Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

SARS-CoV-2: To transgress or Not?

M Varun
Central University of Tamil Nadu
Praveen Ebenezer
Centre for Cellular and Molecular Biology
P S Charan
Bioclues.org
Adithya Joseph
Bioclues.org
Sunil Kunar Polipalli
MAMC &amp, Associated Lok Nayak Hospital
Shyam Prasad Rao
Nitte University
Gyaneshwer Chaubey
Banaras Hindu University
Prashanth Suravajhala
Prash.s@jaipur.manipal.edu

Manipal University Jaipur

Systematic Review

Keywords: SARS-CoV-2, Long non-coding RNAs, Lateral gene transfer, Transgression, Host genome integration, mtDNA, Cytoscape analysis, Viral pathogenesis, Immune evasion, ACE2, TMPRSS2

Posted Date: May 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9240289/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

SARS-CoV-2: to transgress or not?

Varun M[1*] , Praveen Ebenezer[2*], Charan P S V V[2*], Adithya Joseph[3*], Sunil Kunar Polipalli[4] R Shyam Prasad Rao[5], Gyaneshwer Chaubey[6], and Prashanth Suravajhala[7]

1. Department of Biotechnology, Central University of Tamil Nadu, Thiruvarur, India.
2. Bioclues.org, Hyderabad, India
3. St Xavier's College, Kolkata
4. Genetics and Genomics Sequencing Lab, Department of Medical Genetics, MAMC & Associated Lok Nayak Hospital, New Delhi
5. Department of Bioinformatics, NUCSER, NITTE(Deemed to be University), Mangaluru
6. Gyan Cytogenetics Lab, Department of Zoology, Banaras Hindu University, Varanasi
7. Systems Genomics Group, Department of Biosciences, Manipal University Jaipur, Dehmi Kalan, 303007, Rajasthan, India.

*Equal contributing first authors

#Corresponding author Email: Prash.s@jaipur.manipal.edu

Abstract

Over the last three decades, we have witnessed three major outbreaks—SARS, MERS, and COVID-19—resulting in approximately 35 million infections worldwide. SARS-CoV-2, in particular, has developed sophisticated methods to hijack host mechanisms, often involving differentially expressed genes and long non-coding RNAs (lncRNAs), which our lab has studied extensively over the past five years. This study investigates whether the SARS-CoV-2 genome mimics elements of the host genome to facilitate replication, evade the immune system, and hijack host machinery. We achieved this by analyzing sequence similarities between the viral genome and the human genome, including the mitochondrial genome. Our work presents the mechanisms associated with SARS-CoV-2 transgression and explores the plausible implications of lateral gene transfer.

Keywords: SARS-CoV-2, Long non-coding RNAs, Lateral gene transfer, Transgression, Host genome integration, mtDNA, Cytoscape analysis, Viral pathogenesis, Immune evasion, ACE2, TMPRSS2

Introduction

Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) or coronavirus diseases (COVID-19) has created a global pandemic with a larger incidence in those individuals with existing diseases infected more than 778 million globally since 2019(*COVID-19 Cases | WHO COVID-19 Dashboard*, n.d.). While it is widespread in bats around the world, it can be found in many other species as well, including birds, cats, dogs, pigs, mice, horses, whales, and humans and the covid 19 pandemic began with an outbreak of COVID-19 in Wuhan, China in December 2019, later spread to other areas of Asia and worldwide in early 2020 (Santos-López et al, 2021). Over the past two decades corona viruses (CoVs) have caused three large scale outbreaks such as Severe Acute Respiratory syndrome(SARS), Middle Eastern Respiratory Syndrome (MERS) and the recent COVID-19 (Harrison et al., 2020), Coronaviruses are members of *coronaviridae* family and the order *Nidovirales* according to International Committee on Taxonomy of Viruses, the subfamily of Coronaviruses have four genera namely *Alphacoronavirus, Betacoronavirus, Gammacoronavirus and Deltacoronavirus.* (Cui et al, 2019; Ravi et al., 2022)(Akolia et al., 2025). SARS-CoV-2 being a positive sense single strand RNA virus and with one-third of the genome encoding for structural

protein such as surface (S), envelope (E), membrane (M), nucleocapsid (N) and other accessory protein encoded by ORF3a, ORF6, ORF7a, ORF7b, and ORF8 genes (Li et al., 2005), has a vivid genome structure. While the ORF1ab protein is produced through a special translation mechanism known as -1 ribosomal frameshifting, the ribosome encounters a slippery sequence (UUUAAAC) followed by a pseudoknot structure in the RNA, causing it to shift backward by one nucleotide (−1), allowing the translation to continue from ORF1a into ORF1b, resulting in a single extended replicase polyprotein (Kelly et al., 2020), it also produces other proteins responsible for executing key viral processes, including the transcription of mRNA, its translation into proteins, and the encapsulation of the viral genome especially ORF1ab polypeptide cleaved by two proteases. Plpro (papin like protease- part of Nsp3) & 3CLpro (3C like protease- Nsp5)which yields 15 to 16 Nsp include Nsp12 (RNA dependent RNA polymerase) & Nsp13 (Helicase) (Tamayo-Ordóñez et al., 2023;(Sharma et al., 202**1**)

