Genomic evidence for a case of reinfection with SARS-CoV-2

Richard L. Tillett 1 2

Joel R. Sevinsky 3

Paul D. Hartley[4]

Heather Kerwin 5

Natalie Crawford 6

Andrew Gorzalski 7

Chris Laverdure 7

Subhash C. Verma 8

Cyprian C. Rossetto 8

David Jackson 9

Megan J. Farrell 9

Stephanie Van Hooser 7

Mark Pandori 7, 10

Affiliations:

1 Nevada Institute of Personalized Medicine, University of Nevada, Las Vegas

2 University of Nevada, Reno Nevada Center for Bioinformatics

3 Theiagen Consulting LLC

4 Nevada Genomics Center, Research & Innovation, University of Nevada, Reno

5 Division of Epidemiology & Public Health Preparedness Washoe County Health District

6 Renown Health, Reno, NV

7 Nevada State Public Health Laboratory, Reno, NV

8 Department of Microbiology and Immunology, University of Nevada, Reno School of Medicine,

9 Forensic Science Division, Washoe County Sheriff's Office

10 Department of Pathology and Laboratory Medicine, University of Nevada, Reno School of Medicine

Abstract

The degree of protective immunity conferred by infection with SARS-CoV-2 is currently unknown. As such, the possibility of reinfection with this virus is not well understood. Herein, we describe the data from an investigation of two instances of SARS-CoV-2 infection in the same individual. Through nucleic acid sequence analysis, the viruses associated with each instance of infection were found to possess a degree of genetic discordance that cannot be explained reasonably through short-term *in vivo* evolution.

We conclude that it is possible for humans to become infected multiple times by SARS-CoV-2, but the generalizability of this finding is not known.

Introduction

While infection with SARS-CoV-2 leads to a detectable immune response, what is not well understood is how susceptible previously infected individuals are to re-infection with SARS-CoV-2. SARS-CoV-2 infection in humans does result in the generation of neutralizing antibodies[1]. However, the degree to which this indicates a formidable immunity to subsequent infection (with SARS-CoV-2) is not yet elucidated. Evidence from the study of immunity to other coronaviruses has demonstrated that a loss of immunity to such viruses can occur within 1 to 3 years [2–9]. Cases of primary illness due to infection followed by a discrete secondary infection / illness with the same biological agent can best be ascertained as distinct infection events through genetic analysis of the agents associated with each illness event. Herein we describe a case of an individual who has had two distinct COVID illnesses from two genetically distinct SARS-CoV-2 viruses. This strongly supports that reinfection with SARS-CoV-2 can occur.

Materials and Methods

Diagnostic testing: Specimens were taken by nasopharyngeal swab and transported to the Nevada State Public Health Laboratory in viral transport medium (VTM) or Aptima Multiswab Transport Media. Specimens were transported on cold packs and stored by refrigeration (4-8°C) for 72 hours or less prior

to nucleic acid extraction and subsequent real time PCR. For real-time PCR, extraction was performed using Omega Biotek MagBind Viral DNA/RNA 96 Kit per manufacturer's instructions and with an elution volume of 100 µl. Aliquots of eluted RNA (5µl for the CDC 2019-nCoV Real-Time RT-PCR Diagnostic Panel, 10 µl for the Taqpath COVID-19 (EUA) Multiplex assay) (Thermo Scientific, Waltham, MA) were subjected to real time PCR either by the Taqpath COVID-19 Multiplex assay or the CDC Real-Time PCR. Aptima specimens were tested by the Aptima SARS-CoV-2 (Panther System) assay (Hologic, Marlborough, MA). Other than as indicated above, assays were performed according to their respective Emergency Use Authorized procedures.

Viral Genomic Sequencing: Total RNA was extracted from nasopharyngeal swabs as described above. This extracted RNA (70 µl) was treated for 30 minutes at room temperature with QIAGEN DNase I and then cleaned and concentrated with silica spin columns (QIAGEN RNeasy MinElute), with a 12-µl water elution. A portion (7 µl) of this RNA was annealed to an rRNA inhibitor (QIAGEN FastSelect -rRNA HMR), and then reverse transcribed, strand-ligated and isothermally amplified into micrograms of DNA (QIAGEN FX Single Cell RNA Library Kit). A portion (1 µg) of this amplified DNA was sheared and ligated to Illumina-compatible sequencing adapters, followed by 6 cycles of PCR amplification (KAPA HiFi HotStart) to enrich for library molecules with adapters at both ends. Next, these sequencing libraries were enriched for sequence specific to SARS-CoV-2 using biotinylated oligonucleotide baits (myBaits Expert Virus, Arbor Biosciences). An additional 8-16 cycles of PCR were performed post-enrichment, and these SARS-CoV-2 enriched sequencing libraries were pooled and sequenced with an Illumina NextSeq 500 as paired-end 2x75 bp reads.

