IMI2 Project ID 101005077

CARE
Corona Accelerated R&D in Europe

WP6 – From lead to pre-clinical candidate and proof-of-concept in small-animal and non-human primate models

6.3 - Report on reverse genetics system

Lead contributor	P21-EDI-IVI
Other contributors	N/A

Document History

Version	Date	Description
V1.0	29 Sept 2020	Initial Report on Reverse Genetics

The CARE project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 101005077. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and BILL & MELINDA GATES FOUNDATION, GLOBAL HEALTH DRUG DISCOVERY INSTITUTE, UNIVERSITY OF DUNDEE.

1. Summary

Reverse genetics has been an indispensable tool to gain insights into viral pathogenesis and vaccine development. The genomes of large RNA viruses, such as those from coronaviruses, are cumbersome to clone and manipulate in *Escherichia coli* owing to the size and occasional instability of the genome. Therefore, an alternative rapid and robust reverse-genetics platform for RNA viruses would benefit the research community. Here we show the full functionality of a yeast-based synthetic genomics platform to genetically reconstruct diverse RNA viruses, including members of the *Coronaviridae, Flaviviridae* and *Pneumoviridae* families. Viral subgenomic fragments were generated using viral isolates, cloned viral DNA, clinical samples or synthetic DNA, and these fragments were then reassembled in one step in *Saccharomyces cerevisiae* using transformation-associated recombination cloning to maintain the genome as a yeast artificial chromosome. T7 RNA polymerase was then used to generate infectious RNA to rescue viable virus. Using this platform, we were able to engineer and generate chemically synthesized clones of the virus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which has caused the recent pandemic of coronavirus disease (COVID-19), in only a week after receipt of the synthetic DNA fragments. The technical advance that we describe here facilitates rapid responses to emerging viruses as it enables the real-time generation and functional characterization of evolving RNA virus variants during an outbreak.

2. Methods

Cells and general culture conditions

Vero E6 cells (ATCC) were cultured in Dulbecco's modified Eagle's medium (DMEM); BHK-21 and BHK-SARS-N (BHK-21 cells expressing the N protein of SARS), were grown in minimal essential medium (MEM). Both types of medium were supplemented with 10% fetal bovine serum, 1× non-essential amino acids, 100 units ml^{-1} penicillin and 100 µg ml^{-1} streptomycin. BHK-SARS-N cells were grown using MEM supplemented with 5% fetal bovine serum, 1× non-essential amino acids, 100 units ml^{-1} penicillin, and 100 µg ml^{-1} streptomycin, 500 µg ml^{-1} G418 and 10 µg ml^{-1} puromycin and treated with 1 µg ml^{-1} doxycyclin 24 h before electroporation. All cells were maintained at 37 °C and in a 5% CO_2 atmosphere.

Cultured viruses

SARS-CoV-2 (SARS-CoV-2/München-1.1/2020/929) was cultured in Vero E6 cells.

Bacterial and yeast strains

E. coli DH5α (Thermo Scientific) and TransforMax Epi300 (Epicentre) were used to propagate the pVC604 and pCC1BAC-His3 TAR vectors, respectively. The bacteria were grown in lysogeny broth medium supplemented with the appropriate antibiotics at 37 °C overnight. *E. coli* Epi300 cells containing the different synthetic fragments of SARS-CoV-2 in pUC57 or pUC57mini were grown at 30 °C to decrease the risk of instability and/or toxicity. *Saccharomyces cerevisiae* VL6-48N (MATα trp1-Δ1 ura3-Δ1 ade2-101 his3-Δ200 lys2 met14 cir°) was used for all yeast transformation experiments. Yeast cells were first grown in YPDA broth (Takara Bio), and transformed cells were plated on minimal synthetic defined (SD) agar without histidine (SD−His) (Takara Bio). *S. cerevisiae* VL6-48N-derived

clones carrying different YACs were never streaked out together on the same agar dishes as mating switching and resulting recombination might occur at a very low frequency.

Generation of viral sub-genomic fragments for TAR cloning using viral RNA, infectious cDNA clones and synthetic DNA

Viral DNA fragments were obtained by RT–PCR of viral RNA extracted from viral strains, isolates and from clinical specimens, using the SuperScript IV One-Step RT–PCR System following the manufacturer's instructions. Additionally, some fragments were PCR-amplified from vaccinia virus-cloned cDNA, BAC-cloned cDNA and plasmid-cloned synthetic DNA (GenScript), using the CloneAmp HiFi PCR Premix according to the manufacturer's instructions. The accessory sequence, TurboGFP for SARS-CoV-2-GFP was amplified from a plasmid. The fragment encompassing the viral 5′ untranslated regions (UTR) contained the T7 RNA polymerase promoter sequence immediately upstream of the 5′ end of the genome, and the fragment encompassing the 3′ end of the genome contained a unique restriction site downstream of the poly(A) tail.

