Indian J Med Res 152, July & August 2020, pp 70-76
DOI: 10.4103/ijmr.IJMR_2257_20

Quick Response Code:

# Transcriptome & viral growth analysis of SARS-CoV-2-infected Vero CCL-81 cells

Dimpal A. Nyayanit[1,#], Prasad Sarkale[1,#], Shreekant Baradkar[1], Savita Patil[1], Pragya D. Yadav[1], Anita Shete-Aich[1], Kaumudi Kalele[1], Pranita Gawande[1], Triparna Majumdar[1], Rajlaxmi Jain[1] & Gajanan Sapkal[2]

*[1]Maximum Containment Laboratory & [2]Diagnostic Virology Group, ICMR-National Institute of Virology, Pune, Maharashtra, India*

Received May 29, 2020

***Background & objectives*: The genome of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), belonging to the family *Coronaviridae,* encodes for structural, non-structural, and accessory proteins, which are required for replication of the virus. These proteins are encoded by different genes present on the SARS-CoV-2 genome. The expression pattern of these genes in the host cells needs to be assessed. This study was undertaken to understand the transcription pattern of the SARS-CoV-2 genes in the Vero CCL-81 cells during the course of infection.**

***Methods*: Vero CCL-81 cells were infected with the SARS-CoV-2 virus inoculum having a 0.1 multiplicity of infection. The supernatants and cell pellets were harvested after centrifugation at different time points, post-infection. The 50% tissue culture infective dose ($TCID_{50}$) and cycle threshold ($C_t$) values of the *E* and the *RdRp-2* genes were calculated. Next-generation sequencing of the harvested sample was carried out to observe the expression pattern of the virus by mapping to the SARS-CoV-2 Wuhan HU-1 reference sequence. The expressions were in terms of the reads per kilobase million (RPKM) values.**

***Results*: In the inital six hours post-infection, the copy numbers of *E* and *RdRp-2* genes were approximately constant, which raised 10 log-fold and continued to increase till the 12 h post-infection (hpi). The $TCID_{50}$ was observed in the supernatant after 7 hpi, indicating the release of the viral progeny. *ORF8* and *ORF7a*, along with the nucleocapsid transcript, were found to express at higher levels.**

***Interpretation & conclusions*: This study was a step towards understanding the growth kinetics of the SARS-CoV-2 replication cycle. The findings indicated that *ORF8* and *ORF7b* gene transcripts were expressed in higher amounts indicating their essential role in viral replication. Future studies need to be conducted to explore their role in the SARS-CoV-2 replication.**

**Key words** India - next-generation sequencing - RT-PCR - replication cycle - SARS-CoV-2 - transcriptome

[#]Equal contribution

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a positive-sense, single-stranded RNA virus, which has a genome of about 29.9 kb in length[1,2]. It belongs to the genus *Betacoronavirus* (subgenus: *Sarbecovirus)* of the family *Coronaviridae*[3]. The case-fatality ratio of SARS-CoV-2 is lower than that of SARS-CoV-1 and Middle East respiratory syndrome coronavirus (MERS-CoV)[4]. However, ever since its first report from China[5], the SARS-CoV-2 has claimed far more deaths than both SARS-CoV-1 and MERS-CoV combined[6]. Full genome sequences of the first two SARS-CoV-2 viruses were reported from India in February 2020[7], and as on July 20, 2020, there were reports of 1,118,043 confirmed cases with 27,497 deaths from India[8].

The SARS-CoV-2 genome encodes for non-structural proteins [open reading frames (ORF1a, and ORF1b)] and four structural [envelope (E), nucleocapsid (N), membrane (M) and spike (S)] proteins along with six accessory proteins from ORFs 3a, 6, 7a, 7b, 8 and 10[1,2,9], the function of which is yet to be identified. The SARS-CoV employs viral RNA as the template for the replication and transcription of the virus. The expression analysis of SARS-CoV involves the synthesis of sub-genomic RNA by discontinuous transcription along with the genomic positive-sense RNA[9]. The use of next-generation sequencing (NGS) approaches has facilitated the analysis of transcription, as demonstrated in a study that revealed differential filovirus transcription patterns in different cell lines[10].