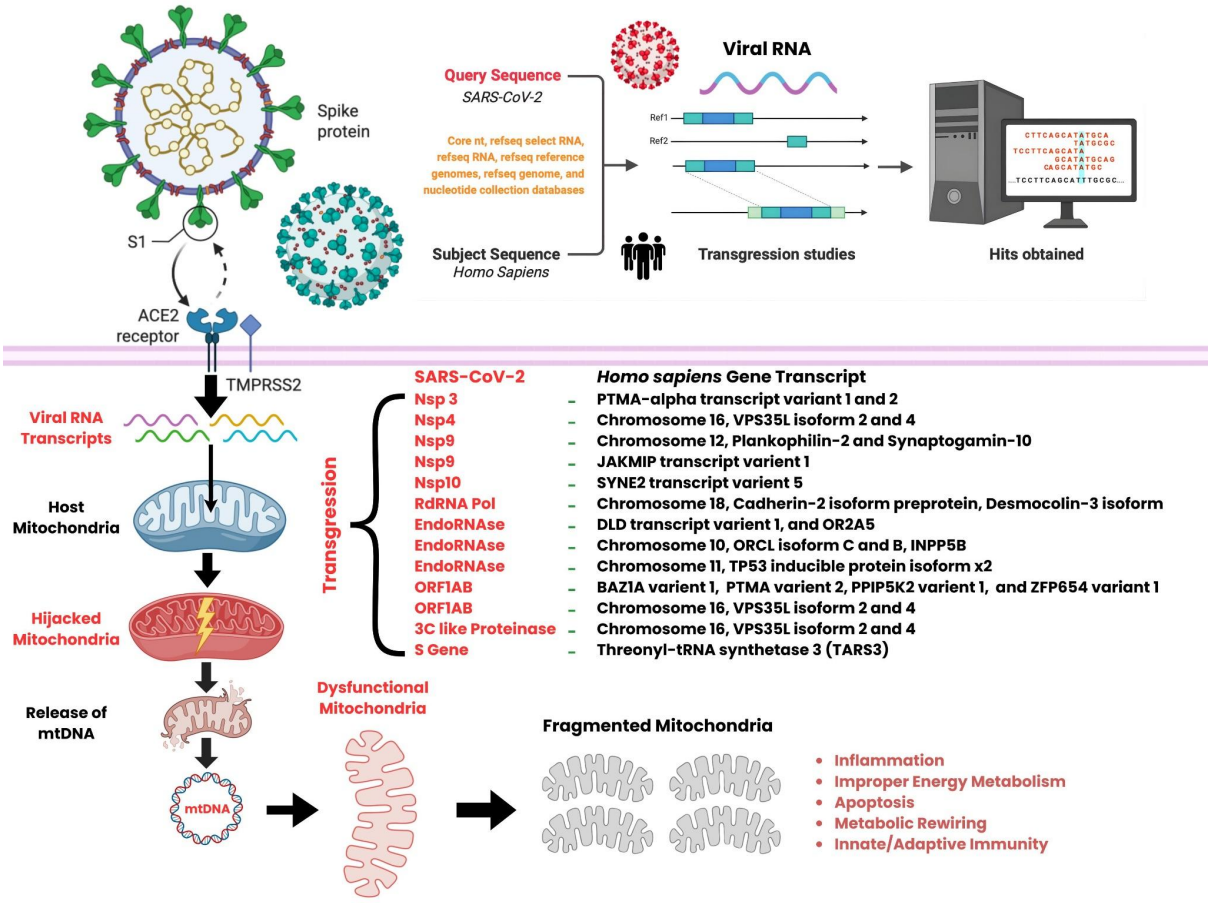
SARS-CoV-2 Pathogenesis: Is it guided?

The viral entry is mediated by the Spike protein(S), a glycoprotein assembled as a homotrimer with multiple copies on the membrane, which gives a crown-like appearance (Jackson et al., 2022). The binding of viral S protein to host cell receptor [angiotensin converting enzyme carboxy peptidase (ACE2), primary role is to convert angiotensin I and angiotensin II into angiotensin 1-9 and regulate blood pressure] where the host protease Transmembrane serine protease (TMPRSS2) cleaves the S protein into S1 and S2 subunits (Sharma et al.,2021). The life cycle of SARS-CoV-2, begins with the virus entering the host cell either by membrane fusion, where the viral envelope merges directly with the host cell membrane, or through endocytosis, where the virus is engulfed into a vesicle. Once inside the host cell, the viral envelope is removed, and the positive-sense RNA genome is released into the cytoplasm. Because the viral RNA is of the positive-sense type, it can function directly as messenger RNA (mRNA). Host ribosomes immediately begin translating the viral genome, particularly the open reading frames ORF1a and ORF1ab, which produce large polyproteins. These polyproteins are not functional in their initial form; they undergo autoproteolytic cleavage, a process mediated by two viral proteases: papain-like protease (PLpro, nsp3) and 3C-like protease (3CLpro, nsp5). This cleavage also yields multiple non-structural proteins (NSPs), which then come together to form the replication-transcription complex (RTC). The RTC uses the incoming viral genome as a template to synthesize complementary negative-sense RNA strands. These newly formed (-) RNA strands serve two major purposes: as templates for the replication of full-length genomic RNA and for the transcription of shorter subgenomic mRNAs, which are essential for expressing viral structural and accessory proteins. By using these (-) sense templates, the virus synthesizes a new batch of positive-sense RNAs. This includes not only full-length copies of the genome, meant to be packaged into new virions, but also multiple subgenomic (+) RNAs that encode the structural proteins—Spike (S), Envelope (E), Membrane (M)—and accessory proteins. The S, E, and M proteins are translated at the endoplasmic reticulum (ER) and are inserted into the ER membrane, while the nucleocapsid (N) protein is translated in the cytoplasm and binds to the new genomic RNAs. The assembly of new virus particles occurs at the ER–Golgi complex. Here, the N protein-genome complexes are packaged into budding vesicles containing the S, E, and M proteins embedded in the membrane, forming mature virions. These newly formed virions are then enclosed within intracellular vesicles and transported through the secretory pathway. Finally, the vesicles fuse with the host cell membrane, releasing the mature virus particles into the extracellular environment by exocytosis, ready to infect neighboring cells and propagate the infection cycle. There is a major research gap on whether or not SARS-CoV-2 pathogenesis is in lieu of fate of the host or intent mitochondrial hijacking. Keeping in view of the aforementioned gaps, in this work,we contemplate the origins of SARS-CoV-2 trespassing.

Materials and Methods

The Global Initiative on Sharing All Influenza Data (GISAID), found in 2008 (https://gisaid.org/ last accessed on October 28, 2025) provides access to genomic data of influenza viruses (Yuelong Shu et al, 2017). Not only Influenza, it also includes genomic data of Covid-19, Pox viruses, Chikungunya

viruses, Dengue viruses, Oropouche viruses, zika viruses. This database helps in real-time surveillance to monitor the emergence of new strains of viruses. The world's largest repository of COVID-19 sequences, and by far the largest SARS CoV 2 sequence database, is maintained by GISAID. Over 10 million sequences have been submitted in early to mid 2022 and over 14.4 million sequences have been submitted by end of 2022 to January 2023 (Figure 1b).