Bioinformatics Analysis of potential reinfection pair: Following the sequencing of each library, FASTQ files were imported into CLC Genomics Workbench v.20.0.4 with the following modules: CLC Microbial Genomics Module, CLC Genome Finishing Module, and Biomedical Genomics Analysis (QIAGEN A/S, Denmark). Briefly, reads were imported, trimmed, and mapped to NBCI SARS-CoV-2 reference sequence

MN908947.3. The alignment was refined using the "InDels and Structural Variants" followed by "Local Realignment" modules. Variants were identified by a minimum coverage of 5, minimum count of 5, and minimum frequency of 70.0%.

To ascertain repeatability of results, a second analysis of the potential reinfection pair of specimens was performed using an independent process and open source tools: the potential reinfection sequence libraries were trimmed using Trimmomatic, version 0.39, with the ILLUMINACLIP adapter-clipping setting "2:30:10:2:keepBothReads" [10]. Sequence pairs were aligned to the SARS-CoV-2 reference genome (MN908947.3) using Bowtie 2, version 2.3.[11]. PCR optical duplicates were flagged using Picard MarkDuplicates, in picard-slim version 2.22.5 [12]. Variants were called for both samples in concert using Freebayes, version 1.0.2, with ploidy settings of 1, minimum allele frequency of 0.75, and minimum depth of four reads for any variant call[13]. The genome sequence of each sample was constructed using coverage statistics from BBtools *pileup* and *applyvariants*, version 38.86, whereby only variants supported by coverage ≥ 4 were written to *bcftools consensus*, v1.10.2, and all positions supported by fewer than four reads, whether reference or alternative, were replaced with Ns[14,15].

Phylogenetic placement of reinfection pair: Phylogenetic analysis of whole genome sequences of the isolates were made in comparison with those of 171 contemporaneous sequences from Nevada, the SARS-CoV-2 reference strain (MN908947.3), and one sequence derived from isolate USA-WA1/2020[16] (Bei Resources, Manassas, VA). After trimming 5' and 3' uncalled bases (Ns), genomic sequences were aligned and related using *NGPhylogeny.fr PhyML+SMS.*[17] There, sequences were aligned using *MAFFT* with automatic flavor selection [18]. Informative regions were selected using *BMGE*, sliding window size 3, maximum entropy 0.5[19]. Unrooted trees were constructed by *PhyML* with Smart Model Selection (SMS), AIC likelihood criteria, and Subtree Pruning and Regrafting (SPR)[20,21] . Newick trees were visualized using the Interactive Tree Of Live (iTOL) v4 and rooted at the Wuhan reference strain [2223]. Major SARS-CoV-2 clade memberships were predicted using Nextclade [24 25].

Identity testing: Specimens (swabs specimens or extracted RNA residual samples) provided to the Washoe County Sheriff's Office were quantified by utilizing 2 µl of the extracted DNA using the Quantifiler Trio DNA Quantification Kit by Applied Biosystems™ on the 7500 Real-Time PCR System by Applied Biosystems™, and analyzed with the 7500 HID software v 1.3.

Amplification of the 24 GlobalFiler™ Short Tandem Repeat (STR) markers was accomplished on the ProFlex™ PCR Instrument through 29 cycles. The 3500xL Genetic Analyzer by Applied Biosystems™ was used for fragment analysis of the amplified STR marker regions in conjunction with HID Data Collection Software v4.0.1 and Genemapper™ *ID-X* Software v1.6. Statistical interpretation of STR data was achieved using the allele frequencies maintained in the National Institute of Standards and Technology (NIST) population database: http://strbase.nist.gov/NISTpop.htm#Autosomal

