Genomic Analysis of HCoV-OC43 in Relation to Intra-Host and Temporal Genetic Variation

James Peng[1,2], Thomas Murphy[2], Manish Sagar, MD[2]

Cedar Falls High School, 2701 W 27th St, Cedar Falls, IA 50613[1]; Department of Medicine, Evans Biomedical Research Center, Boston University School of Medicine, 650 Albany Street, Boston, MA 021182[2].

Introduction

- Human coronavirus OC43 (HCoV-OC43) is a widespread seasonal betacoronavirus
- It has circulated for decades and contributes annually to respiratory infections worldwide.
- Despite its prevalence, there is limited genomic data detailing the natural variation present both within individual hosts and across temporal outbreaks
- Understanding the genetic variability of OC43 is essential to investigating its evolutionary characteristics, viral adaptation mechanisms, and how it compares to other betacoronaviruses such as SARS-CoV-2
- The Nucleocapsid(N) is a protein that packages the genome of coronaviruses. It's the most abundant HCoV translation product, and leads to strong antibody and T cell responses. The nucleocapsid was our best sequencing result.
- This study will focus on sequencing and analyzing the complete genome of OC43 from infected individuals. Current studies use over 100 base pair sections to cover the entire genome. To make the amount of sections smaller we attempted to use significantly less sections, with larger sizes of about 10kb.
- We also aim to investigate intra-host and temporal genetic variations, and how they impact viral evolution as a whole.

Methods

- Saliva samples were collected from individuals infected with HCoV-OC43
- Viral RNA was extracted using the Maxwell RSC nucleic acid extractor
- Extracted RNA was reverse transcribed into complementary DNA (cDNA) using Maxima Reverse Transcriptase (RT).
- OC43 specific primers were used to amplify viral sequences by polymerase chain reaction (PCR).
- PCR amplicons were confirmed by gel electrophoresis.
- The amplicons were sequenced using Oxford Nanopore sequencing technology to generate long-read sequence data.
- Sequence reads were aligned with reference genome, variations were identified and analyzed. Patterns were compared with other temporal OC43 strains and SARS-CoV-2

