JOURNAL OF GENERAL VIROLOGY

INSIGHT REVIEW

Silva *et al.*, *Journal of General Virology* 2020;101:1021–1024

DOI 10.1099/jgv.0.001458

MICROBIOLOGY SOCIETY

OPEN ACCESS

Insights into SARS-CoV-2, the Coronavirus Underlying COVID-19: Recent Genomic Data and the Development of Reverse Genetics Systems

Severino Jefferson Ribeiro da Silva[1], Renata Pessôa Germano Mendes[1], Caroline Targino Alves da Silva[1], Alessio Lorusso[2], Alain Kohl[3,*] and Lindomar Pena[1,*]

Abstract

The emergence and rapid worldwide spread of a novel pandemic of acute respiratory disease – eventually named coronavirus disease 2019 (COVID-19) by the World Health Organization (WHO) – across the human population has raised great concerns. It prompted a mobilization around the globe to study the underlying pathogen, a close relative of severe acute respiratory syndrome coronavirus (SARS-CoV) called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Numerous genome sequences of SARS-CoV-2 are now available and in-depth analyses are advancing. These will allow detailed characterization of sequence and protein functions, including comparative studies. Care should be taken when inferring function from sequence information alone, and reverse genetics systems can be used to unequivocally identify key features. For example, the molecular markers of virulence, host range and transmissibility of SARS-CoV-2 can be compared to those of related viruses in order to shed light on the biology of this emerging pathogen. Here, we summarize some recent insights from genomic studies and strategies for reverse genetics systems to generate recombinant viruses, which will be useful to investigate viral genome properties and evolution.

Coronavirus disease 2019 (COVID-19) [1] is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a recently identified virus (December 2019, Wuhan, Hubei Province, China) [2, 3] belonging to the family *Coronaviridae*, subfamily *Coronavirinae,* genus *Betacoronavirus*, in the species *severe acute respiratory syndrome-related coronavirus* [4]. This pandemic has led to disease and mortality in populations across the globe [1, 5]. On 11 March 2020, the World Health Organization (WHO) declared a SARS-CoV-2 pandemic. As of 31 May 2020, more than 6 million COVID-19 cases have been reported in 188 countries, resulting in 370 000 deaths worldwide. In addition to the genus *Betacoronavirus*, the subfamily *Coronavirinae* is composed of the genera *Alphacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus*. Seven types of CoVs are known to cause human disease. Several alphacoronaviruses (HC 229E and HC NL63) and betacoronaviruses (HC OC43 and HC HKU1) are endemic and cause mild respiratory tract infections [6, 7]. Previous to SARS-CoV-2, two betacoronaviruses had already emerged in the 21st century causing severe respiratory disease: SARS-CoV (same species as SARS-CoV-2) and Middle East respiratory syndrome CoV (MERS-CoV), allowing comparisons to be made between these three viruses and their respective diseases [8–10]. In January 2020 the complete viral genome sequences from five patients in Wuhan during an early stage of the outbreak were published and SARS-CoV-2 was found to be a novel CoV, with just under 80% sequence identity to SARS-CoV [2]. The viruses most closely related to SARS-CoV-2 were coronaviruses isolated from bats, in particular RaTG13 [2, 11]. Therefore, it was hypothesized that bats – a known reservoir of coronaviruses – could serve as a reservoir for this novel coronavirus. However, it is not clear whether transmission of SARS-CoV-2 to humans occurred directly

Received 22 April 2020; Accepted 01 June 2020; Published 24 June 2020

Author affiliations: [1]Department of Virology, Aggeu Magalhães Institute (IAM), Oswaldo Cruz Foundation (Fiocruz), 50670-420, Recife Pernambuco, Brazil; [2]Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise (IZSAM), Teramo, Italy; [3]MRC-University of Glasgow Centre for Virus Research, Glasgow, Scotland G61 1QH, UK.

***Correspondence:** Lindomar Pena, lindomar.pena@cpqam.fiocruz.br; Alain Kohl, Alain.Kohl@glasgow.ac.uk

Keywords: SARS-CoV-2; reverse genetics; COVID-19; genomics.

Abbreviations: COVID-19, Coronavirus disease 2019; IAV, influenza A Virus; MERS-CoV, Middle East respiratory syndrome coronavirus; ORF, open reading frame; pA, poly (A) tail; SARS-CoV, severe acute respiratory syndrome coronavirus; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; T7, T7 RNA polymerase promoter; UTR, untranslated region; WHO, World Health Organization; YAC, yeast artificial chromosome.

