Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2020 doi:10.20944/preprints202004.0009.v1

A flexible genome-scale resource of SARS-CoV-2 coding sequence clones

Dae-Kyum Kim[1,2,3], Jennifer J. Knapp[1,2,3], Da Kuang[1,2,3], Patricia Cassonnet[4,5,6], Payman Samavarchi-Tehrani[3], Hala Abdouni[3], Ashyad Rayhan[1,2,3], Dayag Sheykhkarimli[1,2,3], Étienne Coyaud[7], Sylvie van der Werf[4,5,6], Caroline Demeret[4,5,6], Anne-Claude Gingras[2,3], Brian Raught[8], Yves Jacob[4,5,6,*], Frederick P. Roth[1,2,3,9,*]

[1] Donnelly Centre, University of Toronto, Ontario, Canada

[2] Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

[3] Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Ontario, Canada

[4] Unité de Génétique Moléculaire des Virus à ARN, Département Virologie, Institut Pasteur, Paris, France.

[5] UMR3569, Centre National de la Recherche Scientifique, Paris, France.

[6] Université de Paris, Paris, France.

[7] Univ. Lille, Inserm, CHU Lille, U1192 - Protéomique Réponse Inflammatoire Spectrométrie de Masse - PRISM, F-59000 Lille, France

[8] Department of Medical Biophysics, Princess Margaret Cancer Centre, University of Toronto, Toronto, Ontario, Canada.

[9] Department of Computer Science, University of Toronto, Toronto, Ontario, Canada

* Correspondence: yves.jacob@pasteur.fr (Y.J), fritz.roth@utoronto.ca (F.P.R.)

Abstract

The world is facing a major health crisis, the global pandemic of COVID-19 caused by the SARS-CoV-2 coronavirus, for which no approved antiviral agents or vaccines are currently available. Here we describe a collection of codon-optimized coding sequences for SARS-CoV-2 cloned into Gateway-compatible entry vectors, which enable rapid transfer into a variety of expression and tagging vectors. The collection is freely available via Addgene. We hope that widespread availability of this SARS-CoV-2 resource will enable many subsequent molecular studies to better understand the viral life cycle and how to block it.

Introduction

The world is facing a major health crisis: A global pandemic of the coronavirus disease COVID-19, a severe respiratory illness caused by a novel virus from the family *Coronaviridae* (SARS-CoV-2), has caused a reported 693,224 confirmed cases and 33,106 deaths as of March 30, 2020 (

1). COVID-19 manifestation in patients can range from asymptomatic (no symptoms) to severe pneumonia and death (2). Early analysis of the outbreak in China outlines symptoms including fever (87.9% of patients showing a corresponding symptom), dry cough (67.7%), shortness of breath (18.6%) and myalgia (14.8%) (3). Person-to-person spread through respiratory droplets has been identified as a major source of transmission of the virus (4). Due to its contagious nature, various measures, from social distancing to country-wide lockdowns, have been imposed to contain and control the transmission of SARS-CoV-2 (5). Despite these measures, the number of confirmed COVID-19 cases continue to rise drastically (1), due to the lack of a vaccine or approved antiviral agents. Furthermore, the extrapolations concerning the evolution of the pandemic are particularly alarming (6). It is therefore of intense and pressing interest to better understand this virus and its interaction with host cells on a molecular level. Shortly after the outbreak, the complete genome of one SARS-CoV-2 strain was published (7). Using the genome sequence as a reference, Chan *et al*. identified 12 viral open reading frames (ORFs) (8), including ORF1ab, a large polyprotein which is post-translationally processed into 16 proteins. Progress on molecular characterization has been made on several viral proteins (9, 10), providing valuable insights into host-virus interaction. However, more research is necessary. Broad availability of a collection of SARS-CoV-2 coding sequences (CDSs) has the potential to enable many downstream biochemical and structural studies and thus a better understanding of processes within the viral life cycle, possibly yielding scalable assays for screening drug candidates that could disrupt these processes.

Methods and Results

Based on the published annotation of the genome sequence of the HKU-SZ-005b isolate of SARS-CoV-2 (8), we requested the synthesis of ORF sequences (GenScript), including termination codons and *attB* recombination sequences, with optimization of codon usage to reduce GC content. These sequences were then incorporated into Gateway Entry plasmids: either pDONR207 (Invitrogen) or pDONR223 (12). The Gateway system offers efficient and high-throughput transfer of the CDSs into a large selection of Gateway-compatible destination

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2020 doi:10.20944/preprints202004.0009.v1

vectors used for protein expression in many biological systems, e.g. *Escherichia coli*, *Saccharomyces cerevisiae*, insect, or mammalian cells (13).

