

~~SECRET//NOFORN~~

# SARS-COV-2 GENOME ANALYSIS

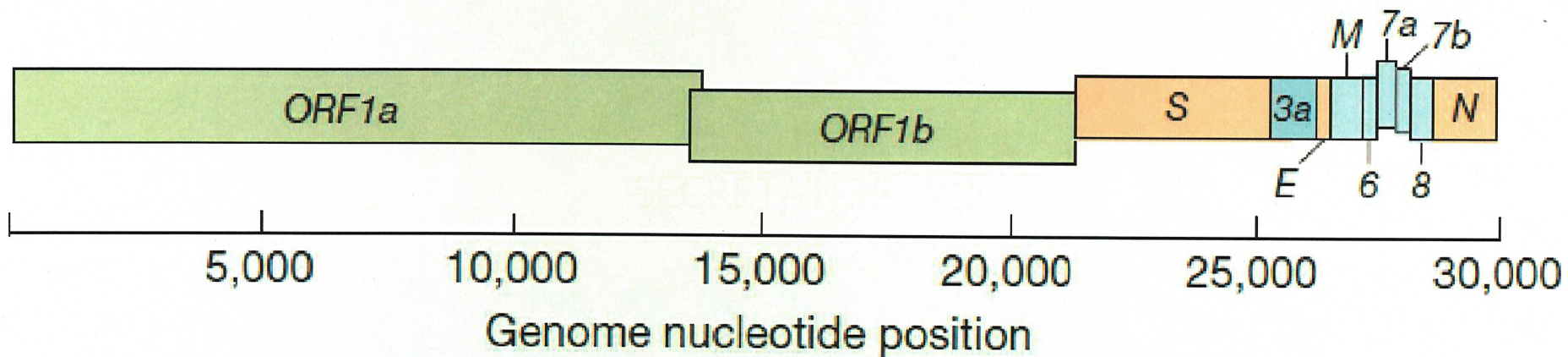
(b)(3);10 USC 424; (b)(6)

June 25, 2020

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# SARS-CoV-2 Genome



The next two pages are DIF citing (b)(1) and (b)(3) and are not provided.

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# Methods to construct a Coronavirus Full-Length Clone

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

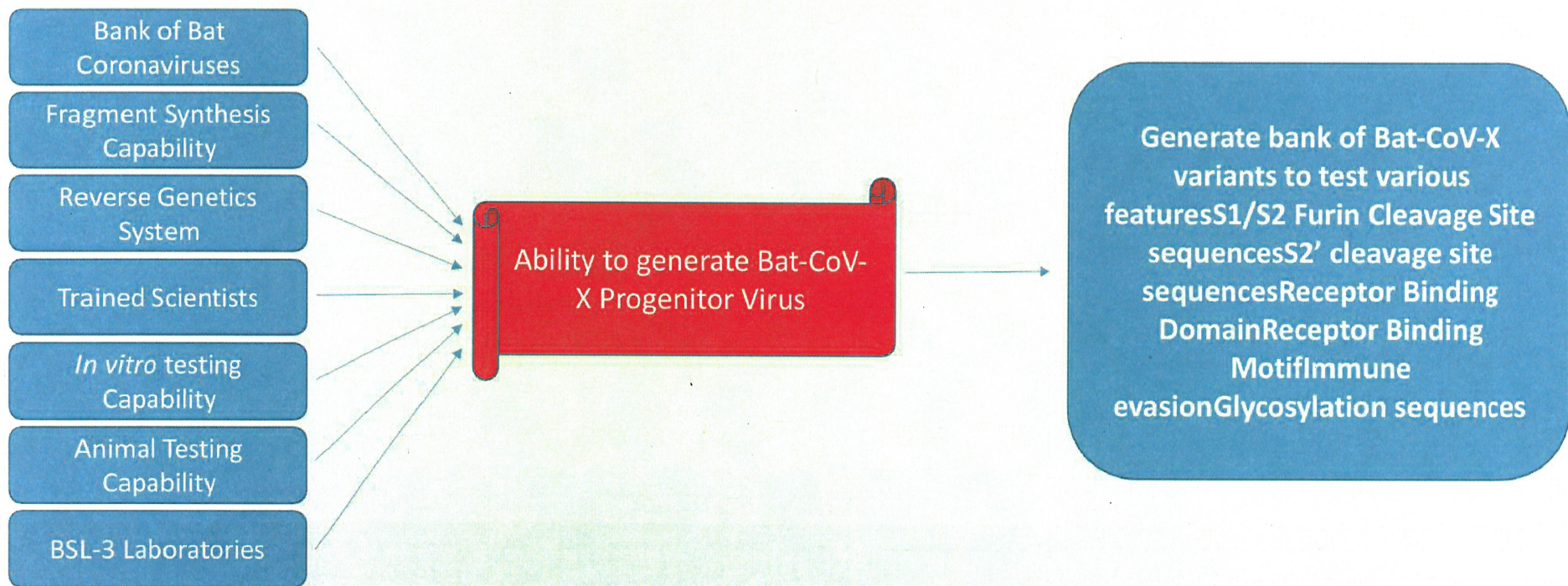
# CORONAVIRUS INFECTIOUS CLONE CONSTRUCTION

- 1) Synthesize or PCR amplify 6-10 segments of a Bat Coronavirus  
Build a 5' transcription initiation fragment  
"stitch" the fragments together using an infectious clone technology  
3' to the transcription initiation fragment  
Restriction-enzyme-based fragment cloning system  
Overlapping Fragment system  
Guided RNA Recombination  
Clone in a suitable host (E. coli, yeast, etc.)  
Sequence verify cloned insert

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# CAPABILITIES NEEDED TO CONSTRUCT A BAT-LIKE CORONAVIRUS INFECTIOUS CLONE



The next page is DIF citing (b)(1) and (b)(3) and is not provided.

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# Type IIS Restriction Enzymes and Golden Gate Assembly System

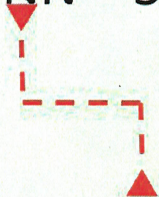
~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

## Type IIS Restriction Enzymes

- Non palindromic recognition siteCuts at sites outside of recognition siteEach digested location has unique nucleotide overhangsExampleBsal

5' – GGTCTC>NNNNN – 3'  
3' – CCAGAG>NNNNN – 5'



~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

## Eleven Tips For Optimizing Your Golden Gate Assembly Reactions

Looking for a reliable, versatile DNA fragment in a single reaction? Here are some tips to keep in mind when planning your Golden Gate Assembly experiment.

- 1 Check your sequences**  
Always check your assembly sequences for the correct orientation, including check for 5' to 3' orientation, and make sure to use the year assembly. When using older Golden Gate assembly kits, check the assembly kit manual for the correct primer or the primer sequence. Always use the correct primer with the correct orientation. Verify the correct orientation of the primer sequence with the correct primer.
- 2 Clean your primers**  
When designing a PCR primer or sequence, always check for any secondary structure, such as hairpins, or any other structure that may interfere with the primer's ability to bind to the target DNA sequence. Always check for any secondary structure that may interfere with the primer's ability to bind to the target DNA sequence.
- 3 Choose the right primer**  
Consider choosing the primer sequence that is closest to the Golden Gate assembly. The correct primer sequence is included in the Golden Gate assembly kit manual and the Golden Gate assembly kit manual. Always check for any secondary structure that may interfere with the primer's ability to bind to the target DNA sequence.
- 4 Choose the right buffer**  
Not every Golden Gate assembly kit uses the same buffer. Always use the correct buffer with the correct Golden Gate assembly kit. Always use the correct buffer with the correct Golden Gate assembly kit.
- 5 Increase your complex assembly efficiency by increasing the Golden Gate cycling time**  
Always use the correct Golden Gate assembly kit. Always use the correct Golden Gate assembly kit.
- 6 Make sure your plasmid prep is DNA clean**  
Always use the correct Golden Gate assembly kit. Always use the correct Golden Gate assembly kit.
- 7 Avoid primer dimers**  
Always use the correct Golden Gate assembly kit. Always use the correct Golden Gate assembly kit.
- 8 Avoid PCR-induced errors**  
Always use the correct Golden Gate assembly kit. Always use the correct Golden Gate assembly kit.

- 9 Decrease insert amount for complex assemblies**  
Always use the correct Golden Gate assembly kit. Always use the correct Golden Gate assembly kit.
- 10 Carefully design EVERY insert's overhang**  
Always use the correct Golden Gate assembly kit. Always use the correct Golden Gate assembly kit.
- 11 Check for a sequence error that your assembly becomes non-functional**  
Always use the correct Golden Gate assembly kit. Always use the correct Golden Gate assembly kit.

**USA**  
New England Biolabs, Inc.  
Telephone: (978) 977-1034  
Fax: (978) 977-1035  
E-mail: [info@neb.com](mailto:info@neb.com)  
[www.neb.com](http://www.neb.com)

**Canada**  
New England Biolabs, Inc.  
Telephone: (514) 375-1034  
[www.neb.com](http://www.neb.com)

**China**  
New England Biolabs (Shanghai) Ltd.  
Telephone: (86) 21 5012 1034  
[www.neb.com](http://www.neb.com)

**France**  
New England Biolabs (France)  
Telephone: (33) 1 69 90 1034  
[www.neb.com](http://www.neb.com)

**Germany & Austria**  
New England Biolabs GmbH  
Telephone: (49) 30 266 1034  
[www.neb.com](http://www.neb.com)

**Japan**  
New England Biolabs (Japan) Ltd.  
Telephone: (81) 3 556 1034  
[www.neb.com](http://www.neb.com)

**Spain**  
New England Biolabs (Spain) S.L.  
Telephone: (34) 91 556 1034  
[www.neb.com](http://www.neb.com)

**United Kingdom**  
New England Biolabs (UK) Ltd.  
Telephone: (44) 1203 51034  
[www.neb.com](http://www.neb.com)

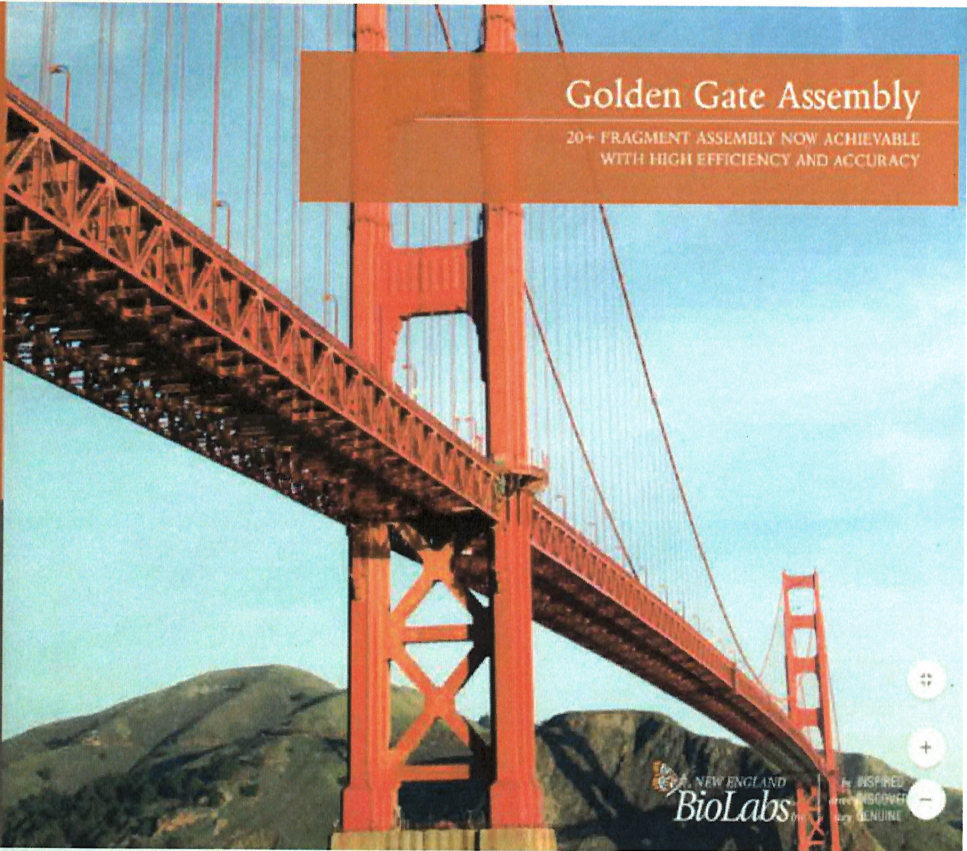
[www.neb.com](http://www.neb.com)

Mixed Sources  
FSC  
www.fsc.org

NEB is a 50% employee-owned company. We are committed to providing a safe and healthy work environment for all our employees. We are also committed to environmental sustainability and to using recycled materials in our products.