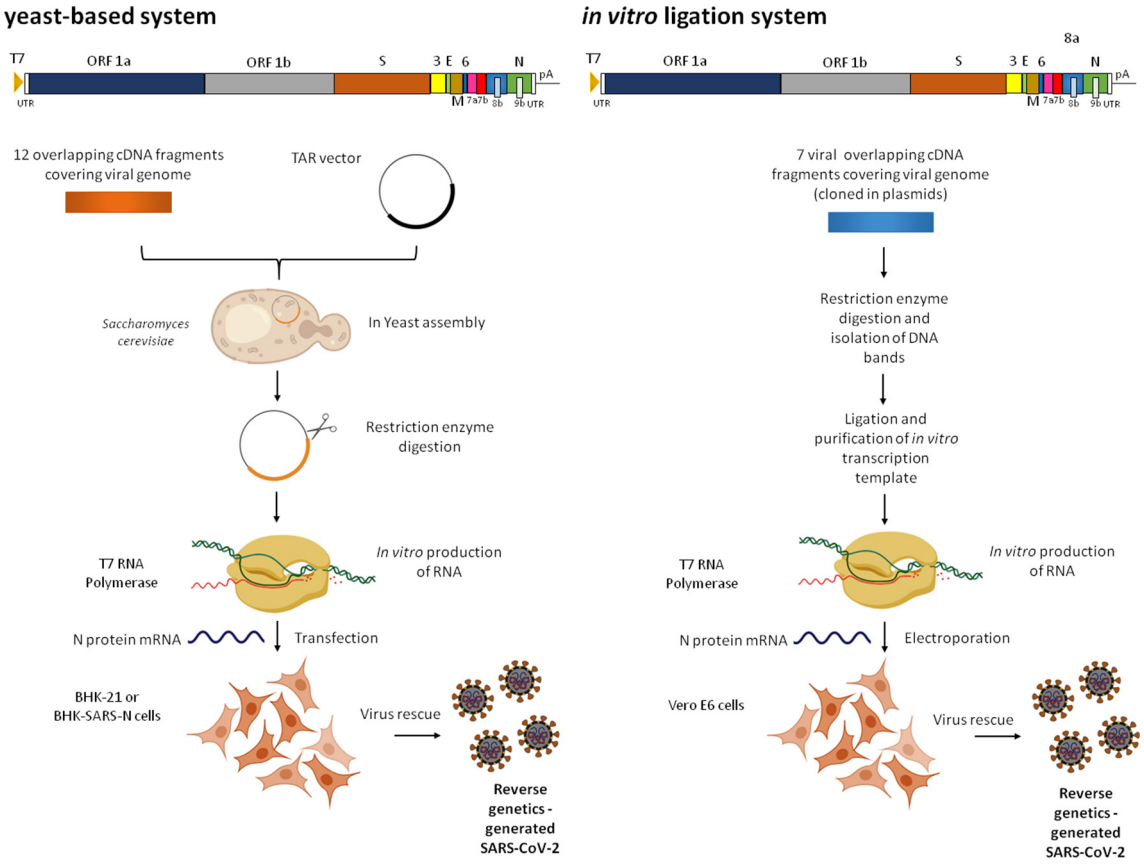

Fig. 1. Reverse genetics systems for SARS-CoV-2. Viral genome and organization are shown at the top of each panel. (a) Description of yeast-based assembly and rescue system. Twelve viral subgenomic cDNA fragments were assembled in *Saccharomyces cerevisiae* using transformation-associated recombination (TAR) cloning to maintain the genome as a yeast artificial chromosome (YAC). *In vitro*-transcribed (by T7 RNA polymerase) viral genome RNAs were electroporated into BHK-21 cells (or BHK-SARS-N) together with an mRNA encoding the SARS-CoV-2 N protein to rescue viable virus. (b) Description of *in vitro* ligation system. In this approach, seven contiguous cDNA fragments covering the entire viral genome were isolated from plasmid vectors and directionally ligated to assemble the full-length viral genome. *In vitro*-transcribed (by T7 RNA polymerase) genome RNA was transfected into Vero E6 cells along with mRNA encoding N protein to recover infectious SARS-CoV-2. A schematic representation of the SARS-CoV-2 genome organization is shown in the upper part of the panels. T7, T7 RNA polymerase promoter; UTR, untranslated region; pA, poly (A) tail. Created with Biorender.com

from bats or through an intermediate host [11–16]. The genome sequences of numerous SARS-CoV-2 strains from across the globe are now publicly available.