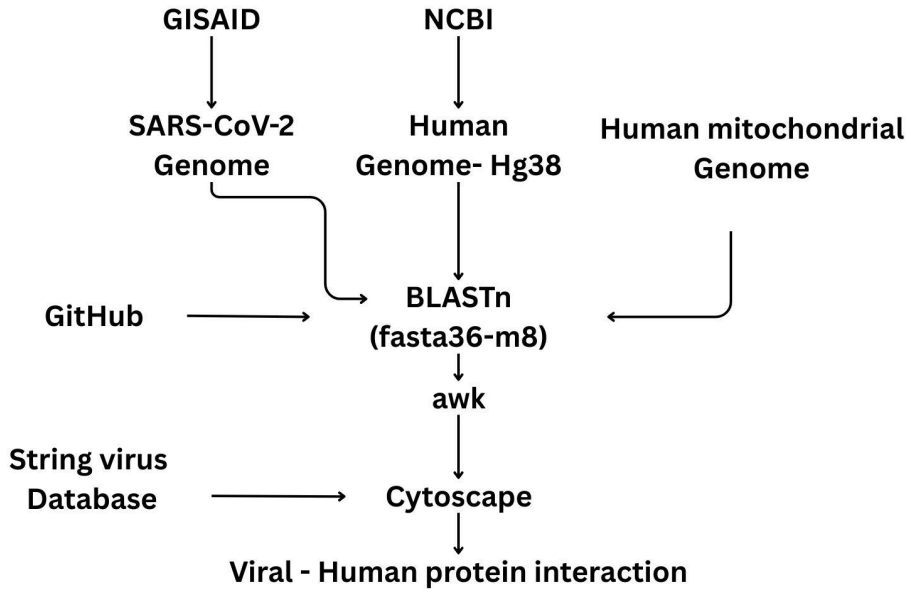


Fig 1 (a): Central SARS-CoV-2 figure, (b): Methodological workflow of pipelines and tools

String Viruses Database: The database which contains the known and predicted protein-protein interactions (http://viruses.string-db.org/ last accessed on November 17, 2025)). These interactions might be direct, which is physical, or indirect, which is a functional association. These interactions are derived from five main sources - Genomic content predictions, High-throughput lab experiments, Co-expression, Automated Textmining, and Previous database knowledge. This String virus database is contributed by Swiss Institute of Bioinformatics, EMBL, University of Zurich, NNF Center for Protein Research.

Cytoscape and Cytohubba: Cytoscape is an open source, Java-based software platform (https://cytoscape.org/ last accessed on October 28, 2025) that helps in building and visualizing molecular interaction networks and biological pathways. Cytoscape also helps in integrating the networks with annotations, gene expression profiles and other state data. Cytoscape also contains additional features available as apps which are available for molecular profiling analyses, new layouts, additional file format support and helps to connect with other databases. This database is funded and maintained by the National Genome Research Institute(NHGRI). From the GISAID (Global Initiative on Sharing All Influenza Data), 500 SARS-CoV-2 samples were collected randomly - 94 from Africa, 103 Asia, 103 from Europe, 100 from South America, 100 from North America and 84 annotated and complete genome sequence samples were collected from NCBI. Along with these samples Human genome Hg38 & Human mitochondrial genome was downloaded from NCBI (National Centre for Biotechnology Information).

Africa	94

Asia	103
Europe	103
South America	100
North America	100
Total	500

Table 2: Number of samples from GISAID.

Australia	01
Brazil	01
China	23
India	02
Italy	01
Korea	01
Nepal	01
Sweden	01
USA	48
Taiwan	03
Vietnam	02
Total	84

Table 3: Samples collected from Genebanks.

Using fasta36 bash liners from GitHub, BLASTn was performed using SARS-CoV-2 genome as a query and Hg38 as subject, similarly with human mitochondrial genome (Pearson, 2015). Wherever applicable, all the files were checked using awk liners using %identity >=90 and variable query coverage from start to end positions. The resultant files were was filtered using awk to filter out the significant hits where **awk ' $3>=70 && $6 <=3 && $11<=0.01' *out** for the BLAST results against hg38 and **awk '$3>=90 && $6<=1 && $11<=0.05' *.out** for the BLAST results against the human mitochondrial genome.

The Cytoscape was used to visualize, analyze and understand the interaction between human proteins and the SARS-CoV-2 proteins.

Results and Discussion

BLAST comparison of SARS-CoV-2 isolates against the Human genome revealed multiple sequence similarities.

SARS-CoV-2 pangenome was constructed from publicly available data in NCBI, and to compare the SARS-CoV-2 genes with the Human genome for potential transgression. GFF files of 84 annotated and complete SARS-CoV-2 genomes were obtained from NCBI. Pangenomes for the 84 isolates were constructed using the Roary tool in the Galaxy web server. The gene sequence of Individual genes from the pan-genome reference file was compared with the Homo sapiens genome using BLASTn (Somewhat similar sequence) (NCBI) across Core nt, RefSeq-select RNA, RefSeq RNA, RefSeq reference genomes and nucleotide collection databases. The results for the hit are provided in the table below (Table 4). It has been observed that six non-structural protein genes of SARS-CoV-2 (Nsps) show similarity with the Human genome. Three SARS-CoV-2 genes; Nsp4, ORF 1ab, and 3c like proteinase, align with vp35 Endosomal protein sorting factor-like isoform 2 and 4 located in chromosome 16. (SARS-CoV-2 genome regions 1001-1071 and 9290-9360). ORF1ab aligns with transcript variants of bromodomain adjacent to zinc finger (BAZ) family genes. BAZ1a and BAZ2b are part of the chromatin remodeling complex. ORF1ab also aligns with Zinc finger protein 654 (ZNF654) (transcript variant), which plays a role in transcription regulation. endoRNAse and Nsp 7b genes align with dihydrolipoamide dehydrogenase transcript variant 1 and IKAROS family zinc finger 5 gene in chromosome 10 respectively.