### Featured Tools

- Download the **NEB App** for iPhone® or iPad®. Scan the augmented reality barcode to find vectors, tutorials and interactive experiments.
- For help designing primers, try the new **NEB Golden Gate Assembly Tool** at [www.neb.com/gga](http://www.neb.com/gga)
- Try our **Ligase Fidelity** tools for the design of high-fidelity Golden Gate assemblies at [www.neb.com/research/ligase-fidelity](http://www.neb.com/research/ligase-fidelity)
  - Ligase Fidelity Viewer™ (v2) - Visualize overhang ligation preferences
  - GenSet™ - Predict high-fidelity junction sets
  - SplitSet™ - Split DNA sequences for seamless high-fidelity assembly



## Golden Gate Assembly

20+ FRAGMENT ASSEMBLY NOW ACHIEVABLE WITH HIGH EFFICIENCY AND ACCURACY

NEW ENGLAND  
**BioLabs**  
INSPIRED  
BY NATURE  
GENUINE

~~SECRET//NOFORN~~



# Push the limits of Golden Gate Assembly

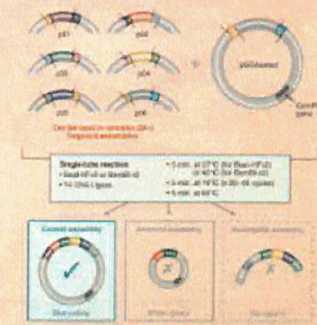
Golden Gate Assembly (1,2) allows for the efficient and seamless assembly of DNA fragments using activities of Type IIS restriction enzymes and T4 DNA Ligase.

With constant advances in both the development of new enzymes and research on maximizing enzyme functionality (e.g., ligase fidelity), NEB is the industry leader in pushing the limits of Golden Gate Assembly and related methods.

## Advantages:

- Clone seamlessly, with no scars remaining after assembly
- Perform single insert cloning in just 5 minutes using our Fast protocols
- Generate libraries with high efficiencies
- Assemble multiple fragments (2-20+) in one, or a single reaction
- Experience high efficiency, even with regions of high GC content and areas of repeats
- Use with a broad range of fragment sizes (<100 bp to >15 kb)

### GOLDEN GATE ASSEMBLY WORKFLOW FOR BOTH SIMPLE AND COMPLEX ASSEMBLIES



In 16 minutes, Golden Gate Assembly requires a Type IIS restriction enzyme, either BsaI or BsmBI, and T4 DNA Ligase. After assembly, there are no scars and no need for a separate ligation step. The assembly is complete in 16 minutes.

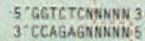
### FEATURED PRODUCTS:

#### Type IIS Restriction Enzymes used in Golden Gate Assembly

Type IIS restriction enzymes recognize asymmetric DNA sequences and cleave outside of their recognition sequence. Type IIS enzymes commonly used in Golden Gate Assembly are listed below. NEB currently offers over 41 Type IIS restriction enzymes. Please visit [www.neb.com](http://www.neb.com) for comprehensive table.

PRODUCT	NEB #	SEQUENCE	SIZE
BsaI	R01951L	GAAGAC (2/6)	307,300 units
BsmBI	R01951L	GAAGAC (2/6)	307,300 units
BsaI	R01951L	GGTCTC (1/5)	300,000 units
BsmBI-v2	R07301L	GGTCTC (1/5)	300,000 units
BsmBI-v2	R07301L	CGTCTC (1/5)	300,000 units
BsaI	R07340L	CGTCTC (1/5)	300,000 units
BsmI	R07355L	CGGATC (3/4)	300,000 units
BsaI	R04461L	GGTCTTC (2/4)	300,000 units
BsmI	R07315L	GGTCTTC (1/4)	300,000 units

### BsmI-HF v2



### What users are saying:

"NEB has provided a reliable set of enzymes and kits that make Golden Gate Assembly easy to use, especially for novices. We have found the Ligase Fidelity assay particularly useful for ensuring assembly with the accuracy required by a precision genetic DNA synthesis. The detailed experimental notes of the kits are the proof of the reliability of the underlying data provided."

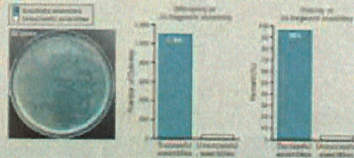
Dr. Glenn Payne  
Genentech, San Francisco, CA

### FEATURED KITS:

#### NEB® Golden Gate Assembly Kits (BsmBI-v2 or BsaI-HF v2)

The absence of internal sites in a sequence determines the choice of which Type IIS restriction enzyme to drive the assembly for your convenience. NEB now offers two kits for Golden Gate Assembly featuring BsmBI-v2 or BsaI-HF v2. Both kits incorporate digestion followed by ligation with T4 DNA Ligase into a single reaction, and can be used to assemble 2-20+ fragments in a single step.

#### COMPLEX GOLDEN GATE ASSEMBLY WITH >95% FIDELITY AND UNPRECEDENTED EFFICIENCY

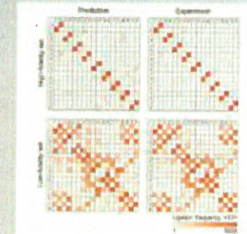


Using NEB Golden Gate Assembly kits (BsmBI-v2 or BsaI-HF v2), assembly of 20 fragments in a single step is possible using the recommended protocol. The efficiency of the assembly is >95% and the fidelity is >95%. The efficiency of the assembly is >95% and the fidelity is >95%. The efficiency of the assembly is >95% and the fidelity is >95%.

### Advances in Ligase Fidelity

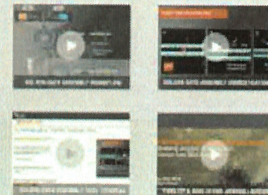
Research at NEB has led to an increased understanding of ligase fidelity, including the development of a comprehensive method for profiling and joining ligase fidelity in order to predict overhang sets with enhanced fidelity (3). This research has enabled complex fragment assemblies with high efficiency and >90% accuracy. More information can be found in the NEB publication, *Comprehensive Profiling of Four EcoRI Overhang Ligase Fidelity by T4 DNA Ligase and Application to DNA Assembly (1)* or in our website, *Fidelity and Bias in High-Fidelity Assemblies*, including complex, multi-fragment Golden Gate DNA Assembly. We also encourage you to try our first Ligase Fidelity Tools.

All of these are available at [www.neb.com/GoldenGate](http://www.neb.com/GoldenGate).



Product areas observed highest fidelity in Golden Gate assembly of 10 high fidelity and 10 low fidelity fragments. The fidelity of the low fidelity products is the lowest of products if they products observed in a single reaction (200) sequencing assembly. Assembly of 10 high fidelity products. Product fidelity of products are based on the ability to assemble 20 fragments in a single reaction. Product fidelity of products are based on the ability to assemble 20 fragments in a single reaction. Product fidelity of products are based on the ability to assemble 20 fragments in a single reaction.

Visit [www.neb.com/GoldenGate](http://www.neb.com/GoldenGate) to learn more and view related videos



### Expanded "assembly standards" for MoClo, GoldenBraid 2.0 and Other Modular Golden Gate Assembly Methods

MoClo (and GoldenBraid 2.0) uses 3 levels of successive assembly. The community has agreed upon a set of common standard overhangs for each level. Utilizing gathered ligase fidelity information, NEB has expanded each level of assembly overhang without sacrificing fidelity. The expanded sets are:

#### Expanded MoClo Standard Assembly Overhangs\*

- Level 0 (Basic primers)
  - ACAT, TTGT, ACTG, GCTA, CCCA, AATA, ATTC, GTGA, CGCC, AAGA, AARE, AACG, CTGC, GACC, AAGA, AARC, AACG, CTGC, GACC, CTA, ATCC, TACA, GGA, CAA, AGAG (95% fidelity)
- Level 1 (Transcription start)
  - GGAG, TACT, GCAT, AATG, AGGT, TTCC, GCTT, GATA, CGCT, GAAA, TCAA, ATAA, GGA, CGCT, GAAA, AACA, AAT, GCAC, CTTA, TCCA (92% fidelity)
- Level 2 (Multiple contexts)
  - TGCC, GACA, ACTA, TTAC, CABA, TGTG, GAGC, GGA, GATA, CTTC, ATCC, ATAG, GCAT, AAT, ACCG, AAAA, AGAG, AGG, TAAA, ATGA (95% fidelity)

\* These standard overhangs are part of the MoClo Toolkit released on the page at [www.neb.com](http://www.neb.com). They are the standard overhangs at MoClo Toolkit. They are the standard overhangs at MoClo Toolkit. They are the standard overhangs at MoClo Toolkit.

References  
1. Gupta, C., Boudon, E., and Malhotra, A. (2010) *Proc Natl Acad Sci U S A* 107: 11841-11846.  
2. Gupta, C., et al. (2010) *Nature* 463: 473-478.  
3. Payne, G. et al. (2010) *PLoS One* 5: e12000.

The next page is DIF citing (b)(1) and (b)(3) and is not provided.

# WIV Bat Coronavirus Collection Efforts

- WIV possesses a large bank of Bat Coronaviruses isolated from various bat species in Yunnan Province China Ge et al., 2013 Yang et al., 2016 Hu et al., 2017 Five-year longitudinal study to isolate Bat Coronaviruses (April 2011 – October 2015) Only a few sequences have been published

Table 1. Summary of SARS-CoV detection in bats from a single habitat in Kunming, Yunnan.

Sampling time	Sample type	Sample Numbers			SARS-CoV + bat species (No.)
		Total	CoV +	SARS-CoV +	
April, 2011	anal swab	14	1	1	<i>R. sinicus</i> (1)
October, 2011	anal swab	8	3	3	<i>R. sinicus</i> (3)
May, 2012	anal swab & feces	54	9	4	<i>R. sinicus</i> (4)
September, 2012	feces	39	20	19	<i>R. sinicus</i> (16) <i>R. formosiquinum</i> (3)
April, 2013	feces	52	21	16	<i>R. sinicus</i> (16)
July, 2013	anal swab & feces	115	9	8	<i>R. sinicus</i> (8)
May, 2014	feces	131	8	4	<i>A. stoliczkanus</i> (3) <i>R. affinis</i> (1)
October, 2014	anal swab	19	4	4	<i>R. sinicus</i> (4)
May, 2015	feces	145	3	0	
October, 2015	anal swab	25	6	5	<i>R. sinicus</i> (5)
<b>Total</b>		<b>602</b>	<b>84</b>	<b>64</b>	<b>R (61) A (3)</b>

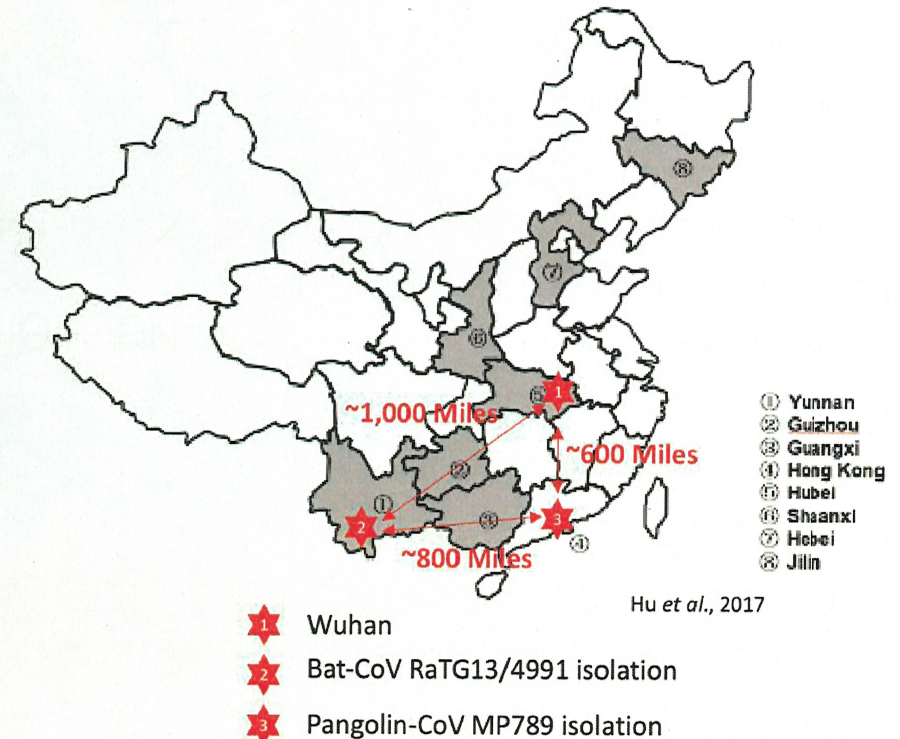


The next page is DIF citing (b)(1) and (b)(3) and is not provided.

~~SECRET//NOFORN~~

## Isolation Locations for RaTG13 and MP789

- RaTG13/4991 isolated from a cave in Yunnan Province  
RaTG13/4991 is a RdRp lineage 1 BetaCoV  
MP789 was isolated from diseased Pangolins in Guangdong Province  
MP789 is a RdRp lineage 2 BetaCoV  
~800 miles separate these two locations  
WIV also collected CoV's from Guangdong and may have a MP789-related virus in their bank  
"All the genomic constituents of SARS-CoV including the hypervariable regions S and ORF8 were discovered from different bat SARSr-CoVs in the same cave in Yunnan, with evidence of recombination events detected between these bat SARSr-CoVs..." (Yu et al., 2019)  
Question: How would a Pangolin RBD from 800 miles away in Guangdong Province recombine into a BatCoV in Yunnan Province?