Each SARS-CoV-2 CDS clone was isolated from a single colony, with its inserted CDS sequenced by Sanger sequencing. All clones with a pDONR223 backbone were sequenced with M13F (5´-GTAAAACGACGGCCAGT-3´) and M13R (5´-CAGGAAACAGCTATGAC-3´) primers (TCAG DNA sequencing facility, Toronto, Canada). Clones with a pDONR207 backbone were sequenced with customized forward (5'-TCGCGTTAACGCTAGCATGGATCT-3') and reverse (5'-GTAACATCAGAGATTTTGAGACAC-3') primers.

Clones for a total of 24 code-determining sequences (CDSs) are currently included in the collection (Table 1), with two more clones (N and S proteins) still in production. Upon completion, this collection will cover 26 Gateway-compatible clones out of 27 total CDSs in the SARS-CoV-2 genome. NSP11 was omitted because of its short length (36 base pairs, 12 amino acids), which makes this CDS incompatible with the Gateway cloning system (14). Other viral ORF lists have been published (7, 11), however, for consistency, we adapted the 27 viral coding sequences described in Chan *et al* (8).

To promote open-access dissemination of the collection, all clones have been deposited to Addgene (15) [Addgene deposit number: 77998]. Supplementary Table 1 summarizes all CDSs in the collection, together with their nucleotide and amino acid lengths, function annotation and direct links to Addgene.

We hope that this SARS-CoV-2 CDS-clone collection will be a valuable resource for many applications, including study of how coronaviruses can exploit host cellular processes for the viral replication cycle, e.g., (16), and understanding virus-host protein-protein interactions (11, 17), production of recombinant virus proteins for structural studies (18), mapping of protein subcellular localization using N-terminal fluorescent reporters (19), or development of vaccines or other therapeutics (20, 21). We are working to extend this SARS-CoV-2 Entry clone collection to include clones without termination codons, thus enabling C-terminal fusions.

 doi:10.20944/preprints202004.0009.v1

Acknowledgement

This work was supported by Canadian Institute for Health Research (CIHR) Foundation Grant (F.P.R.) and the Canada Excellence Research Chairs Program (F.P.R.), the LabEx Integrative Biology of Emerging Infectious Diseases (grant 10-LABX-0062) (Y.J.) and PREPARE, E-U GRANT N° 602525 (Y.J.). D.-K.K. was supported by a Banting Postdoctoral Fellowship through the Natural Sciences and Engineering Research Council (NSERC) of Canada and by the Basic Science Research Program through the National Research Foundation (NRF) of Korea funded by the Ministry of Education [2017R1A6A3A03004385].

Conflict of Interest

The authors declare that there is no conflict of interest.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2020 doi:10.20944/preprints202004.0009.v1

Table 1. The genome-scale SARS-CoV-2 coding sequence clone collection.

Gene Symbol	CDS Name	Putative Function/Domain	Protein Length	Status
ORF1AB	NSP1	Suppress antiviral host response	180	✓
	NSP2	Unknown	639	✓
	PLPRO (NSP3)	Putative PL-pro domain	1,946	✓
	NSP4	Complex with NSP3 & 6 for DMV (double-membrane vesicle) formation	501	✓
	NSP5	3CL-pro domain	307	✓
	NSP6	Complex with NSP 3 & 4 for DMV formation	291	✓
	NSP7	DNA primase subunits	84	✓
	NSP8		199	✓
	NSP9	RNA/DNA binding activity	114	✓
	NSP10	Complex with NSP14: Replication fidelity	140	✓
	RNA-pol (NSP12)	RNA-dependent RNA polymerase	919	✓
	Heli (NSP13)	Helicase	602	✓
	NSP14	ExoN: 3'-5' exonuclease	528	✓
	NSP15	XendoU: poly(U)-specific endoribonuclease	347	✓
	NSP16	2'-O'-MT: 2'-O-ribo methyltransferase	299	✓
S	S	Spike glycoprotein trimer that binds to host cell receptors (e.g. ACE2)	1,273	In process
ORF3A	3A	Induce inflammatory response and apoptosis	275	✓

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 April 2020 doi:10.20944/preprints202004.0009.v1

ORF3B	3B	Induce inflammatory response and inhibit the expression of IFNβ	58	✓
E	E	Envelope protein pentamer	75	✓
M	M	Membrane protein	222	✓
ORF6	6	Antagonize STAT1 function and IFN signalling, and induce DNA synthesis	61	✓
ORF7A	7A	Induce inflammatory response and apoptosis	121	✓
ORF7B	7B	Induce inflammatory response	43	✓
ORF8	8	Induce apoptosis and DNA synthesis	121	✓
N	N	Facilitate viral RNA packaging	419	In process
ORF9B	9B	Induce apoptosis	98	✓

✓ indicates that clone is currently at Addgene

References

1. World Health Organization (2020) COVID-19 Situation Reports. *World Health Organization*.

2. Huang,C., Wang,Y., Li,X., Ren,L., Zhao,J., Hu,Y., Zhang,L., Fan,G., Xu,J., Gu,X., *et al.* (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*, **395**, 497–506.