The SARS-CoV-2 synthetic DNA fragments were delivered cloned into pUC57 or pUC57mini by GenScript. Fragments 1.1, 1.2, 1.3 and 12 contained homologous sequences to pCC1BAC-His3. Each fragment was sequence verified using Sanger sequencing after plasmid isolation using QIAGEN Midiprep kit (QIAGEN). Fragments were released from the vector using restriction enzymes. Restricted fragments were subsequently gel-purified using standard methods. DNA concentrations and purities of all fragments to be used for TAR cloning were determined using NanoDrop 2000/2000c Spectrophotometer.

In-yeast cloning of viral genomes using TAR

In general, we used overlapping DNA fragments for TAR cloning with overlaps ranging from 45 to 500 bp. As all of our cloning experiments worked well, we did not assess whether the lengths of the overlap affected homologous recombination efficacy. The vectors pVC604 and pCC1BAC-His3 were used for TAR cloning. These vectors were amplified by PCR using primers containing at least 45-bp overlaps to fragments encompassing the 5′ or 3′ ends of different viral genomes. Amplification was performed using KOD Hot Start DNA polymerase (Merck Millipore) according to the manufacturer's instructions. TAR cloning was also used to reconstruct the full-length synthetic fragments 5 and 7 in yeast.

Yeast transformation was done using the high-efficiency lithium acetate/SS carrier DNA/PEG method as described elsewhere. In brief, yeast cells were grown in rich YPDA medium (Takara Bio) at 30 °C with agitation until an optical density at 600 nm of 1.0 was reached. Then, 3 ml of yeast culture was used per transformation event. DNA mixtures were prepared beforehand and contained 100–200 fmol of 3′ and 5′ open ends for all fragments. Transformation mixtures were plated onto SD−His plates (Takara Bio) and incubated at 30 °C for 48 h. Colonies were resuspended in 20 µl of SD−His broth, and DNA was extracted following the GC prep method. Extracted DNA was used as template for screening by multiplex PCR using the QIAGEN Multiplex PCR kit (QIAGEN) according to the manufacturer's instruction. One or two multiplex PCRs were designed to encompass different subsets of primer pairs, and cover all desired recombination junctions. Clones tested positive for all junctions were grown in SD−His until late logarithmic phase, and plasmids were extracted from 500 ml culture using the QIAGEN Maxiprep Kit (QIAGEN) with modifications.

In brief, 10 ml of Buffer P1 was supplemented with 1 ml of zymolyase solution (10 mg ml^{-1} Zymolyase 100-T; 50 mM Tris-HCl pH 7.5; 50% (v/v) glycerol) and 100 µl of β-mercapthoethanol. The mixture was incubated for 1 h at 37 °C before the addition of buffer P2. The rest of the protocol followed the manufacturer's instructions. DNA preparations were successfully used as templates to generate *in vitro* transcribed viral RNA even if they contained traces of yeast genomic DNA. In parallel, isolated YACs containing full-length synthetic fragments 5 and 7, as well as SARS-CoV-2 and SARS-CoV-2-GFP viral genomes, were successfully transformed into *E. coli* TransforMax Epi300 electrocompetent cells (Epicentre).

Stability testing of the YAC containing entire RNA virus genomes in yeast

The stability of viral genomes maintained as YACs in *S. cerevisiae* was tested for the clones containing MHV-GFP or MERS-CoV for 1 week. A single colony was grown in 20 ml of SD−His liquid medium, 1 ml aliquots were removed and expanded in fresh medium every 12 h. The generation time for each of the clones was estimated to range from 150 to 160 min. After 15–17 passages, each YAC clone was isolated and subjected to sequencing by MinION (Oxford Nanopore Technologies) to obtain the entire YAC sequence. Individual regions for which MinION sequencing did not reveal a clear sequence were resequenced by Sanger sequencing (Microsynth).

Virus rescue

The YAC containing viral cDNA was cleaved at the unique restriction site located downstream of the 3′ end poly(A) tail. In brief, 1–2 µg of phenol–chloroform-extracted and ethanol-precipitated restricted DNA was resolved in nuclease-free water and used for *in vitro* transcription using the T7 RiboMAX Large Scale RNA production system (Promega) with m7G(5′)ppp(5′)G cap provided as described previously. Additionally, a similar protocol was performed on a PCR product of the *N* gene from corresponding coronaviruses, producing a capped mRNA that encodes the N protein. Then, 1–10 µg of *in vitro* transcribed viral RNA was electroporated together with 2 µg of the *N* gene transcript into BHK-21 cells and/or BHK-21 cells expressing the corresponding coronavirus N protein. Electroporated cells were co-cultured with susceptible Vero E6 cells to rescue rSARS-CoV-2, rSARS-CoV-2-GFP and synSARS-CoV-2-GFP. Progeny viruses that were collected from the supernatant immediately after electroporation were termed passage 0 viruses and were used to produce stocks for subsequent analysis. Virus-infected cells were monitored, and images were acquired using an EVOS fluorescence microscope equipped with a 10× air objective. Brightness and contrast were adjusted using FIJI. Figures were assembled using the FigureJ plugin.