SARS-CoV-2 Gene	*Homo sapiens* Gene transcript	Number of SARS-CoV-2 isolates	Isolate demography
Nsp 3	Prothymosin alpha(PTMA) transcript variant 1 and 2	2	AUS, CHN
Nsp 4	Chromosome 16,vp35 Endosomal protein sorting factor-like isoform 2 and 4	2	AUS, CHN
Nsp 9	Chromosome 12, 5' of plankophilin-2 and 3' of synaptogamin-10	2	AUS, CHN
Nsp 9	Janus Kinase and microtubule interacting protein 2 transcript variant 1	2	AUS, CHN
Nsp 10	Spectrin repeat containing nuclear envelope protein 2 (SYNE2) transcript variant 5	2	AUS, CHN
RNA dependent RNA polymerase	Chromosome 18, 5' of cadherin-2 isoform preprotein, 3' of desmocolin-3 isoform preprotein	3	AUS, CHN
endoRNAse	Dihydrolipoamide dehydrogenase transcript variant 1, Olfactory receptor family 2 subfamily A member 5	3	AUS, CHN
endoRNAse	Chromosome 10, inositol polyphosphate 5-phosphatase ocrl isoform c and b	3	AUS, CHN
endoRNAse	Chromosome 11, 5' of tumor protein p53-inducible protein 11 isoform x2 and 3' of pr domain containing	3	AUS, CHN

	protein 11 isoform x4		
ORF1ab	PTMA transcript variant 2, bromodomain adjacent to zinc finger domain 1A variant 1, diphosphoinositol pentakisphosphate kinase 2 variant 1, protein phosphate 4 regulatory subunit 2 variant 1, bromodomain adjacent to zinc finger domain 2B variant 1, Zinc finger protein 654 variant 1,	ALL	
ORF1ab	Chromosome 16,vp35 Endosomal protein sorting factor-like isoform 2 and 4	ALL	
3c like proteinase	Chromosome 16,vp35 Endosomal protein sorting factor-like isoform 2 and 4	3	AUS, CHN
S gene	Threonyl-tRNA synthetase 3 (TARS3)	ALL	
Nsp 7b	Chromosome 10, l-seryl-trna kinase isoforms 1 and x1, IKAROS family zinc finger 5	10	AUS, CHN, USA
Nsp 7b	Chromosome 3, 5' of metabotropic glutamate receptor 7 isoform x3 and 3' of lim and cysteine rich domains protein 1 isoform 1	10	AUS, CHN, USA
Nsp 7b	Nuclear factor of activated T cells 2, transcript variant 1	10	AUS, CHN, USA

Table 4: Represents SARS-CoV-2 Gene from respective isolates that aligns with Human Gene transcripts.

Sequence Searches of SARS-CoV-2 against the Human genome yielded distinct patterns.

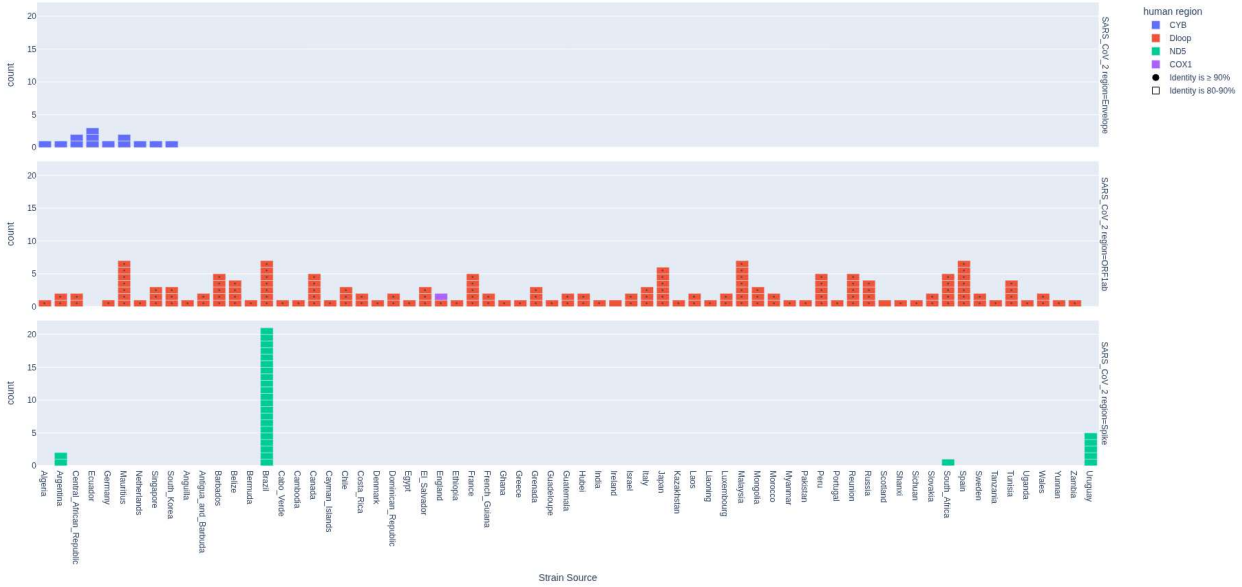
When these genomes were blasted against the human genome (Hg38, taxid:9606) using fasta36 -m8. After filtering the results using the command awk '$3>=70 && $6<=3 && $11<=0.01', we identified 459 hits. Of these, 438 hits show homology between the SARS-CoV-2 ORF1ab region and the BAZ1A gene on chromosome 14 (NC_000014.9, positions 34764711 to 34764795). Notably, sample EPI_ISL_19633144 from India deviates from this pattern. It shows homology with the ORF1ab region at the NUBPL gene (positions 31687788–31687860). Furthermore, this specific sample produced 27 unique hits across various regions of Hg38, which are detailed in the table below. Aside from the 438 hits on chromosome 14 shared by the other samples, all remaining 27 hits were unique to this Indian sample. This sample requires further investigation to determine why it exhibits these diverse hits that were not present in the other 499 samples.