~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# Postulated WIV Bat-CoV-X Full-length Clone Construction Process

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

## Quote from Zeng *et al.*, 2016

- From Materials and Methods, Virus and cells section  
“All experiments using live virus was conducted under biosafety level 2 (BSL2) conditions.”  
From the first paragraph of the Discussion  
“In this study, we have developed a fast and cost-effective method for reverse genetics of coronaviruses by combining two approaches developed by others (29, 30). Our method allows the genomes of coronaviruses to be split into multiple fragments and inserted into a BAC plasmid with a single step. Recombinant viruses can then be efficiently rescued by direct transfection of the BAC construct. As the genomes can be divided into multiple short fragments, mutations can be introduced into individual fragments easily (31). Using this method, we successfully rescued three recombinant viruses derived from SL-CoV WIV1 (rWIV1, rWIV1-DX, and rWIV1-GFP-DX).”

The next page is DIF citing (b)(1) and (b)(3) and is not provided.

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# WIV SARS-CoV Reverse Genetics System

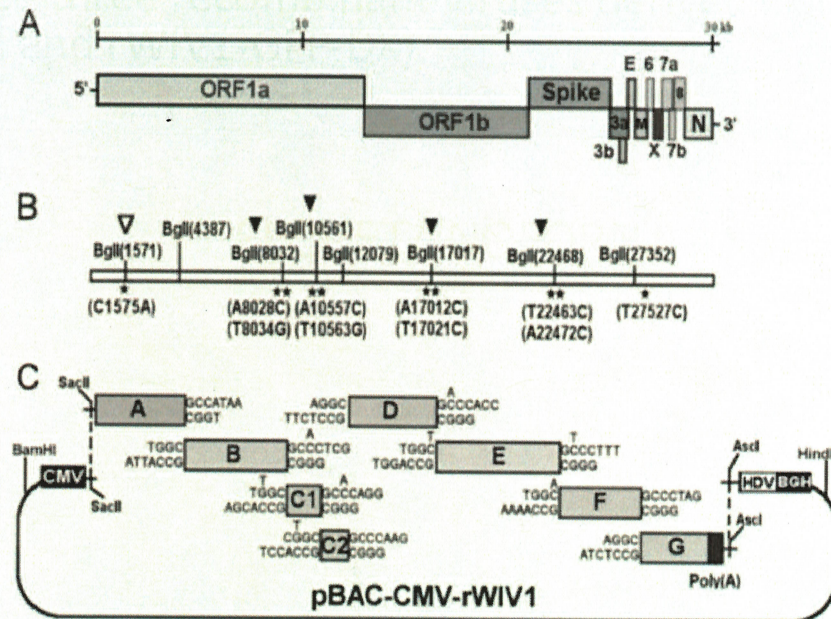
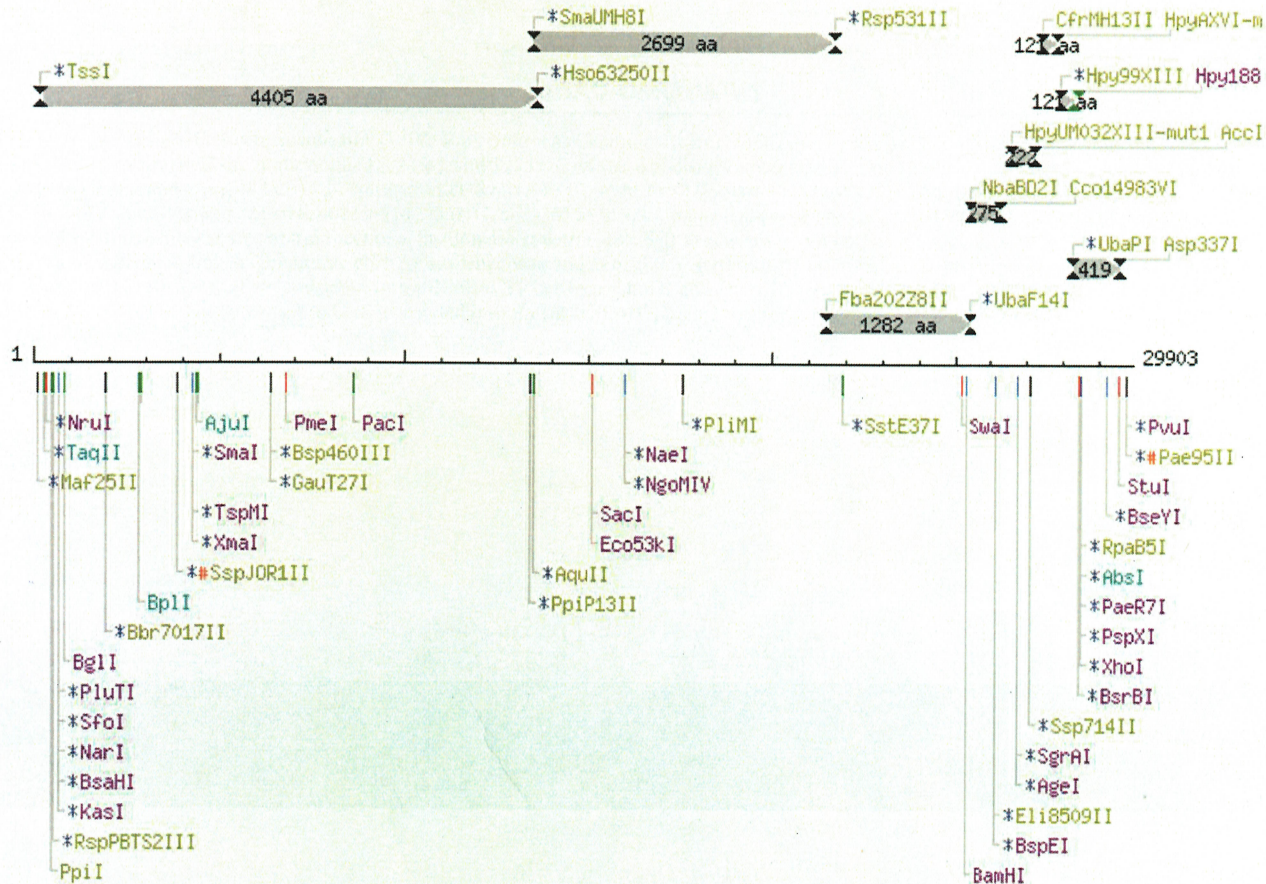


FIG 1 Strategy for construction of an infectious WIV1 BAC clone. (A) Genomic structure of WIV1. (B) The mutations are indicated under the stars. C1575A was used to ablate a natural BglI site at nucleotide 1571 (▽), and T27527C was used to disrupt a potential T7 stop site. The others were for introducing BglI sites (▼). (C) The WIV1 genome was split into eight contiguous cDNAs (A to G): A, nt 1 to 4387; B, nt 4388 to 8032; C1, nt 8033 to 10561; C2, nt 10562 to 12079; D, nt 12080 to 17017; E, nt 17018 to 22468; F, nt 22469 to 27352; G, nt 27353 to 30309. Unique BglI sites were introduced into the fragments by synonymous mutations to make these fragments capable of unidirectional ligation along with native BglI sites in the genome. The original nucleotides are shown above the flanking sequences of corresponding fragments. A poly(A) sequence was added to the 3' terminus of fragment G. A CMV promoter, HDV ribozyme, and BGH transcriptional terminal signal were inserted into pBeloBAC11 between BamHI and HindIII sites. SacII and Ascl sites were introduced between the CMV promoter and ribozyme. Fragments A to G were inserted into the pBAC-CMV plasmid in a single step.

~~SECRET//NOFORN~~

# SARS-COV-2 GENOME RESTRICTION MAP



# Type IIS Restriction Enzymes

- BsmBI (Plus strand)

5' - CGTCTCNNNNN-3'

3' - GCAGAGNNNNN - 5'

5' - GGTCTCNNNNN-3'

3' - CCAGAGNNNNN - 5'

have any SacII or AscI restriction sites

- BsmBI (Minus strand)

5' - NNNNNGAGACG-3'

3' - NNNNNCTCTGC-

5' BsaI (Plus strand)

5' BsaI (Minus strand)

5' NNNNNGAGACG-3'

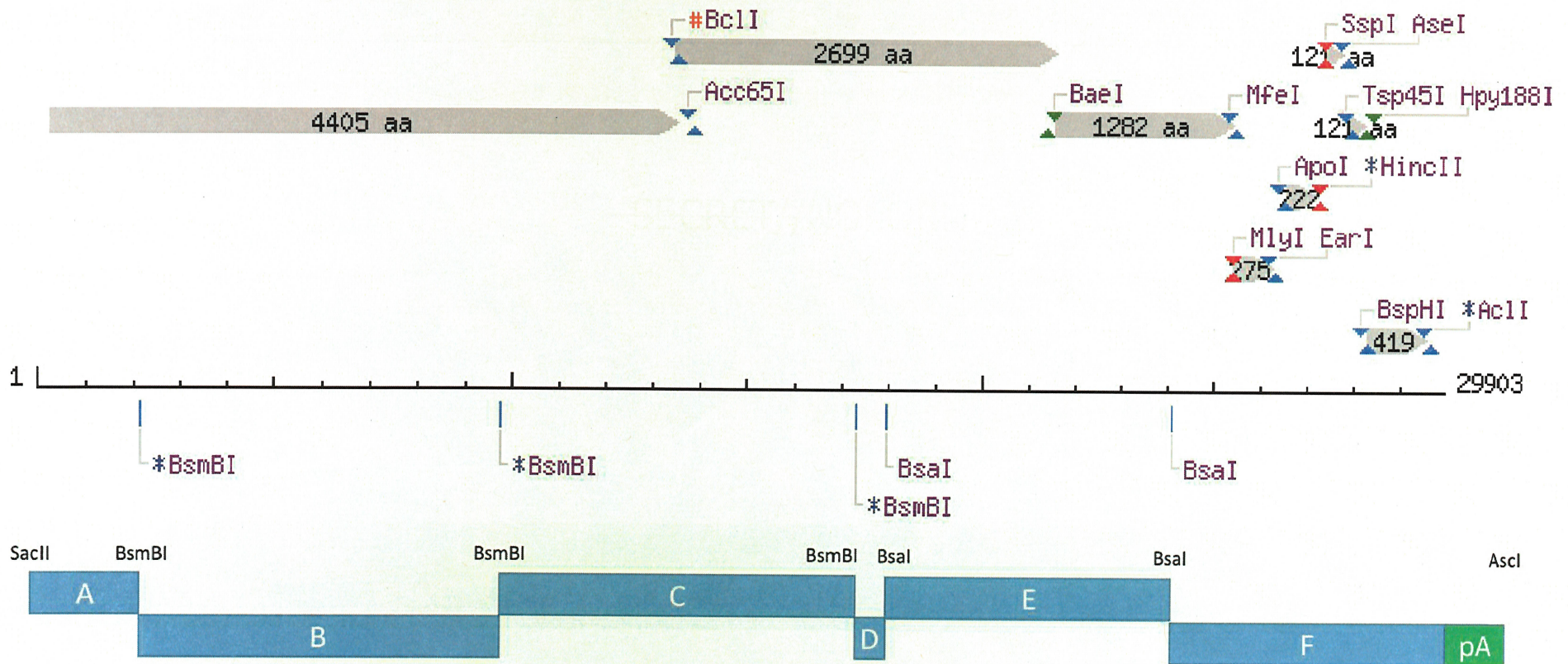
3' NNNNNCTCTGG - 5'

(b)(1); (b)(3) 50 USC 3024(i); Sec. 1.4(c); Sec. 1.4(e)



