Golestan University of Medical Sciences
Journal of Clinical and Basic Research
Online ISSN: 2538-3736

Analysis of the RNA-Seq data of Solanum tuberosum revealed viral sequence reads of a severe laboratory-developed strain of SARS-CoV-2 containing novel substitutions

Alireza Mohebbi [1*]

1. Department of Virology, School of Medicine, Iran University of Medical Sciences, Tehran, Iran
Correspondence: Alireza Mohebbi, Department of Virology, School of Medicine, Iran University of Medical Sciences, Tehran, Iran.
Tel: +987068568868; Email: Alirezaa2s@gmail.com

Abstract

Background: Metagenomics is a promising approach to discovering novel sequences of microorganisms in environmental samples. A recently published RNA-Seq data of Solanum tuberosum from China was used for a metavirome study.

Methods: RNA-seq data of a BioSample project of S. tuberosum containing sequence read archive (SRA) of six plant samples were imported into the Galaxy server. Transcriptome data were de novo assembled for viral sequences with rnaviralSPAdes. The contig files were further organized by VirHunter and Kraken tools. Raw SRA data were trimmed and assembled for severe acute respiratory syndrome coronavirus 2 (SARS‐CoV‐2) genome by coronaSPAdes. The scaffolds were arranged by pairwise alignment against the SARS-CoV-2 reference genome (NC_045512.2). Coronavirus Typing Tool, Nextstrain, and Pangolin platforms were used to further investigate the SARS-CoV-2 genotype, phylogenetic analysis, and mutation estimations.

Results: Several environmentally related non-intact virus sequences reads from forest animals, moths, bacteria, and amoeba were detected. Further investigation resulted in non-indigenous sequences of SARS-CoV-2 genomes of lineage B with novel substitutions. Three polymorphisms, including A22D and A36V in the envelope protein, and Q498H in the spike (S) glycoprotein that were recently reported from a mice-adopted strain of SARS-CoV-2 with enhanced virulence were detected in all samples. Further novel substitutions at ORF1ab were also uncovered. These were L1457V, D4553N, W6538S, I1525T, D1585Y, D6928G, N3414K, and T3432S. Two unexpected frameshifts, ORF1a:2338-4401 and ORF1a:3681-4401, were also detected.

Conclusion: The findings of the presented study highlight the threats of the emerged potentially severe genotypes of SARS-CoV-2 bearing substitutions that are not yet clinically reported.

Article History

Received: 20 February 2023
Received in revised form: 25 March 2023
Accepted: 30 May 2023
Published online: 30 June 2023

DOI: 10.29252/JCBR.6.4.13

Keywords:
SARS-CoV-2
Solanum tuberosum
Metagenomics
RNA-Seq
Metavirome

Article Type: Original Article

Highlights:

What is current knowledge?

The virosphere is a valuable resource for studying viral population genetics, microbial evolution, and the discovery of novel sequences.
Metagenomics, including high-throughput sequencing (HTS), has made the virome accessible and revolutionized virology by detecting unknown viruses.

What is new here?

The presence of SARS-CoV-2 in the potato samples raises questions about the source of contamination and the potential for environmental persistence of the virus.

Introduction

The virosphere, the largest community of genetically diverse viruses, is a valuable source of studying viral population genetics and dynamics, microbial evolution, and the discovery of novel sequences (1). It offers the development of unique technologies for life science, industrial, medical, and diagnostic applications. These populations of viruses constitute the virome of different ecosystems or organisms. The development of different advanced next-generation sequencing technologies, such as high-throughput sequencing (HTS) that was exploited by metagenomics, made the virome pretty accessible (2).

Metagenomics, including deep-sequencing or HTS of RNA or DNA, provides transcriptome or genomic data for unhinging the complexity of a certain virome. De novo assembly of the virome of different metagenomic data revolutionized the field of virology by providing evidence for the detection of unknown viruses. In this regard, Yin et al. discovered two new RNA viruses with metagenomic analysis of symptomatic marigold plants (3). Furthermore, a recent metagenomic study by Lappe et al. uncovered novel RNA viruses belonging to the Betaflexiviridae, Tombusviridae, and Geminiviridae families (4). A different geographical region may possess genetically diverse sequence reads of viruses. Accordingly, a study assessed 73 grape samples for novel RNA viruses with HTS in Russia. The study revealed two novel RNA and DNA viruses namely grapevine umbra-like virus and grapevine pararetrovirus, respectively.

