Longitudinal Genetic Variation in Human Coronavirus OC-43 Identified via Novel Long-Read Sequencing

Daniel Lee[1,2], Thomas Murphy[2], Manish Sagar[2]

Oakton High School, 2900 Sutton Rd, Vienna, VA 22181[1]; Department of Medicine, Evans Biomedical Research Center, Boston University School of Medicine, 650 Albany Street, Boston, MA 021182[2].

Introduction

- Human coronavirus OC-43 (HCoV-OC43) is a seasonal **betacoronavirus** that evolves over time.
- **Genomic changes** in OC-43 remain **poorly characterized**.
- Most prior studies have relied on **short-read sequencing** (<100 tiles), which may miss key variations.
- Our study utilizes **long-read sequencing** via Oxford Nanopore to identify **year-to-year nucleotide changes** in the OC-43 genome.
- We hypothesize that OC-43 undergoes **significant genetic changes over time**, particularly in genes affecting host interaction (e.g. nucleocapsid).
- Scientists traditionally utilize **static reference genomes** in clinical assays.
- Revealing how the genome changes over time can emphasize the need to **update reference sequences** to reflect circulating strains.

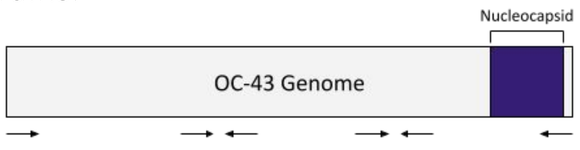

Figure 1: Schematic of the HCoV-OC43 Genome. Nucleocapsid gene is highlighted in purple.

Methods

- Patient respiratory samples collected across **multiple years**.
- Viral RNA extracted and reverse-transcribed into **complementary DNA (cDNA)**.
- Targeted **PCR amplification** of nucleocapsid region.
- Verification of amplification via **gel electrophoresis**.
- Amplified DNA subjected to **Oxford Nanopore long-read sequencing**.
- Sequencing data aligned and compared to assess **nucleotide variation** between individual and years.

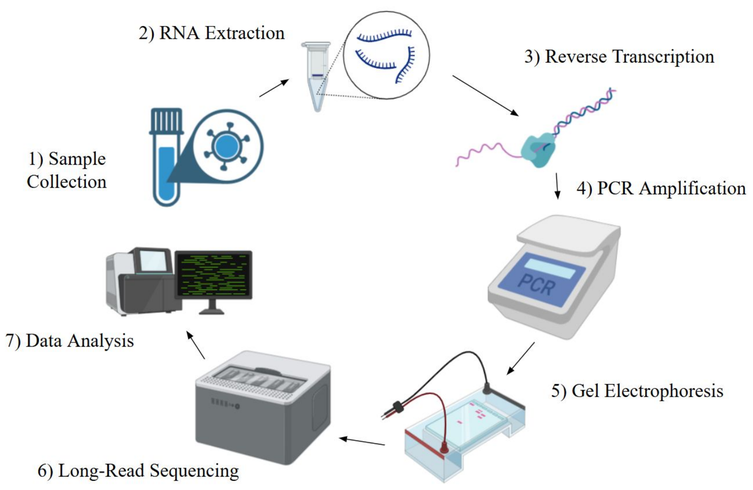

Figure 2: Workflow for Longitudinal Genetic Analysis of HCoV-OC43. Process flow is clockwise.

Results

Troubleshooting:

- Significant time was devoted to optimizing **primer–reverse transcription (RT) combinations** for specific amplification.
- Due to inconsistent amplification across the full genome, we focused on the **nucleocapsid gene**, which showed more reliable results **(Figure 3e)**.