~~SECRET//NOFORN~~

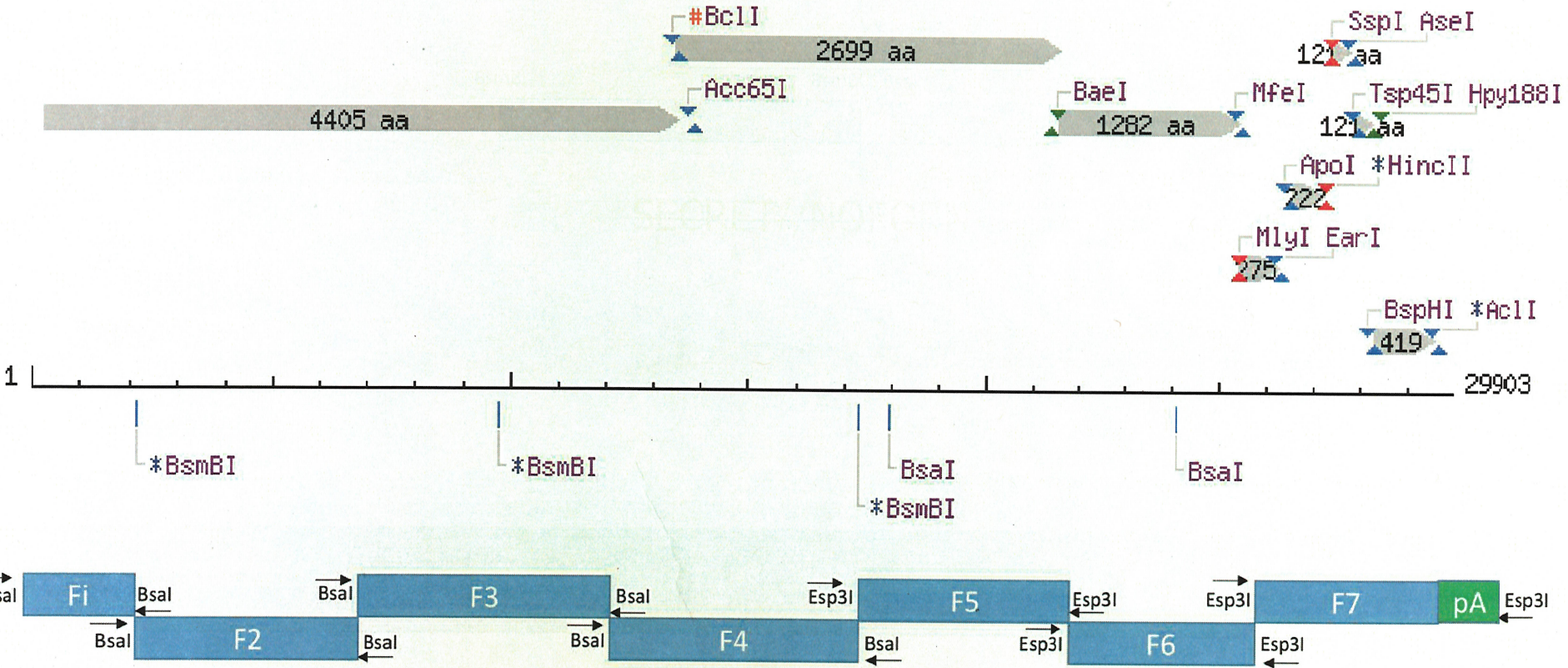
# RE-based Fragment Build Option – *BsmBI/BsaI* (4 nt overhangs)



~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# RE-based Fragment Build Option – *Bsal*/*Esp31* (Invisible restriction sites)



~~SECRET//NOFORN~~

# Xie et al., 2020 SARS-CoV-2 FLC Assembly

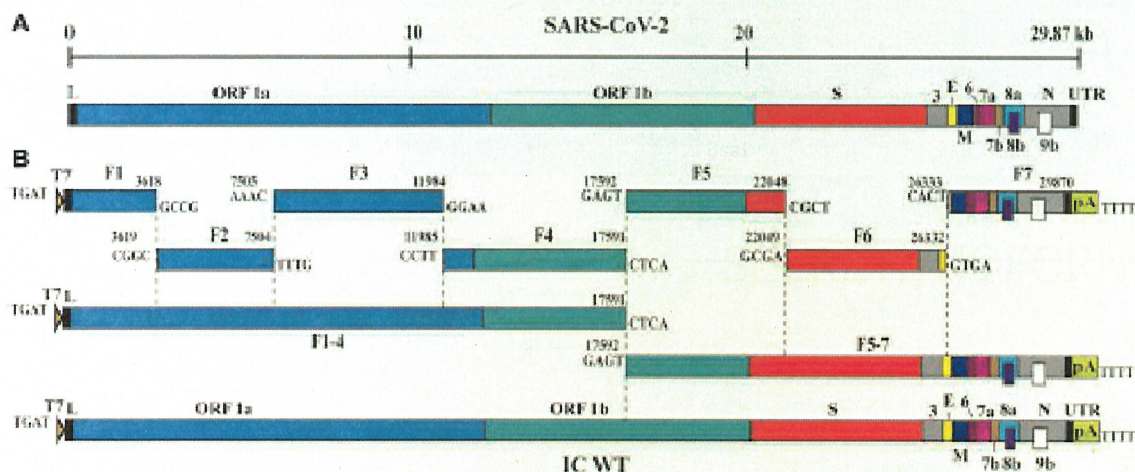


Figure 1. Assembly of a Full-Length SARS-CoV-2 Infection cDNA Clone

(A) Genome structure SARS-CoV-2. The open reading frames (ORFs) from the full genome are indicated.

(B) Strategy for *in vitro* assembly of an infectious cDNA clone of SARS-CoV-2. The nucleotide sequences and genome locations of the cohesive overhangs are indicated. The WT full-length (FL) cDNA of SARS-CoV-2 (IC WT) was directionally assembled using *in vitro* ligation.

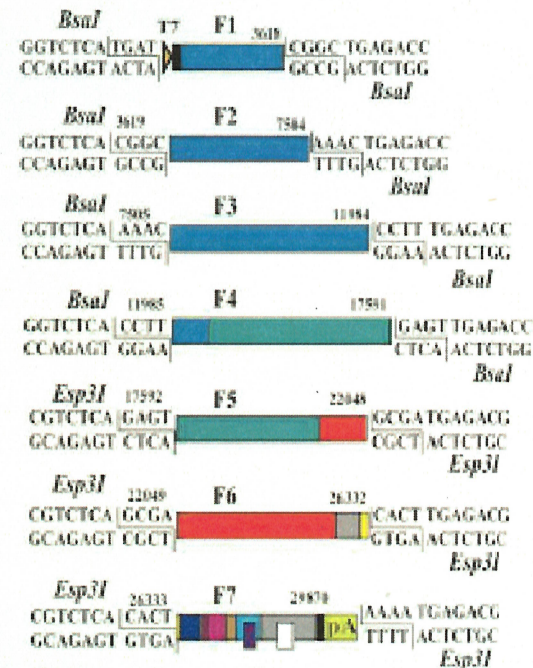
(C) Diagram of the terminal sequences of each cDNA fragment recognized by BsaI and Esp3I.

(D) Gel analysis of the seven purified cDNA fragments. Individual fragments (F1-F7) were digested from corresponding plasmid clones and gel purified. Seven purified cDNA fragments (50-100 ng) were analyzed on a 0.8% native agarose gel. The 1-kb DNA ladders are indicated.

(E) Gel analysis of cDNA ligation products. About 400 ng of purified ligation product was analyzed on a 0.8% native agarose gel. Triangle indicates the FL cDNA product. Circles indicate the intermediate cDNA products.

(F) Gel analysis of RNA transcripts. About 1 µg of *in vitro*-transcribed (IVT) RNAs were analyzed on a 0.8% native agarose gel. DNA ladders are indicated. Because this is a native agarose gel, the DNA size is not directly correlated to the RNA size. Triangle indicates the genome-length RNA transcript. Circles show the shorter RNA transcripts.

## C Fragment Terminal Sequence



(b)(3);50 USC 3024(i); Sec. 1.4(c); Sec. 1.4(e)

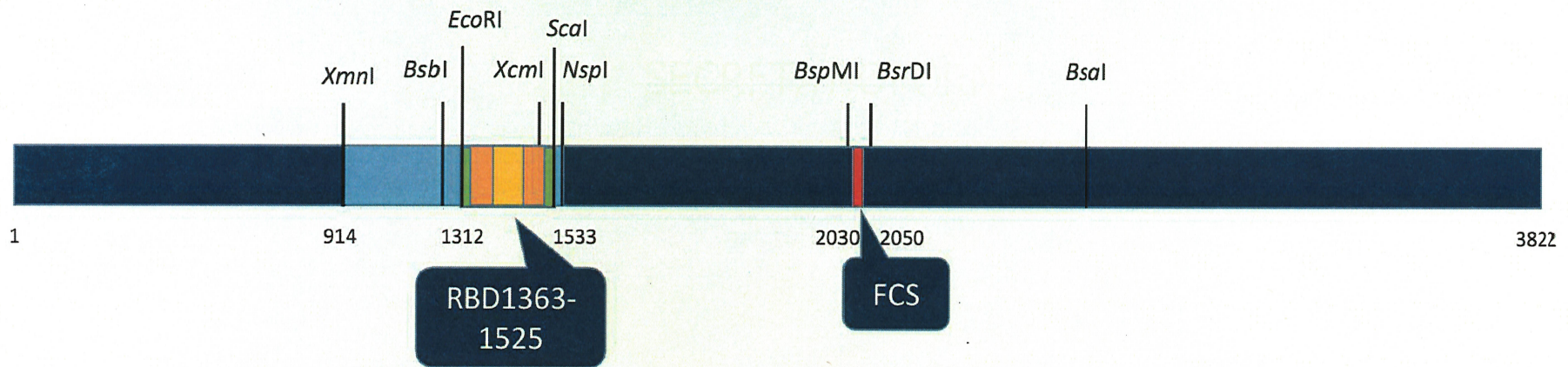
~~SECRET//NOFORN~~

# Spike Gene

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# SARS CoV-2 SPIKE GENE



Highest homology to RaTG13 Pangolin CoV

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# Spike Gene Swapping Using the WIV SARS-CoV Reverse Genetics System

- Hu et al., 2017 swapped out the WIV1 spike gene for the spike gene of the following: Rs4231Rs73 27Rf4075Rs4081Rs40 85Rs4235As6526Rp3