All work involving the rescue and characterization of recombinant SARS-CoV-2 was performed in a biosafety level 3 laboratory at the Institute of Virology and Immunology, Mittelhäusern, Switzerland under appropriate safety measures with respect to personal and environmental protection.

Virus growth kinetics

In brief, 24 h before infection with SARS-CoV-2, Vero E6 cells were seeded in a 24-well plate at a density of 3.6×10^5 cells per ml. Cells were washed once with PBS and inoculated with viruses (multiplicity of infection (MOI) = 0.01). After 2 h, the virus-containing supernatant was

removed, and cells were washed three times with PBS and supplied with medium as described above. Cell-culture supernatants were collected at the indicated time points after infection. Statistical significance was determined by two-sided unpaired Student's *t*-test without adjustments for multiple comparisons.

Plaque assay and $TCID_{50}$

A $TCID_{50}$ assay was performed for SARS-CoV-2 and SARS-CoV-2-GFP in Vero E6 cells. In brief, cells were seeded 24 h before infection in a 96-well plate at a density of 2×10^6 cells per plate. Viruses were serially diluted at 1:10 dilution from 10^{-1} to 10^{-8}. After 72 h of incubation, the medium was removed and cells were fixed and stained with crystal violet. The $TCID_{50}$ ml^{-1} titre was determined using the Spearman–Kaerber method.

The PFU ml^{-1} of SARS-CoV-2 and SARS-CoV-2-GFP was determined by plaque assay using Vero E6 cells in a 6-well format. In brief, 24 h before infection, Vero E6 cells were seeded at a density of 2×10^6 cells per plate. At the time of infection, cells were washed with PBS and inoculated with viruses serially diluted in cell-culture medium at 1:10 dilution. Cells were washed with PBS 1 h after inoculation and overlaid with 2.4% Avicel mixed at 1:1 with 2× DMEM supplemented with 20% fetal bovine serum, 200 units ml^{-1} penicillin and 200 µg ml^{-1} streptomycin. After 48 h of incubation, the overlay was removed and cells were fixed and stained with crystal violet.

Sequencing and computational analysis

Full-length sequences of the SARS-CoV-2 and SARS-CoV-2-GFP cDNAs cloned in yeast were confirmed by Sanger sequencing (Microsynth). All other virus genomes cloned in yeast were confirmed using the Nanopore sequencer MinION from Oxford Nanopore Technologies according to standard protocols. The operating software MinKNOW performed data acquisition and real-time base calling, generating data as fast5 and/or fastq files. Subsequently, the Python command line qcat (Mozilla Public License 2.0., copyright 2018 Oxford Nanopore Technologies, v1.1.0, http://www.github.com/nanoporetech/qcat) was run to demultiplex Nanopore reads from fastq files. Alignment of demultiplexed reads to reference sequences was carried out using the Minimap2 program, producing a fasta file. Mutations of consensus sequences and regions for which the sequences were not clear were verified by Sanger sequencing (Microsynth).

rSARS-CoV-2 and SARS-CoV-2-GFP RNA was sequenced by next-generation sequencing using poly(A)-purified RNA. In brief, 1×10^6 Vero E6 cells were infected with rSARS-CoV-2 clones 1.1, 2.2, 3.1 and rSARS-CoV-2-GFP clones 4.1, 5.2, 6.2 (all passage 1) at an MOI = 0.001. Cellular RNA was prepared using NucleoSpin RNA Plus (Macherey-Nagel) according to the manufacturer's recommendation. The quantity and quality of the extracted RNA was assessed using a Thermo Fisher Scientific Qubit 4.0 fluorometer with the Qubit RNA BR Assay Kit (Thermo Fisher Scientific, Q10211) and an Advanced Analytical Fragment Analyzer System using a Fragment Analyzer RNA Kit (Agilent, DNF-471), respectively. Sequencing libraries were produced using an Illumina TruSeq Stranded mRNA Library Prep kit (Illumina, 20020595) in combination with TruSeq RNA UD Indexes (Illumina, 20022371) according to Illumina's guidelines. Pooled cDNA libraries were paired-end sequenced using an Illumina NovaSeq 6000 S Prime Reagent Kit (300 cycles; Illumina, 20027465) on an Illumina NovaSeq 6000 instrument, generating an average of 69 million reads per sample.

