Intrahost evolution and forward transmission of a novel SARS-CoV-2 Omicron BA.1 subvariant

Ana S. Gonzalez-Reiche¹, Hala Alshammary², Sarah Schaefer³, Gopi Patel⁴, Jose Polanco²,₄, Angela A. Amoako²,₄, Aria Rooker²,₄, Christian Cognigni²,₄, Daniel Floda¹, Adriana van de Guchte¹, Zain Khalil¹, Keith Farrugia¹, Nima Assad⁵, Jian Zhang¹, Bremy Alburquerque¹,₅, Levy Sominsky²,₄, Komal Srivastava²,₄, Robert Sebra¹,₈,⁹,¹₀, Juan David Ramirez⁶,¹¹, Radhika Banu⁶, Paras Shrestha⁶, Alberto Paniz-Mondolfi⁶, Emilia Mia Sordillo⁶ †, Viviana Simon²,₃,⁴,⁶,⁷, †, Harm van Bakel¹,₆,⁸,†

1. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
2. Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
3. Division of Infectious Diseases, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
4. Center for Vaccine Research and Pandemic Preparedness (C-VARPP), Icahn School of Medicine at Mount Sinai, New York, NY, USA
5. Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
6. Department of Pathology, Molecular, and Cell-Based Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
7. The Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
8. Icahn Genomics Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
9. Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA
10. Sema4, a Mount Sinai venture, Stamford CT, 06902

Corresponding authors/senior authors:
† Dr. Emilia Mia Sordillo (Emilia.Sordillo@mssm.org), Dr. Viviana Simon (viviana.simon@mssm.edu) and Dr. Harm van Bakel (harm.vanbakel@mssm.edu)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Persistent SARS-CoV-2 infections have been reported in immune-compromised individuals and people undergoing immune-modulatory treatments. It has been speculated that the emergence of antigenically diverse SARS-CoV-2 variants such as the Omicron variant may be the result of intra-host viral evolution driven by suboptimal immune responses, which must be followed by forward transmission. However, while intrahost evolution has been documented, to our knowledge no direct evidence of subsequent forward transmission is available to date.

Here we describe the emergence of an Omicron BA.1 sub-lineage with 8 additional amino acid substitutions within the spike (E96D, L167T, R346T, L455W, K458M, A484V, H681R, A688V) in an immune-compromised host along with evidence of 5 forward transmission cases. Our findings show that the Omicron BA.1 lineage can further diverge from its exceptionally mutated genome during prolonged SARS-CoV-2 infection; highlighting an urgent need to employ therapeutic strategies to limit duration of infection and spread in vulnerable patients.
Emergence of antigenically diverse SARS-CoV-2 variants of concern (VOC) poses a challenge to SARS-CoV-2 vaccine-induced immunity. Prolonged evolution in immunocompromised hosts is thought to have played a role in the emergence of several SARS-CoV-2 VOC such as Alpha and Omicron\textsuperscript{1,2}. However, although persistent SARS-CoV-2 replication with stepwise acquisition of mutations in spike within immunocompromised COVID-19 patients has been documented\textsuperscript{3-7}, clear evidence for forward transmission has been lacking.

The Mount Sinai Pathogen Surveillance Program (MS-PSP) has profiled the evolving landscape of SARS-CoV-2 in New York City (NYC) since the beginning of the pandemic\textsuperscript{8,9}. We routinely perform complete viral genome sequencing from at least 10% of randomly selected contemporaneous SARS-CoV-2 positive specimens collected from patients seeking care at our health system.

We describe a persistent SARS-CoV-2 Omicron BA.1 infection in an immunocompromised individual during a 12-week period and document the accumulation of eight additional amino acid substitutions in the already antigenically distinct Omicron BA.1 spike protein. In addition, we identified five subsequent cases harboring the same genotype; three of which were identified in the same health system. The presence of the unique combination of mutations in all six cases is indicative of forward transmission of this novel Omicron BA.1 sub-lineage. Although the majority of amino acid changes occur at positions known to confer either immune escape\textsuperscript{10,11} or altered viral fusogenicity\textsuperscript{12-14}, some of the mutations have rarely been seen in other lineages, and their overall constellation is unique.
**MATERIALS AND METHODS**