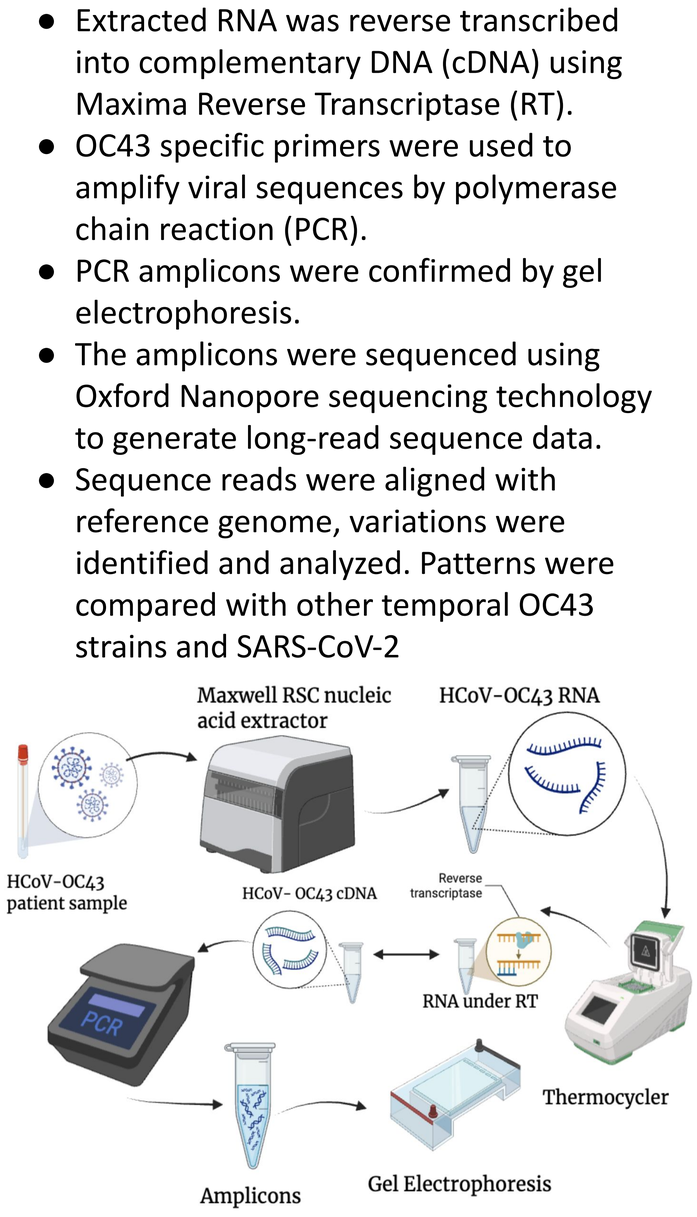

Figure 1: OC43 Extraction and analysis method

Results

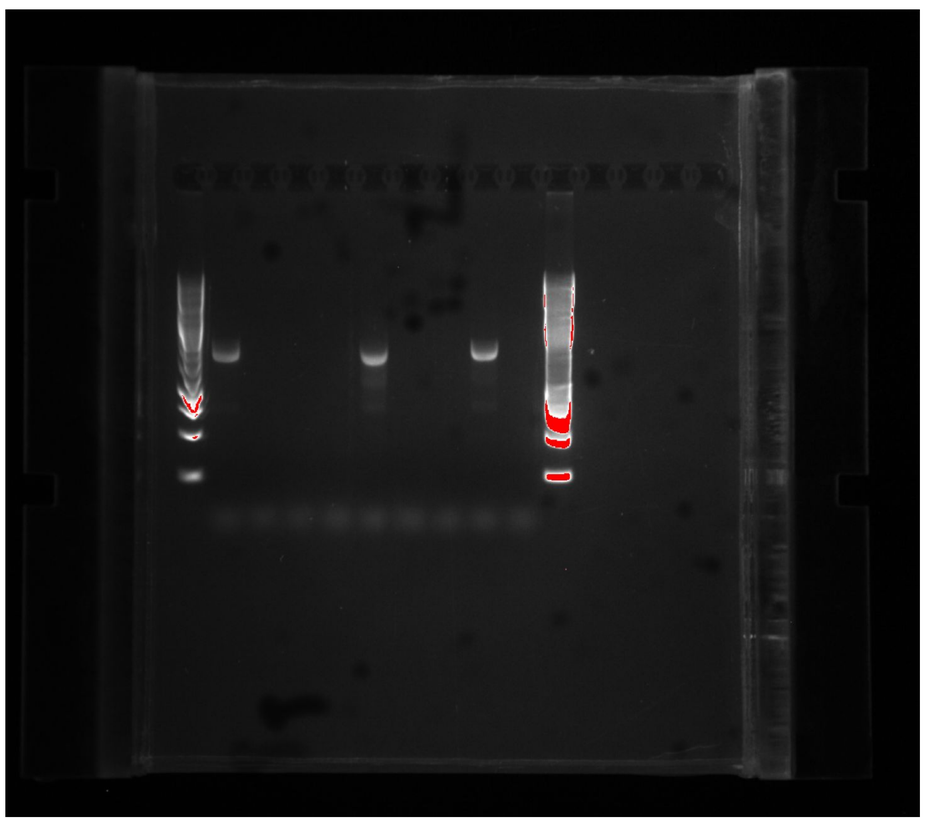

Figure 2: Analysis of patient sample OC43 nucleocapsid by gel electrophoresis. The molecular weight ladder indicates fragments of size 2kb. Horizontal bands correspond to the reference genome and lanes 2, 6, and 9 which were later sequenced.