Application of metagenomics in clinical specimen vector-borne viruses (5, 6), respiratory viral infections (7), and blood-borne viruses have been studied (8). However, despite the potential of metagenomics studies in the discovery of novel viruses and revealing the hiding nods of the viral evolution, its robustness for surveillance on emerging and reemerging viral outbreaks, and bioterrorism remain mostly untouched. Because of the strength of the metagenomic studies in identifying novel viral sequences and the growing open-access data on biosamples in sequence read archive (SRA) depository, the present study aimed to hunt viral reads and make assemblies with accurate contigs to introduce novel viral intact sequences. Unexpectedly, the findings of the present study demonstrated the role of metagenomics investigations on plant specimens to uncover viral contaminations with potential public health threads.

Methods

Obtaining SRA data

RNA-Seq data of a BioSample project (SAMN32036489) of Solanum tuberosum raw sequence reads from a study in Qinghai University, Xining, China were obtained from NCBI's SRA database. RNA-Seq data have been generated by Illumina NovaSeq 6000 instrument by pair-end sequencing strategy. The BioSample project consisted of six samples namely sample_B3 (SRX18492426), sample_B2 (SRX18492425), sample_B1 (SRX18492424), sample_A3 (SRX18492423), sample_A2 (SRX18492422), and sample A1 (SRX18492421).

Genome assembly and virome detection

As the BioSample data consisted of transcriptome data of plant stolon samples of S. tuberosum, it was thought that it might provide some evidence of novel plant virome in the organism. For this, SRA data were retrieved by using the Galaxy platform (9). Paired-end sequence assembly of each. fastq data of samples was performed by de novo assembler for transcriptomes, metatranscriptomes, and metaviromes (rnaviralSPAdes) package v3.15.3 (10–12). The assembled contigs generated by rnaviralSPAdes were further classified by a deep learning method that uses convolutional neural networks and a random forest classifier to identify viruses in datasets, named VirHunter (13). Kraken v1 package (14) was used to assign taxonomic labels for sequence reads after classifying viral reads by

virHunter. Briefly, the fasta scaffolds of each assembled genome of the plant samples were selected separately, and labels were assigned based on the viral_2020 database. Further options were kept at default.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genome assembly

Surprisingly, a signature of SARS-CoV-2 was observed among sequencing reads. Therefore, it was aimed to uncover whether it was belonging to an intact virus or not. For this purpose, fastq files were trimmed for low-quality reads and further adaptors with Trimmomatic tools (15). In addition, the genome of S. tuberosum was mapped out in each sample with bowtie2 v2.5.0 (16, 17). Unaligned reads of mapping were further assembled by the SARS-CoV-2 de novo genome assembler, coronaSPAdes (10-12, 18). All the assembler options were kept as default. The resulting sequence reads in each scaffold were classified by Kraken package v1.1.1 (14). In this regard, classified reads were selected for further study as the main output based on the viral_2020 Kraken database. The matching query cover for each contig was also double-checked by NCBI's Nucleotide BLAST tool. For this purpose, the megablast algorithm was selected and a query of contigs was aligned to the SARS-CoV-2 (taxid: 2697049) database.

Genotyping and mutation analysis of SARS-CoV-2

The scaffolds of each sample were arranged based on the reference SARS-CoV-2 Wuhan-Hu-1 (NC_045512.2) with RagTag (19). Accordingly, the operation mode was set as homology-based misassembly correction. Coronavirus Typing Tool was used to further investigate the SARS-CoV-2 genotype, phylogenetic analysis, and mutation estimations among samples (20). The lineage and distribution of samples were determined by Nextstrain (https://clades.nextstrain.org/) or Pangolin (https://cov-lineages.org/resources/pangolin.html) web servers (21-24).

Results

The RNA sequencing data of S. tuberosum stolon samples were used to detect viral RNA reads. This could be attributed to viral RNA genome or DNA viral transcripts. Accordingly, the contigs of the assembled genome were scavenged by the rnaviralSPAdes package. A range of sequence reads was assigned to different viral taxonomic classes.

The virome of S. tuberosum

A combined reads of six S. tuberosum stolon specimen are shown in Figure 1. From the Caulimoviridae family and the order Ortevirales, the tobacco vein clearing virus was the only detected species. Other plant-associated viruses were dahlia mosaic virus and Cavemovirus. In the next frequent taxonomic rank, Nucleocytoviricota phylum, BeAn 58058 virus, which was originally isolated from a wild forest rodent was the most prevalent species. Further members of this phylum were constituted of the Megaviricetes class that mainly infects free-living amoeba. In addition, Naldaviricetes class was the third most frequent viral taxonomic rank. Choristoneura fumiferana granulovirus was the most frequent species in this rank followed by Thysanoplusia orichalcea nucleopolyhedrovirus and Spodoptera frugiperda granulovirus, which all infect moths.

