journal=Arch+Virol.&volume=165&pages=2341-8)

16. Haque F, Lillie P, Haque F, Maraveyas A. Deficient DNA mismatch repair and persistence of SARS-CoV-2 RNA shedding: a case report of Hereditary Nonpolyposis Colorectal Cancer with COVID-19 infection. *BMC Infect Dis.* (2020) 21:854. doi: 10.1186/s12879-021-06500-1

    PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/34418963) | CrossRef Full Text (https://doi.org/10.1186/s12879-021-06500-1) | Google Scholar (http://scholar.google.com/scholar_lookup?author=F+Haque&author=P+Lillie&author=F+Haque&author=A+Maraveyas+&publication_year=2020&title=Deficient+DNA+mismatch+repair+and+persistence+of+SARS-CoV-2+RNA+shedding%3A+a+case+report+of+Hereditary+Nonpolyposis+Colorectal+Cancer+with+COVID-19+infection&journal=BMC+Infect+Dis.&volume=21&pages=854)

17. V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev.* (2021) 19:155–70. doi: 10.1038/s41579-020-00468-6

    PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/33116300) | CrossRef Full Text (https://doi.org/10.1038/s41579-020-00468-6) | Google Scholar (http://scholar.google.com/scholar_lookup?author=P+V'kovski&author=A+Kratzel&author=S+Steiner&author=H+Stalder&author=V+Thiel+&publication_year=2021&title=Coronavirus+biology+and+replication%3A+implications+for+SARS-CoV-2&journal=Nat+Rev&volume=19&pages=155-70)

18. Graham RL, Baric RS. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J Virol.* (2010) 84:3146. doi: 10.1128/JVI.01394-09

    PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/19906932) | CrossRef Full Text (https://doi.org/10.1128/JVI.01394-09) | Google Scholar (http://scholar.google.com/scholar_lookup?author=RL+Graham&author=RS+Baric+&publication_year=2010&title=Recombination,+reservoirs,+and+the+modular+spike%3A+mechanisms+of+coronavirus+cross-species+transmission&journal=J+Virol.&volume=84&pages=3146)

19. Sola I, Almazan F, Zuniga S, Enjuanes L. Continuous and discontinuous RNA synthesis in coronaviruses. *Annu Rev Virol.* (2015) 2:265–88. doi: 10.1146/annurev-virology-100114-055218

    PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/26958916) | CrossRef Full Text (https://doi.org/10.1146/annurev-virology-100114-055218) | Google Scholar (http://scholar.google.com/scholar_lookup?author=I+Sola&author=F+Almazan&author=S+Zuniga&author=L+Enjuanes+&publication_year=2015&title=Continuous+and+discontinuous+RNA+synthesis+in+coronaviruses&journal=Annu+Rev+Virol.&volume=2&pages=265-88)

20. Rakosy-Tican E, Lörincz-Besenyei E, Molnár I, Thieme R, Hartung F, Sprink T, et al. New Phenotypes of potato co-induced by mismatch repair deficiency and somatic hybridization. *Front Plant Sci.* (2019) 10:3. doi: 10.3389/fpls.2019.00003

    PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/30723483) | CrossRef Full Text (https://doi.org/10.3389/fpls.2019.00003) | Google Scholar (http://scholar.google.com/scholar_lookup?author=E+Rakosy-Tican&author=E+Lörincz-Besenyei&author=I+Molnár&author=R+Thieme&author=F+Hartung&author=T+Sprink+&publication_year=2019&title=New+Phenotypes+of+potato+co-induced+by+mismatch+repair+deficiency+and+somatic+hybridization&journal=Front+Plant+Sci.&volume=10&pages=3)

21. Harmsen T, Klaasen S, van de Vrugt H, te Riele H. DNA mismatch repair and oligonucleotide end-protection promote base-pair substitution distal from a CRISPR/Cas9- induced DNA break. *Nucl Acids Res.* (2018) 46:2945–55. doi: 10.1093/nar/gky076

    PubMed Abstract (https://pubmed.ncbi.nlm.nih.gov/29447381) | CrossRef Full Text (https://doi.org/10.1093/nar/gky076) | Google Scholar (http://scholar.google.com/scholar_lookup?author=T+Harmsen&author=S+Klaasen&author=H+van+de+Vrugt&author=H+te+Riele+&publication_year=2018&title=DNA+mismatch+repair+and+oligonucleotide+end-protection+promote+base-pair+substitution+distal+from+a+CRISPR%2FCas9-+induced+DNA+break&journal=Nucl+Acids+Res&volume=46&pages=2945-55)

Keywords: SARS-CoV-2 spike, furin cleavage site, MSH3 gene, sequence homology, recombinability

Citation: Ambati BK, Varshney A, Lundstrom K, Palú G, Uhal BD, Uversky VN and Brufsky AM (2022) MSH3 Homology and Potential Recombination Link to SARS-CoV-2 Furin Cleavage Site. *Front. Virol.* 2:834808. doi: 10.3389/fviro.2022.834808

Received: 13 December 2021; **Accepted:** 19 January 2022;
Published: 21 February 2022.