*Molecular SARS-CoV-2 diagnostics*

SARS-CoV-2 molecular diagnostic testing was performed in the Molecular Microbiology Laboratories of the MSHS Clinical Laboratory by nucleic acid amplification tests (NAAT) that have been validated for nasopharyngeal, anterior nares swabs and saliva specimens. All but one positive sample included in this study were tested using the PerkinElmer® New Coronavirus Nucleic Acid Detection Kit which provides qualitative detection of nucleic acid from SARS-CoV-2. It includes two SARS-CoV-2 targets (ORF1ab, N) and one internal positive control (IC; bacteriophage MS2). Cycle threshold (Ct) values are generated for all three targets and provide a quantitative estimate of the viral load. The only other positive sample that was run on another testing platform was the first specimen of the index case (P1). It was tested as a point of care test (POC) using the LIAT SARS-CoV-2 assay, and the biospecimen was discarded prior to transfer to the MS-PSP. In addition, two other negative specimens registered for patient P3 during the same time window were run in the LumiraDx assay, which does not provide Ct values.

*The Mount Sinai Pathogen Surveillance Program (MS-PSP)*

Residual nasopharyngeal and anterior nares (AN) swab specimens were collected after completion of the diagnostic process, as part of the Mount Sinai Pathogen Surveillance Program (IRB approved HS#13-00981).

*Complete SARS-CoV-2 genome sequencing*

RNA was extracted using the Chemagic™ Viral DNA/RNA 300 Kit H96 (PerkinElmer, CMG-1033-S) on the automated Chemagic™ 360 instrument (PerkinElmer, 2024-0020) from 300uL of viral transport media per the manufacturer’s protocol. cDNA synthesis followed by whole genome amplification with two custom primer panel sets targeting 1.5 and 2kb regions across the SARS-CoV-2 genome was performed as previously described with addition of new oligonucleotides to minimize amplicon dropout for
Omicron lineages derived from PANGO lineage B.1.1.529 (available on request).

Paired-end (2x150bp) Nextera XT (Illumina, cat. FC-131-1096) libraries prepared from amplicons were sequenced on a MiSeq instrument. SARS-CoV-2 genomes were assembled using a custom reference-based pipeline as previously described. The final sequencing depth per genome ranged between ~135k to ~415k reads.

**Phylogenetic analysis**

Global background SARS-CoV-2 sequences were downloaded from the GISAID EpiCOV database (as of 2022-05-13). The GISAID database was queried for the novel mutations observed in P1-P4 sequences in order to identify their closest matches. Sequence hits were confirmed by their Mash distance. For this, a genome sketch was generated from the sanitized alignment of the global sequences filtered for lineages BA.1.* as produced by Nextstrain v11 for SARS-CoV-2 with default parameters ([https://github.com/nextstrain/ncov](https://github.com/nextstrain/ncov)). This allowed pairwise comparisons of our data with high-quality global sequences with available metadata. Phylogenetic inferences of the MS-PSP SARS-CoV-2 genomes, including the closest sequence matches and all other BA.* lineage sequences from the MS-PSP surveillance program were done using an Omicron BA.* and New York State-focused background with proximity subsampling scheme with Nextstrain. Clade and lineage assignments were done with Nextclade CLI v3.2 (2021-11-04) and with PANGO-v1.8 (pangolin v3.1.17, pangoLEARN v.2022-04-22).

**Viral cultures**

Replication-competent SARS-CoV-2 was obtained as previously described. Briefly, Vero-E6 cells expressing TMPRSS2 were cultured in Dulbecco’s Modified Eagle Medium containing 10% heat-inactivated fetal bovine serum and 1% Minimum Essential Medium (MEM) Amino Acids Solution, supplemented with 100 U/ml penicillin, 100 μg/ml streptomycin, 100 μg/ml normocin, and 3 μg/ml puromycin. 200ul of viral transport media from the nasopharyngeal or anterior nares swab specimen was added to Vero-
E6-TMPRSS2 cells in culture media supplemented with 0.5 μg/ml Amphotericin B. Cultures were maintained for 10 days. Culture supernatants were collected and clarified by centrifugation (4000g, 5min) upon the appearance of cytopathic effect (CPE). The viral stocks were sequence-verified. Viral cultures were successful for the initial specimens obtained from P2, P3 and P4, but failed for all P1 specimens.