A recent study by Kim *et al*[11] looked upon the transcriptome and the epi-transcriptome of the SARS-CoV-2 through the sub-genomic RNAs, and the modification on the viral transcripts in SARS-CoV-2-infected Vero CCL-81 cells at 12 h post-infection (hpi). The purpose of the present study was to understand the growth kinetics of the SARS-CoV-2 and the variation in the transcription pattern of SARS-CoV-2 in a Vero CCL-81 cell line at different time points. This study also compared the quasispecies present in the clinical sample and the fourth passage (P) of the virus isolate.

## Material & Methods

The study was conducted during April to May 2020 in the Maximum Containment Complex of the ICMR-National Institute of Virology (ICMR-NIV), Pune, India, with prior approval of the protocol.

*Cell culture*: SARS-CoV-2 used was isolated from the throat/nasal swab samples of Italian tourists (GISAID number: EPI_ISL_420545) in India during March 2020[12]. The monolayers of Vero CCL-81 cells grown in 24-well tissue culture plates were infected with SARS-CoV-2 with a 50% tissue culture infective dose ($TCID_{50}$) of $10^{6.5}$/ml, Vero CCL-81 P-3.

The study was performed with a multiplicity of infection (MOI) of 0.1 to infect the Vero CCL-81 cells in triplicate sets. After centrifugation, the supernatant and cell pellet were harvested at indicated time points (15, 30 and 45 min and 1, 2, 3, 4, 5, 6, 7, 8, 10 and 12 hpi) and stored at −80°C till further use. The supernatants treated with RNase A (10 µg/ml) for one hour at 37°C were used as control. Studies using infectious viruses were performed in the biosafety level 4 (BSL-4) facility of the ICMR-NIV. SARS-CoV-2 was inactivated using Trizol reagent (Thermo Fisher Scientific, USA) before removal to lower containment laboratories for further analysis. The $TCID_{50}$ was calculated for both the supernatant and the cell pellet. The cycle threshold ($C_t$) values for each sample were determined by *E* and *RdRp-2* genes-specific SARS-CoV-2 real-time reverse transcription-polymerase chain reaction (RT-PCR) using the protocol described elsewhere[7,13].

*Next-generation sequencing and data analysis*: Total RNA was extracted from 0.5 ml of the infected tissue culture lysate, containing $10^5$ cells, and 1 ml of the cell supernatant using the TriPure reagent (Sigma-Aldrich, USA) followed by column purification with a Qiagen viral RNA extraction kit (Qiagen, Germany) as per manufacturer's instructions. Poly-A-containing mRNAs were purified from total RNA using oligo-dT beads provided in the Illumina TruSeq Stranded mRNA LT Sample Preparation Kit (Illumina, USA). The cDNA libraries were prepared from the purified mRNAs using the same library preparation kit. The libraries were sequenced using the Illumina MiniSeq platform (Illumina, USA) using the mid-output cartridge (300 cycles, 150×2 paired-end sequencing). The uninfected Vero CCL-81 cells (used as control) and the RNase A-treated control supernatants were processed using the above-described steps. RNase A-treated control supernatants were used after depleting ribosomal RNA (rRNA) from the sample with the NEBNext® rRNA Depletion Kit (New England Biolabs, USA).

Complementary DNA (cDNA) libraries were made from extracted RNAs using the same protocol as above. Total reads generated from the libraries made were mapped with the reference SARS-CoV-2 isolate Wuhan HU-1 genome (accession number: NC_045512.1).

Transcription of each gene was quantified using the reads per kilobase million (RPKM) using the reads generated from each libraries. The RPKM values normalize the total reads for sequencing depth as well as gene length, which, in turn, gives the number of reads mapped to each gene. The read coverage and RPKM values were determined using the QIAGEN CLC Genomics Workbench 20.0[14] (QIAGEN, Aarhus, Denmark). RPKM values were calculated for each time point at which the infected cells were harvested to determine gene expression over time. Insertion and deletion variation analysis was performed with a maximum cut-off set to one per cent to identify variants present in the genes using the QIAGEN CLC Genomics Workbench 20.0. The methods used to determine the read coverage of the genome, RPKM values of the expressed genes, and the variants detected are limited by the algorithms implemented in the software used for the analysis.