Chromosome	**Hits**	**Gene/ Region**
NC_000014.9	438	BAZ1A, NUBPL
NC_000003.12	4	Unknown
NC_000004.12	3	Unknown

NC_000005.10	3	Unknown
NC_000023.11	3	Unknown
NC_000011.10	2	PAX6-AS1, DISC1FP1
NC_000012.12	2	LINC02456
NC_000006.12	2	Unknown
NC_000018.10	2	DCC
NC_000009.12	1	KDM4C
NC_000010.11	1	Unknown
NC_000013.11	1	Unknown
NC_000019.10	1	Unknown
NW_003315921.1	1	Unknown

Table 5: SARS-CoV-2 against Human genome.

Human mitochondrial genome comparison yielded typical hits in non-coding regions against SARS-CoV-2 genome.



Using the initial filter (awk '$3>=90 && $6<=1 && $11<=0.05' *.out), we identified 181 hits showing homology between the SARS-CoV-2 ORF1ab region and the mitochondrial D-loop (regions 273 to 289). As the D-loop is a non-coding region containing essential transcription factors and the origin of

replication for the leading strand, it is critical for mitochondrial genome transcription and replication. By adjusting the filter to a percentage identity of 80% while keeping other parameters constant, we identified additional homologies, with 29 hits between the SARS-CoV-2 spike region and the ND5 (NADH Dehydrogenase) gene and 13 hits between the SARS-CoV-2 membrane region and the CYTB (Cytochrome B) gene. These findings support the hypothesis that SARS-CoV-2 may hijack host mitochondria for survival, potentially leading to mitochondrial depletion, cell damage, and cell death.

SARS-CoV-2 region	Number of hits	Mitochondrial region
ORF1ab	181	Dloop
Spike	29	ND5
Membrane	13	CYTB

Table 6: SARS-CoV-2 genome against Human mitochondrial genome.

Reverse blast was performed where human mitochondrial genome was blasted against all the 584 SARS-CoV-2 samples(both part A and part B) for the search of best directional hits using the same fasta36 -m8 command which resulted into 134 hits of D-loop region of mitochondria against the ORF1ab region of SARS-CoV-2 along with that 2 hits from mitochondrially encoded tRNA glutamine (TRNQ), 1 hit from cytochrome c oxidase subunit II against the ORF1ab region of SARS-CoV-2, 19 hits from ND5 (NADH dehydrogenase subunit 5) of mitochondrial genome against the spike region of SARS-CoV-2, 13 hits from CYTB(Cytochrome B) region of mitochondrial genome against the membrane region of SARS-CoV-2.

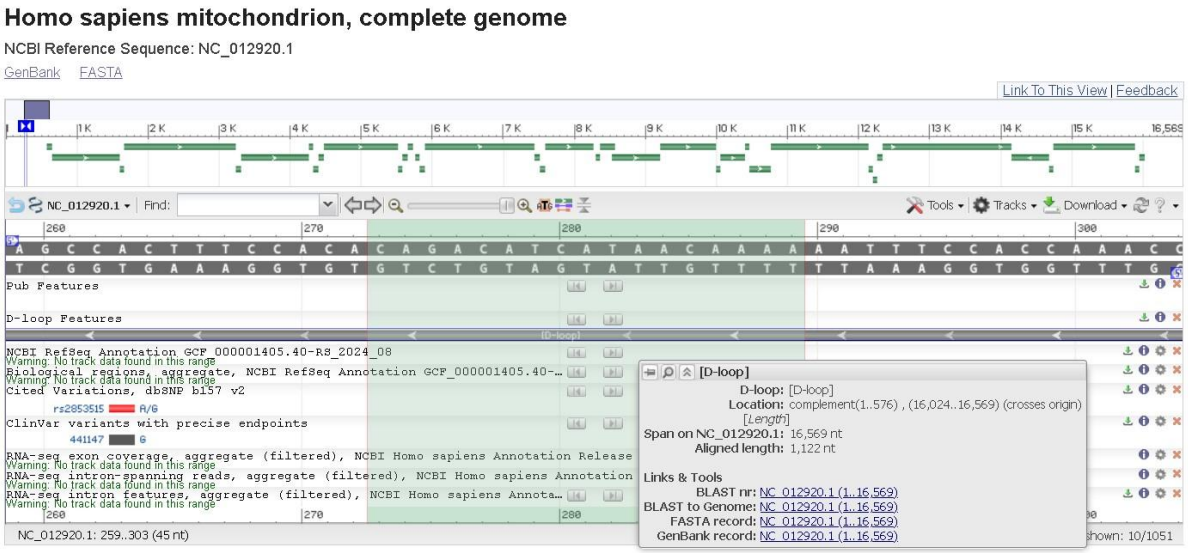


Fig. 5: Graphics overview of D-Loop region (aligned length: 1122 nt) of *Homo sapiens* (taxid:9606) mitochondrion, complete genome (NC_012920.1)

Hijacking host mitochondria: New origins or old foes?

Angiotensin Converting Enzyme 2 (ACE2) serves as the primary entry point for SARS-CoV-2 due to the strong binding affinity between the viral spike protein and the human ACE2 receptor (Haibo Zhang et al., 2020). Recent studies suggest that strong localization signals within the 5' and 3' untranslated regions (UTRs) of SARS-CoV-2 RNA may direct the viral genome and subgenomic RNAs (sgRNAs) to the mitochondrial matrix and nucleolus. This localization appears to induce mitochondrial stress and the formation of mitochondrial-derived vesicles (MDVs), facilitating communication between the mitochondria and the endoplasmic reticulum (ER). Similar to how the virus utilizes ER-derived double-membrane vesicles for replication and immune evasion, SARS-CoV-2 may also direct the production of MDVs to support its lifecycle (Singh et al., 2020). The mtDNA acts as danger-associated molecular patterns (DAMPs) and so we sought to ask whether in SARS-CoV-1, the ORF9b protein targets the mitochondrial antiviral signaling (MAVS) protein complex, which is essential for initiating antiviral responses. By hijacking the host E3 ubiquitin ligase, the virus triggers the degradation of MAVS, as well as TRAF3 and TRAF6 (Shi et al., 2014). This degradation constrains the host cell's interferon (IFN) responses, weakening the innate immune defense and leading to mitochondrial dysfunction, oxidative stress, and eventual cell death. Furthermore, these viral proteins interact with mitochondrial permeability transition pore (mPTP) complexes. This interaction alters mitochondrial morphology, disrupts the Ca2+ cycle, and inhibits voltage-gated calcium channel (VDCC) activity, resulting in ion homeostasis disorders and compromised cell viability (Chen et al., 2023).