Edited by:
Xin Yin (https://loop.frontiersin.org/people/1218113/overview), Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences (CAAS), China

Reviewed by:
Jitao Chang (https://loop.frontiersin.org/people/1601192/overview), Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences (CAAS), China

***Correspondence:** Kenneth Lundstrom, **lundstromkenneth@gmail.com (mailto:lundstromkenneth@gmail.com)**

---

11 Comments - Christophe Brun (https://loop.frontiersin.org/people/1690435/overview), John Smith (https://loop.frontiersin.org/people/1693447/overview), Martijn Weterings (https://loop.frontiersin.org/people/1693818/overview), Vincent Joseph (https://loop.frontiersin.org/people/19516/overview), Seth Crosby (https://loop.frontiersin.org/people/1706555/overview), Fernando Cardona (https://loop.frontiersin.org/people/103175/overview) and **Jon Burg (https://loop.frontiersin.org/people/1715535/overview)**.

(https://loop.frontiersin.org/people/1690435/overview) **Christophe Brun (https://loop.frontiersin.org/people/1690435/overview)** Really interesting but in the introduction, the sentence "An FCS addition enhanced the infectivity of SARS Co-V-2 in 2019 (4)" makes me think the study in reference 4 date from 2019 which does not seem to be the case in any way. The article date from 2020, it relates to previously published articles of course but... It can create a false sense that SARS Co-V-2 was studied in 2019.

10:38 AM, 22 February 2022

(https://loop.frontiersin.org/people/1693447/overview) **John Smith (https://loop.frontiersin.org/people/1693447/overview)** Probability theory states that P(A and B) = P(A) x P(B) only if events A and B are independent events.

The calculations shown in Figure 2 of the article are:

P(19nt sequence in humans appears in the virus's 30,000nt genome)
= (30,000 - 18) x 10^-7

These two probabilities are multiplied, but there is no evidence given in the article that a sequence appearing in humans is independent from it appearing in the virus. In fact, the opposite may be true. The virus is not a random sequence. The virus is being studied because it matches the sequence in humans, making it more able to infect humans. These are not shown to be independent events, and thus it is incorrect to multiply their probabilities.

P(19nt sequence in humans appears in one of the 24712 sequences of 3300nt patented by Moderna)
= 24712 x P2 x (1-P2)^24711

The article does not provide any evidence that these are independent events either. On the contrary, a pharmaceutical company which aims to treat illnesses in humans is probably more likely to patent sequences from the human genome. Thus, these are likely not independent events either, and their probabilities cannot be multiplied.

These issues may increase the final probability of 3.21 ×10^-11 shown in the article, making the article inaccurate. If possible, would you be able to offer an explanation?

4:05 AM, 24 February 2022

User
Comment deleted on 10:47 AM, 24 February 2022

(https://loop.frontiersin.org/people/1693818/overview) **Martijn Weterings (https://loop.frontiersin.org/people/1693818/overview)** John Smith points out how the probability theory is a simplification, but even if we would consider the nucleotides completely random and independent then still the conventional biostatistical analysis contains several mayor errors. Some of them increase the probability others decrease the probability.

PROBLEM 1

The probabilities P1 and P2: the occurrence of an m-length nucleotide sequence in a random nucleotide sequence of length n. The computation computes the probability of the sequence of length m in a particular sequence (0.25)^m and multiplies this with n-(m-1).

(n-m+1)*(0.25)^m

We can very clearly show that this is wrong by using a smaller problem. What is the probability that a nucleotide T (length m=1) occurs in a random genome of length n=4. It is *not* 4*0.25^1 = 1. That is the expectation value for the number of occurances. But the probability for 1 or more occurances is more like:

1-(1-(0.25)^m)^(n-m+1)

And... that is getting closer, but *also* wrong. This considers the probability as n-m+1 independent flips of a coin with probability (0.25)^m. But the flips are not independent.