## Results

*Copy number and growth kinetics*: The viral RNA copy number in the cells and the supernatants harvested at defined intervals post-infection were measured using *E* and *RdRp-2* genes of SARS-CoV-2-specific real-time RT-PCR. The average of the three replicates performed to determine the *E* and *RdRp-2* genes copy numbers in the cell pellet, and the supernatant is depicted in Figure 1A and B. It was observed that the viral load was approximately constant for the first six hours in the cell pellet. At seven hours post-infection, *E* and *RdRp-2* copy numbers increased by 10 log-fold further increasing exponentially till the 12 hpi. The copy numbers for *E* and *RdRp-2* genes at the 12 hpi were found to be $4.1\times10^{9}$ and $1.8\times10^{8}$ RNA copies per $10^{5}$ cells, respectively. The *RdRp-2* copies started appearing within the cell supernatants by 7 hpi, which increased 10-fold by the 8 hpi. Towards the 12 hpi, the viral RNA load further increased, and the final *E* and *RdRp-2* copy numbers were $1.9\times10^{7}$ and $6.6\times10^{6}$ per 100 µl, respectively.

The number of infectious viral particles in the cell supernatant was measured by calculating the $TCID_{50}$ at different times post-infection, depicted in Figure 2. It was observed that the supernatant did not have any detectable level of virus particles until 6 hpi. Infectious viral progeny was observed in the supernatant with a titre of $1\times10^{1.5}$ $TCID_{50}$/100 µl starting from 7 hpi. An increased titre of approximately 1.5-fold ($1\times10^{3.5}$ $TCID_{50}$/100 µl) was observed after 8 hpi, which increased until the 12 hpi ($1\times10^{5.5}$ $TCID_{50}$/100 µl). NGS analysis of the cell supernatant showed a parallel increase of viral reads through the course of infection (data not shown).

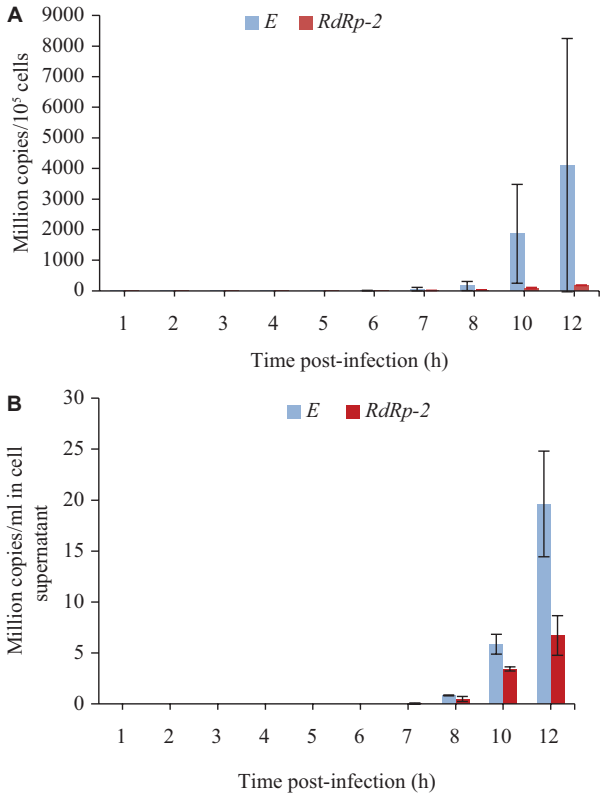


**Fig. 1.** Copy number of *E* and *RdRp-2* genes in the (**A**) cell pellet and (**B**) supernatant of SARS-CoV-2 infected Vero CCL-81 cells: Cell pellet and supernatant of the SARS-CoV-2-infected Vero CCL-81 cells were collected at defined time points, and the viral load was determined using real-time RT-PCR. The average copy number±SD of three replicates is plotted.

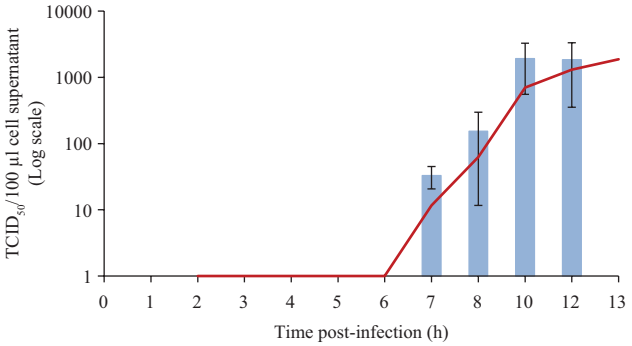


**Fig. 2.** 50% tissue culture infective dose ($TCID_{50}$) of SARS-CoV-2-infected Vero CCL-81 cell line: Vero CCL-81 cells in the 24-well culture plates were infected with the SARS-CoV-2 strain (GISAID accession number EPI_ISL_420545). The supernatant was collected at defined times. The average of $TCID_{50}$/100 µl is obtained from the three replicates along with standard deviation as the error bars. The maroon line depicts the moving average of three periods.