Transgress or Trespass?

Upon viral infection, expressed lncRNAs may contribute to the host's antiviral immune response. Because viral replication requires significant energy, viruses have developed mechanisms to hijack host metabolism, a process in which these differentially expressed lncRNAs may play a role (Shukla et al., 2021). It has been hypothesized that the viral genome preferentially copies the host genome, specifically non-coding RNA, for survival and replication. Supporting the concept of transgression, similarity searches between SARS-CoV-2 and the human genome show similarities to annotated lncRNAs (such as NONHSAT239888, highly expressed in the prostate) and certain non-annotated genomic regions. Furthermore, a comparison of the SARS-CoV and SARS-CoV-2 genomes reveals that ORF3a, which codes for the APA3 viroporin (a pro-apoptosis inducing protein), is missing in SARS-CoV-2.

Cytoscape visualization provided hubs for prioritizing the genes for ranking

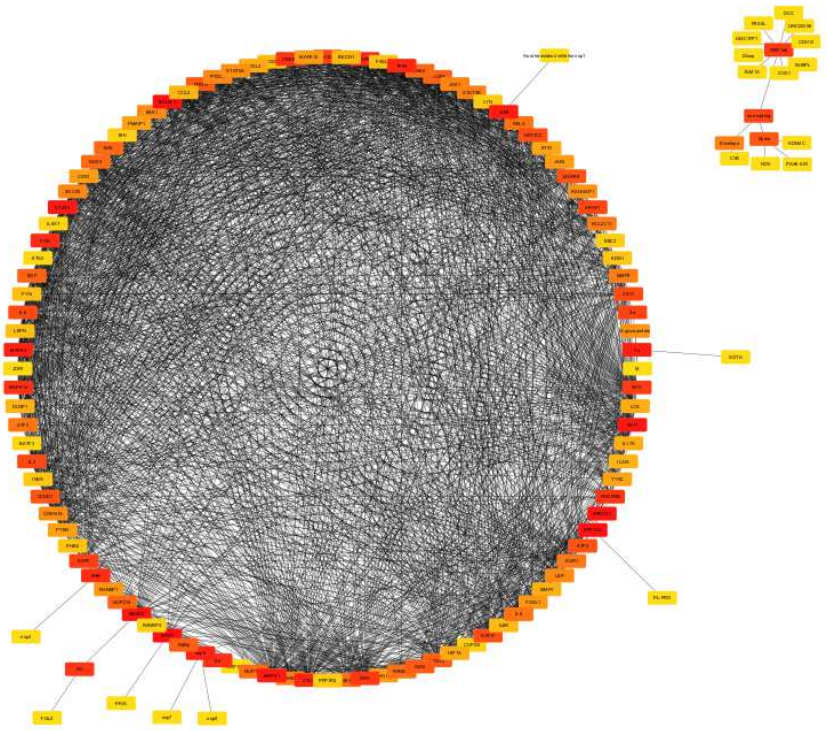


Fig. 6: The interaction between SARS_CoV_2 and human genes from string Database.

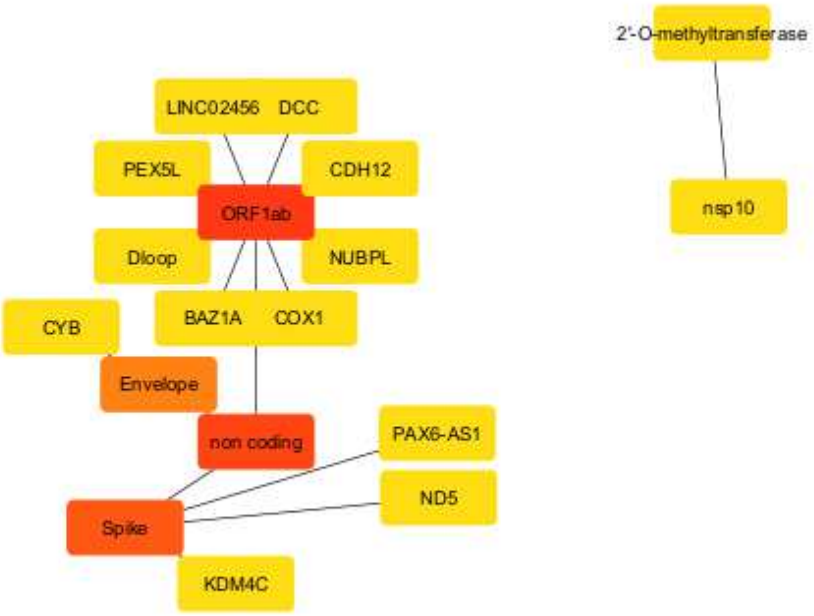


Fig. 7: The interaction between SARS-CoV-2 and human genes from STRING database.