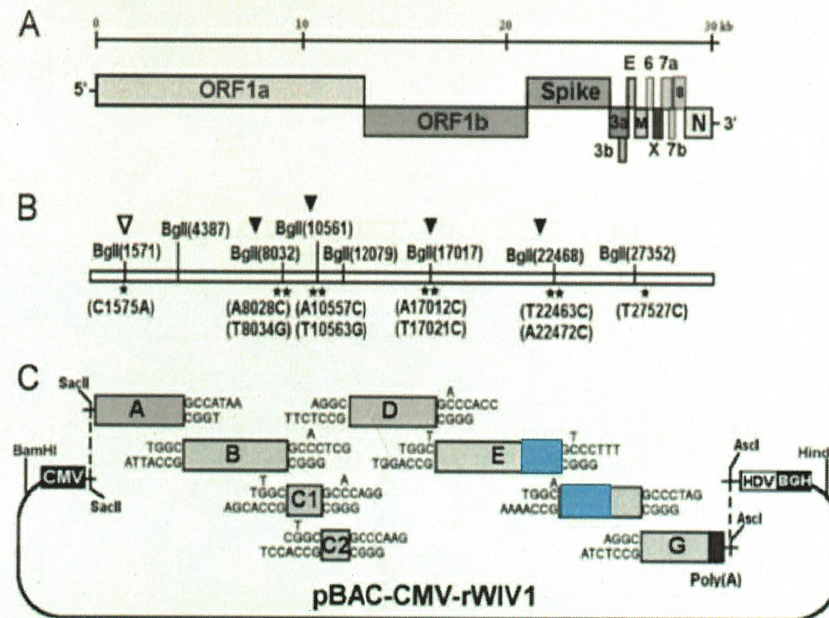


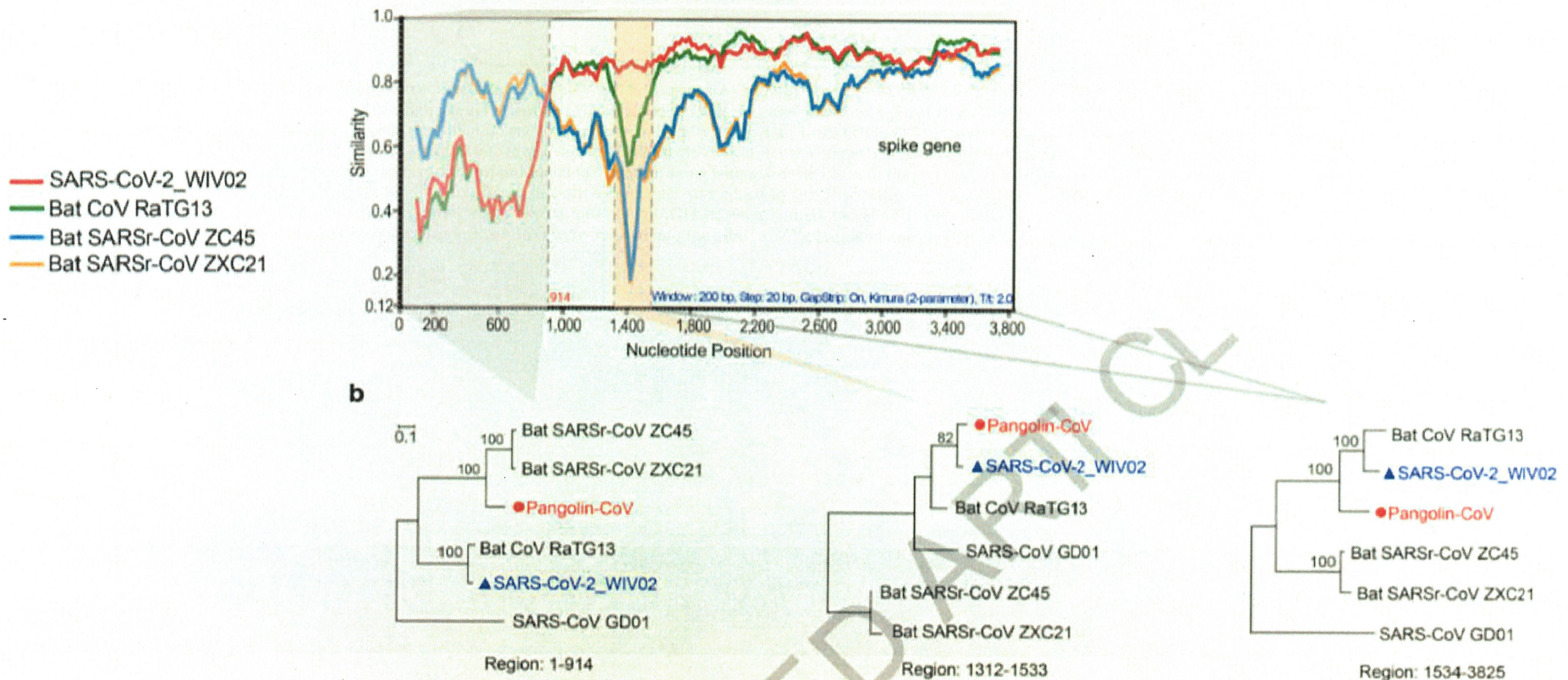
FIG 1 Strategy for construction of an infectious WIV1 BAC clone. (A) Genomic structure of WIV1. (B) The mutations are indicated under the stars. C1575A was used to ablate a natural BglI site at nucleotide 1571 (▽), and T27527C was used to disrupt a potential T7 stop site. The others were for introducing BglI sites (▼). (C) The WIV1 genome was split into eight contiguous cDNAs (A to G): A, nt 1 to 4387; B, nt 4388 to 8032; C1, nt 8033 to 10561; C2, nt 10562 to 12079; D, nt 12080 to 17017; E, nt 17018 to 22468; F, nt 22469 to 27352; G, nt 27353 to 30309. Unique BglI sites were introduced into the fragments by synonymous mutations to make these fragments capable of unidirectional ligation along with native BglI sites in the genome. The original nucleotides are shown above the flanking sequences corresponding fragments. A poly(A) sequence was added to the 3' terminus of fragment G. A CMV promoter, HDV ribozyme, and BGH transcriptional terminal signal were inserted into pBeloBAC11 between BamHI and HindIII sites. SacII and Ascl sites were introduced between the CMV promoter and ribozyme. Fragments A to G were inserted into the pBAC-CMV plasmid in a single step.

Zeng et al., 2016 Hu et al., 2017

~~SECRET//NOFORN~~

# SARS-COV-2 SPIKE GENE SEGMENTS

## QUERY: PANGOLIN-COV

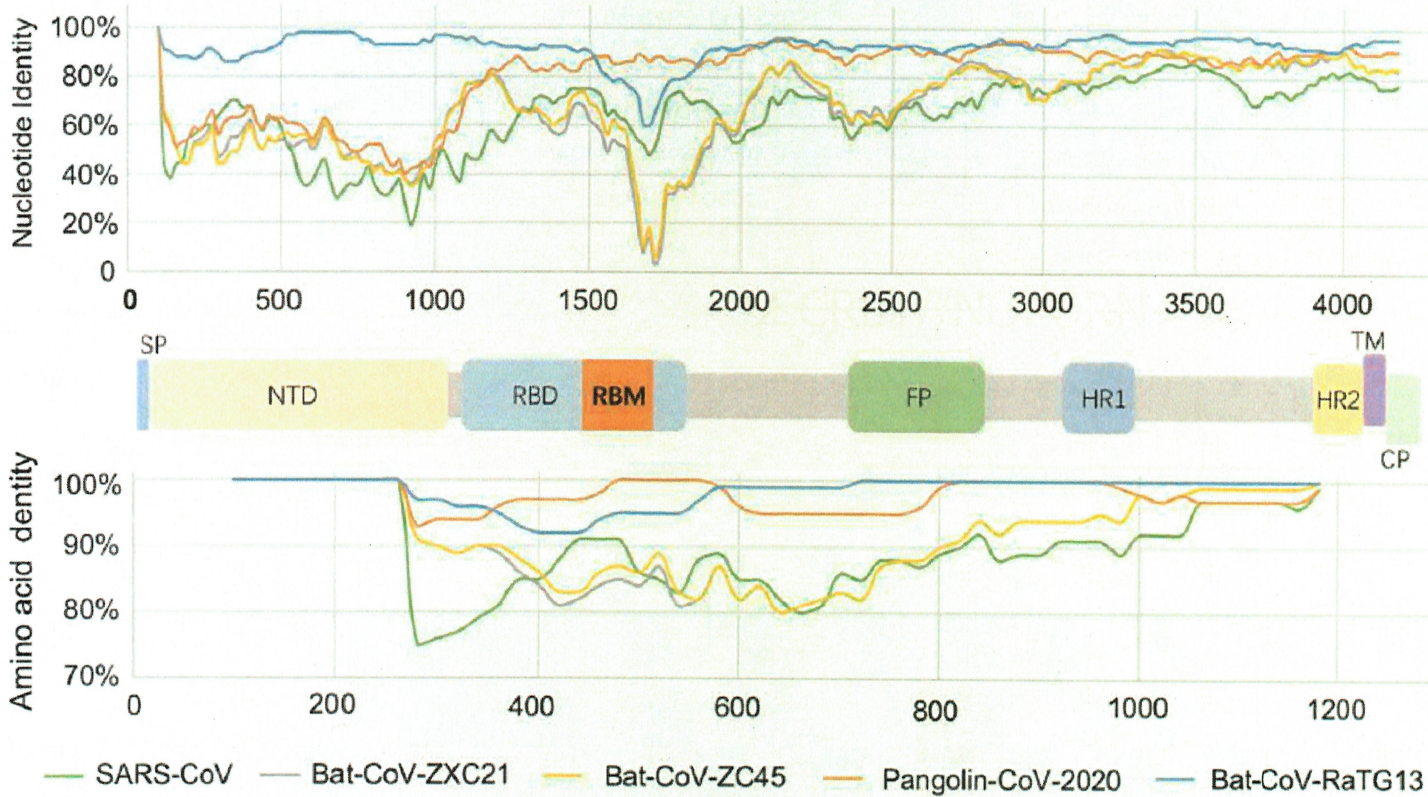


Xiao *et al.*, 2020

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# SARS-CoV-2 SimPLOT



- "A recent study found that a human ACE2-binding ridge in SARS-CoV-2 RBD takes a more compact conformation compared with the SARS-CoV RBD; moreover, several residue changes in SARS-CoV-2 RBD may also enhance its human ACE2-binding affinity [13]. The core residues in RBM which may related to higher human ACE2-binding affinity than SARS-CoV are 100% identical between SARSCoV-2 and CoV-Pangolin-2020. Therefore, pangolin-CoV-2020 (CoV-pangolin/GD) potentially recognizes human ACE2 better than the SARS-CoV."

~~SECRET//NOFORN~~



~~SECRET//NOFORN~~

Ren *et al.*, 2008

JOURNAL OF VIROLOGY, Feb. 2008, p. 1899–1907  
0022-538X/08/\$08.00+0 doi:10.1128/JVI.01085-07  
Copyright © 2008, American Society for Microbiology. All Rights Reserved.

Vol. 82, No. 4

## Difference in Receptor Usage between Severe Acute Respiratory Syndrome (SARS) Coronavirus and SARS-Like Coronavirus of Bat Origin<sup>∇</sup>

Wuze Ren,<sup>1†</sup> Xiuxia Qu,<sup>2†</sup> Wendong Li,<sup>1‡</sup> Zhenggang Han,<sup>1</sup> Meng Yu,<sup>3</sup> Peng Zhou,<sup>1</sup> Shu-Yi Zhang,<sup>4</sup>  
Lin-Fa Wang,<sup>3\*</sup> Hongkui Deng,<sup>2</sup> and Zhengli Shi<sup>1\*</sup>

*State Key Laboratory of Virology, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, China<sup>1</sup>; Key Laboratory of Cell Proliferation and Differentiation of the Ministry of Education, College of Life Sciences, Peking University, Beijing, China<sup>2</sup>; CSIRO Livestock Industries, Australian Animal Health Laboratory and Australian Biosecurity Cooperative Research Center for Emerging Infectious Diseases, Geelong, Australia<sup>3</sup>; and School of Life Science, East China Normal University, Shanghai, China<sup>4</sup>*

Received 20 May 2007/Accepted 15 November 2007

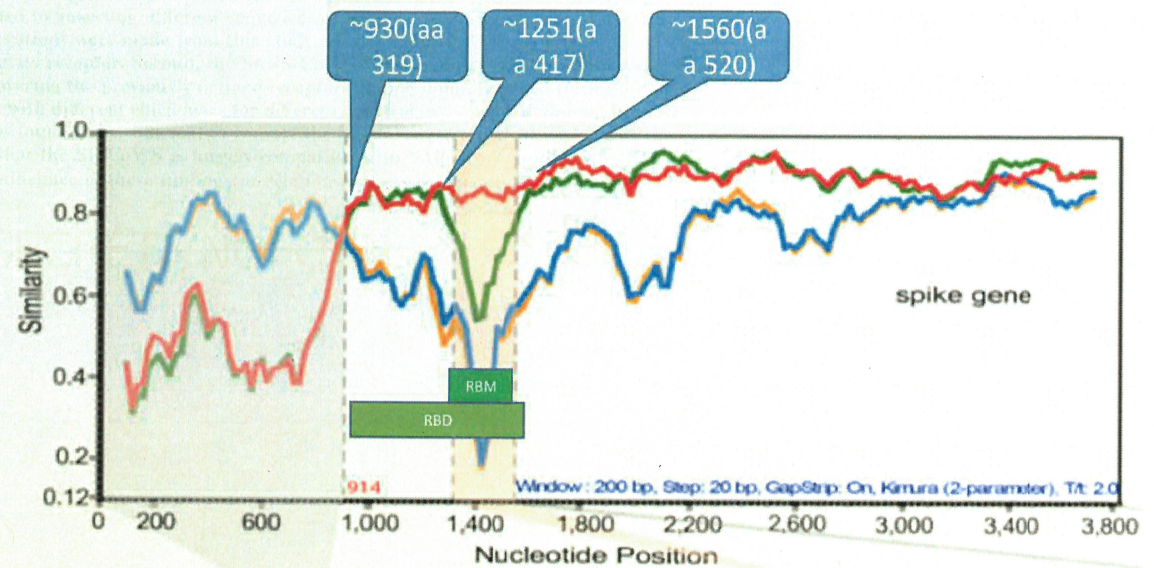
Severe acute respiratory syndrome (SARS) is caused by the SARS-associated coronavirus (SARS-CoV), which uses angiotensin-converting enzyme 2 (ACE2) as its receptor for cell entry. A group of SARS-like CoVs (SL-CoVs) has been identified in horseshoe bats. SL-CoVs and SARS-CoVs share identical genome organizations and high sequence identities, with the main exception of the N terminus of the spike protein (S), known to be responsible for receptor binding in CoVs. In this study, we investigated the receptor usage of the SL-CoV S by combining a human immunodeficiency virus-based pseudovirus system with cell lines expressing the ACE2 molecules of human, civet, or horseshoe bat. In addition to full-length S of SL-CoV and SARS-CoV, a series of S chimeras was constructed by inserting different sequences of the SARS-CoV S into the SL-CoV S backbone. Several important observations were made from this study. First, the SL-CoV S was unable to use any of the three ACE2 molecules as its receptor. Second, the SARS-CoV S failed to enter cells expressing the bat ACE2. Third, the chimeric S covering the previously defined receptor-binding domain gained its ability to enter cells via human ACE2, albeit with different efficiencies for different constructs. Fourth, a minimal insert region (amino acids 310 to 518) was found to be sufficient to convert the SL-CoV S from non-ACE2 binding to human ACE2 binding, indicating that the SL-CoV S is largely compatible with SARS-CoV S protein both in structure and in function. The significance of these findings in relation to virus origin, virus recombination, and host switching is discussed.