The quality-control assessments, generation of libraries and sequencing run were all performed at the Next Generation Sequencing Platform, University of Bern, Switzerland. For analysis, the adaptor sequences were trimmed using TrimGalore software (v.0.6.5) and reads shorter than 20 nucleotides in length and/or with a Phred score of less than 20 were removed. Paired-end trimmed reads were mapped to the SARS-CoV-2 genome (GenBank accession MT108784; synthetic construct derived from SARS-2 BetaCoV/Wuhan/IVDC-HB-01/2019) using the Spliced Transcripts Alignment to a Reference (STAR) aligner (v.2.7.0a) with default parameters. Before mapping, STAR was also used to generate a genome index for SARS-CoV-2 with the parameters --genomeSAindexNbases 7 and --sjdbOverhang 149. SAMtools (v.1.10) was used to calculate mapped read depth from the resulting mapped read pairs at each position in the genome and subsequently visualized using a variety of software packages in R. Calculations were performed on UBELIX (http://www.id.unibe.ch/hpc), the HPC cluster at the University of Bern. Sequencing data have been deposited in the Sequence Read Archive (SRA) of the NCBI (http://www.ncbi.nlm.nih.gov/sra).

Apart from MinION and next-generation sequencing data handling, other sequence analyses were performed using Geneious Prime v.2019.2.3. Results from virus growth kinetics were analysed and graphically presented using GraphPad Prism v.8.3.0 for Windows. All figures were created with Adobe Illustrator and Biorender.com.

Identification of leader–body junctions of viral mRNAs

To identify reads that mapped discontinuously to the SARS-CoV-2 genome and determine the location of potential transcription regulatory sites (TRS), we pooled reads that mapped to the viral genome as well as unmapped reads and searched for the sequence TTCTCTAA**ACGAAC** (nucleotides 62–75 of MT108784; leader TRS is indicated in bold). We then filtered for reads that had at least 18 nucleotides 3′ of the aforementioned sequence and evaluated whether these reads were compatible with any of the SARS-CoV-2 mRNA sequences. Reads matching these criteria were used as input for the generation of a consensus sequence for each TRS site and analysed using a combination of SAMtools (v.1.10), R and the Integrative Genomics Viewer (IGV). Mapped read depth was also calculated for the discontinuously mapped reads as explained in the previous section.

5′-RACE

Recombinant SARS-CoV-2 and SARS-CoV-2-GFP poly(A)-purified RNA used for next-generation sequencing was also used to determine the genome 5′ ends by 5′-RACE. M-MLV reverse transcription (Promega) was performed according to the manufacturer's instructions using the gene-specific primer pWhSF-ORF1a-R18-655 and 10 U RNase Inhibitor RNasin plus (Promega) per 25 µl reaction volume. Following reverse transcription, 1 µl RNase H (5 U μl^{-1}, New England Biolabs) per 25 µl reaction was added, and the mixture was incubated at 37 °C for 20 min. The cDNA was immediately purified with the High Pure PCR product purification kit (Roche) according to the manufacturer's instructions. A poly(A) tail was added to the cDNA with Terminal Transferase (New England Biolabs) according to the manufacturer's instructions. Subsequently, a PCR reaction with the tailed cDNA was performed with the primer pair pWhSF-ORF1a-R18-655 and TagRACE_dT16 using the HotStarTaq Master Mix (QIAGEN) according to the manufacturer's instructions with a touchdown cycling protocol: 95 °C for 15 min; 15 cycles of 94 °C for 30 s, 65 °C touchdown to 50 °C for 1 min, 72 °C for 1 min; 25 cycles of 94 °C for 30 s, 50 °C for 1 min, 72 °C for 1 min. Subsequently, 1 µl of this reaction was used for a nested re-amplification with the primer pair

pWhSF-5utr-R17-273 and TagRACE in a final volume of 50 µl following the same cycling protocol as described above. The PCR fragment was purified using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel) according to the manufacturer's instructions, and the purified PCR fragment was sent to Microsynth for Sanger sequencing with the primer pWhSF-5utr-R17-273. Sequencing raw data were assessed using the SeqManTM II sequence analysis software (DNASTAR).

Remdesivir experiment

Remdesivir (MedChemExpress) was dissolved in DMSO and stored at −80 °C in 20 mM stock aliquots. One day before the experiment, Vero E6 cells were seeded in 24-well plates at a density of 8×10^4 cells per well. Cells were infected with synSARS-CoV-2-GFP (passage 1) at MOI = 0.01 or mock-infected as control. Innocula were removed at 1 h after infection, and replaced with medium containing remdesivir (0.2 µM or 2 µM) or the equivalent amount of DMSO. At 48 h after infection, cells were washed once with PBS and incubated in fresh PBS. Images were acquired using an EVOS fluorescence microscope equipped with a 10× air objective. Brightness and contrast were adjusted identically for each condition and their corresponding control using FIJI. Figures were assembled using the FigureJ plugin.