Further investigation demonstrated a vast number of bacteriophages Figure 2, all belonging to the Heunggongvirae kingdom. In this rank, the Lambdavirus species was the most frequent (52% of the kingdom sequence reads). The assessment of the transcriptome reads of the Orthornavirae kingdom revealed further plant- and insect-associated viral species, including pepper chlorotic spot orthotospovirus and Himetobi P virus, respectively. Bat-associated circovirus 4 reads (100% reads) in the genus Circovirus was the other finding of the viral transcriptome analysis of potato samples.

Further interesting findings were related to the finding of the sequence reads of bat coronavirus BM48-31/BGR/2008, which comprised 2% of the classified viral reads. Concerning the current pandemic of SARS-CoV-2, the study aimed to reassemble the S. tuberosum transcriptome in a reference-sequence-based manner.

Signs of SARS-CoV-2 genome in potato

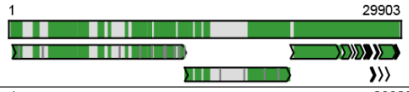
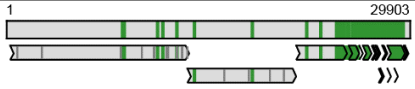
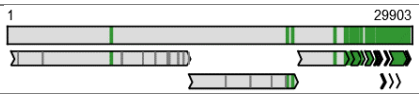
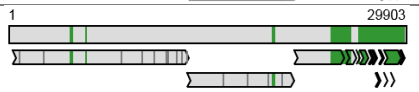
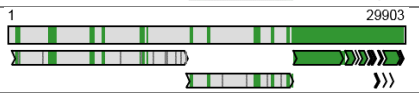
The fingerprints of the SARS-CoV-2 genome were found in all six samples. However, only partial genomic contigs were managed to construct. Accordingly, the range of SARS-CoV-2 query coverage of the sample contigs with the viral reference genome is shown in Table 1.

Genotyping of SARS-CoV-2 contaminations among potato specimens

The assembled data consisted of low contigs coverage but reliable information on the SARS-CoV-2 genome. Accordingly, all samples were clustered in the international A_B diversity. Furthermore, all samples belonged to lineage B and Clade 19A. The results also demonstrated some nucleotide mutations in all samples. Accordingly, Sample_A1 contained nine substitutions including, C3264T (ORF1ab: T1000I), T4634G (ORF1ab: L1457V), C11152A, C13019T, C18888T, A21917C (S: I119L), A23056C, C26309A (E: A22D), and C26351T (E: A36V). Unexpectedly, an unknown frameshift was observed at ORF1a:3681-4401, leading to deleted codon range. Additional mutations were observed in Sample_A2. Accordingly, four protein mutations, including G13921A (ORF1ab/ORFb: D4553N/D152N), C25609T (ORF3: L73F), C26309A (E: A22D), and C26351T (E: A36V) were detected. Also, three protein mutations were detected in Sample_A3. These were an unexpected premature stop codon at G21082T (ORF1ab/ORF1b: E6940*/E2539*), and C26309A (E: A22D) and C26351T (E: A36V).

Further results on Sample_B1 demonstrated two protein mutations at G19877C (ORF1ab/ORF1b: W6538S/W2137S) and C26351T (E: A36V). More mutations were also observed in the genomic data of Sample_B2 with 12 mutations containing both codon and protein mutations. The mutations were G943A, C3264T (ORF1ab/ORF1a: T1000I), T4839C (ORF1ab/ORF1a: I1525T), G5018T (ORF1ab/ORF1a: D1585Y), A10126T, A20895G, A21047G (ORF1ab/ORF1b: D6928G), T21847C, A23056C (S: Q498H), C23997G (S: P812R), C26309A (E: A22D), and C26351T (E: A36V). Moreover, nine nucleotide substitutions were found in Sample_B3 including C3264T (ORF1ab/ORF1a: T1000I), C10507A (ORF1ab/ORF1a: N3414K), A10559T (ORF1ab/ORF1a: T3432S), C18888T, G22817A (S: A419T), A23056C (S: Q498H), C23997G (S: P812R), C26309A (E: A22D), and C26351T (E: A36V). Furthermore, a deletion 22814_22816delATT (I418del) was detected in the S coding sequence. Similar to the above-mentioned results, an unexpected frameshift was observed in the same position (ORF1a:2338-4401), leading to a deleted codon range.