Results

In April, 2020, a twenty-five year old resident of Reno, NV tested positive for SARS-CoV-2 through a community-based testing event held by the Washoe County Health District (collection date: 4/18/2020). The patient indicated symptoms consistent with viral infection (sore throat, cough, headache, nausea, diarrhea; onset: 3/25/20). During isolation, the patient indicated resolution of symptoms (4/27/20). The patient was subsequently tested by two nucleic acid amplification tests and was found negative for the presence of SARS-CoV-2 RNA for specimens collected on 5/9/2020 (by transcription-mediated amplification (TMA)) and again on 5/26/2020 (by real-time PCR (RT-PCR)). The patient continued to feel well until 5/28/20. On 5/31/20, the patient sought care with self-reported fevers, headache, dizziness, cough, nausea, and diarrhea. A chest x-ray was performed and he was discharged home. Five days later, on 6/5/20, the patient presented to a family care doctor and was found to be hypoxic and was instructed to go to the emergency department after provision of oxygen. The patient was hospitalized

that day and was assessed for SARS-CoV-2 infection by RT-PCR testing. The patient required ongoing oxygen support and reported symptoms that included myalgia, cough and shortness of breath. A chest x-ray was performed on 6/5/20 and compared to that of 5/31/20 with the development of new patchy bilateral interstitial opacities suggestive of a viral or atypical pneumonia. RT-PCR results were positive for the presence of SARS-CoV-2. On 6/6/20, the patient was tested for IgG/IgM for SARS-CoV-2 and was positive. A summary of specimens tested with result details is shown in Table 1.

With consideration of two episodes of symptoms consistent with COVID-19, and two specimens found reactive for SARS-CoV-2 specimens separated by: symptomatic recovery; a period of 48 days, and two non-reactive (negative) SARS-CoV-2 test results, we performed nucleic acid sequencing of the viruses associated with the positive cases.

Sequence data indicated that the specimen collected in April of 2020 (specimen "A") was found to be a member of clade 20C by way of genomic sequence analysis identifying all five mutation positions and bases that describe the clade. The second reactive (positive) specimen, collected in June of 2020 (specimen "B") was also found to be a member of clade 20C, presenting the clade-defining mutations C3037T, C14408T, A23403G, C1059T and G25563T. In addition to possessing hallmark mutations of the 20C clade, case A was determined to possess five further single nucleotide variants (SNVs) compared to the reference genome. Sequence data from case B show 6 additional SNVs and a mutation at position 14407, being adjacent to C14408T, recorded as a dinucleotide multi-nucleotide variant (MNV) at positions 14407-14408 of the genome (Figure 1 and Table 2). Six SNVs were shared between case A and case B (Table 2A)–the five which define the 20C clade, and C241T. Case A had four SNVs that are absent from the later case (Table 2B), while case B had seven SNVs (Table 2C) absent in the former. A visualization of the relationship of the sequence data sets between cases A and B is shown in Figure 1. There were an additional three deletions and one insertion in Case B sequence relative to the reference

sequence (Supplemental Table 1), as called using CLC Genomics Workbench. To confirm that these findings would be obtained independently of the software tools used, we performed additional analyses on the FASTQ files generated from cases A and B. Using the Bowtie 2 aligner and Freebayes variant caller with 75% allele frequency stringency, each case-specific and shared SNPs and MNPs with the exception of locus 4113 in case A were verified. Predictions of insertions and deletions were less stable. The additional analysis also predicted in case B of the deletion at loci 2084 and an insertion at 6018, but not others detected by CLC Genomic Workbench based analysis. Freebayes analysis did detect one deletion at 22832 in case B that was not called by CLC Genomics Workbench (Supplemental Table 2). INDEL predictions from short read alignments are known to be less reliable than SNV predictions and are presented here for completeness.[26]

Each of the specimens A and B were members of a cohort of specimens that were sequenced from the State of NV (174) collected from March 5 through June 5, 2020. A phylogenetic diagram demonstrating the relatedness of A and B to each other and their comparative distance among these additional positive specimens is shown in Figure 2.

To rule out the possibility of specimen mishandling or mislabeling errors during RNA extractions, we investigated the source and intermediate materials of specimens A and B by forensic identity testing. The original collected swab/transport media specimens for A and B, the residual extracted nucleic acid derived from A and B, and the residual aliquots of extracted nucleic acid supplied to the sequencing core facility were subjected to Short Tandem Repeat (STR) Analysis for identity comparison by the Washoe County Sheriff's Forensics Laboratory. Analysis of each of the specimens, residual extractions, and aliquot residuals were in agreement, that the A and B specimens and samples were derived from the same individual with a 1 in 53.48 septillion (53.48×10^{24}) chance of the specimens being from different persons.