Viral replication takes place in the cytoplasm. The genome organization of the virus is summarized in Fig. 1 (alongside descriptions of reverse genetics systems; see below), with the overall structure and key elements of the genome being comparable to those of related coronaviruses. Viral replicase activities are directed by 16 non-structural proteins that are produced following the proteolytic cleavage of 2 replicase polyproteins [open reading frames (ORFs) 1a/b]. The genome also encodes the structural proteins N, S, E and M, as well as accessory proteins [17–24]. The genome termini play critical roles in replication and transcription; ORFs 1a/b are translated from the genome, but subgenomic mRNAs mediate the expression of the remaining viral proteins [25, 26].

Clearly COVID-19 is different from the disease caused by SARS-CoV-2's close relative, SARS-CoV. Indeed the case fatality rates for COVID-19 are lower, and the disease can be mild or asymptomatic [27]. Investigating the differences between the two related viruses is thus of critical importance for future investigations. Analysis of SARS-CoV-2 sequences showed that the spike protein (S) has a furin(-like) cleavage site that is absent in related coronaviruses, and this was experimentally confirmed [28, 29]. Intriguingly, the loss of this cleavage site was shown upon passaging the virus in cell culture [30]. Processing by furin cleavage could have implications for virulence and/or adaptation. Moreover, roles for nonstructural proteins nsp2, nsp3, ns7b and ns8 in the pathogenesis of SARS-CoV-2 have been proposed following analysis of selective pressure on ORFs [31, 32]. Recent analysis of over 200 SARS-CoV-2 sequences classified the virus into 5 main groups based on high-frequency mutations (mutant allele

frequency >5%), with group 4 most frequently being found outside mainland China. Intriguingly, analysis of group 4 SARS-CoV-2 genome data from France showed that viruses carrying a mutation in 3a also often have a mutation in S (ORF3a:c.752gGt>gTt and S:c.1099Gtc>Ttc), although the biological relevance of this is not clear [33]. Moreover, comparative analysis between SARS-CoV-2 and RaTG13 has suggested some evidence of limited positive selection, although this cannot be interpreted as evidence for adaptation to humans [11]. More recently, it has been proposed that the D614G mutation in the S glycoprotein increases the transmissibility of SARS-CoV-2, as evidenced by sequence analysis [34]. However, care should be taken when inferring function from sequence information alone. In the context of another respiratory virus, influenza A virus (IAV), the PB1-F2 protein was identified as an important virulence factor and the N66S mutation in this protein was associated with the high lethality of the 1918 and other IAVs. However, reverse genetics studies found that the role of PB1-F2 (and the N66S polymorphism) in the virulence of different IAVs was host- and strain-dependent, ranging from increased virulence to no effect and even attenuation [35, 36]. Together, these examples illustrate how phenotype inference from sequence information needs experimental confirmation – importantly, in relevant systems.

Reverse genetics is a powerful technique for the generation of an infectious virus from the cloned full-length cDNA/synthetized DNA of a given virus. Manipulation of the DNA by well-established molecular biology methods allows modification of the sequence before virus production, if this is desired. It allows unequivocal identification of molecular markers for a given virus, including the genome features, virulence, host range, etc. of a given virus, and as such it is a key component in the study of coronaviruses [37]. Ultimately, such systems will be required to answer questions on various aspects of SARS-CoV-2 biology and genomics. The generation of coronaviruses entirely from full-length DNA can be challenging. This is mainly because of the large genome size of coronaviruses (~30 Kb) as well as the instability of genome sequences during cloning. However, such systems, as well as minireplicons (replication/transcription active, propagation-incompetent viral RNAs), have been developed successfully in the past for coronaviruses, including SARS-CoV. This was achieved through the use of bacterial artificial chromosomes, *in vitro* ligation of DNA fragments covering the full-length genome, or vaccinia virus-based expression vectors [38]. The systems included reverse genetics for SARS-CoV as well as MERS-CoV [39–43]. A ligation-based methodology combining individual stretches of DNA covering the SARS-CoV-2 genome (originally cloned into plasmids), followed by T7 RNA polymerase-based transcription to produce viral RNA, was used successfully to recover this virus [44]. Similarly, a yeast artificial chromosome-based system was developed recently to propagate the full-length SARS-CoV-2 genome assembled from DNA. Again, viral RNA was transcribed by T7 RNA polymerase before transfection [45]. Both of these recent methodologies are summarized in Fig. 1. An mRNA encoding the SARS-CoV-2 N protein was co-transfected in both approaches to enhance the infectivity of viral RNA transcripts. Both approaches allowed the successful generation of fluorescent protein-expressing recombinant viruses. The availability of SARS-CoV-2 reverse genetics systems will allow effective manipulation of its genome and unravel questions over entry, gene expression, replication, tropism, etc highlighted by genomic analysis as discussed above. These technologies can also be applied to the generation of vaccine candidates and the discovery of antivirals against this devastating human pathogen. The rapid development of such systems for SARS-CoV-2 using different approaches is testimony to the pioneering work carried out with related viruses and underlines the need for continuous research on these and other pathogens. They will greatly enhance our ability to investigate this novel pathogen, as well as coronaviruses that may emerge in the future.