Immunofluorescence assay

One day before infection, Vero E6 cells were seeded in a 12-well removable chamber glass slide (Ibidi) at a density of 4×10^4 cells per well. Cells were infected with rSARS-CoV-2 clone 3.1 (passage 2) or mock-infected as control. At 6 and 24 h after infection, cells were washed twice with PBS and fixed with 4% (v/v) neutral-buffered formalin. Cells were washed twice with PBS before permeabilization with 0.1% Triton X-100 and blocking with PBS supplemented with 50 mM NH_4Cl, 0.1% (w/v) saponin and 2% (w/v) BSA (confocal buffer) for 60 min. Primary antibodies (anti-dsRNA, J2, English and Scientific Consulting, 10010500; and anti-SARS-CoV Nucleocapsid (N), Rockland, 200-401-50) and secondary antibodies (donkey anti-rabbit 594, Jackson ImmunoResearch 711-585-152; and donkey anti-mouse 488, Jackson ImmunoResearch 715-545-150) were diluted in confocal buffer. Slides were covered with 0.17-mm thick, high-performance (1.5H) glass coverslips and mounted using ProLong Diamond Antifade mountant containing 4′,6-diamidino-2-phenylindole (DAPI) (Thermo Fisher Scientific). Images were acquired using an EVOS FL Auto 2 Imaging System equipped with a coverslip-correct 40× air objective. Brightness and contrast were adjusted identically for each condition and their corresponding control using FIJI. Figures were assembled using the FigureJ plugin.

Serum neutralization assay

One day before the experiment, Vero E6 cells were seeded in a 96-well clear-bottom, black plate at a density of 2×10^6 cells per well. Serum 2 has been described in another study as patient serum ID7 (convalescent human anti-SARS-CoV-2 serum). Serum 4 has been described previously as patient serum CSS 2 (convalescent human anti-SARS-CoV serum). Sera 1 and 3 were control sera. In brief, all sera were inactivated for 30 min at 56 °C and diluted at 1:10 in OptiMEM. A twofold serial dilution was performed in OptiMEM in a final volume of 50 µl in a separate 96-well plate (dilutions 1:10 to 1:1,280). Then, 50 µl of synSARS-CoV-2-GFP containing 250 $TCID_{50}$ was added to the diluted sera. The serum–virus mixture was incubated at 37 °C for 60 min, and subsequently added to Vero E6 cells. After 1 h of incubation, supernatants were removed and replaced with medium as described above. At 48 h after infection, expression of GFP and cytopathogenic effects were monitored,

and images were acquired using an EVOS fluorescence microscope equipped with a 10× air objective. Brightness and contrast were adjusted identically for each condition and their corresponding control using FIJI. Figures were assembled using the FigureJ plugin.

3. Results

The detection of a new coronavirus in China at the end of 2019 prompted us to test the applicability of our synthetic genomics platform to reconstruct virus based on the genome sequences released on 10–11 January 2020 (Fig. 1). We divided the genome into 12 overlapping DNA fragments. In parallel, we aimed to generate a SARS-CoV-2 clone that expressed GFP, as this could facilitate the screening of antiviral compounds and be used to establish diagnostic assays (for example, virus neutralization assays). This was achieved by dividing fragment 11 into three subfragments and GFP was inserted in-frame of ORF7a, replacing nucleotides 40–282. We noticed that nucleotides 3–5 at the 5′ end of the reported SARS-CoV-2 sequence (5′-AU**UAA**AGG; GenBank MN996528.1; nucleotides that are different are highlighted in bold) differed from SARS-CoV (5′-AU**AUU**AGG; GenBank AY291315) and from the more closely related bat SARS-related CoVs ZXC21 and ZC45 (5′-AU**AUU**AGG). We therefore designed three 5′-end versions, and each version was combined with the remaining SARS-CoV-2 genome (constructs 1–3) or a corresponding SARS-CoV-2-GFP genome (constructs 4–6). Constructs 1 and 4 contained the 5′ end modified by three nucleotides according to the bat SARS-related CoVs (5′-AU**AUU**AGG), constructs 2 and 5 contained the 124 5′-terminal nucleotides of SARS-CoV, and constructs 3 and 6 contained the reported SARS-CoV-2 sequence (5′-AU**UAA**AGG; according to MN996528.1). Notably, differences between SARS-CoV-2 and SARS-CoV within the 5′-terminal 124 nucleotides are in agreement with the predicted RNA secondary structures.

Fig. 1: Timeline of the reconstruction and recovery of rSARS-CoV-2 in relation to key events of the COVID-19 pandemic.

