



HAL
open science

A Comprehensive, Flexible Collection of SARS-CoV-2 Coding Regions

Dae-Kyum Kim, Jennifer Knapp, Da Kuang, Aditya Chawla, Patricia Cassonnet, Hunsang Lee, Dayag Sheykhkarimli, Payman Samavarchi-Tehrani, Hala Abdouni, Ashyad Rayhan, et al.

► **To cite this version:**

Dae-Kyum Kim, Jennifer Knapp, Da Kuang, Aditya Chawla, Patricia Cassonnet, et al.. A Comprehensive, Flexible Collection of SARS-CoV-2 Coding Regions. *G3*, 2020, 10 (9), pp.3399-3402. 10.1534/g3.120.401554 . pasteur-02939727

HAL Id: pasteur-02939727

<https://pasteur.hal.science/pasteur-02939727v1>

Submitted on 15 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

A Comprehensive, Flexible Collection of SARS-CoV-2 Coding Regions

Dae-Kyum Kim,^{*,†,‡,1} Jennifer J. Knapp,^{*,†,‡,1} Da Kuang,^{*,†,‡,1} Aditya Chawla,^{*,†,‡} Patricia Cassonnet,^{§,**,††} Hunsang Lee,^{*,†} Dayag Sheykhkarimli,^{*,†,‡} Payman Samavarchi-Tehrani,[‡] Hala Abdouni,[‡] Ashyad Rayhan,^{*,†,‡} Roujia Li,^{*,†,‡} Oxana Pogoutse,^{*,†,‡} Étienne Coyaud,^{**} Sylvie van der Werf,^{§,**,††} Caroline Demeret,^{§,**,††} Anne-Claude Gingras,^{†,‡} Mikko Taipale,^{*,†,§§} Brian Raught,^{***} Yves Jacob,^{§,**,††,2} and Frederick P. Roth^{*,†,‡,†††,2}

^{*}Donnelly Centre, [†]Department of Molecular Genetics, ^{**}Department of Medical Biophysics, Princess Margaret Cancer Centre, ^{†††}Department of Computer Science, University of Toronto, Ontario, Canada M5S 2E4, [‡]Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Ontario, Canada M5G 1X5, [§]Unité de Génétique Moléculaire des Virus à ARN, Département Virologie, Institut Pasteur, Paris, France 75015, ^{**}UMR3569, Centre National de la Recherche Scientifique, Paris, France, ^{††}Université de Paris, France, 75016 ^{**}Univ. Lille, Inserm, CHU Lille, U1192 - Protéomique Réponse Inflammatoire Spectrométrie de Masse - PRISM, F-59000 Lille, France, and ^{§§}Molecular Architecture of Life Program, Canadian Institute for Advanced Research, Toronto, Ontario, Canada M5G 1M1
ORCID ID: 0000-0003-3554-0464 (D.K.)

ABSTRACT The world is facing a global pandemic of COVID-19 caused by the SARS-CoV-2 coronavirus. 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. 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.

KEYWORDS
SARS-CoV-2
coding sequence
collection
Gateway-compatible
TEV (tobacco etch virus)
sequence

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 infected millions and caused hundreds of thousands of deaths (World Health Organization 2020a). COVID-19

manifestation in patients can range from a lack of symptoms to severe pneumonia and death (Huang *et al.* 2020). Person-to-person spread through respiratory droplets has been identified as a major source of transmission of the virus (Yu *et al.* 2020). Various measures, from social distancing to nationwide lockdowns, have been imposed to contain and control the transmission of SARS-CoV-2 (Cohen and Kupferschmidt 2020). Despite these measures, the number of confirmed COVID-19 cases has continued to rise (World Health Organization 2020a), highlighting the need for an effective vaccine and antiviral agents. Furthermore, the extrapolations concerning the evolution of the pandemic are particularly alarming (Ferguson *et al.* 2020). 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 two SARS-CoV-2 strains were published (Chan *et al.* 2020; Wu *et al.* 2020). Using the genome sequence as a reference, Chan *et al.* identified 12 viral open reading frames (ORFs), including one encoding

Copyright © 2020 Kim *et al.*
doi: <https://doi.org/10.1534/g3.120.401554>
Manuscript received July 10, 2020; accepted for publication July 22, 2020; published Early Online August 6, 2020.
This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Supplemental material available at figshare: <https://doi.org/10.25387/g3.12725096>.
¹These authors contributed equally to this work.
²Corresponding authors: Yves Jacob. Dept de Virologie- Institut Pasteur-28 rue du Dr Roux -75724. Phone: (33) 1 45 68 87 53. E-mail: yves.jacob@pasteur.fr.
Frederick P. Roth. Donnelly Centre, 160 College Street, Toronto, Ontario, Canada M5S 3E1. Phone: (1) 416-946-5130. E-mail: fritz.roth@utoronto.ca.