Funding information
L.P. is supported by grants from FACEPE, CNPq and IDRC. A.K. is funded by the UK Medical Research Council (MC_UU_12014/8).

Acknowledgements
We are grateful to Dr Claire L. Donald for critical reading of the manuscript.

Author contributions
Conceptualization: S.J.R.D.S., R.P.G.M., C.T.A.D.S., A.L., A.K., L.P.; writing original draft preparation: S.J.R.D.S., R.P.G.M., C.T.A.D.S., A.L., A.K., L.P.; writing- review and editing: S.J.R.D.S., R.P.G.M., C.T.A.D.S., A.K., L.P.; visualization: A.K., L.P.; funding: A.K., L.P.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Jin Y, Yang H, Ji W, Wu W, Chen S *et al*. Virology, epidemiology, pathogenesis, and control of COVID-19. *Viruses* 2020;12:372.
2. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L *et al*. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;579:270–273.
3. Lu R, Zhao X, Li J, Niu P, Yang B *et al*. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 2020;395:565–574.
4. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 2020;5:536–544.
5. He F, Deng Y, Li W. Coronavirus disease 2019: what we know? *J Med Virol*;2020.
6. Li S-W, Lin C-W. Human coronaviruses: clinical features and phylogenetic analysis. *Biomedicine* 2013;3:43–50.
7. Corman VM, Muth D, Niemeyer D, Drosten C. Hosts and sources of endemic human coronaviruses. *Adv Virus Res* 2018;100:163–188.
8. de Wit E, van Doremalen N, Falzarano D, Munster VJ. Sars and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol* 2016;14:523–534.
9. Song Z, Xu Y, Bao L, Zhang L, Yu P *et al*. From SARS to MERS, Thrusting coronaviruses into the spotlight. *Viruses* 2019;11:59
10. Weston S, Frieman MB. COVID-19: knowns, unknowns. *and Questions. mSphere* 2020;5.