Table 1: Coronavirus genotyping results

Specimen	# Contigs	Est. # Reads	Coverage (%)	NT position	Identity (%)		Sub-clustering	Genome Coverage
					NT	AA		
Sample_A1	17	~140	74.8	199-29858	100	99.7	International A_B Diversity	
Sample_A2	11	~70	23.6	8558-29551	99.9	99.4	International A_B Diversity	
Sample_A3	9	~50	19.6	7562- 29869	99.9	99.4	International A_B Diversity	
Sample_B1	5	~50	19.6	4564- 29861	100	99.6	International A_B Diversity	
Sample_B2	18	~160	54.2	207-29859	99.9	99.6	International A_B Diversity	
Sample_B3	13	~120	41.2	551- 29894	99.9	99.4	International A_B Diversity	

The phylogenic analysis (Figure 3) of the SARS-CoV-2 assembled genome revealed an origination of the Sample_A1 to the Netherlands_19 sequence that both are genetically associated closely to the reference sequence, SARS-CoV-2 isolate Wuhan-Hu-1 (NC_045512). Further isolates, Sample_A2, Sample_B3, Sample_A3, Sample_B2, and Sample_B1 were slightly distant genetically from the origin.

Figure 1: Krona pie chart on combined viral sequence reads. The reads are taxonomically arranged in different viral groups. The order of Ortevirales comprised 46% of viral reads. The reads from Nucleocytoviricota phylum, Naldaviricetes class, and Heunggongvirae kingdom with 23%, 10%, and 10% frequency were the next important viral reads

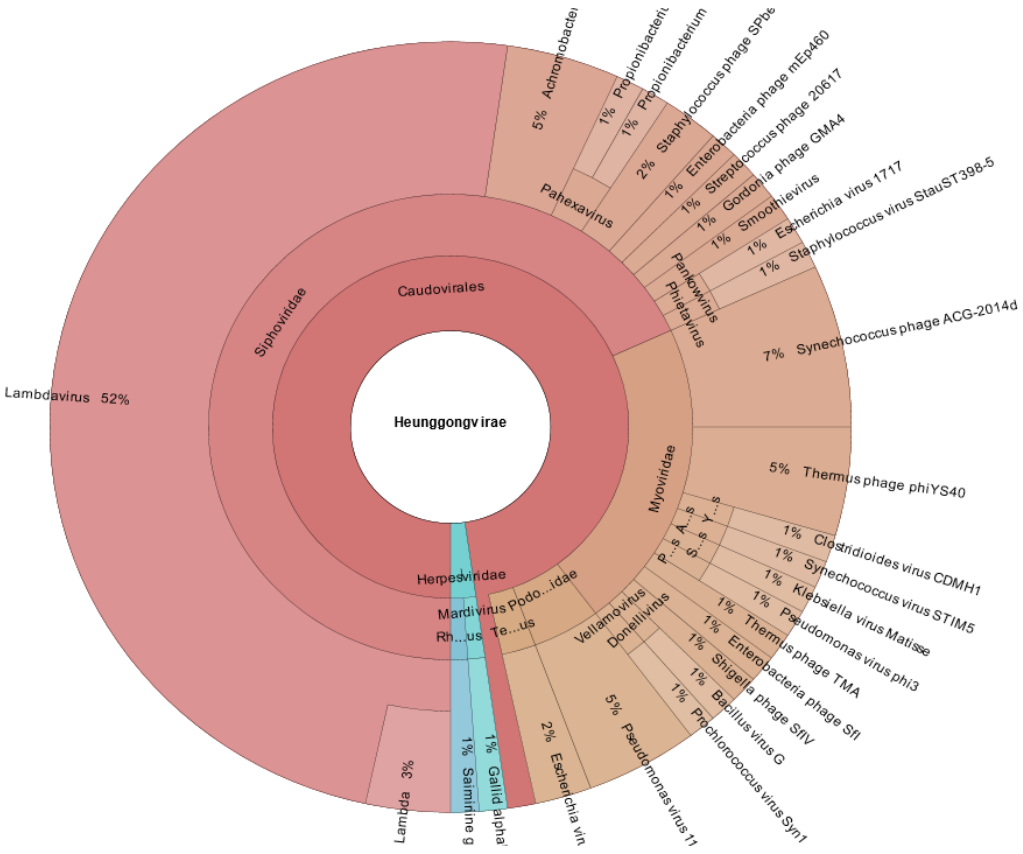


Figure 2: Krona pie chart of the kingdom of Heunuggongvirae. Different bacteriophage transcriptome is identified among the stem samples of S. tuberosum.