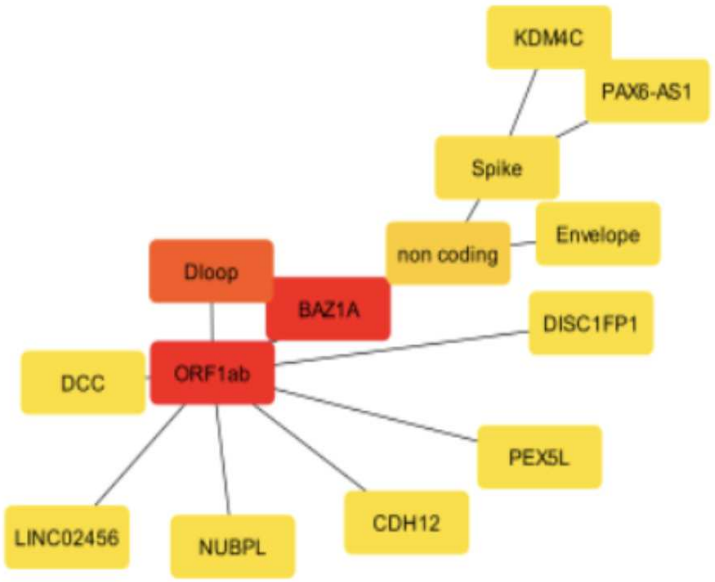


Fig. 8: The interaction between D-Loop and ORF1ab from the STRING database.

We identified interactions between the SARS-CoV-2 genome and various human genomic regions, including coding, non-coding, and long non-coding areas. Our analysis, where node color intensity reflects betweenness centrality, shows that ORF1ab interacts with several regions, including the mitochondrial genome. Additionally, the spike and envelope regions of SARS-CoV-2 demonstrate significant interactions. These results suggest that SARS-CoV-2 manipulates the host system at both the nuclear and mitochondrial levels through a process called transgression, *viz.* similarity between ORF1ab and BAZ1A suggests the virus may target transcriptional regulation. On the other hand, interactions between the mitochondrial D-loop and SARS-CoV-2 may disrupt mitochondrial replication and immune signaling wherein homology with mitochondrial regions such as D-loop, ND5, and CYTB indicates that the virus interacts with these elements to disrupt cellular energy and oxidative phosphorylation. While our preliminary analysis of global data does not show strong evidence for variant-specific transgression, a unique sample from India (EPI_ISL_19633144) shows multiple hits beyond BAZ1A. This suggests the possibility of variant-specific or population-specific transgression patterns that require further exploration. Our findings indicate that non-coding regions, including regulatory RNAs and the mitochondrial D-loop, serve as strategic targets for viral survival by influencing transcription and immune signaling. We are now investigating whether these transgression patterns differ across various SARS-CoV-2 variants and whether sequence changes influence genomic homology with human elements.

Taken together the results suggested that the transgression is not only restricted to a single type of genomic element, but rather spreads in parts of coding, noncoding, nuclear and mitochondrial sequences. Its biological consequences likely depend on the viral variant, genomic context. Future studies integrating variant stratification, splicing analysis and mitochondrial functional assay will be crucial to fully map the role of transgression in SARS-CoV-2 biology. One of the impending questions we asked was about taking this in the wet-lab by taking up the cellline-mediated assays for transgressed regions and checking their expression in SARS-CoV-2, *viz. Calu-3 (Pulmonary), Caco-2 (Intestinal), Huh-7 (Hepatic) and HEK 293T (Renal)*

Conclusions

We investigated whether the SARS-CoV-2 genome exhibits sequence similarity or interaction with host nuclear and mitochondrial genomes. Our findings reveal significant homology between the SARS-CoV-2 ORF1ab region and segments of the human BAZ1A gene, as well as the D-loop region of the human mitochondrial genome. This suggests a potential viral strategy to manipulate host regulatory mechanisms. Additionally, hits with the ND5 and CYTB genes indicate that SARS-CoV-2 may interfere with mitochondrial function, potentially leading to energy dysregulation and immune modification. Cytoscape network analysis further highlights that viral proteins—specifically ORF1ab, spike, and membrane proteins—serve as central hubs with high betweenness centrality. These nodes connect viral factors to multiple human genes and regulatory regions. The high betweenness of these viral nodes provides key insights into SARS-CoV-2 survival strategies and its manipulation of the host environment.

Competing interests: None
Funding: None
Authors' contributions: VM, PE, CP, AJ contributed equally to the work, RSP, and SKP supervised the work with GC and PS. PS proofread the manuscript before all authors agreeing to it
Ethics clearance: Not Applicable
Data availability: All data is available as files. Scripts and extraneous data may be made available upon reasonable request

References

- Chen TH, Chang CJ, Hung PH. Possible Pathogenesis and Prevention of Long COVID: SARS-CoV-2-Induced Mitochondrial Disorder. Int J Mol Sci. 2023 Apr 28;24(9):8034. doi: 10.3390/ijms24098034. PMID: 37175745; PMCID: PMC10179190.
- Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol. 2019 Mar;17(3):181-192. doi: 10.1038/s41579-018-0118-9. PMID: 30531947; PMCID: PMC7097006.
- Harrison AG, Lin T, Wang P. Mechanisms of SARS-CoV-2 Transmission and Pathogenesis. Trends Immunol. 2020 Dec;41(12):1100-1115. doi: 10.1016/j.it.2020.10.004. Epub 2020 Oct 14. PMID: 33132005; PMCID: PMC7556779.
- GISAID: https://en.wikipedia.org/wiki/GISAID
- Jackson CB, Farzan M, Chen B, Choe H. Mechanisms of SARS-CoV-2 entry into cells. Nat Rev Mol Cell Biol. 2022 Jan;23(1):3-20. doi: 10.1038/s41580-021-00418-x. Epub 2021 Oct 5. PMID: 34611326; PMCID: PMC8491763.
- Kelly JA, Olson AN, Neupane K, Munshi S, San Emeterio J, Pollack L, Woodside MT, Dinman JD. Structural and functional conservation of the programmed -1 ribosomal frameshift signal of SARS coronavirus 2 (SARS-CoV-2). J Biol Chem. 2020 Jul 31;295(31):10741-10748. doi: 10.1074/jbc.AC120.013449. Epub 2020 Jun 22. PMID: 32571880; PMCID: PMC7397099.
- Khailany RA, Safdar M, Ozaslan M. Genomic characterization of a novel SARS-CoV-2. Gene Rep. 2020 Jun;19:100682. Doi: 10.1016/j.genrep.2020.100682. Epub 2020 Apr 16. PMID: 32300673; PMCID: PMC7161481.
- Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science. 2005 Sep 16;309(5742):1864-8. doi: 10.1126/science.1116480. PMID: 16166518.
- Paul O, Akolia IK, Qin Tao J, Jain N, Louneva N, Montone KT, Fisher AB, Rajapakse CS, Bermudez C, Chatterjee S. Reactive oxygen species in endothelial signaling in COVID-19: Protective role of the novel peptide PIP-2. PLoS One. 2024 May 21;19(5):e0289854. doi: 10.1371/journal.pone.0289854. PMID: 38771750; PMCID: PMC11108150.