**RESULTS AND DISCUSSION**

In this study we report the genomic analysis of serially collected nasopharyngeal (NP) and anterior nares (AN) samples from an immunocompromised patient with persistent SARS-CoV-2 Omicron BA1 replication between December 2021 and March 2022. Over a 12-week period, we documented the accumulation of eight amino acid substitutions in the spike N-terminal domain (NTD), the receptor binding domain (RBD), and cleavage site region (CSR) ([Figure 1](#)) within the same patient. The first three mutations R346T, K458M and E484V were detected simultaneously 40 days after initial SARS-CoV-2 diagnosis and were fixed. Two weeks later (day 64), L167T and the furin cleavage site mutation P681Y were detected in addition to three mutations listed. During the following weeks (day 72 and day 81), two different viral populations emerged each carrying shared (L455W) as well as additional distinct signature mutations (E96D at day 72, S477D at day 81). The majority of the amino acid substitutions that accumulated in spike have been associated with neutralization escape^{10,11} or improved viral fusogenicity^{12}. All but one of the mutations observed in the consensus sequences outside the spike gene were non-synonymous, further supporting positive intrahost selection of spike changes with competitive advantage.

Background health system-wide SARS-CoV-2 genomic surveillance by MS-PSP conducted during the same time period identified three other patients with COVID-19 harboring SARS-CoV-2 Omicron variants that were similar to that of the index case and consistent with multiple forward transmissions based on the timeline of infection ([Figure 1](#), and see **Table 1** and **Figure 2** for details and the CT values for the individual specimens). These three forward transmissions were detected in immunocompromised individuals. In one case, infection persisted for more than three weeks resulting in
acquisition of an additional mutation in the spike RBD (V445A) (Figure 1). In some specimens we detected minority alleles at variant positions suggesting ongoing intra-host selection (Figure 1 and Figure 3). Although viral isolation failed for specimens available for the index case, we successfully cultured SARS-CoV-2 from the initial positive specimens of the three forward transmission cases, confirming the transmission of replication-competent virus.

Lastly, two additional related SARS-CoV-2 genomes from the NYC area were identified from the GISAID database (Figure 1). Based on the metadata provided, these two cases differed by age and gender from our cases, consistent with limited community spread of this new Omicron BA1 sub-variant. All transmission cases contained one synonymous change in Orf1a (T6001C) previously only observed at day 72 in the index case (Figure 4), narrowing the time window of the first transmission to a few weeks following day 64. As of mid-May 2022, the most recent case was detected in mid-April 2022 and the last SARS-CoV-2 positive specimen from a forward transmission was collected in early May 2022.

In conclusion, our data demonstrate that intra-host evolution of SARS-CoV-2 occurring during persistent infection in immunocompromised individuals can drive the emergence and spread of novel (sub)variants with extensive mutations in key antigenic regions, even in the context of the already highly divergent Omicron lineage. Our data suggest that genomic monitoring of persistent long-term infections is warranted and further underscore the need to limit the duration of viral infection. Improved early detection of novel SARS-CoV-2 variants, a better understanding of the selection processes driving SARS-CoV-2 evolution, and therapies that limit virus persistence and shedding are essential to reduce emergence of highly mutated viral variants in the future.

**ACKNOWLEDGMENTS**

We thank the Mount Sinai Hospital and School Leadership, in particular Dr. David Reich and Dr. Dennis Charney, for their ongoing support of the MS-PSP throughout the pandemic. We thank the laboratory technicians in the Molecular Microbiology Laboratories and the Rapid Response Laboratories of the Mount Sinai Health System...
since without their assistance and help, none of this surveillance work would be possible.

We gratefully acknowledge the authors and originating and submitting laboratories of sequences from GISAID’s EpiCoV (www.gisaid.org) that were used as background for our phylogenetic inferences.

**Funding Sources**

The work reported was in part supported by a contract from the National Institute of Allergy and Infectious Diseases (75N93021C00014, Option 12A) awarded to the Center for Research on Influenza Pathogenesis and Transmission – (a Center of Excellence for Influenza Research and Response) as part of the SARS-CoV-2 Assessment of Viral Evolution (SAVE) Program, a contract from the National Institute of Allergy and Infectious Diseases (HHSN272201400008C) awarded to the Center for Research on Influenza Pathogenesis – (a Center of Excellence for Influenza Research and Surveillance), philanthropic donations from the JPB Foundation, a research grant (2020-215611 [5384]) from the Open Philanthropy Project, and awards (S10OD026880 and S10OD030463) from the NIH Office of Research Infrastructure Programs.