The individual associated with these cases possesses no significant conditions of an immunological nature that would imply facilitation of re-infection. They were not utilizing any immunosuppressive medications. The individual was negative for HIV by antibody and RNA testing (data not shown) and had no obvious cell count abnormalities. The secondary positive case (B) occurred simultaneously to a positive case of a co-habitant (parent), also positive by NAAT (TMA) on 6/5/2020. Sequencing is being attempted on this case to ascertain its potential role in case B. However, the infected co-habitant's positive specimen was collected and tested in the Hologic Aptima format, which did not align with the procedures established at our sequencing laboratory at the time of submission. The positive case provides a possible source for secondary exposure and (re)infection.

Discussion

The data herein support an instance of reinfection with SARS-CoV-2. For case A to experience mutations to become case B, the virus would have had to exhibit a rate of 83.64 substitutions per year, a rate that markedly exceeds that of 23.12, currently observed [30]. However, of enormous significance, four of the discordant loci seen between case A and case B would be reversions specific to the ancestral genotype. The odds of this occurring are vanishingly remote and virtually assure that these are two distinct viral infection events. Of course, if such a remarkable event of base change *did* occur in that timeframe, then the remarkable nature of cases A and B would shift from a case of possible reinfection to a case of high-rate evolution within an infected individual. Both Case A and B were found to be in clade 20C, which was the predominant major clade observed in northern NV at collection time (Fig 2). An implication of this finding is that initial exposure to the SARS-CoV-2 virus may not result in a level of immunity that is 100% protective for all individuals. With regard to vaccination, this is an established understanding, with influenza regularly demonstrating the challenges of effective vaccine design[31]. It is crucial to note that the frequency of such a phenomenon is not defined by a singular case study. This

may represent a rare event. The lack of comprehensive genomic sequencing of positive cases in the United States and worldwide limits the sophistication of public health surveillance required to find these cases. Certainly, limitations in screening / testing availability for SARS-CoV-2 exacerbate the poor surveillance efforts being undertaken not only to diagnose COVID-19 but also to obtain actionable genetic tracking of this agent.

Acknowledgements

We thank Nevada IDEA Network of Biomedical Research (INBRE) for supporting this work and the publication was made possible by grants from the National Institute of General Medical Sciences (GM103440 and GM104944) from the National Institutes of Health. We also thank the Washoe County Health District, Washoe County Sheriff's Department for helping to identify and confirm these findings.

Data Availability

The CLC workflow used for analysis, SARS-CoV-2_Illumina_WF-0.1.cpw, the combined mapping report, *Combined_Mapping_Reports.pdf,* detailed parameters for all CLC modules, *CLC_Workflow_History_and_Settings.pdf*, BAM alignments and VCF-format files can be found at https://zenodo.org/record/3988783; DOI:10.5281/zenodo.3988782.

Figure Legends

Figure 1: Variant mapping of reinfection cases. *Identifies variant 14407 in Case A and both variants 14407 and 14408 in Caes B.

Figure 2: Phylogenetic placement of infection cases within Nevada isolates and global clades.

References

1. Ju, B. *et al.* Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* **584**, 115–119 (2020).
2. Callow, K. A., Parry, H. F., Sergeant, M. & Tyrrell, D. A. The time course of the immune response to experimental coronavirus infection of man. *Epidemiol. Infect.* **105**, 435–446 (1990).
3. Chang, S.-C. *et al.* Longitudinal analysis of Severe Acute Respiratory Syndrome (SARS) coronavirus-specific antibody in SARS patients. *Clin. Diagn. Lab. Immunol.* **12**, 1455–1457 (2005).
4. Huang, A. T. *et al.* A systematic review of antibody mediated immunity to coronaviruses: antibody kinetics, correlates of protection, and association of antibody responses with severity of disease. *medRxiv* (2020) doi:10.1101/2020.04.14.20065771.
5. Liu, W. *et al.* Two-year prospective study of the humoral immune response of patients with severe acute respiratory syndrome. *J. Infect. Dis.* **193**, 792–795 (2006).
6. Mo, H. *et al.* Longitudinal profile of antibodies against SARS-coronavirus in SARS patients and their clinical significance. *Respirology* **11**, 49–53 (2006).
7. Reed, S. E. The behaviour of recent isolates of human respiratory coronavirus in vitro and in volunteers: evidence of heterogeneity among 229E-related strains. *J. Med. Virol.* **13**, 179–192 (1984).
8. Woo, P. C. Y. *et al.* Longitudinal profile of immunoglobulin G (IgG), IgM, and IgA antibodies against the severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein in patients with pneumonia due to the SARS coronavirus. *Clin. Diagn. Lab. Immunol.* **11**, 665–668 (2004).