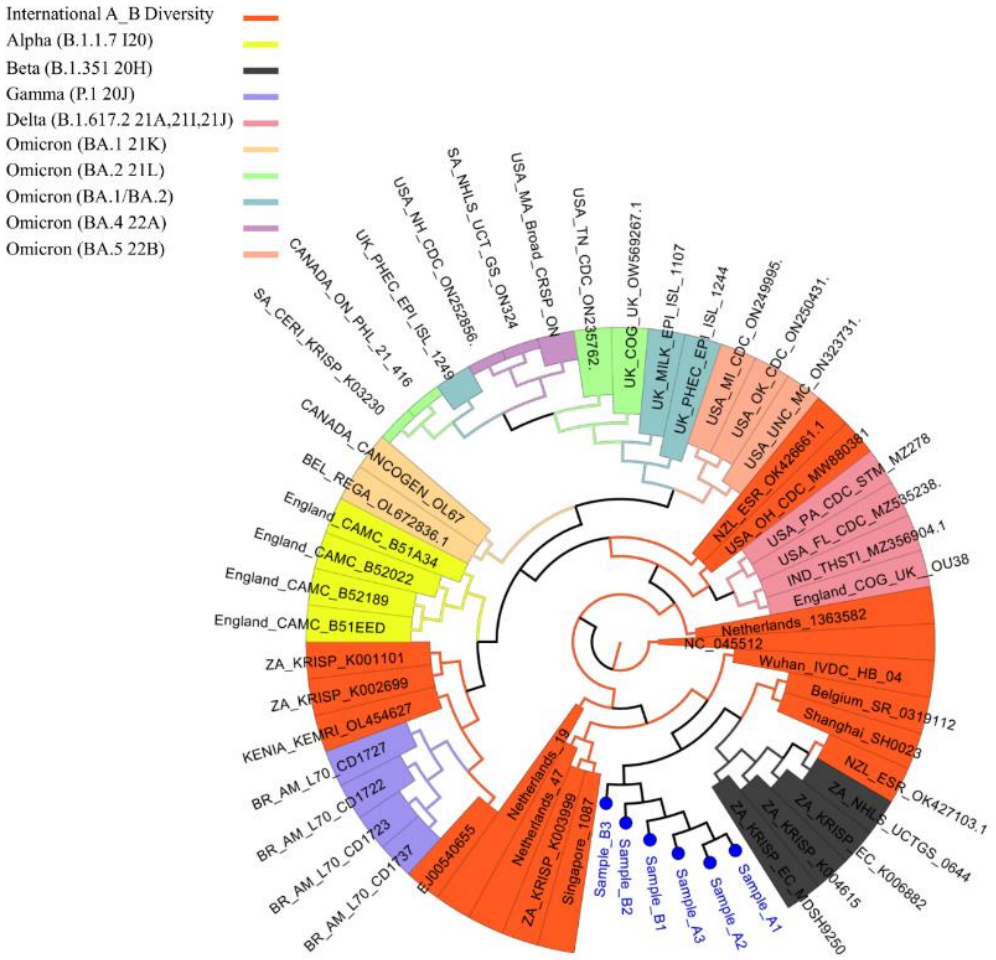


Figure 3: Phylogenetic tree representing sub-clustering S. tuberosum samples. Sample_A1 with the fewest gaps in the genome assembly characterizes close genetic distance with the origin of the SARS-CoV-2

Discussion

Metagenomic studies of plants, animals, and environmental samples are always enriched with virosphere containing undiscovered viruses. In the present study, a bioproject of six biosamples of the transcriptome of S. tuberosum stolon specimen was assessed for viral sequences. Investigation of the plants' virome revealed sequence reads of the viruses infecting mainly the plant itself or other passing through organisms like moths, forest rodents, and bacteriophages infecting widespread bacteria. Since stolon is a part of the plant growing outside the ground, it is exposed to virus-bearing organisms, and it would be a good source of unknown viral sequences. Further investigation of the transcriptome of the S. tuberosum did not result in finding any intact viral genome. Although successful mapping of different contigs of the transcriptome to the bacteriophage reference genome was expected, no intact viral genomes were obtained (data are not shown).

One of the interesting findings of the present study was the contamination of the samples with bat viruses, including bat coronavirus BM48-31/BGR/2008 and bat-associated circovirus 4. Since the sampling was recent, it was decided to serendipitously map the transcriptome with the human SARS-CoV-2. Surprisingly, all six samples contained different-sized scaffolds with SARS-CoV-2 genome contigs. The contigs were arranged based on the SARS-CoV-2 Wuhan isolate to see the exact overlapping of the contigs to the full-length viral genome. The reason for masking the SARs-CoV-2 sequence reads in the taxonomic classification could be the Kraken viral database (Viral-2020) that may not contain enough information on SARS-CoV-2. However, the results demonstrated relatively full overlaps of the SARS-CoV-2 genome with assemblies of Sample_A1 (22379 nt) and Sample_B2 (16202 nt). Some gaps were observed in the genomes, which is might be due to the low coverage or degradation of some parts of the genome owing to the sampling or environmental nucleases.