**Competing Interests**

The Icahn School of Medicine at Mount Sinai has filed patent applications relating to SARS-CoV-2 serological assays listing Viviana Simon as co-inventor. Mount Sinai has spun out a company, Kantaro, to market serological tests for SARS-CoV-2. Robert Sebra is VP of Technology Development and a stockholder at Sema4, a Mount Sinai Venture. This work, however, was conducted solely at Icahn School of Medicine at Mount Sinai.
REFERENCES


**Table 1: Spike mutations detected in the sequenced cases. Mutations are shown relative to BA.1, the lineage of the earliest sample. Mutations in bold correspond to changes at positions already mutated in BA.1.**

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Sex</th>
<th>Age</th>
<th>Study Day</th>
<th>GISAI Accession</th>
<th>Source</th>
<th>Mutations in Spike relative to earliest sample (BA.1)</th>
</tr>
</thead>
</table>
Figure 1. Intrahost emergence and forward transmission of an Omicron BA.1 subvariant

A) Maximum likelihood phylogenetic subtree with SARS-CoV-2 (BA.1) sequences from the prolonged-infection case and forward transmissions, in a global background of sequences available in GISAID. Patients are depicted in different colors and the number of days after the first SARS-CoV-2 positive specimen is indicated in brackets. Sibling clusters of BA.1 and other sublineages are collapsed for easier visualization. The x-axis shows the number of nucleotide substitutions relative to the root of the phylogenetic tree.

B) Nucleotide and amino acid substitutions in the spike protein observed in sequential specimens obtained from patient 1 (P1) with prolonged infection with BA.1 and transmission cases (P2, P3 and P4). P1 shows progressive accumulation of mutations in the N-terminal domain (NTD), receptor binding domain (RBD), and cleavage site region (CSR); The same constellation of mutations was subsequently detected in three documented transmission cases (P2, P3 and P4) and in two sequences from GISAID (EPI_ISL_11628400 and EPI_ISL_11696379). The number of days since the first positive test in P1 is shown on the left, with the number of days after the first positive test for each patient between brackets. The BA.1 substitutions are shown in blue and novel substitutions are shown in red, relative to the reference genome sequence NC_045512.2. Positions with minor single nucleotide variants at the relevant positions are indicated.
Figure 2: Overview of the longitudinal SARS-CoV-2 viral load estimates and the clinical encounters for the four cases captured by the MS-PSP.

A) PerkinElmer® Real-time RT-PCR SARS-CoV-2 cycle threshold (Ct) values for nasopharyngeal and anterior nares specimens. The horizontal dashed line represents the cutoff value for a positive result. Vertical lines represent the sampling days of the cases recorded in GISAID relative to P1’s first positive test. For P2, the
negative tests by alternative diagnostic method (LumiraDx™) with no Ct value
available, are indicated with empty markers. NTGT: N gene, ORF1: ORF1ab gene,
IC: Internal Control.

**B) Timeline of the encounters of P1, P2, P3 and P4 with the health system. Each tick
represents an individual hospital encounter while the horizontal bars denote hospital
admissions (duration is indicated by the length of the bar).**
Figure 3: Summary of the SARS-CoV-2 minority viral variant distribution within diagnostic nasal biospecimen.

The distribution of nucleotide variants below the consensus level for positions that changed over the course of the infection of the index and transmitted infections are shown. Single nucleotide variants (SNVs) are colored by strain. The dotted bars indicate positions with coverage below the threshold of 100X coverage for calling minority variants. The positions with minority variants present at frequencies > 0.1% are indicated with triangles for their respective samples.
Figure 4: Overview of genome-wide SARS-CoV-2 mutations found in the diagnostic nasal biospecimen.

The mutations present across the complete SARS-CoV-2 genome are shown. Nucleotide substitutions observed in sequential specimens obtained from patient 1 (P1) with prolonged infection with BA.1 and forward transmission cases (P2, P3 and P4). There is accumulation and fixation of new SNVs in the spike region in P1.

The same constellation of mutations was subsequently detected in three documented transmission cases (P2, P3 and P4) and in two sequences from GISAID (EPI_ISL_11628400 and EPI_ISL_11696379). The only shared synonymous SNV outside of the spike (ORF1a:T6001C) is indicated with an arrow.

The number of days since the first positive test in P1 is shown on the left, with the number of days after the first positive test for each patient between brackets. The BA.1 substitutions are shown in blue and novel substitutions are shown in red, relative to the reference genome sequence NC_045512.2.