*Next-generation sequencing and data analysis*: The transcriptional data obtained for the different intervals were mapped against the SARS-CoV-2 isolate Wuhan HU-1. During the initial time points, only a few reads mapped to the *N* and *ORF10* genes near the 3' end of the genome. The mapped reads corresponding to the infected cells gradually increased towards the 5' end. The expression of all viral genes was detected at 6 hpi in the infected cells. The supernatant of infected cells, which was used as a control in this experiment, did not show differential gene expression as that observed in the cell samples (Fig. 3).

Figure 4 depicts the RPKM values for structural, non-structural, and accessory genes at different times post-infection in the cell pellets. As expected, the cell pellets expressed a higher amount of *N* gene compared to any other genes. The presence of *N* gene transcripts at higher levels is suggestive of its importance in the viral replication and expression. The expressions of other genes are normalized with respect to the *N* gene so that their values lie between 0-1. The expression of *N* transcript was followed by the other two accessory proteins, *ORF8* and *ORF7a* (Fig. 4B). The *ORF1ab* was the least expressed gene at all the periods post-infection (Fig. 4A). *ORF7b* transcripts were observed to be expressed towards the end of the replication cycle.

*Variant and quasispecies*: The consensus nucleotide sequences from the clinical sample and isolate (P-0 and P-4) were compared, and a few common mutations were observed between them at 5' UTR [genomic position (GP): 241]: *ORF1ab* (GP: 3037, 4809 and 14408) and *S* gene (GP: 23403). Nucleotide substitutions at GPs: 4809, 14408 and 23403 are non-synonymous changes causing mutation in the amino acid S→F, P→L, and D→G, respectively. The consensus nucleotide of the P-4 isolate had an additional mutation at the *S* gene

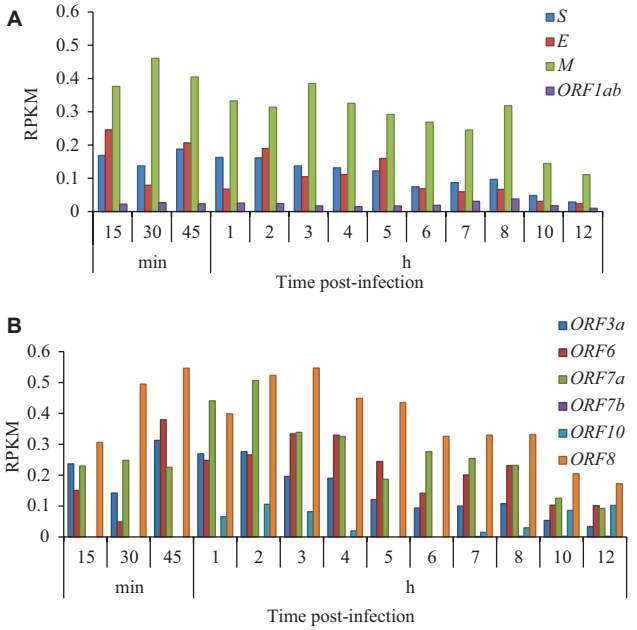


**Fig. 4.** Reads per kilobase million (RPKM) values of cell pellet of (**A**) structural, (**B**) non-structural, and accessory transcripts of SARS-CoV-2 gene expressed in Vero CCL-81, collected at a defined time point. This value is normalized against the nucleocapsid (*N*) RPKM values at every point. *S*, Spike; *E*, envelope; *M*, membrane.

(GP: 23607) compared to the P-1 and clinical sequence (P-0).