To get a more precise results we can consider the problem: A fair die is rolled 1,000 times. What is the probability of rolling the same number 5 times in a row? https://stats.stackexchange.com/questions/492000/a-fair-die-is-rolled-1-000-times-what-is-the-probability-of-rolling-the-same-nu

The solution of this problem can be computed in several ways. The fastest method is to compute it with a Markov Chain. Below is a program in R that does this computation:

create Markov Chain matrix
M = matrix(rep(0,20*20),20)
M[1,1:19] = 3/4
for (i in 1:19) {M[i+1,i] = 1/4}
M[20,20] = 1

compute probability
start <- c(1,rep(0,19))
matrix.power(M,3300) %*% start

The result is 9.0 * 10^-9 (8.955794e-09) which is a *smaller* probability than the 11.9 * 10^-9 from the article, but not a big difference.

PROBLEM 2

The article computes the probability that "the probability of this sequence randomly being present in a 30,000-nucleotide viral genome". But this is a red herring. It is not the relevant probability under consideration.

Namely, the article computes the probability for a *specific sequence* to occur in the viral genome and in the database of 24712 sequences.

This is like computing the probability of rolling two six sided dices with the same number as p1*p2 = 1/6*1/6 = 1/36. But every fan of board games knows that the probabiliy of the same number on the two dice is 1/6 instead of 1/36. We can roll two ones, two twos, two threes, two fours, two fives or two sixes. Each situation has 1/36 probability, and together the probabiltiy is 1/6.

PROBLEM 3

There is a problem of multiple comparisons.

The article focuses on a sequence CCTCGGCGGGCA and finds a match in a database for a piece of length 19 around it with 2 pairs in front and 5 pairs behind it CT-CCTCGGCGGGCA-CGTAG. The computation of te probability is for this specific match of the sequence CT-CCTCGGCGGGCA-CGTAG, but we could have also had other matches of length 19 TAATTCT-CCTCGGCGGGCA, AATTCT-CCTCGGCGGGCA-C, ATTCT-CCTCGGCGGGCA-CG, TTCT-CCTCGGCGGGCA-CGT, TTCT-CCTCGGCGGGCA-CGT, TCT-CCTCGGCGGGCA-CGTA, CT-CCTCGGCGGGCA-CGTAT, T-CCTCGGCGGGCA-CGTATG, CCTCGGCGGGCA-CGTATGT, as well as the sequences in reverse order.

This is a complicated multiple comparisons problem which needs to be computed more exactly (it is complicated due to the correlations between the individiual cases) but it is certain that it will considerably *increase* the probability.

Conclusion

The problem 1 will decrease the probability, but not by much. The problems 2 and 3 will increase the probability and by a lot.

Exact computations need to be done for problem 3 but it is probably gonna be some order times of the probability P2c (which in the altered computation is the slightly less 0.0002212666 instead of 0.00029).

10:47 AM, 24 February 2022

(https://loop.frontiersin.org/people/19516/overview) **Vincent Joseph (https://loop.frontiersin.org/people/19516/overview)** Relevant comments here: ----- https://sciencebasedmedicine.org/the-return-of-the-revenge-of-covid-19-mrna-vaccines-permanently-alter-your-dna-and-lab-leak ----- Frontiers is sadly becoming something it should not be.

7:59 PM, 03 March 2022

(https://loop.frontiersin.org/people/1706555/overview) **Seth Crosby (https://loop.frontiersin.org/people/1706555/overview)** That stretch of 19 bases (CTACGTGCCCGCCGAGGAG) is not unique to SarsCoV2 and Moderna; it occurs in a number of microorganisms. Using BLAST, a simple online tool, I compared it to all known genomes (excluded SARSCoV2 to see what other organisms contain it). Dozens of perfect hits. Try it yourself! https://blast.ncbi.nlm.nih.gov/Blast.cgi

2:16 PM, 04 March 2022

(https://loop.frontiersin.org/people/103175/overview) **Fernando Cardona (https://loop.frontiersin.org/people/103175/overview)** "The absence of CTCCTCGGCGGGCACGTAG from any eukaryotic or viral genome in the BLAST database makes recombination in an intermediate host an unlikely explanation for its presence in SARS-CoV-2". The last assertion, taken from the discussion in that paper, is the result of a poorly use of the BLAST tool. Indeed, more careful searches with this 19-nucleotide long sequence in BLAST and using stringent parameters (100% identity and 100% coverage) and restricted to different phyla/genus/species (virus excluding SARS-CoV-2, birds, bacteria, bats or reference RNA sequences) provide numerous hits. This demonstrates the existence of this exact sequence in numerous other eukaryotes, prokaryotes or viruses. The results of our basic BLAST search certainly questions the idea that this could only happen when a coronavirus infected a human being, as the paper seem to suggest.