9. Wu, L.-P. *et al.* Duration of antibody responses after severe acute respiratory syndrome. *Emerg. Infect. Dis.* **13**, 1562–1564 (2007).

10. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

11. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

12. Broad Institute. Picard Toolkit. *Picard* https://broadinstitute.github.io/picard/.

13. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv [q-bio.GN]* (2012).

14. Bushnell, B. *BBMap: A Fast, Accurate, Splice-Aware Aligner*. https://www.osti.gov/biblio/1241166-bbmap-fast-accurate-splice-aware-aligner (2014).

15. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

16. Paul D. Hartley, Richard L. Tillett, Yanji Xu, David P. AuCoin, Joel R. Sevinsky, Andrew Gorzalski, Mark Pandori, Cyprian C. Rossetto, Subhash C. Verma. A mutation in RdRp (nsp12) of SARS-CoV-2 detected at high frequency from patient specimens in Northern Nevada by next-generation sequencing. *tbd* (2020).

17. Lemoine, F. *et al.* NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res.* **47**, W260–W265 (2019).

18. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

19. Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.*

10, 210 (2010).

20. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
21. Lefort, V., Longueville, J.-E. & Gascuel, O. SMS: Smart Model Selection in PhyML. *Mol. Biol. Evol.* **34**, 2422–2424 (2017).
22. Junier, T. & Zdobnov, E. M. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* **26**, 1669–1670 (2010).
23. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
24. Nextclade. https://clades.nextstrain.org.
25. Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* vol. 34 4121–4123 (2018).
26. Abnizova, I., te Boekhorst, R. & Orlov, Y. L. Computational Errors and Biases in Short Read Next Generation Sequencing. *Journal of Proteomics & Bioinformatics* vol. 10 (2017).
27. Hill, V. & Rambaut, A. Phylogenetic analysis of SARS-CoV-2 genomes. *Virological* https://virological.org/t/phylodynamic-analysis-of-sars-cov-2-update-2020-03-06/420.
28. Mercatelli, D. & Giorgi, F. M. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Front. Microbiol.* **11**, 1800 (2020).
29. Pachetti, M. *et al.* Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J. Transl. Med.* **18**, 179 (2020).
30. Hadfield, J. *et al.* Nextstrain Clock, 4339 genomes, December 3, 2019 - August 7, 2020. *Nextstrain* https://nextstrain.org/ncov/global?c=region&l=clock.
31. Osterholm, M. T., Kelley, N. S., Sommer, A. & Belongia, E. A. Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis. *Lancet Infect. Dis.* **12**, 36–44 (2012).

Table 1. Summary of Laboratory Results

Specimen Designation:	A			B	
Collection Date:	*4/18/2020*	*5/9/2020*	*5/26/2020*	*6/5/2020*	6/6/2020
Test Result	Positive	Negative	Negative	Positive	Positive
Test methodology*	RT-PCR	TMA	RT-PCR	RT-PCR	IA (antibody)
quantitative result**	Ct: 35.24	RLU: 299	--	Ct: 35.31	N/A

*RT-PCR: N (Taqpath) or N1 (CDC Assay) region amplification reaction, Real-time PCR; TMA: Transcription-mediate
IA: Immunoassay (IgG/IgM Detection)
**Ct: Cycle threshold result, for RT-PCR; RLU: relative light units, (Hologic Panther)

Table 2A. Shared variants of A and B versus reference genome.