Quasispecies were observed in the sequencing data of the P-4 when compared with P-1 (GISAID number: EPI_ISL_420546) data. These nucleotide variations may have aroused among the virus population due to adaptation to host cells. Further, it was observed that the nucleotide reads at GP position 28254 (*ORF8*) of the P-4 isolate had a single-nucleotide deletion, leading to a frameshift (Fig. 5). The frameshift led to an increase in the length of the protein encoded by the *ORF8* gene by four amino acids. Three of these

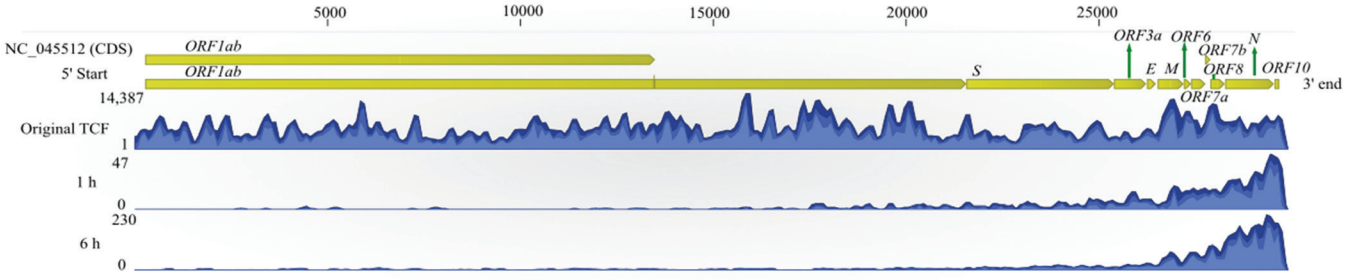


**Fig. 3.** Expression of SARS-CoV-2 virus genes in Vero CCL-81 cell line at a defined time point with respect to control. The CDS encoding different genes is depicted in the upper panel. The second panel determines the viral RNA present in the original supernatant used to infect the Vero CCL-81 cell line. The bottom two panels are the 1st and the 6th time point Vero CCL-81 expression data.

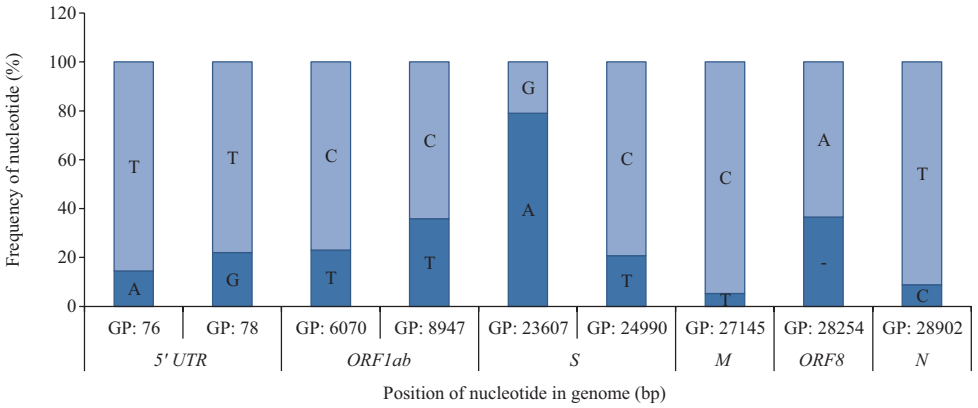


**Fig. 5.** Per cent nucleotide variations observed in the SARS-CoV-2-infected Vero CCL-81 cell pellet retrieved after a 12 h time point of P-4. The variant was obtained using the basic variant tool present in the QIAGEN CLC Genomics Workbench 20.0. Dark blue colour shows the per cent of alleles present in the isolate sequences, whereas the light blue colour indicates the nucleotide observed in the reference sequence. The per cent is expressed as the variation in the length of the different nucleotides observed.

quasispecies changes at GPs 8947, 27145 and 28902 are non-synonymous changes.

## Discussion

This study was aimed to analyze the growth kinetics and the transcription pattern of SARS-CoV-2 in the Vero CCL-81 cells. The determination of viral titres suggested that one complete virus replication took 7 hpi, while the expression of the complete gene set was observed at 6 hpi. The final concentration of the viral particle in the cell and supernatant was approximately two-fold more than the first release at 7 hpi. The $TCID_{50}$ value for 1 MOI demonstrated a similar trend (data not shown).