3. World Health Organization (2020) Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19).

4. Yu,P., Zhu,J., Zhang,Z., Han,Y. and Huang,L. (2020) A familial cluster of infection associated with the 2019 novel coronavirus indicating potential person-to-person transmission during the incubation period. *J. Infect. Dis.*, 10.1093/infdis/jiaa077.

5. Cohen,J. and Kupferschmidt,K. (2020) Countries test tactics in “war” against COVID-19. *Science*, **367**, 1287–1288.

6. Ferguson,N., Laydon,D., Nedjati Gilani,G., Imai,N., Ainslie,K., Baguelin,M., Bhatia,S., Boonyasiri,A., Cucunuba Perez,Z., Cuomo-Dannenburg,G., *et al.* (2020) Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. 10.25561/77482.

7. Wu,F., Zhao,S., Yu,B., Chen,Y.-M., Wang,W., Song,Z.-G., Hu,Y., Tao,Z.-W., Tian,J.-H., Pei,Y.-Y., *et al.* (2020) A new coronavirus associated with human respiratory disease in China. *Nature*, **579**, 265–269.

8. Chan,J.F.-W., Kok,K.-H., Zhu,Z., Chu,H., To,K.K.-W., Yuan,S. and Yuen,K.-Y. (2020) Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.*, **9**, 221–236.

9. Walls,A.C., Park,Y.-J., Tortorici,M.A., Wall,A., McGuire,A.T. and Veesler,D. (2020) Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, 10.1016/j.cell.2020.02.058.

10. Zhang,L., Lin,D., Sun,X., Curth,U., Drosten,C., Sauerhering,L., Becker,S., Rox,K. and Hilgenfeld,R. (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science*, 10.1126/science.abb3405.

11. Gordon,D.E., Jang,G.M., Bouhaddou,M., Xu,J., Obernier,K., O’Meara,M.J., Guo,J.Z., Swaney,D.L., Tummino,T.A., Huttenhain,R., *et al.* (2020) A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing. *bioRxiv*, 10.1101/2020.03.22.002386.

12. Rual,J.-F., Hirozane-Kishikawa,T., Hao,T., Bertin,N., Li,S., Dricot,A., Li,N., Rosenberg,J., Lamesch,P., Vidalain,P.-O., *et al.* (2004) Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Res.*, **14**, 2128–2135.

13. Walhout,A.J.M., Temple,G.F., Brasch,M.A., Hartley,J.L., Lorson,M.A., van den Heuvel,S.

and Vidal,M. (2000) [34] GATEWAY recombinational cloning: Application to the cloning of large numbers of open reading frames or ORFeomes. In Thorner,J., Emr,S.D., Abelson,J.N. (eds), *Methods in Enzymology*. Academic Press, Vol. 328, pp. 575–IN7.

14. ThermoFisher Gateway Recombination and Seamless Cloning Support. *ThermoFisher*.

15. Kamens,J. (2015) The Addgene repository: an international nonprofit plasmid and data resource. *Nucleic Acids Res.*, **43**, D1152–7.

16. de Wilde,A.H., Snijder,E.J., Kikkert,M. and van Hemert,M.J. (2018) Host Factors in Coronavirus Replication. In Tripp,R.A., Tompkins,S.M. (eds), *Roles of Host Gene and Non-coding RNA Expression in Virus Infection*. Springer International Publishing, Cham, pp. 1–42.

17. Lasso,G., Mayer,S.V., Winkelmann,E.R., Chu,T., Elliot,O., Patino-Galindo,J.A., Park,K., Rabadan,R., Honig,B. and Shapira,S.D. (2019) A Structure-Informed Atlas of Human-Virus Interactions. *Cell*, **178**, 1526–1541.e16.

18. Edavettal,S.C., Hunter,M.J. and Swanson,R.V. (2012) Genetic construct design and recombinant protein expression for structural biology. *Methods Mol. Biol.*, **841**, 29–47.

19. Tanz,S.K., Castleden,I., Small,I.D. and Millar,A.H. (2013) Fluorescent protein tagging as a tool to define the subcellular distribution of proteins in plants. *Front. Plant Sci.*, **4**, 214.

20. McDonald,W.F., Huleatt,J.W., Foellmer,H.G., Hewitt,D., Tang,J., Desai,P., Price,A., Jacobs,A., Takahashi,V.N., Huang,Y., *et al.* (2007) A West Nile virus recombinant protein vaccine that coactivates innate and adaptive immunity. *J. Infect. Dis.*, **195**, 1607–1617.

21. Jing,L., Haas,J., Chong,T.M., Bruckner,J.J., Dann,G.C., Dong,L., Marshak,J.O., McClurkan,C.L., Yamamoto,T.N., Bailer,S.M., *et al.* (2012) Cross-presentation and genome-wide screening reveal candidate T cells antigens for a herpes simplex virus type 1 vaccine. *J. Clin. Invest.*, **122**, 654–673.