- Pearson WR. Finding Protein and Nucleotide Similarities with FASTA. Curr Protoc Bioinformatics. 2016 Mar 24;53:3.9.1-3.9.25. doi: 10.1002/0471250953.bi0309s53. PMID: 27010337; PMCID: PMC5072362.
- Ravi V, Saxena S, Panda PS. Basic virology of SARS-CoV 2. Indian J Med Microbiol. 2022 Apr-Jun;40(2):182-186. doi: 10.1016/j.ijmmb.2022.02.005. Epub 2022 Mar 14. PMID: 35300895; PMCID: PMC8919811.
- Ren B, Guan MX, Zhou T, Cai X, Shan G. Emerging functions of mitochondria-encoded noncoding RNAs. Trends Genet. 2023 Feb;39(2):125-139. doi: 10.1016/j.tig.2022.08.004. Epub 2022 Sep 19. PMID: 36137834.
- Santos-López G, Cortés-Hernández P, Vallejo-Ruiz V, Reyes-Leyva J. SARS-CoV-2: basic concepts, origin and treatment advances. Gac Med Mex. 2021;157(1):84-89. English. doi: 10.24875/GMM.M21000524. PMID: 34125824.
- Sharma HN, Latimore COD, Matthews QL. Biology and Pathogenesis of SARS-CoV-2: Understandings for Therapeutic Developments against COVID-19. Pathogens. 2021 Sep 19;10(9):1218. doi: 10.3390/pathogens10091218. PMID: 34578250; PMCID: PMC8470303.
- Shi CS, Qi HY, Boularan C, Huang NN, Abu-Asab M, Shelhamer JH, Kehrl JH. SARS-coronavirus open reading frame-9b suppresses innate immunity by targeting mitochondria and the MAVS/TRAF3/TRAF6 signalosome. J Immunol. 2014 Sep 15;193(6):3080-9. doi: 10.4049/jimmunol.1303196. Epub 2014 Aug 18. PMID: 25135833; PMCID: PMC4179872.
- Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 2017 Mar 30;22(13):30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494. PMID: 28382917; PMCID: PMC5388101.
- Shukla N, Prasad A, Kanga U, Suravajhala R, Nigam VK, Kishor PBK, Polavarapu R, Chaubey G, Singh KK, Suravajhala P. SARS-CoV-2 transgressing LncRNAs uncovers the known unknowns. Physiol Genomics. 2021 Oct 1;53(10):433-440. doi: 10.1152/physiolgenomics.00075.2021. Epub 2021 Sep 7. PMID: 34492207; PMCID: PMC8562947.
- Singh KK, Chaubey G, Chen JY, Suravajhala P. Decoding SARS-CoV-2 hijacking of host mitochondria in COVID-19 pathogenesis. Am J Physiol Cell Physiol. 2020 Aug 1;319(2):C258-C267. doi: 10.1152/ajpcell.00224.2020. Epub 2020 Jun 8. PMID: 32510973; PMCID: PMC7381712.
- Tamayo-Ordóñez MC, Rosas-García NM, Ayil-Gutiérrez BA, Bello-López JM, Tamayo-Ordóñez FA, Anguebes-Franseschi F, Damas-Damas S, Tamayo-Ordóñez YJ. Non-Structural Proteins (Nsp): A Marker for Detection of Human Coronavirus Families. Pathogens. 2023 Sep 21;12(9):1185. doi: 10.3390/pathogens12091185. PMID: 37764993; PMCID: PMC10537875.
- van de Leemput J, Han Z. Understanding Individual SARS-CoV-2 Proteins for Targeted Drug Development against COVID-19. Mol Cell Biol. 2021 Aug 24;41(9):e0018521. doi: 10.1128/MCB.00185-21. Epub 2021 Aug 24. PMID: 34124934; PMCID: PMC8384068.
- West AP, Shadel GS. Mitochondrial DNA in innate immune responses and inflammatory pathology. Nat Rev Immunol. 2017 Jun;17(6):363-375. doi: 10.1038/nri.2017.21. Epub 2017 Apr 10. PMID: 28393922; PMCID: PMC7289178.
- Yadav R, Chaudhary JK, Jain N, Chaudhary PK, Khanra S, Dhamija P, Sharma A, Kumar A, Handu S. Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19. Cells. 2021 Apr 6;10(4):821. doi: 10.3390/cells10040821. PMID: 33917481; PMCID: PMC8067447.
- Yan C, Duanmu X, Zeng L, Liu B, Song Z. Mitochondrial DNA: Distribution, Mutations, and Elimination. Cells. 2019 Apr 25;8(4):379. doi: 10.3390/cells8040379. PMID: 31027297; PMCID: PMC6523345.
- Zhang H, Penninger JM, Li Y, Zhong N, Slutsky AS. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. Intensive

Care Med. 2020 Apr;46(4):586-590. doi: 10.1007/s00134-020-05985-9. Epub 2020 Mar 3. PMID: 32125455; PMCID: PMC7079879.

- Zhang J, Shang J, Wang F, Huo X, Sun R, Ren Z, Wang W, Yang M, Li G, Gao D, Liu R, Bai P, Wang S, Wang Y, Yan X. Decreased mitochondrial D-loop region methylation mediates an increase in mitochondrial DNA copy number in CADASIL. Clin Epigenetics. 2022 Jan 4;14(1):2. doi: 10.1186/s13148-021-01225-z. PMID: 34983647; PMCID: PMC8725280.