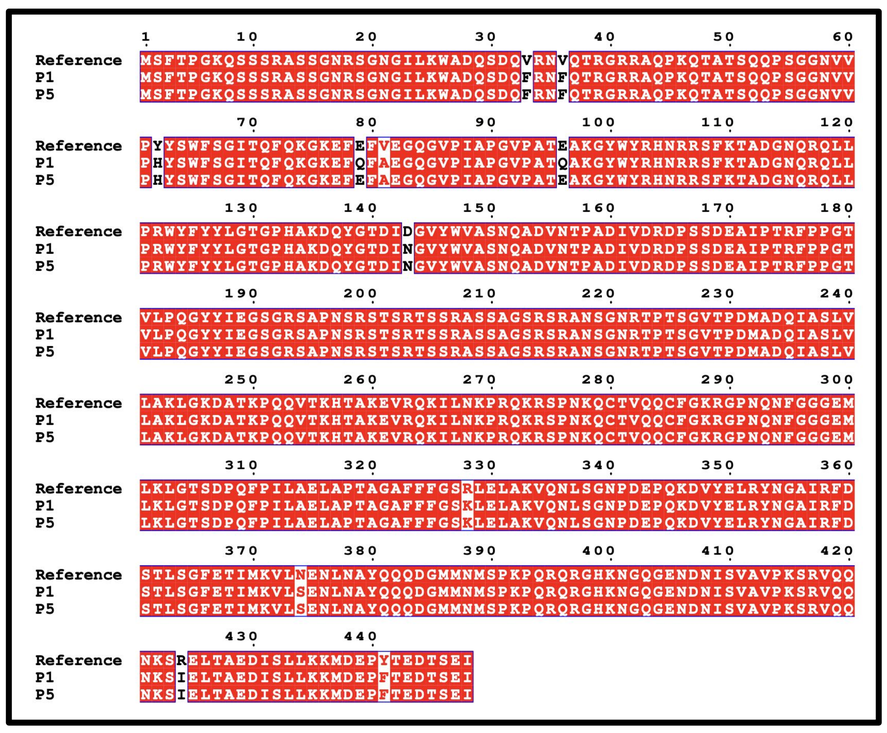

Figure 3: Amino acid sequence comparison. The reference OC43 nucleocapsid amino acid sequence was compared to patient 1 and 5's nucleocapsid amino acid sequences. Note a lack of red highlighting in areas with point mutations.

	Ref	1	5
Ref	1		
1	98.3%(29)	1	
5	98.4%(28)	99.5%(9)	1

Figure 4: Pairwise identities between the nucleocapsid nucleotide sequences of the reference genome, patient 1, and patient 5

.

	Ref vs. P1	Ref vs. P5
Ds	0.0106	0.0086
Dn	0.378	0.0413
Ds/Dn	0.028	0.208

Figure 5: Comparison of dS and dN values in Reference and Patient Samples

.

Discussion/ Conclusions

- Due to the large size of the genome fragments, sequencing the entire genome was not feasible.
- The nucleocapsid region of OC43 was the only fragment that amplified successfully, with sizes around 2 kb.
- After processing patient-derived OC43 samples, two samples yielded successful amplification, corresponding to lanes 2 (Patient 1) and 6 (Patient 5).
- As shown in Figure 4, the dS values for both patients are relatively low, indicating few synonymous mutations. However, the dN values reveal that Patient 1 exhibits higher nonsynonymous substitutions, suggesting more amino acid-altering mutations.
- Despite this, the overall low dS/dN ratios indicate prevailing negative (purifying) selection acting on these sequences. Such selection likely imposes constraints on the virus, as random mutations may reduce viral fitness.
- Figure 3 shows that all three sequences share very high similarity percentages. Nevertheless, the differences observed in dN rates emphasize that even minor sequence variations can result in different selective pressures, potentially impacting viral protein function.
- One possible explanation is that given the nucleocapsid role in immune recognition, it may be subject to stronger functional constraints and therefore less evolutionary change compared to other genomic regions, such as the spike protein.
- Finally, Patient 1's sample (January 2025) and Patient 5's sample (April 2025) were collected closer in time which may explain their higher similarity compared to the reference.
- **Future work:** There are two primary goals for this project. The first is to tile the entire genome with 2-3kb fragments. The second is to expand this to other coronaviruses such as HKU-1, NL63, 229E

References

- Kim, Mi Il, and Choongho Lee. "Human Coronavirus OC43 as a Low-Risk Model to Study COVID-19." Viruses vol. 15,2 578. 20 Feb. 2023, doi:10.3390/v15020578
- A.D. López-Muñoz, J.J.S. Santos, & J.W. Yewdell, Cell surface nucleocapsid protein expression: A betacoronavirus immunomodulatory strategy, Proc. Natl. Acad. Sci. U.S.A. 120 (28) e2304087120, https://doi.org/10.1073/pnas.2304087120 (2023).
- Xavier Robert, Patrice Gouet, Deciphering key features in protein structures with the new ENDscript server, Nucleic Acids Research, Volume 42, Issue W1, 1 July 2014, Pages W320–W324, https://doi.org/10.1093/nar/gku316
- Korber, B. (2000). HIV Signature and Sequence Variation Analysis. In A. G. Rodrigo & G. H. Learn (Eds.), Computational Analysis of HIV Molecular Sequences (Chapter 4, pp. 55-72). Kluwer Academic Publishers.

Acknowledgements

I'd also like to thank Riley Aiken and Dr. Xianbao He for their help, assistance, and support conducting research.