11. Cagliani R, Forni D, Clerici M, Sironi M. Computational inference of selection underlying the evolution of the novel coronavirus, SARS-CoV-2. *J Virol* 2020.
12. Ye Z-W, Yuan S, Yuen K-S, Fung S-Y, Chan C-P *et al*. Zoonotic origins of human coronaviruses. *Int J Biol Sci* 2020;16:1686–1697.
13. Cui J, Li F, Shi Z-L. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* 2019;17:181–192.
14. Banerjee A, Kulcsar K, Misra V, Frieman M, Mossman K. Bats and coronaviruses. *Viruses* 2019;11:41.
15. Khan S, Siddique R, Shereen MA, Ali A, Liu J *et al*. The emergence of a novel coronavirus (SARS-CoV-2), their biology and therapeutic options. *J Clin Microbiol* 2020;58:e00187–20.
16. Zhang Y-Z, Holmes EC. A genomic perspective on the origin and emergence of SARS-CoV-2. *Cell* 2020;181:223–227.
17. Snijder EJ, Decroly E, Ziebuhr J. The nonstructural proteins directing coronavirus RNA synthesis and processing. *Adv Virus Res* 2016;96:59–126.
18. de Wilde AH, Snijder EJ, Kikkert M, van Hemert MJ. Host factors in coronavirus replication. *Curr Top Microbiol Immunol* 2018;419:1–42.
19. Forni D, Cagliani R, Clerici M, Sironi M. Molecular evolution of human coronavirus genomes. *Trends Microbiol* 2017;25:35–48.
20. Neuman BW, Buchmeier MJ. Supramolecular architecture of the coronavirus particle. *Adv Virus Res* 2016;96:1–27.
21. Liu DX, Fung TS, Chong KK-L, Shukla A, Hilgenfeld R. Accessory proteins of SARS-CoV and other coronaviruses. *Antiviral Res* 2014;109:97–109.
22. Neuman BW, Chamberlain P, Bowden F, Joseph J. Atlas of coronavirus replicase structure. *Virus Res* 2014;194:49–66.
23. Enjuanes L, Almazán F, Sola I, Zuñiga S. Biochemical aspects of coronavirus replication and virus-host interaction. *Annu Rev Microbiol* 2006;60:211–230.
24. Enjuanes L, Almazán F, Sola I, Zúñiga S, Alvarez E *et al*. Biochemical aspects of coronavirus replication. *Adv Exp Med Biol* 2006;581:13–24.
25. Yang D, Leibowitz JL. The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res* 2015;206:120–133.
26. Sawicki SG, Sawicki DL, Siddell SG. A contemporary view of coronavirus transcription. *J Virol* 2007;81:20–29.
27. Mahase E. Coronavirus covid-19 has killed more people than SARS and MERS combined, despite lower case fatality rate. *BMJ* 2020;368:m641.
28. Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG *et al*. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res* 2020;176:104742.
29. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT *et al*. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 2020;181:281–292.
30. Davidson AD, Williamson MK, Lewis S, Shoemark D, Carroll MW *et al*. Characterisation of the transcriptome and proteome of SARS-CoV-2 using direct RNA sequencing and tandem mass spectrometry reveals evidence for a cell passage induced in-frame deletion in the spike glycoprotein that removes the furin-like cleavage site. *bioRxiv* 2020.
31. Angeletti S, Benvenuto D, Bianchi M, Giovanetti M, Pascarella S *et al*. COVID-2019: the role of the Nsp2 and NSP3 in its pathogenesis. *J Med Virol* 2020:584–588.
32. Fahmi M, Kubota Y, Ito M. Nonstructural proteins NS7b and NS8 are likely to be phylogenetically associated with evolution of 2019-nCoV. *Infect Genet Evol* 2020;81:104272.
33. Wang M, Li M, Ren R, Li L, Chen E-Q *et al*. International expansion of a novel SARS-CoV-2 mutant. *J Virol* 2020;94:JVI.00567-20..
34. Brufsky A. Distinct viral clades of SARS-CoV-2: implications for modeling of viral spread. *J Med Virol* 2020
35. Schmolke M, Manicassamy B, Pena L, Sutton T, Hai R *et al*. Differential contribution of PB1-F2 to the virulence of highly pathogenic H5N1 influenza A virus in mammalian and avian species. *PLoS Pathog* 2011;7:e1002186.
36. Pena L, Vincent AL, Loving CL, Henningson JN, Lager KM *et al*. Restored PB1-F2 in the 2009 pandemic H1N1 influenza virus has minimal effects in swine. *J Virol* 2012;86:5523–5532.
37. Johnson BA, Graham RL, Menachery VD. Viral metagenomics, protein structure, and reverse genetics: key strategies for investigating coronaviruses. *Virology* 2018;517:30–37.
38. Almazán F, Sola I, Zuñiga S, Marquez-Jurado S, Morales L *et al*. Coronavirus reverse genetic systems: infectious clones and replicons. *Virus Res* 2014;189:262–270.
39. Scobey T, Yount BL, Sims AC, Donaldson EF, Agnihothram SS *et al*. Reverse genetics with a full-length infectious cDNA of the middle East respiratory syndrome coronavirus. *Proc Natl Acad Sci U S A* 2013;110:16157–16162.
40. Yount B, Curtis KM, Fritz EA, Hensley LE, Jahrling PB *et al*. Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. *Proc Natl Acad Sci U S A* 2003;100:12995–13000.
41. Menachery VD, Yount BL, Debbink K, Agnihothram S, Gralinski LE *et al*. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat Med* 2015;21:1508–1513.
42. Eriksson KK, Makia D, Thiel V. Generation of recombinant coronaviruses using vaccinia virus as the cloning vector and stable cell lines containing coronaviral replicon RNAs. *Methods Mol Biol* 2008;454:237–254.
43. Almazán F, Dediego ML, Galán C, Escors D, Alvarez E *et al*. Construction of a severe acute respiratory syndrome coronavirus infectious cDNA clone and a replicon to study coronavirus RNA synthesis. *J Virol* 2006;80:10900–10906.
44. Xie X, Muruato A, Lokugamage KG, Narayanan K, Zhang X *et al*. An infectious cDNA clone of SARS-CoV-2. *Cell Host Microbe* 2020;27:841–848.
45. Thao TTN, Labroussaa F, Ebert N, V'kovski P, Stalder H *et al*. Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* 2020.