Locus (base)	Type	Reference *	Allele	Reference allele	Coverage (A/B)	Allele Frequency (A/B)	Forward/reverse balance (A/B)**	Average quality (A/B)†
241	SNV	C	T	No	67/6	100.0/100.0	0.37/0.38	35.6/36.0
1059	SNV	C	T	No	144/55	100.0/92.7	0.48/0.26	35.6/35.4
3037	SNV	C	T	No	89/425	100.0/99.8	0.42/0.19	35.6/35.5
14408#	SNV	C	T	No	73/1145	100.0/99.6	0.40/0.43	35.7/35.6
23403	SNV	A	G	No	6859/10484	99.9/99.9	0.19/0.46	35.7/35.6
25563	SNV	G	T	No	421/757	100.0/99.1	0.45/0.48	35.2/35.4

Table 2B. Case A specific variants versus reference genome.

Locus (base)	Type	Reference *	Allele	Reference allele	Coverage	Allele Frequency	Forward/reverse balance**	Average quality†
539	SNV	C	T	No	141	99.3	0.45	35.6
4113	SNV	C	T	No	159	70.4	0.38	35.6
7921	SNV	A	G	No	182	98.9	0.49	35.7
16741	SNV	G	T	No	173	99.4	0.47	35.6

Table 2C. Case B specific variants versus reference genome.

Locus (base)	Type	Reference *	Allele	Reference allele	Coverage	Allele Frequency	Forward/reverse balance**	Average quality†
8140	SNV	C	T	No	1046	85.0	0.43	35.6
11102	SNV	C	T	No	1713	99.9	0.44	35.5
14407#	SNV	C	T	No	1145	99.7	0.43	35.6
15190	SNV	G	C	No	139	90.6	0.33	35.7
15981	SNV	C	T	No	224	100.0	0.38	35.5
26013	SNV	C	T	No	1415	99.2	0.38	35.5
29466	SNV	C	T	No	86	98.8	0.07	35.8

*Wuhan-Hu-1, GenBank: MN908947.3
**Ratio of forward to reverse reads covering the locus
†Phred score
#CLC Genomics classified this as a MNV. The two variants were split in this table for clarity.

Fig1

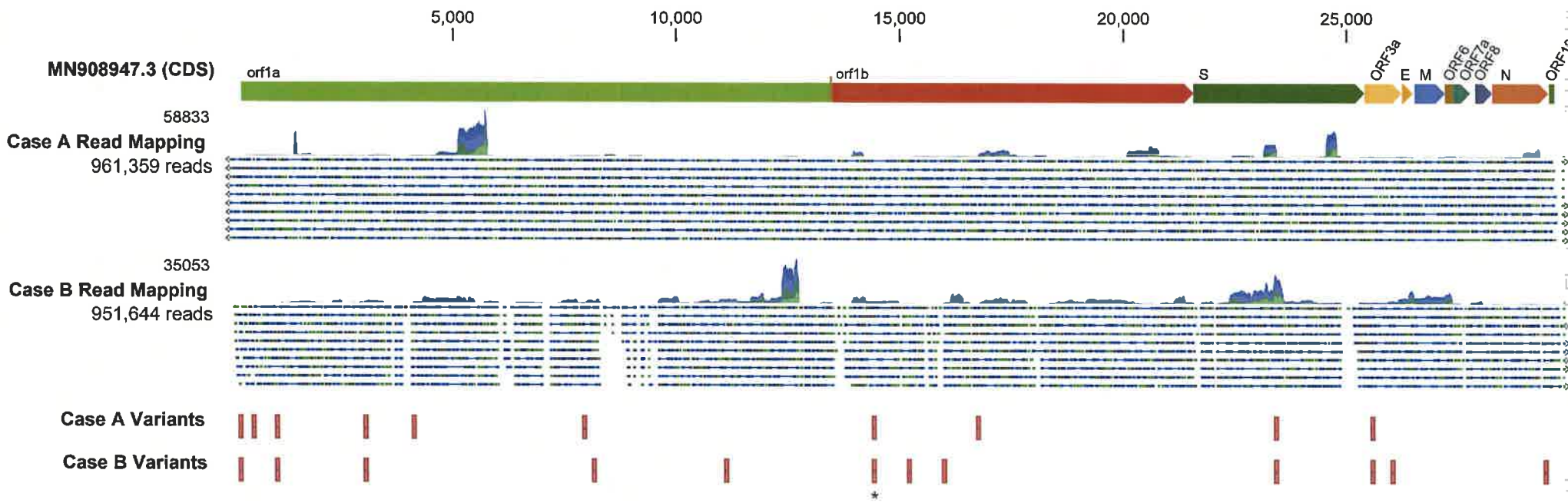

Fig 2