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# Minimal Receptor Binding Domain Cassette

- WIV scientists previously defined the minimal Receptor Binding Domain cassette that could functionally transfer ACE2 binding capability from one Spike protein to another SARS Nucleotide: 930-1554 SARS Amino Acid: 310-518 Receptor Binding Motif SARS Nucleotide: 1251-1482 SARS Amino Acid: 417-494 Homology cut points of SARS-CoV-2 coincide with WIV-identified borders of RBD and RBM



— SARS-CoV-2\_WIV02  
— Bat CoV RaTG13  
— Bat SARSr-CoV ZC45  
— Bat SARSr-CoV ZXC21

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# Furin Cleavage Site

SARS-CoV 1980 CATGTCGACACTTCTTATGAGTGCGACATTCCTATTGGAGCTGGCATTGTGCTAGTTAC  
H V D T S Y E C D I P I G A G I C A S Y SARS-CoV-2  
CATGTCAACAACACTCATATGAGTGTGACATACCCATTGGTGCAGGTATATGCGCTAGTTAT 2022  
H V N N S Y E C D I P I G A G I C A S Y BCoV RaTG13  
CATGTCAATAACTCGTATGAGTGTGACATACCTATTGGTGCAGGAATATGCGCCAGTTAT 2022  
H V N N S Y E C D I P I G A G I C A S Y SARS-CoV CATAAGTTTCTTTATT-----  
ACGTAGTACTAGCCAAAAATCTATTGTGGCT 2028 H T V S L L R S T S Q K S I  
V ASARS-CoV-2 CAGACTCAGACTAATTCTCCTCGGCGGGCACGTAGTGTAGCTAGTCAATCCATCATTGCC  
2082 Q T Q T N S P R R A R S V A S Q S I I ABCoV RaTG13  
CAGACTCAAACACTAATTC-----ACGTAGTGTGGCCAGTCAATCTATTATTGCC 2070 Q T Q  
T N S R S V A S Q S I I AFurin Cleavage SiteNmeAIII Restriction Site



A unique restriction site facilitates identifying the correct *E. coli* clone

~~SECRET//NOFORN~~

# SARS-CoV-2 Furin Cleavage Site GC Content

- The percent GC of the furin cleavage site insert is 77% compared to ~40% of the surrounding DNA. Contains an NmeAIII restriction site. The other CoV's with FCS have a %GC of <55%

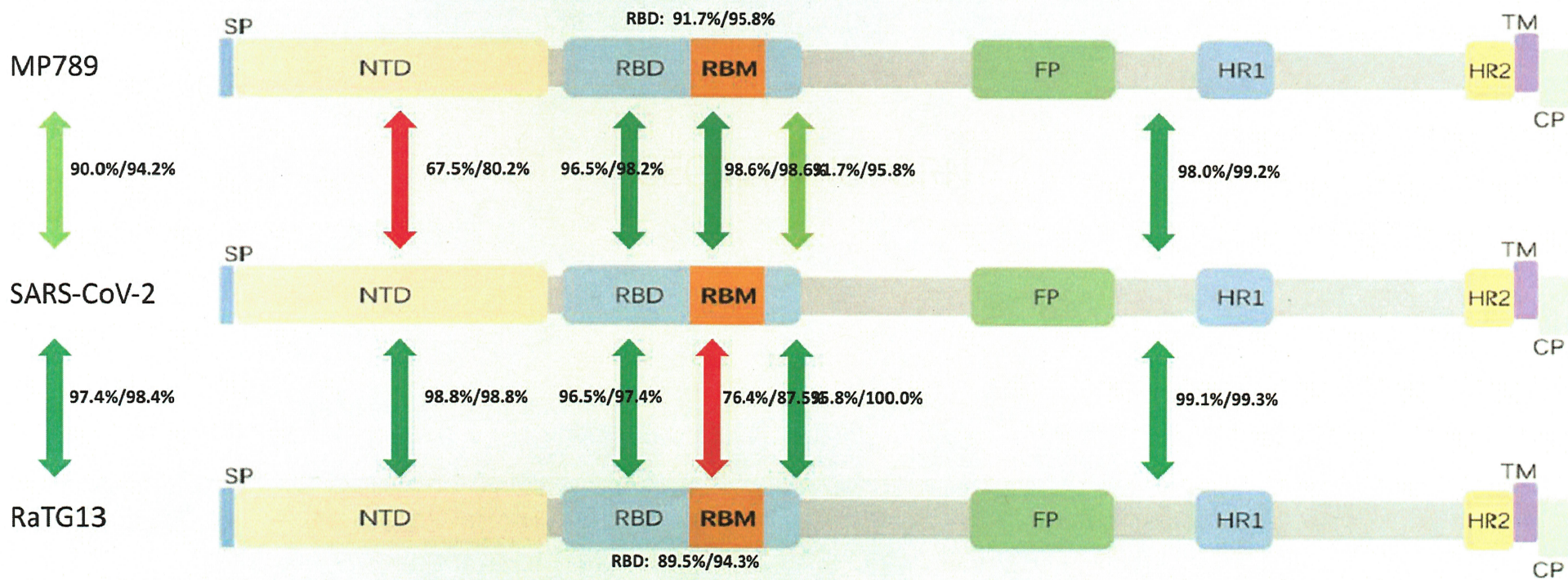
Virus	Nucleotide and Amino Acid Sequences																																																																																						
SARSCoV2 (47%)	G	G	T	A	T	A	T	G	C	G	C	T	A	G	T	T	A	T	C	A	G	A	C	T	C	A	G	A	C	T	A	A	T	T	C	T	C	T	C	G	G	C	G	G	G	C	A	C	G	T	A	G	T	G	T	A	G	C	T	A	G	T	C	A	A	T	C	C	A	T	C	C	A	T	T	G	C	C	T	A	C	A	C	T	A	T	G
	G	I	C	A	S	Y	Q	T	Q	T	N	S	P	R	R	A	R	S	V	A	S	Q	S	I	I	A	Y	T	M																																																										
	%GC: 14/35 = 40%										%GC: 10/13 = 77%										%GC: 17/39 = 44%																																																																		
MERS-CoV (54%)	C	T	C	T	G	T	G	C	T	C	T	T	C	C	T	G	A	C	A	C	A	C	C	T	A	G	T	A	C	T	C	T	C	A	C	A	C	C	T	C	G	C	A	G	T	G	T	G	C	G	C	T	C	T	G	T	T	C	C	A	G	G	T	G	A	A	A	T	G	C	G	C	T	T	G	G	C	A	T	C	C	A	T	T	G	C	T
	L	C	A	L	P	D	T	P	S	T	L	T	P	R	S	V	R	S	V	P	G	E	M	R	L	A	S	I	A																																																										
	%GC: 18/35 = 51%										%GC: 8/13 = 62%										%GC: 21/39 = 54%																																																																		
BatCoV-HKU5 (47%)	G	G	T	C	A	A	T	C	A	C	T	T	T	G	T	G	C	T	A	T	T	C	C	A	C	A	A	C	T	A	C	T	T	C	T	T	C	A	C	G	C	G	T	T	C	G	A	C	G	T	G	C	T	A	C	T	T	C	T	G	G	T	G	C	A	T	C	T	G	A	T	G	T	G	T	T	T	C	A	A	T	C	G	C	C		
	G	Q	S	L	C	A	I	P	P	T	T	S	S	R	V	R	R	A	T	S	G	A	S	D	V	F	Q	I	A																																																										
	%GC: 15/35 = 43%										%GC: 7/13 = 54%										%GC: 18/39 = 49%																																																																		
IBV-Beaudette (34%)	G	A	A	E	A	G	S	I	A	C	A	T	C	A	C	T	A	A	G	A	A	G	G	T	A	G	A	C	T	T	C	A	T	T	A	D	A	A	C	C	T	G	A	A	A	T	T	C	C	T	T	T	G	T	T																																
	E	V	S	F	V	K	T	P	P	S	F	N	N	P	Y	V																																																																							
	%GC: 12/35 = 31%										%GC: 5/13 = 38%										%GC: 13/39 = 33%																																																																		

Influenza viruses convert from low path to high path by addition of a poly basic cleavage site by virtue of RNA Polymerase stuttering which adds preferentially A's and T's – this is not the case with SARS-CoV-2

~~SECRET//NOFORN~~

% Identity:>95

# Spike Protein Regions



~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# SARS-CoV-2 Spike RBD Alignment: Possible RBM Cassette Insertion

```
SARSCoV2          TRFQTLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPBCoV_RaTG13
TRFQTLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPMP789          TKFRTLLTIHRGDPMP---
NNGWTVFSAAYVGYLAPRTFMLNENGTITDAVDCALDP          NTD<>RBDSARSCoV2
LSEKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRCoV_RaTG13
LSEKCTLKSFTVEKGIYQTSNFRVQPTDSIVRFPNITNLCPFGEVFNATTFASVYAWNRP789/Manis
LSEAKCTLKSLTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATTFASVYAWNRSARSCoV2
KRISNCVADYSVLYNSASFSTFKCYGVSPTKLNLDLCTNVDYADSFVIRGDEVRQIAPGQTCoV_RaTG13
KRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNLDLCTNVDYADSFVITGDEVRQIAPGQTMP789/Manis
KRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNLDLCTNVDYADSFVVRGDEVRQIAPGQT
>RBMSARSCoV2          GKIADYNYKLPDDFTGCVIAWNSNNLDSKVGNGYNYLYRFLFRKSNLKPFFERDISTEIQABCoV_RaTG13
GKIADYNYKLPDDFTGCVIAWNSKHIDAKEGGNFNYLYRFLFRKANLKPFFERDISTEIQAMP789/Manis
GRIADYNYKLPDDFTGCVIAWNSNNLDSKVGNGYNYLYRFLFRKSNLKPFFERDISTEIQAB
RBM<          RBD<SARSCoV2
GSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHAPATVCGPKKSTNLVKBCoV_RaTG13
GSKPCNGQTLNLCYPLRYGFYPTDGVGHQPYRVVLSFELLNAPATVCGPKKSTNLVKMP789/Manis
GSTPCNGVEGFNCYFPLQSYGFHPTNGVGYQPYRVVLSFELLNAPATVCGPKQSTNLVKSARSCoV2
NKCVMFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLLEILDITPCSFQGGVBCoV_RaTG13
NKCVMFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLLEILDITPCSFQGGVMP789
NKCVMFNFNGLTGTGVLTESSKKFLPFQQFGRDIADTTDAVRDPQTLLEILDITPCSFQGGV
```