The ratio of gene expression was determined at 96 hpi in the infected cells with an MOI of 1 (data not shown) in the Vero CCL-81 cells. It was observed that *M:E:S* was expressed in the ratio of 0.22:0.09:0.08 with respect to *N* gene, whereas non-structural genes *ORF8:ORF7a:ORF6:ORF3a* had a ratio of 0.35:0.22:0.17:0.12 at the 7 hpi in the supernatant. The data revealed comparatively higher expression of *ORF8, ORF7a,* and *M* gene transcripts. *ORF8* gene of SARS-CoV-2 is reported to diverge from the *ORF8b* gene of SARS-CoV, which is involved in the activation of innate immune response by inducing the NLRP3 (NOD-, LRR- and pyrin domain-containing protein 3) inflammasome[15]. The presence of a nucleotide deletion in the *ORF8* transcript of the P-4 isolate reads was intriguing. The deletion in the *ORF8* needs further investigation on its role played in eliciting host cell innate response and pathogenicity of the virus to the host cells. The role of *ORF7a* of SARS-CoV-2, which is 17.9 per cent diverged from the *ORF7a* gene of SARS-CoV, also needs further investigation. *ORF7a* of the SARS-CoV is reported to be a potent inhibitor of the bone marrow stromal antigen 2, which is involved in restricting the virus release from the cell membrane[16].

Transcription of coronavirus occurs in both continuous and discontinuous methods[17]. Transcription regulatory sequences (TRSs) present at the 5' end of each gene play an important role in the discontinuous method of transcription of sRNAs (sg mRNAs). The TRSs act to slow down the replication complex by controlling the transcription of sg mRNAs[18]. Alignment of the SARS-CoV-1 (accession number: NC_004718), SARS-CoV-2 (accession number: NC_045512), and MERS-CoV (accession number: NC_019843) reference sequences revealed the conservation of TRS leader and TRS sequences between different genes of SARS-CoV-1 and SARS-CoV-2. This possibly indicates a comparable level of transcriptional control for both SARS-CoV-1 and SARS-CoV-2. The higher expression of N transcript can be related to the different functions it plays in SARS-CoV-1 as RNA chaperone[19], cell cycle regulation, pathogenesis, signal transduction, assembly, and stress[20]. The higher expression of an accessory protein, such as *ORF8* and *ORF7a*, next to *N*, needs to be evaluated. Su *et al*[21] demonstrated reduced replication fitness of the SARS-CoV-2 lined to the 382-nt deletion in the region spanning the *ORF8* and its TRS.

The emergence of virulent or non-virulent subpopulation strains from the existing strains occurs in a process where a virus is under the influence of selection to adapt to a particular host[22]. The current

study indicated the need for analyzing the different quasispecies present or developed during the passages of the virus in the cell line along with the changes in the clinical samples of a different host. The identification of quasispecies is important in a step towards designing drugs and vaccines for restricting the virus[23] and understanding the wider range of variants present that need to be checked[24].

The time period of growth kinetics needs to be enlarged to understand the kinetics of the virus which is one of the study limitations. The time point expression analysis of the subgenomic RNA needs to be looked upon to identify early expressing RNAs.

In conclusion, our findings for the growth kinetics indicate that SARS-CoV-2 completes its replication cycle in seven hours. Considering this observation, it could be hypothesized that the replication cycle of the SARS-CoV-2 would approximately be the same in humans. The quasispecies present in the clinical samples needs to be assessed to understand the changes in the virus population that occur due to host adaptation.

***Acknowledgment*:** Authors thank Dr Priya Abraham, Director, ICMR-National Institute of Virology, Pune, and acknowledge the encouragement and support extended by Prof. (Dr) Balram Bhargava, Secretary to the Government of India, Department of Health Research, Ministry of Health and Family Welfare, and Director-General, Indian Council of Medical Research (ICMR), and Drs Raman Gangakhedkar and Nivedita Gupta, ICMR, New Delhi.

***Financial support & sponsorship*:** This work was supported by intramural funding of the ICMR-National Institute of Virology, Pune.