It is well known that the SARS-CoV-2 genome might endure in outdoor environments for days. Unfortunately, I am not aware of the sampling time and the sequencing, or the process of sample preparation to correctly decide whether the samples were infected during sampling by the researchers or whether the viral genome was present before sampling. The former hypothesis might be true as the samples were collected from China, and this country still suffers from the coronavirus disease 2019 pandemic. Moreover, all six genomes of SARS-CoV-2 were belonging to lineage B, which is one of the two original haplotypes of SARS-CoV-2 that were recorded in late 2019. Additionally, the phylogenetic analysis demonstrated that Sample_A1, which contained the largest intact genome, had a close genetic distance from the firstly discovered viral isolates. These results may support that the earlier quasispecies of SARS-CoV-2 are still in circulation. On the other hand, the distribution of genomic mutation of the SARS-CoV-2 genome in the samples was not identical, indicating that there was no cross-contamination either during the sampling and experimental process or through contaminated sequencing instruments.

Since the contigs were strongly accurate, the investigation for genetic changes was performed. The alignment resulted in some protein-coding and codon mutations only at ORF1ab, ORF1a, ORF1b, S, ORF3a, and E. Accordingly, amino acid substitution A36V (26351C>T) in E was observed in all samples (6/6), and it was followed by A22D (26309C>A) (5/6). One more frequent substitution was T1000I at ORF1ab/ORF1a (3/6). Two substitutions at the E coding sequence are novel and recently reported by a study in China due to the several passages of a laboratory strain of two lethal mouse-adapted SARS-CoV-2 variants (25). This finding is very important due to the current severity of SARS-CoV-2 in China. The authors also reported the association of these two mutations with an elevated level of virus severity in mice. Here, it can be suggested that the serial passage of the earlier isolate of SARS-CoV-2 (Wuhan) has led to the development of a more severe strain of the virus bearing two important substitutions at E protein. This also implies the leak of an animal-adopted strain of SARS-CoV-2 since the Q498H substitutions were recognized in Sample_A1 and Sample_B3. Accordingly, Yan et al. reported Q498H at the receptor-binding domain of S protein is necessary for receptor switching from human angiotensin-converting enzyme 2 (ACE2) to mice ACE2, and it also grows the virulence of the virus due to the increased affinity of the viral ligand to its cognate receptor. Another mutation at the S protein was I119L, which co-existed with the Q498H in Sample_A1, which has not

been reported before, and for the first time, it is found in an environmental sample. It may also contribute to the severity of the virus, as its incidence coincided with Q498H. Further study is required to fully uncover the role of I119L in the S protein. One deletion, I418del (22814_22816delATT) in the S protein was observed in Sample_B3. It is reported that I418 is involved in the virus interaction with the ACE2 receptor (26, 27). However, there is no knowledge of the deletion at this position on the SARS-CoV-2 genome. In this study, L73F was the only amino acid substitution at the ORF3a and envelope protein, which was found in Sample_A2. This substitution is involved in the stabilization of viral protein due to a higher number of interactive atoms (28). This deletion was associated with three more amino acid substitutions in Sample_B3, including A419T, Q498H, and P812R. Again, A419T has not been reported before; however, P812R was also developed due to the passage of the virus in Vero cells at the S2' cleavage site (29, 30).

Due to the overlapping frames of ORF1ab in SARS-CoV-2 with ORF1a and ORF1b, I only present the amino acid changes based on the amino acid position at ORF1ab and the position at the overlapping frames present between parentheses. The T1000I that was observed in three samples along with L1457V in Sample_A1, D4553N in Sample_A2, E6940* (ORF1b: E2539*), in Sample_A3, W6538S (ORF1b: W2137S) in Sample_B1, I1525T, D1585Y, and D6928G (ORF1b: D2527G) in Sample_B2, and N3414K in Sample_B3 have not been reported before. Only T3432S in Sample_B3 was reported in an Omicron sub-lineage isolate in Ecuador (31). Interestingly, two novel unknown frameshifts were detected in two samples, including ORF1a:2338-4401 and ORF1a:3681-4401, both leading to a deleted codon range in ORF1a. Unfortunately, I was not able to investigate the role of these frameshifts in viral protein size changes, but it is warranted to study these novel genetic changes in SARS-CoV-2.