10:03 AM, 07 March 2022

User
Comment deleted on 2:52 PM, 07 March 2022

User
Comment deleted on 2:52 PM, 07 March 2022

(https://loop.frontiersin.org/people/103175/overview) **Fernando Cardona (https://loop.frontiersin.org/people/103175/overview)** CTCCTCGGCGGGCACGTAG
-Virus, excluding SARS-CoV-2, RefSeq representative genomes
100%identity and 100% coverage: phages, herpesvirus, poxvirus, halovirus, adenovirus, ... (2372 sequences)
e-value<1: Pandoravirus (2), Orf virus, Murid herpesvirus 1, Murid betaherpesvirus 1
-Virus, RefSeq representative genomes, excluding all phages (8 types)
-RefSeq representative genomes, birds. 100%identity and 100% coverage
Chaetura, Aquila, Ansa, Oxyura, Aythya, Cygnus, Serinus, Gallus
- RefSeq representative genomes, Bacteria. 100%identity and 100% coverag: 68 sequences
-Reference RNA sequences. 100%identity and 100% coverage: Saprolegnia parasítica (Protista), Saprolegnia diclina (Protista), Chaetura pelágica (bird).
-Bats. 100%identity and 100% coverage: Pipistrellus, Molossus, Phyllostomus, Rousettus, Pteropus (14 sequences).

2:53 PM, 07 March 2022

(https://loop.frontiersin.org/people/1715535/overview) **Jon Burg (https://loop.frontiersin.org/people/1715535/overview)** Is anyone else floored that the reviewers of this manuscript allowed authors to make these claims and did not require them to follow-up with any actual lab work?

6:22 PM, 10 March 2022

Write a comment...

Add

TABLE OF CONTENTS

Campbell, Anguo Liu, Andrew Pekosz, Irina Burd and Sabra L. Klein (https://loop.frontiersin.org/people/265733/overview)

Editorial: Innovative Approaches in Diagnosis of Emerging/Re-emerging Infectious Diseases (/articles/10.3389/fmicb.2020.619498/full)
Aleksandra Barac (https://loop.frontiersin.org/people/405750/overview), Mario Poljak (https://loop.frontiersin.org/people/684064/overview) and David S. Y. Ong (https://loop.frontiersin.org/people/540667/overview)

Saliva for COVID-19 Testing: Simple but Useless or an Undervalued Resource? (/articles/10.3389/fviro.2021.778790/full)
Sara Pijuan-Galito, Francesco Saverio Tarantini, Hannah Tomlin, Harry Jenkins, Jamie Louise Thompson (https://loop.frontiersin.org/people/1521690/overview), Danielle Scales, Amy Stroud, Ana Tellechea Lopez, James Hassall, Philip G. McTernan (https://loop.frontiersin.org/people/1540438/overview), Andy Coultas (https://loop.frontiersin.org/people/1517919/overview), Asta Arendt-Tranholm, Caroline Reffin, Ian Hill, I-ning Lee, Siyu Wu, Joanne Porte (https://loop.frontiersin.org/people/1486297/overview), Joseph Chappell, Katarzyna Lis-Slimak, Kazuyo Kaneko, Lara Doolan, Mairead Ward, Martin Stonebridge, Mohammad Ilyas, Patrick McClure (https://loop.frontiersin.org/people/905581/overview), Patrick Tighe, Penny Gwynne, Ralph Hyde, Jonathan Ball, Claire Seedhouse, Andrew V. Benest, Moira Petrie and Chris Denning (https://loop.frontiersin.org/people/1483525/overview)

Genomic Characterization of the Emerging SARS-CoV-2 Lineage in Two Districts of Campania (Italy) Using Next-Generation Sequencing (/articles/10.3389/fviro.2022.814114/full)
Marianna Scrima (https://loop.frontiersin.org/people/441265/overview), Alessia Maria Cossu (https://loop.frontiersin.org/people/954364/overview), Egildo Luca D'Andrea, Marco Bocchetti (https://loop.frontiersin.org/people/778873/overview), Ylenia Abruzzese, Clara Iannarone, Cinzia Miarelli, Piera Grisolia, Federica Melisi, Lucia Genua (https://loop.frontiersin.org/people/1601967/overview), Felice Di Perna, Paolo Maggi, Giovanbattista Capasso (https://loop.frontiersin.org/people/580711/overview), Teresa Maria Rosaria Noviello (https://loop.frontiersin.org/people/820419/overview), Michele Ceccarelli (https://loop.frontiersin.org/people/534474/overview), Alessandra Fucci and Michele Caraglia (https://loop.frontiersin.org/people/775927/overview)

Frontiers in Virology: An Innovative Platform for Integrative Virus Research (/articles/10.3389/fviro.2021.665473/full)
Akio Adachi (https://loop.frontiersin.org/people/14796/overview)

(https://www.frontiersin.org/search?tab=top-results&origin=https://www.frontiersin.org/articles/10.3389/fviro.2022.834808/full)