***Conflicts of Interest*:** None.

## References

1. Chan JF, Kok KH, Zhu Z, Chu H, To KK, Yuan S, *et al*. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect* 2020; *9* : 221-36.
2. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, *et al*. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe* 2020; *27* : 325-8.
3. Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, *et al*. The species severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 2020; *5* : 536-44.
4. Petrosillo N, Viceconte G, Ergonul O, Ippolito G, Petersen E. COVID-19, SARS and MERS: Are they closely related? *Clin Microbiol Infect* 2020; *26* : 729-34.
5. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, *et al*. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020; *382* : 727-33.
6. Mahase E. Coronavirus: COVID-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. *BMJ* 2020; *368* : m641.
7. Yadav PD, Potdar VA, Choudhary ML, Nyayanit DA, Agrawal M, Jadhav SM, *et al*. Full-genome sequences of the first two SARS-CoV-2 viruses from India. *Indian J Med Res* 2020; *151* : 200-9.
8. World Health Organization. *Coronavirus disease (COVID-19) Situation Report - 182*. Geneva: WHO; 2020.
9. Hussain S, Pan J, Chen Y, Yang Y, Xu J, Peng Y, *et al*. Identification of novel subgenomic RNAs and noncanonical transcription initiation signals of severe acute respiratory syndrome coronavirus. *J Virol* 2005; *79* : 5288-95.
10. Albariño CG, Wiggleton Guerrero L, Chakrabarti AK, Nichol ST. Transcriptional analysis of viral mRNAs reveals common transcription patterns in cells infected by five different filoviruses. *PLoS One* 2018; *13* : e0201827.
11. Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. The architecture of SARS-CoV-2 transcriptome. *Cell* 2020; *181* : 914-21.
12. Sarkale P, Patil S, Cherian SS, Sapkal G, Baradkar S, Lakra R, *et al*. First isolation of SARS-CoV-2 from clinical samples in India. *Indian J Med Res* 2020; *151* : 244-50.
13. World Health Organization. *Laboratory Guidance; 2020*. Available from: *https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/laboratory-guidance*, accessed on March 3, 2020.
14. CLC Genomics Workbench 20.0 (QIAGEN). Available from: *http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Introduction_CLC_Genomics_Workbench.html*, accessed on April 7, 2020.
15. Yuen KS, Ye ZW, Fung SY, Chan CP, Jin DY. SARS-CoV-2 and COVID-19: The most important research questions. *Cell Biosci* 2020; *10* : 40.
16. Taylor JK, Coleman CM, Postel S, Sisk JM, Bernbaum JG, Venkataraman T, *et al*. Severe acute respiratory syndrome coronavirus ORF7a inhibits bone marrow stromal antigen 2 virion tethering through a novel mechanism of glycosylation interference. *J Virol* 2015; *89* : 11820-33.
17. Sola I, Almazán F, Zúñiga S, Enjuanes L. Continuous and discontinuous RNA synthesis in coronaviruses. *Annu Rev Virol* 2015; *2* : 265-88.
18. Alonso S, Izeta A, Sola I, Enjuanes L. Transcription regulatory sequences and mRNA expression levels in the coronavirus transmissible gastroenteritis virus. *J Virol* 2002; *76* : 1293-308.

19. Zúñiga S, Sola I, Moreno JL, Sabella P, Plana-Durán J, Enjuanes L. Coronavirus nucleocapsid protein is an RNA chaperone. *Virology* 2007; *357* : 215-27.

20. McBride R, van Zyl M, Fielding BC. The coronavirus nucleocapsid is a multifunctional protein. *Viruses* 2014; *6* : 2991-3018.

21. Su YCF, Anderson DE, Young BE, Linster M, Zhu F, Jayakumar J, *et al*. Discovery and genomic characterization of a 382-nucleotide deletion in *ORF7b and ORF8* during the early evolution of SARS-CoV-2. *mBio* 2020; *11* : e01610-20.

22. Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. *Microbiol Mol Biol Rev* 2012; *76* : 159-216.

23. Karamitros T, Papadopoulou G, Bousali M, Mexias A, Tsiodras S, Mentis A. SARS-CoV-2 exhibits intra-host genomic plasticity and low-frequency polymorphic quasispecies. *bioRxiv* 2020. doi: 10.1101/2020.03.27.009480.

24. Russell GC, Zadoks RN, Willoughby K, Bachofen C. Bovine viral diarrhoea virus loses quasispecies diversity rapidly in culture. *Microb Genom* 2020; *6* : e000343.

*For correspondence*: Dr Pragya D. Yadav, Maximum Containment Laboratory, ICMR-National Institute of Virology, Pashan, Pune 411 021, Maharashtra, India
e-mail: hellopragya22@gmail.com