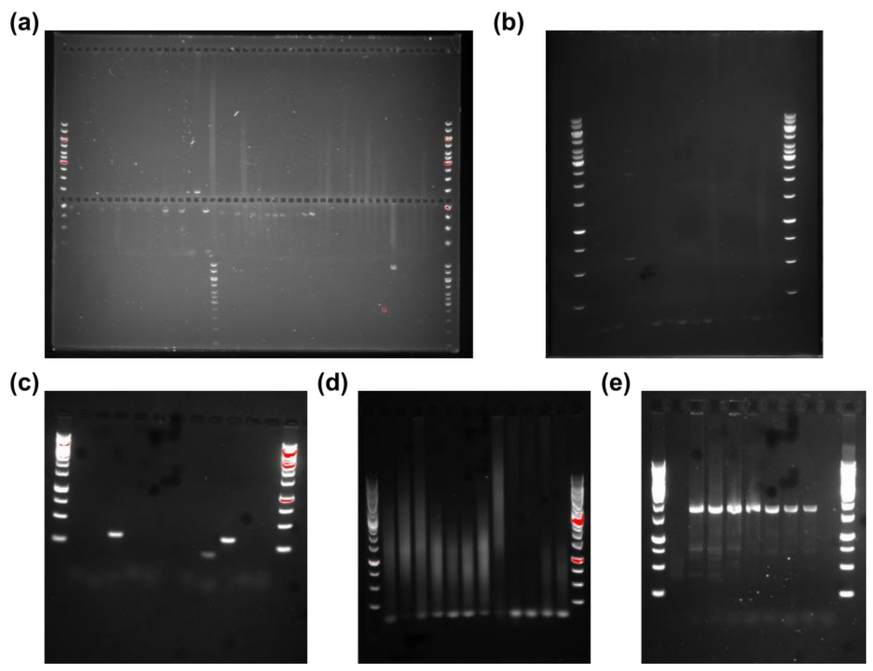

Figure 3: Sequential Gel Images Demonstrating Primer–RT Troubleshooting. (a) Initial primers (JEB, 1R, 2F, 2R, 3F, 3R) and RT/hexamer combinations tested at 61–71 °C. **(b)** Updated primer and RT/hexamer combinations. **(c)** Updated primers with shorter amplicons and RT/hexamer combinations. **(d)** Long-read amplification using Maxima RT. **(e)** Amplification of nucleocapsid gene.

Sequencing:

- Patient samples of the **nucleocapsid gene** were run on a gel.
- Of the seven patients, only **Patients 1 and 5** showed correct amplification and were subsequently sequenced and compared to the reference genome.

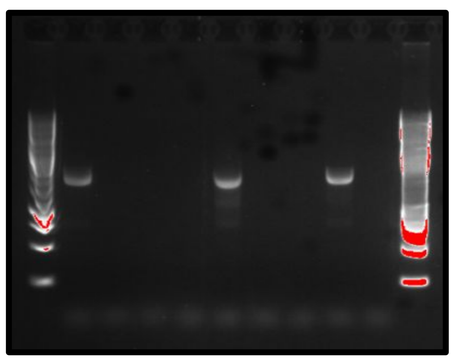

Figure 4: PCR results for seven patient samples. Bands observed for Patients 1 and 5 indicate successful amplification.

Figure 5: Sequence Alignment of Patient 1 & 5 with Reference Genome. All observed mutations were point mutations. **Patient 1**: 29 nucleotide changes and 11 AA substitutions (98.3% alignment). **Patient 5**: 28 nucleotide changes and 9 AA substitutions (98.4% alignment). Patient 1 and 5 had 99.5% sequence alignment.

Discussion

- Long-read sequencing **successfully** identified point mutations in the nucleocapsid gene.
- **Homopolymer repeats** in the OC-43 genome contribute to strong, non-specific primer binding and off-target amplification.
- Multiple amino acid changes suggest, but do not confirm, potential **functional differences** in host–virus interactions.
- The **strong sequence alignment (99.5%)** between the two patient samples suggests that the observed mutations are not random but reflect evolutionary changes in circulating strains.
- The absence of bands in the remaining patient samples suggests two possibilities: **sample degradation** or **extensive sequence divergence** preventing primer binding.
- Large, immune-targeted genes like **spike** likely harbors more mutations, meaning nucleocapsid results may not capture full functional variation.
- More consistent amplification of the nucleocapsid gene is likely due to its **shorter size** (~2kb) and **proximity to the 3' end** of the genome, where reverse transcription begins.
- The divergence between circulating and reference strains supports **updating reference genomes** to match current variants.
- Future studies should optimize primer and reverse transcription conditions to enable long-read sequencing of the **full genome**.

References

Lau, Susanna K., et al. “Molecular epidemiology of human coronavirus OC43 reveals evolution of different genotypes over time and recent emergence of a novel genotype due to natural recombination.” *Journal of Virology*, vol. 85, no. 21, Nov. 2011, pp. 11325–11337, https://doi.org/10.1128/jvi.05512-11.

Wang, Chunyan, et al. “Antigenic structure of the human coronavirus OC43 spike reveals exposed and occluded neutralizing epitopes.” *Nature Communications*, vol. 13, no. 1, 25 May 2022, https://doi.org/10.1038/s41467-022-30658-0.

Acknowledgements

I would like to thank **Riley Aiken** and **Dr. Xianbao He** for their guidance in introducing me to lab techniques and overseeing experiments.