Conclusion

This study was designed to computationally investigate the virome of potatoes to discover an intact viral genome that led to the finding of potentially lab-leaked SARS-CoV-2 sequence reads. This was also confirmed by the phylogenic investigation and finding of two substitutions, A22D and A36V, which were developed in mice. The laboratory strains were developed in Changchun, China, while the potato samples were from Xining. Since there was no report of such substitutions, it is hard to conclude if the virus strain having both mutations emerged as an evolutionary incidence. However, more studies are required for uncovering the role of introduced novel mutations in the virus. Furthermore, contamination of environmental raw materials such as plants with SARS-coV-2 still poses a threat to public health.

Acknowledgements

None.

Funding source

The author received no financial support for the research, authorship, and/or publication of this article.

Ethical statement

Not applicable.

Conflict of interest

The author declares there is no conflict of interest regarding the publication of this article.

Author contributions

All stages of the study from design to manuscript preparation was done by A.M.

References

1. Zhang YZ, Shi M, Holmes EC. Using Metagenomics to Characterize an Expanding Virosphere. Cell. 2018;172(6):1168-72. [View at Publisher] [Google Scholar] [DOI] [PMID]
2. Aevarsson A, Kaczorowska AK, Adalsteinsson BT, Ahlqvist J, Al-Karadaghi S, Altenbuchner J, et al. Going to extremes - a metagenomic journey into the dark matter of life. FEMS Microbiol Lett. 2021;368(12). [View at Publisher] [Google Scholar] [DOI] [PMID]
3. Yin H, Dong Z, Wang X, Lu S, Xia F, Abuduwaili A, et al. Metagenomic Analysis of Marigold: Mixed Infection Including Two New Viruses. Viruses. 2021;13(7):1254. [View at Publisher] [Google Scholar] [DOI] [PMID]
4. Lappe RR, Elmore MG, Lozier ZR, Jander G, Miller WA, Whitham SA. Metagenomic identification of novel viruses of maize and teosinte in North America. BMC Genomics. 2022;23(1):1-15. [View at Publisher] [Google Scholar] [DOI] [PMID]
5. Damian D, Maghembe R, Damas M, Wensman JJ, Berg M. Application of viral metagenomics for study of emerging and reemerging tick-borne viruses. Vector Borne Zoonotic Dis. 2020;20(8):557-65. [View at Publisher] [Google Scholar] [DOI] [PMID]
6. Souza JVC, Santos H de O, Leite AB, Giovanetti M, Bezerra R dos S, Carvalho E de, et al. Viral Metagenomics for the Identification of Emerging Infections in Clinical Samples with Inconclusive Dengue, Zika, and Chikungunya Viral Amplification. Viruses. 2022;14(9):1933. [View at Publisher] [Google Scholar] [DOI] [PMID]
7. Mohsin H, Asif A, Fatima M, Rehman Y. Potential role of viral metagenomics as a surveillance tool for the early detection of emerging novel pathogens. Arch Microbiol. 2021;203(3):865-72. [View at Publisher] [Google Scholar] [DOI] [PMID]
8. Slavov SN. Viral Metagenomics for Identification of Emerging Viruses in Transfusion Medicine. Viruses. 2022;14(11):2448. [View at Publisher] [Google Scholar] [DOI] [PMID]
9. Erratum: Correction to "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update". Nucleic Acids Res. 2022; 50(15):8999. [View at Publisher] [Google Scholar] [DOI]
10. Antipov D, Korobeynikov A, McLean JS, Pevzner PA. HYBRIDSPADES: an algorithm for hybrid assembly of short and long reads. Bioinformatics. 2016;32(7):1009-15. [View at Publisher] [Google Scholar] [DOI] [PMID]
11. Prjibelski AD, Vasilinetc I, Bankevich A, Gurevich A, Krivosheeva T, Nurk S, et al. ExSPAnder: a universal repeat resolver for DNA fragment assembly. Bioinformatics. 2014;30(12):i293-301. [View at Publisher] [Google Scholar] [DOI] [PMID]
12. Vasilinetc I, Prjibelski AD, Gurevich A, Korobeynikov A, Pevzner PA. Assembling short reads from jumping libraries with large insert sizes. Bioinformatics. 2015;31(20):3262-8. [View at Publisher] [Google Scholar] [DOI] [PMID]
13. Sukhorukov G, Khalili M, Gascuel O, Candresse T, Marais-Colombel A, Nikolski M. VirHunter: A Deep Learning-Based Method for Detection of Novel RNA Viruses in Plant Sequencing Data. Front Bioinform. 2022;2:38. [View at Publisher] [Google Scholar] [DOI] [PMID]
14. Wood DE, Salzberg SL. Kraken: Ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15(3):1-2. [View at Publisher] [Google Scholar] [DOI] [PMID]
15. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20. [View at Publisher] [Google Scholar] [DOI] [PMID]
16. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9. [View at publisher] [Google scholar] [DOI] [PMID]
17. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):1-10. [View at Publisher] [Google Scholar] [DOI] [PMID]
18. Meleshko D, Hajirasouliha I, Korobeynikov A. coronaSPAdes: from biosynthetic gene clusters to RNA viral assemblies. Bioinformatics. 2021;38(1):1-8. [View at Publisher] [Google Scholar] [DOI] [PMID]
19. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, et al. RaGOO: Fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 2019;20(1):1-7. [View at Publisher] [Google Scholar] [DOI] [PMID]
20. Maan H, Mbareche H, Raphenya AR, Banerjee A, Nasir JA, Kozak RA, et al. Genotyping SARS-CoV-2 through an interactive web application. Lancet Digit Heal. 2020;2(7):e340-1. [View at Publisher] [Google Scholar] [DOI] [PMID]
21. O'Toole Á, Hill V, Pybus OG, Watts A, Bogoch II, Khan K, et al. Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2 with grinch. Wellcome Open Res. 2021;6:121. [View at Publisher] [Google Scholar] [DOI] [PMID]
22. O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol. 2021;7(2):veab064. [View at Publisher] [Google Scholar] [DOI] [PMID]
23. Rambaut A, Holmes E, O'Toole Á, VH-N, 2020 undefined. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020;5(11):1403-7. [View at Publisher] [Google Scholar] [DOI] [PMID]
24. Aksamentov I, Roemer C, Hodcroft EB, Neher RA. Nextclade: clade assignment, mutation calling and quality control for viral genomes. JOSS. 2021;6(67):3773. [View at Publisher] [Google Scholar] [DOI]