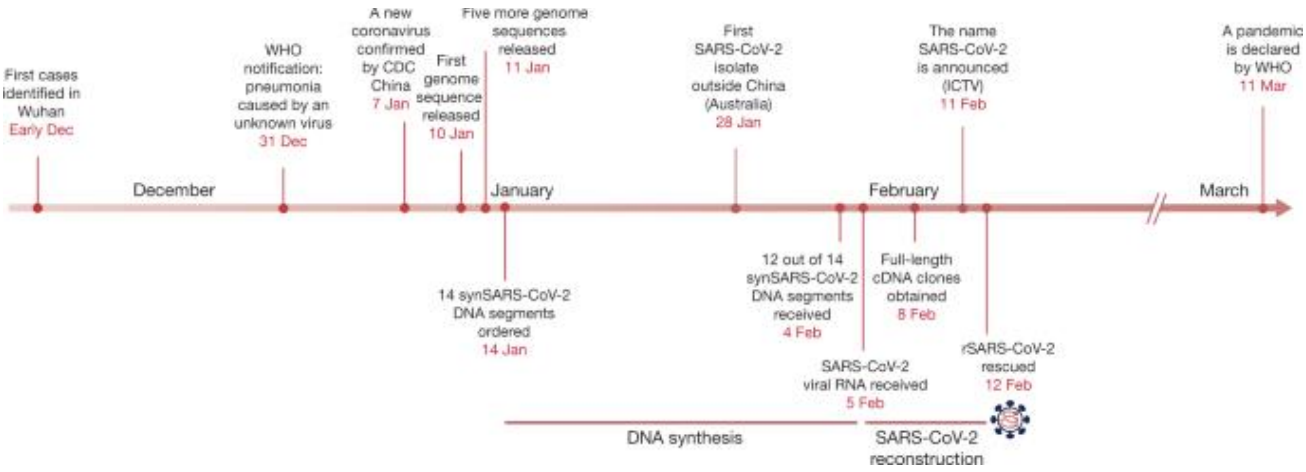
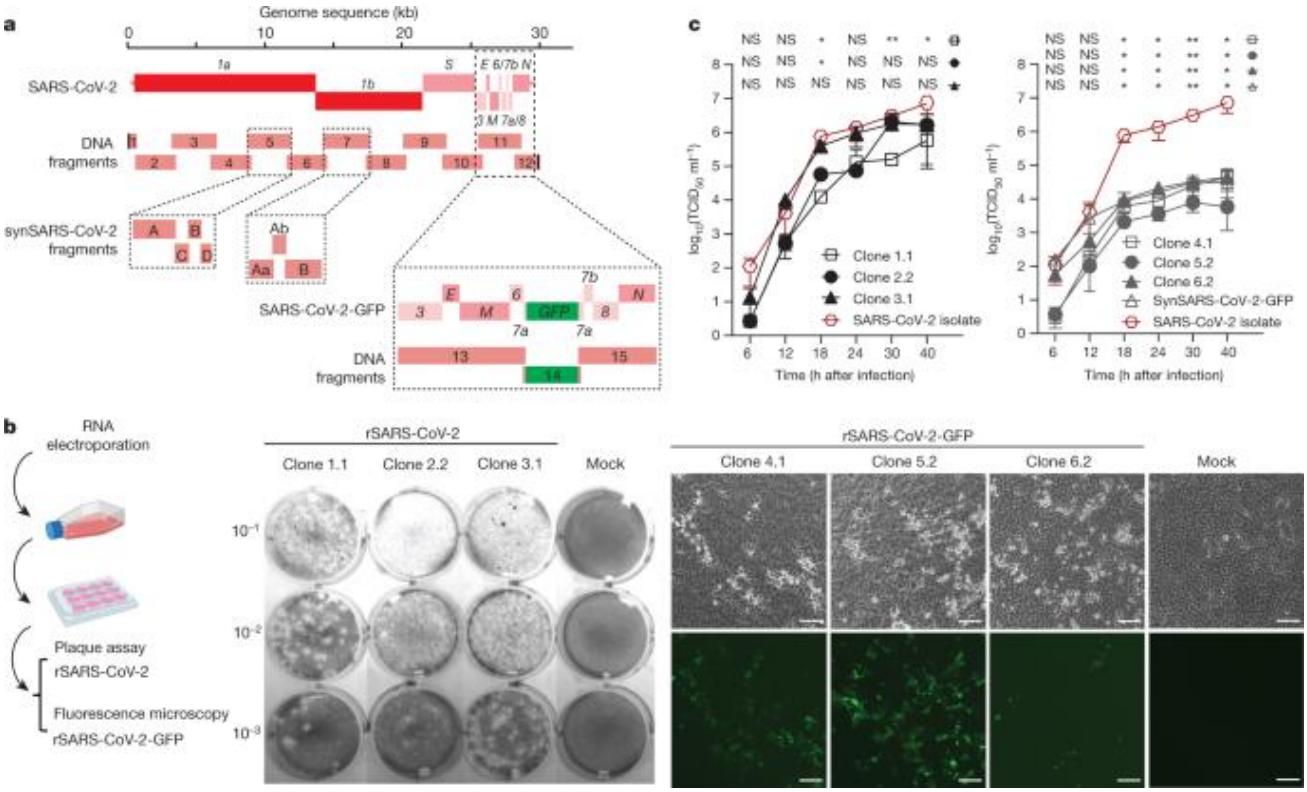


Illustration of the rapidity of rSARS-CoV-2 reconstruction along with the timeline of key events of the COVID-19 pandemic. CDC, Center for Disease Control and Prevention; ICTV, International Committee on Taxonomy of Viruses; WHO, World Health Organization.