ACE2 Critical ContactACE2

~~SECRET//NOFORN~~

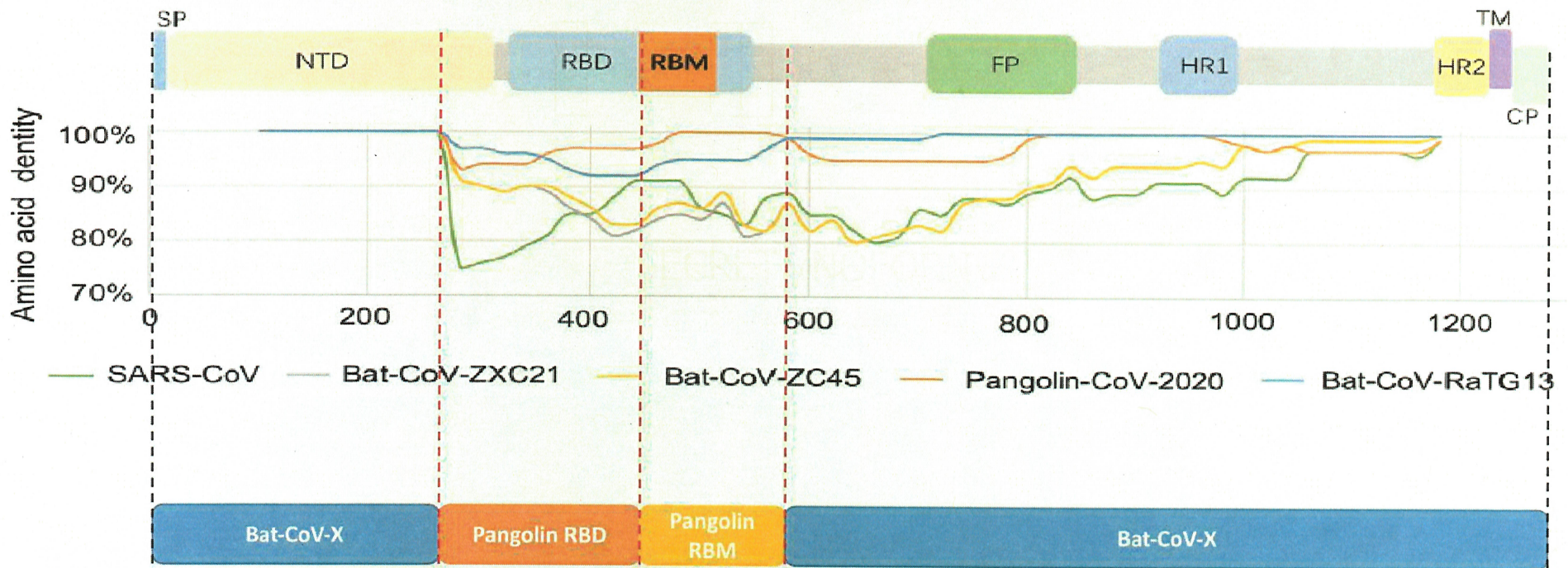
# SARS-COV-2 RBD vs PCoV MP789 RBD

>RBM  
D Y N Y K L P D D F T G C V I A W N S N MP789  
GACTATAATTATAAACCCTCGATGATTTACAGGTTGTGTAATAGCTTGGAAATCTAAC 1305SARSCOV2  
GATTATAATTATAAATACCAGATGATTTACAGGCTGCGTTATAGCTTGGAAATCTAAC 1317 D  
Y N Y K L P D D F T G C V I A W N S N N L D  
S K V G G N Y N Y L Y R L F R K S MP789  
AACCTTGATTCTAAGGTTGGTGGTAATTATAACTACCTTTATAGATTGTTAGAAAAGTCC 1365SARSCOV2  
AATCTTGATTCTAAGGTTGGTGGTAATTATAATTACCTGTATAGATTGTTAGGAAGTCT 1377 N  
L D S K V G G N Y N Y L Y R L F R K S N L K  
P F E R D I S T E I Y Q A G S T P MP789  
AACCTCAAACCTTTTGAACGAGACATTTCTACAGAAATATACCAAGCTGGTAGTACACCC 1425SARSCOV2  
AATCTCAAACCTTTGAGAGAGATATTTCAACTGAAATCTATCAGGCCGGTAGCACACCT 1437 N  
L K P F E R D I S T E I Y Q A G S T P C N G  
V E G F N C Y F P L Q S Y G F Q P MP789  
TGCAATGGGGTTGAAGTTTTAACTGTTACTTTCCTCTACAATCTTATGGTTCCACCCT 1485SARSCOV2  
TGTAATGGTGTGAAGTTTTAATIGTTACTTTCCTTTACAATCATATGGTTTCCAACCC 1497 C  
N G V E G F N C Y F P L Q S Y G F Q P  
RBM< T N G V G Y Q P Y R V V V L S F E L L H  
MP789 ACTAATGGTGTGGTTACCAACCTTATAGAGTAGTAGTATGTCATTTGAACTTTTAAAA  
1545SARSCOV2  
ACTAATGGTGTGGTTACCAACCATAACAGAGTAGTAGTACTTTCTTTTGAACCTTCTACAT 1557 T  
N G V G Y Q P Y R V V V L S F E L L K  
RBD< A P A T V C G P K K S T N MP789  
GCACCTGCTACTGTTTGTGGACCTAAAACAGTCCACTAACCTAGTTAAAAACAATGTGTC 1605SARSCOV2  
GCACCAGCAACTGTTTGTGGACCTAAAAGTCTACTAATTTGGTTAAAAACAATGTGTC 1617 A  
P A T V C G P K K S T N

- 38 codon differences  
First: 4  
Second: 0  
Third: 31  
First and third: 31 results in an amino acid change  
Pangolin RBD cassette appears to be a codon optimized insert

~~SECRET//NOFORN~~

# SARS-CoV-2 Spike Appears to be a Chimera



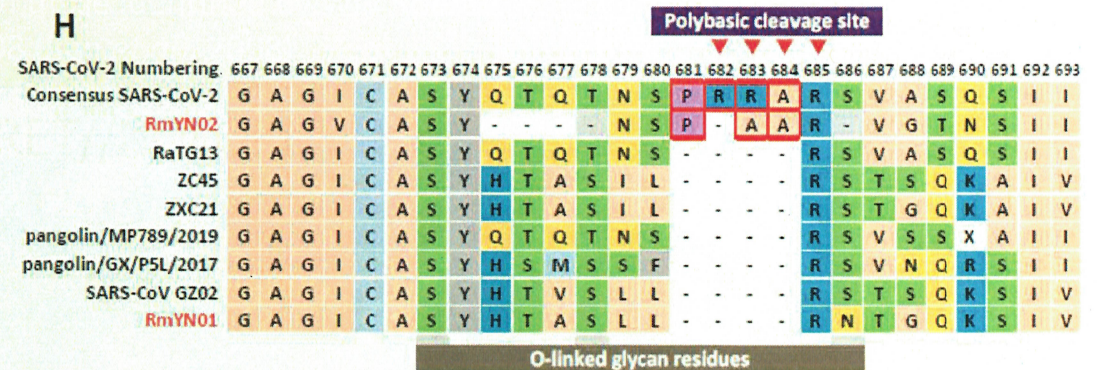
Break points align with those identified by WIV Scientists in 2008 (Ren et al., 2008)

~~SECRET//NOFORN~~



# RmYN02 - A Red Herring?

- Zhou et al., 2020 publish paper describing Bat CoV rmYN02. Next generation sequencing was done on pooled bat samples to develop two genome sequences – RmYN01 and RmYN02. Claim that RmYN02 contains inserted nucleotides at the S1/S2 cleavage site. Assert that the SARS-CoV-2 FCS is therefore of natural origin. No virus is available for peer confirmation.





~~SECRET//NOFORN~~

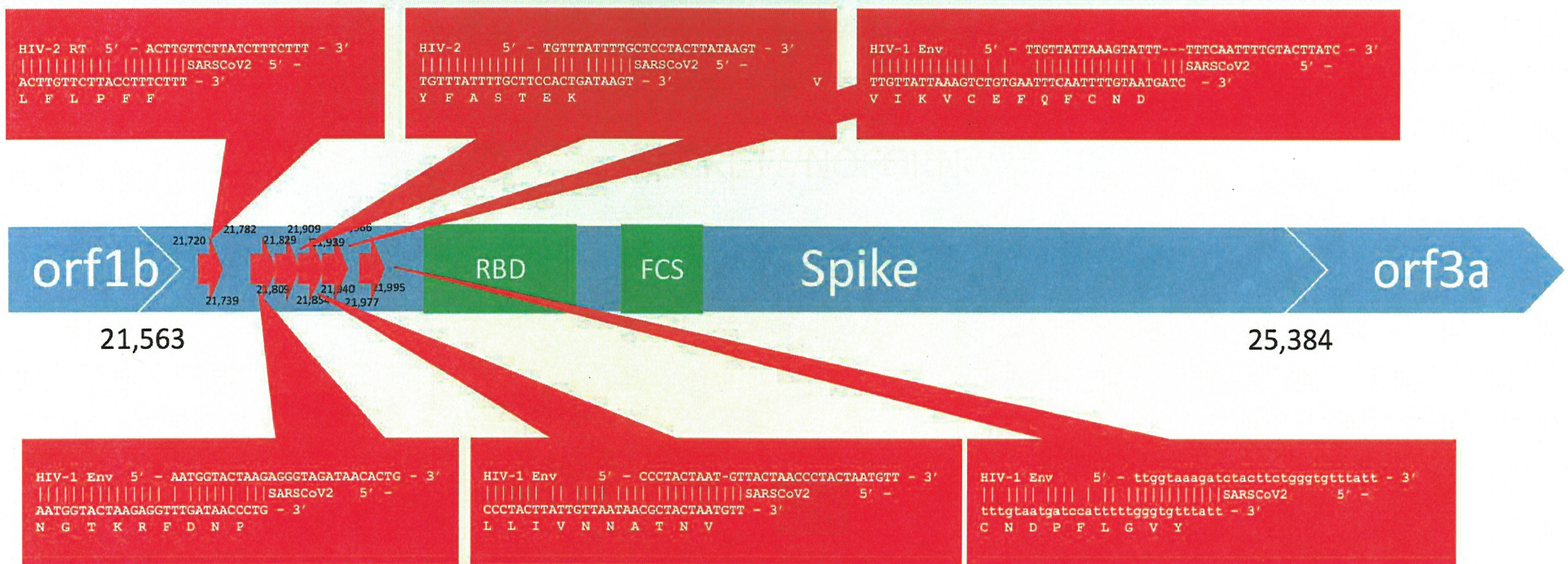
# HIV Epitopes

Perez, 2020

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# HIV Sequences in the SARS-CoV-2 Spike Gene



Adapted from Perez, 2020

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

## Perez, 2020 Scientific Challenges

- None of the six proposed regions are identical at either the nucleotide or amino acid level with the corresponding HIV/SIV segments. None of the six peptides are related to identified immunosuppressive regions of HIV and SIV (Retroviral ISU Domains). The HIV gp41 Immunosuppressive (ISU) Domains sequence is KQLQARILAVERYLKDQQLLGG - this sequence does not match any of the six. Four of the six regions either perfectly or almost perfectly match corresponding peptides in multiple Pangolin CoVs - Perez did not account for Pangolin genomes in the paper. Several are only found in Pangolin CoV Spike sequences and not in Bat CoV Spike sequences, indicating that the SARS-CoV-2 Spike NTD region originated from a Pangolin CoV template.

The next 2 pages are DIF citing (b)(1) and (b)(3) and are not provided.

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# Alternative Scenario

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# Hypothetical Laboratory Origin of SARS-CoV-2

- WIV conducted a longitudinal studies to isolate a large number of bat Coronaviruses from multiple locations in China (2011-2015)WIV Developed Reverse Genetic System, assembled WIV1 full-length infectious clone, and created chimeric viruses exchanging the WIV1 spike gene with the spike gene from other bat Coronaviruses (2015-2017)WIV and other Chinese scientists conduct gain of function studies on SARS, MERS, IBV, and PEDV to insert furin cleavage sites demonstrating increased virulence of the chimeric virusesWIV conducted in vivo and in vitro studies to characterize the bank of bat CoronavirusesWIV conducted the live bat Coronavirus studies under BSL2 conditionsChinese BSL2 and US BSL2 conditions are differentChinese labs have had a history of virus escapes from BSL2 laboratoriesHypothesis: Between 2017 and 2019, WIV created a full-length infectious clone in pBAC-CMV using an unpublished bat Coronavirus genome as template (BatCoVX)Hypothesis: Between 2017 and 2019, WIV created chimeric Bat-CoV-X viruses using the pBAC-CMV-BCoVX backbone and swapping out key cassettes with other bat Coronaviruses (RBD, RBM, etc.) and adding additional features such as a furin cleavage siteHypothesis: In 2018-2019, WIV conducted in vitro and in vivo studies to characterize the BatCoVX chimeric viruses under BSL2 conditionsHypothesis: In mid-2019, one of the not fully characterized Bat-CoV-X chimeric viruses escaped from the WIV facilities and begins infecting civilians in the city of WuhanHypothesis: Starting in mid-2019 through present, WIV and other Chinese laboratories conduct studies to characterize the Chimeric BCoVX virus that escaped (now called SARS-CoV-2)WIV (Zhou et al., 2020) publishes the 2019-nCoV genome sequence showing relatedness to RaTG13 (a previously unpublished genome)BatCoVX likely highly related to RaTG13Hypothesis: Beginning in early 2020, WIV and other government controlled agencies begin to publish obfuscation information to drive the narrative that SARS-CoV-2 is of natural origin and resulted from natural recombinationRaTG13RMYN02Pangolin CoV's

~~SECRET//NOFORN~~

~~SECRET//NOFORN~~

# CONCLUSION

The next page is DIF citing (b)(1) and (b)(3) and is not provided.

~~SECRET//NOFORN~~



~~SECRET//NOFORN~~

## Concluding Points

- WIV possesses a bank of Bat Coronavirus isolatesWIV has scientists experienced in Coronavirology and Coronavirus Infectious Clone generationWIV Scientists generated chimeric SARS CoV and Bat CoV Spike genes to identify minimal Spike Receptor Binding Domain cassette that could transfer receptor binding specificity (Ren et al., 2008)WIV possesses an existing and published Coronavirus Reverse Genetics System (Zeng et al., 2016) utilizing their pBAC-CMV plasmidWIV has utilized the pBAC-CMV-WIV1 Full-length clone to generate chimeras with Bat CoV spike genes (Hu et al., 2017)WIV has BSL2/BSL3/BSL4 animal facilitiesWIV has multiple in vitro assays (apoptosis, IFN-B induction, etc.) to characterize their Bat Coronaviruses and chimeric Bat CoronavirusesWIV and other Chinese researchers have conducted Gain of Function studies in SARS, MERS, IBV, and PEDV to add Furin Cleavage Sites to CoV Spike proteinThe absence of a published progenitor virus for SARS-CoV-2 only indicates that it has not been published, not that it does not existThe genomic sequence of SARS-CoV-2 has Type IIS restriction sites that are consistent with being generated by the Golden Gate Cloning system utilizing the published pBAC-CMV plasmidThe SARS-CoV-2 genome has several break points where homology jumps from Bat Coronaviruses to Pangolin Coronaviruses which is consistent with a synthesized chimeric virusThe SARS-CoV-2 Spike protein similarity with RaTG13 and Pangolin CoV Spike proteins may also be explained by use of cassettes swapped into the base virus – these break points align with those identified by WIV scientists (Ren et al., 2008)The Pangolin RBD cassette is 100% identical at the amino acid level while the DNA sequence appears to be codon optimizedThere are no other published SARS lineage Betacoronaviruses that possess a Furin Cleavage Site in their Spike protein (RmYN02 does not have an insertion) and the SARS-CoV-2 FCS does not appear to be inserted via the same mechanism that drives Influenza virus insertions of polybasic cleavage sitesZeng et al., 2016 stated that “All experiments using live virus was conducted under biosafety level 2 (BSL2) conditions” which would make an accidental release of a pathogenic Bat CoV capable of binding human ACE2 more likelyA chimeric virus comprised of segments from natural Bat CoV genomes would appear like a recombined virus