25. Yan F, Li E, Wang T, Li Y, Liu J, Wang W, et al. Characterization of Two Heterogeneous Lethal Mouse-Adapted SARS-CoV-2 Variants Recapitulating Representative Aspects of Human COVID-19. Front Immunol. 2022;13: 821664. [View at Publisher] [Google Scholar] [DOI] [PMID]
26. Shannon A, Fattorini V, Sama B, Selisko B, Feracci M, Falcou C, et al. A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase. Nat Commun. 2022;13(1):621. [View at Publisher] [Google Scholar] [DOI] [PMID]
27. Watanabe C, Okiyama Y, Tanaka S, Fukuzawa K, Honma T. Molecular recognition of SARS-CoV-2 spike glycoprotein: quantum chemical hot spot and epitope analyses. Chem Sci. 2021;12(13):4722-39. [View at Publisher] [Google Scholar] [DOI] [PMID]
28. Mou K, Abdalla M, Wei DQ, Khan MT, Lodhi MS, Darwish DB, et al. Emerging mutations in envelope protein of SARS-CoV-2 and their effect on thermodynamic properties. Inform Med Unlocked. 2021;25:100675. [View at Publisher] [Google Scholar] [DOI] [PMID]
29. Bamford CGG, Broadbent L, Aranday-Cortes E, McCabe M, McKenna J, Courtney DG, et al. Comparison of SARS-CoV-2 Evolution in Paediatric Primary Airway Epithelial Cell Cultures Compared with Vero-Derived Cell Lines. Viruses. 2022;14(2):325. [View at Publisher] [Google Scholar] [DOI] [PMID]
30. Dieterle ME, Haslwanter D, Bortz RH, Wirchnianski AS, Lasso G, Vergnolle O, et al. A Replication-Competent Vesicular Stomatitis Virus for Studies of SARS-CoV-2 Spike-Mediated Cell Entry and Its Inhibition. Cell Host Microbe. 2020;28(3):486-96. [View at Publisher] [Google Scholar] [DOI] [PMID]
31. Carrazco-Montalvo A, Herrera-Yela A, Alarcón-Vallejo D, Gutiérrez-Pallo D, Armendáriz-Castillo I, Andrade-Molina D, et al. Omicron sublineages current status in Ecuador. Preprints [V1]. 2022. [View at Publisher] [Google Scholar] [DOI]

How to Cite:
Mohebbi A. Analysis of the RNA-Seq data of Solanum tuberosum revealed viral sequence reads of a severe laboratory-developed strain of SARS-CoV-2 containing novel substitutions. *JCBR*. 2022;6(4):13-18.