Fig. 2: Reconstruction, rescue and characterization of rSARS-CoV-2, rSARS-CoV-2-GFP and synSARS-CoV-2-GFP.



a, Schematic representation of the SARS-CoV-2 genome organization and DNA fragments used to clone rSARS-CoV-2, rSARS-CoV-2-GFP and synSARS-CoV-2-GFP. Inserts show synthetic subfragments comprising fragments 5 (A–D) and 7 (Aa, Ab, B), and the fragments used to insert the *GFP* gene (fragments 13–15). **b**, Left, schematic of the experiment. Middle, rescue of rSARS-CoV-2 from yeast clones 1.1, 2.2 and 3.1. Supernatants (10^{-1}, 10^{-2} and 10^{-3} ml) of cells infected with the indicated clones or mock-infected cells were transferred to Vero E6 cells to detect plaques (rSARS-CoV-2). Right, rescue of rSARS-CoV-2-GFP from yeast clones 4.1, 5.2 and 6.2. Supernatants (1 ml) from individual rescue experiments were transferred to Vero E6 cells to detect green fluorescence (rSARS-CoV-2-GFP). Mock, uninfected cells. Scale bars, 100 µm. **c**, Replication kinetics of rSARS-CoV-2 clones 1.1, 2.2, 3.1 (left) and rSARS-CoV-2-GFP clones 4.1, 5.2, 6.2 and synSARS-CoV-2-GFP (right) compared with the SARS-CoV-2 isolate. Vero E6 cells were infected (MOI = 0.01), and supernatants were collected at the indicated time points after infection and titrated (50% tissue culture infectious dose ($TCID_{50}$) assay). Data represent the mean ± s.d. of three independent biological replicates. Statistical significance was determined for each clone against the SARS-CoV-2 isolate by two-sided unpaired Student's *t*-test without adjustments for multiple comparisons. *P* values (from left to right): left, top, NS, $P = 0.0851$; NS, $P = 0.1775$; *$P = 0.0107$; NS, $P = 0.0648$; **$P = 0.0013$; *$P = 0.0373$; middle, NS, $P = 0.0851$; NS, $P = 0.1713$; *$P = 0.0133$; NS, $P = 0.0535$; NS, $P = 0.0909$; NS, $P = 0.0632$; bottom, NS, $P = 0.1119$; NS, $P = 0.1641$; NS, $P = 0.0994$; NS, $P = 0.4921$; NS, $P = 0.3336$; NS, $P = 0.0790$; right, top, NS, $P = 0.0858$; NS, $P = 0.1429$; *$P = 0.0104$; *$P = 0.0466$; **$P = 0.0011$; *$P = 0.0287$; second, NS, $P = 0.0872$; NS, $P = 1360$; *$P = 0.0102$; *$P = 0.0461$; **$P = 0.0011$; *$P = 0.0282$; third, NS, $P = 0.4810$; NS, $P = 0.1758$; *$P = 0.0106$; *$P = 0.0478$; **$P = 0.0011$; *$P = 0.0287$; bottom, NS, $P = 0.3739$; NS, $P = 0.6817$; *$P = 0.0106$; *$P = 0.0473$; **$P = 0.0011$ *$P = 0.0285$.

Fourteen synthetic DNA fragments were ordered as sequence-confirmed plasmids and all but fragments 5 and 7 were delivered. As we received SARS-CoV-2 viral RNA from an isolate

of a Munich patient (BetaCoV/Germany/BavPat1/2020) at the same time, we amplified the regions of fragments 5 and 7 by RT–PCR. TAR cloning was immediately initiated, and for all six SARS-CoV-2 and SARS-CoV-2-GFP constructs we obtained correctly assembled molecular clones. Because sequence verification was not possible within this short time frame, we randomly selected two clones for each construct, isolated the YAC DNA and performed *in vitro* transcription. The resulting RNAs were electroporated together with an mRNA that encodes the SARS-CoV-2 N protein into BHK-21 and, in parallel, into BHK-SARS-N cells that expressed the SARS-CoV N protein. Electroporated cells were seeded on Vero E6 cells and two days later we observed green fluorescent signals in cells that received the GFP-encoding SARS-CoV-2 RNAs. Indeed, we could rescue infectious viruses for almost all rSARS-CoV-2 and rSARS-CoV-2-GFP clones. As shown in Fig. 2, for rSARS-CoV-2 clones 1.1, 2.2, and 3.1, plaques were readily detectable, demonstrating that infectious virus has been recovered irrespectively of the 5′-terminal sequences. Sequencing of the YACs and corresponding rescued viruses revealed that almost all DNA clones and viruses contained the correct sequence, except for some individual clones that contained mutations within fragments 5 and 7 that were probably introduced by RT–PCR. Nevertheless, we obtained at least one correct YAC clone for all constructs except for construct 6. To correct this, we reassembled construct 6 by replacing the RT–PCR-generated fragments 5 and 7 with four and three shorter synthetic double-stranded (ds)DNA fragments, respectively. The resulting molecular clone was used to rescue the synthetic SARS-CoV-2-GFP (synSARS-CoV-2-GFP) virus without any mutations exclusively from chemically synthesized DNA.