The molecular biology capabilities of WIV and the genome assessment are consistent with the hypothesis that SARS-CoV-2 was a lab-engineered virus that was part of a bank of chimeric viruses in Zhen-Li Shi’s laboratory at WIV that escaped from containment

~~SECRET//NOFORN~~

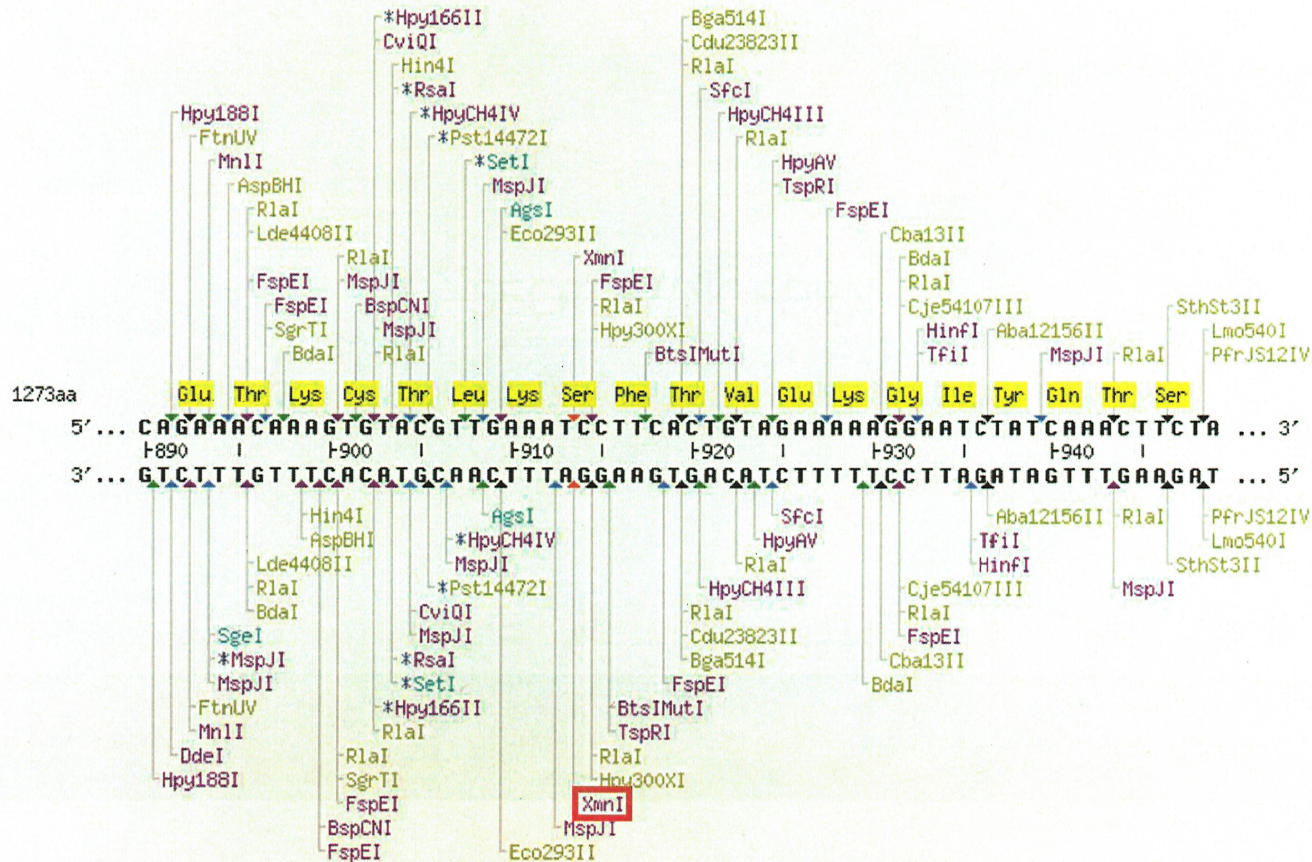
~~SECRET//NOFORN~~

# BACK-UP SLIDES

~~SECRET//NOFORN~~

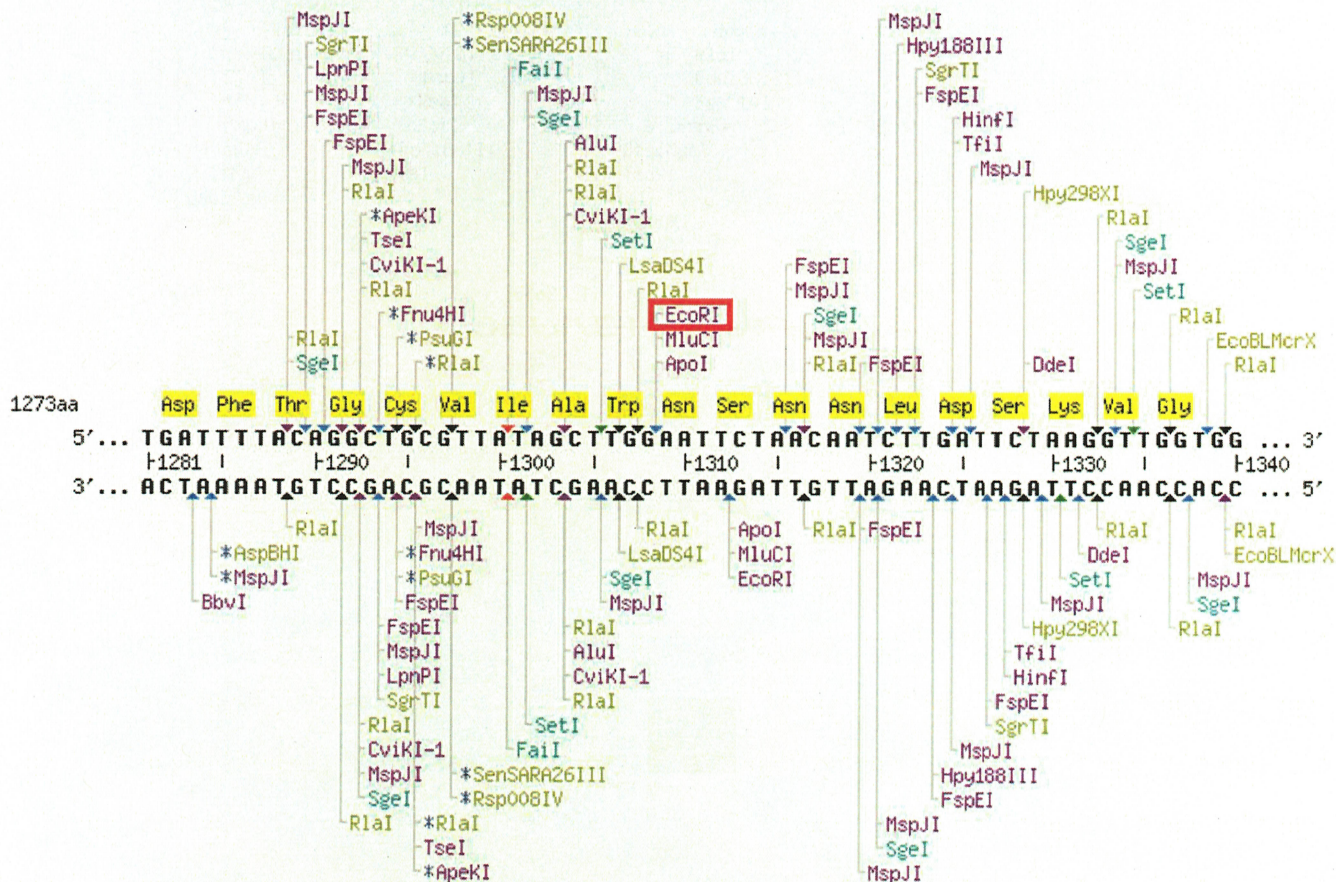
~~SECRET//NOFORN~~

# Nucleotide 914 Region



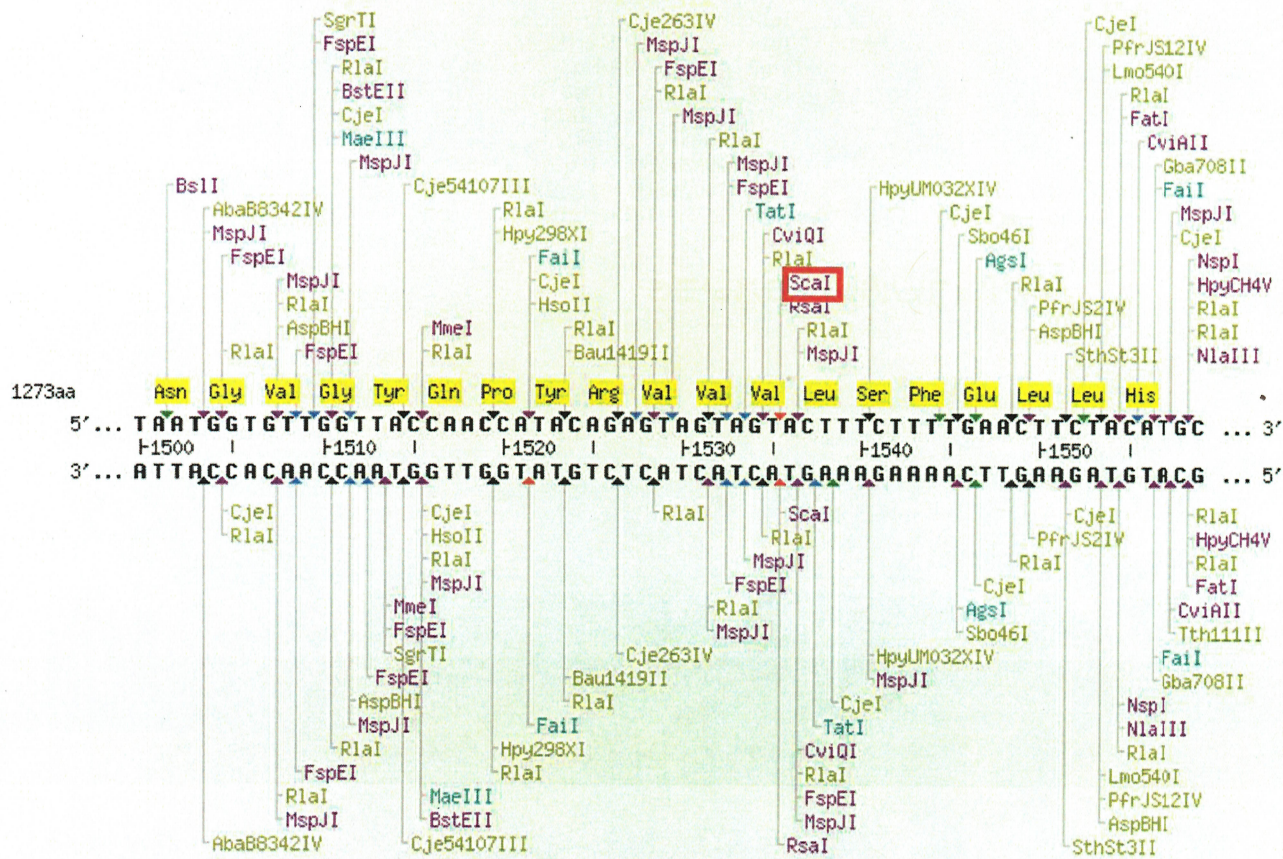
~~SECRET//NOFORN~~

# Nucleotide 1312 Region



~~SECRET//NOFORN~~

# Nucleotide 1535 Region



~~SECRET//NOFORN~~

SARS-CoV 1980 CATGTCGACACTTCTTATGAGTGCGACATTCCTATTGGAGCTGGCATTGTGCTAGTTAC  
H V D T S Y E C D I P I G A G I C A S Y SARS-CoV-2  
CATGTCAACAACACTCATATGAGTGTGACATACCCATTGGTGCAGGTATATGCGCTAGTTAT 2022  
H V N N S Y E C D I P I G A G I C A S Y BCoV RaTG13  
CATGTCAATAACTCGTATGAGTGTGACATACCTATTGGTGCAGGAATATGCGCCAGTTAT 2022  
H V N N S Y E C D I P I G A G I C A S Y SARS-CoV CATAACAGTTTCTTTATT-----  
ACGTAGTACTAGCCAAAAATCTATTGTGGCT 2028 H T V S L L R S T S Q K S I  
V ASARS-CoV-2 CAGACTCAGACTAATTCTCCTCGGCGGGCACGTAGTGTAGCTAGTCAATCCATCATTGCC  
2082 Q T Q T N S P R R A R S V A S Q S I I ABCoV RaTG13  
CAGACTCAAACACTAATTC-----ACGTAGTGTGGCCAGTCAATCTATTATTGCC 2070 Q T Q  
T N S R S V A S Q S I I AFurin Cleavage SiteBspMI Restriction SiteNmeAIII Restriction SiteBsrDI  
Restriction Site

~~SECRET//NOFORN~~