Next we assessed the 5′ end of the recombinant viruses and the Munich virus isolate and confirmed the published 5′ end sequence of SARS-CoV-2 (5′-AU**UAA**AGG; GenBank MN996528.3). Full-length sequencing of the viral genomes and 5′ rapid amplification of cDNA end (5′-RACE) analysis of the recombinant viruses confirmed the identity of each virus, and showed that the 5′ end variant of each virus retained the cloned 5′ terminus. This demonstrates that the 5′ ends of SARS-CoV and bat SARS-related CoVs ZXC21 and ZC45 are compatible with the replication machinery of SARS-CoV-2. Sequencing results also revealed the identity of leader–body junctions of SARS-CoV-2 subgenomic mRNAs, which are identical to those of SARS-CoV. We also analysed rSARS-CoV-2 clone 3.1 for protein expression and demonstrated the presence of the SARS-CoV-2 nucleocapsid protein in dsRNA-positive cells. The replication kinetics of rSARS-CoV-2 clone 3.1, which contains the authentic 5′ terminus, was indistinguishable from replication of the SARS-CoV-2 isolate, while clones 1.1 and 2.2 showed slightly reduced replication (Fig. 2, left). All rSARS-CoV-GFP clones and synSARS-CoV-GFP displayed similar growth kinetics but they were significantly reduced compared with the SARS-CoV-2 isolate, suggesting that the insertion of GFP and/or the partial deletion of ORF7a affects replication. Despite the reduced replication, green fluorescence was readily detectable and we demonstrated the use of the synSARS-CoV-GFP clone for antiviral drug screening by testing remdesivir, a promising compound for the treatment of COVID-19. Similarly, the simple readout of green fluorescence greatly facilitates the demonstration of virus neutralization with human serum.

4. Discussion

Our results demonstrate the full functionality of the SARS-CoV-2 reverse-genetics system and we expect that this fast, robust and versatile synthetic genomics platform will provide new insights into the molecular biology and pathogenesis of a number of emerging RNA viruses. Although homologous recombination in yeast has already been used for the

generation of a number of molecular virus clones in the past, we present a thorough evaluation of the feasibility of this approach to rapidly generate full-length cDNAs for large RNA viruses that have a known history of instability in *E. coli*. We show that one main advantage of the TAR cloning system is that the viral genomes can be fragmented to at least 19 overlapping fragments and reassembled with remarkable efficacy. This facilitated the cloning and rescue of rSARS-CoV-2 and rSARS-CoV-2-GFP within one week. It should be noted that we see considerable potential to reduce the time of DNA synthesis. Currently, synthetic DNA fragments get routinely cloned in *E. coli*, which turned out to be problematic for SARS-CoV-2 fragments 5 and 7. We, however, used shorter synthetic dsDNA parts to assemble these fragments by TAR cloning and to generate the molecular clone synSARS-CoV-2-GFP by using exclusively chemically synthesized DNA, which is an additional proof of the superior cloning efficiency of yeast- versus *E. coli*-based systems.

5. Conclusion

The COVID-19 pandemic emphasizes the need for preparedness to rapidly respond to emerging virus threats. The rapidity of our synthetic genomics approach to generate SARS-CoV-2 and the applicability to other emerging RNA viruses make this system an attractive alternative to provide infectious virus samples to health authorities and diagnostic laboratories without the need of having access to clinical samples. As the COVID-19 pandemic is ongoing, we expect to see sequence variations and possibly phenotypic changes of the evolving SARS-CoV-2 virus in the human host. With this synthetic genomics platform, it is now possible to rapidly introduce such sequence variations into the infectious clone and to functionally characterize SARS-CoV-2 evolution in real time.

Outlook on generation and testing of mouse adapted (MA) SARS-CoV-2 clones

Based on published sequence data from experimentally mouse-adapted virus mutants following *in vivo* passaging SARS-CoV and SARS-CoV-2 in mice, we have rescued 3 independent types MA mutant clones. Preliminary data using males and females of various inbred and outbred mouse strain backgrounds with the first generated MA SARS-CoV-2 clone have been initiated. The data indicate successful virus replication in the respiratory tract. Additional work is ongoing to identify the most appropriate virus/mouse combination for future compound testing in the context of this project.

Repository for primary data

A large proportion of the work towards this deliverable was undertaken prior to the kickoff of the CARE IMI2 project and has been published in the following publication:-

Thi Nhu Thao, T., Labroussaa, F., Ebert, N. *et al.* Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* **582,** 561–565 (2020). https://doi.org/10.1038/s41586-020-2294-9