Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak

Tao Zhang, Qunfu Wu, Zhigang Zhang

Digital version, contact details and references.

Created by Jonas Schuck, Jessica Bender

Abstract

The outbreak of the 2019 novel coronavirus (SARS-CoV-2) in the city of Wuhan in China and its extensive impact on human health requires further research on affected animal species with CoV-like viruses. Phylogenetic analysis is essential to examine key information about CoV-like viruses that could potentially help combat the spread of COVID-19. To discover evolutionary relatives, it is necessary to determine similarities in the genomic organization, for example the presence and similarity of the spike (S) protein as well as specific cleavage motifs used for host cell entry. However, apart from BatCoV RaTG13, there are more known species affected by potentially highly similar CoV-like viruses, with their affinity to the SARS-CoV-2 virus being still unexplored. Here they show that Pangolin-CoV is the second-closest known relative to SARS-CoV-2, right behind BatCoV RatG13, with an identity of 91.02 % on the whole genome level. Additionally, they analyzed the identity of the (S) protein, which is around 97.5 %, indicating functional similarities between the Pangolin-CoV and the SARS-CoV-2 virus. Moreover, they found that the highly conserved nucleocapsid (N) protein is also present in the Pangolin-CoV. The (N) protein is often used as a marker in diagnostic assays, while the (S) protein is essential for the cell entry process of the virus, giving information about the pathogenic potential of Pangolin-CoV. With these findings, the pangolin species can be classified as a natural reservoir of SARS-CoV-2-like CoVs. Further experiments should be performed to deepen the knowledge about the interconnection of potentially related viruses to be able to block interspecies transmission.

Finding Potential Intermediate Hosts

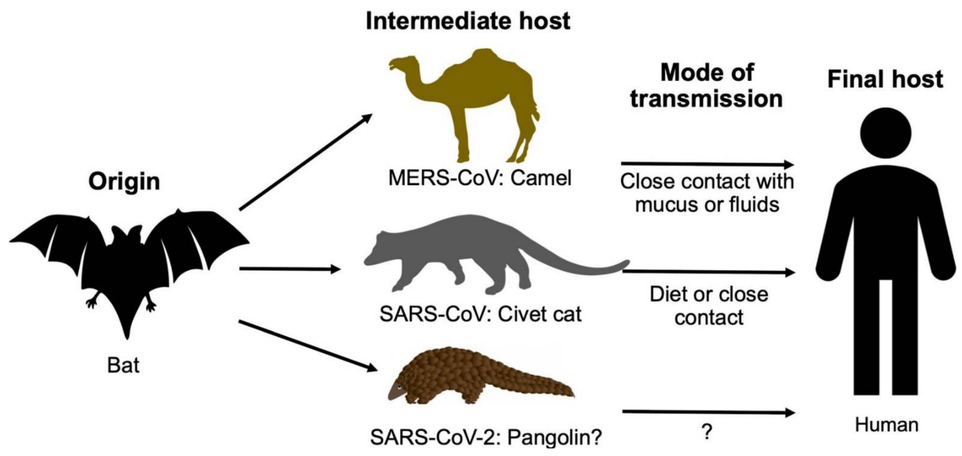

Figure 1: The origins and intermediate hosts of SARS-CoV-2, SARS-CoV, and MERS-CoV [2].

Why Pangolins?

- Pangolins are the most poached and trafficked mammal in the world [3]
- **All eight species are classified as endangered**
- Low immunity and poor health condition
- Known infection with SARS-CoV-like viruses
- Potential intermediate host for SARS-CoV-2
- **Infection through meat or contact with faeces**

Figure 2: Adult pangolin.

Zoonotic diseases propose threat for probable future pandemics

Workflow And Results

Data collection & preprocessing

- Raw RNA reads from collected pangolin lung samples
- Quality control & contamination removal [Tool: Trimmomatic]
- Identify virus reads with **reference-based mapping** [Tool: Bowtie2]

De novo assembly & gene prediction

- **De novo assembly & taxonomical annotation** of the assembled contigs
- **Draft genome** from annotated contigs
  - **Reference-guided scaffolding**
- Align draft genome with SARS-CoV-2 to discover ORF's [Tool: MUSCLE]
- **Compare ORF-identities** [PangoCoV/BatCoV and PangoCoV/SARS-CoV]

Phylogenetic analysis

- Generate phylogenetic trees [Tool: MegaX]
- Based on whole genome, orf1a, orf1b, (M) gene, (S) gene, RdRp gene
- **SARS-CoV-2 cluster** [BatCoV RatG13, SARS-CoV-2, PangoCoV]

Main results summarized

- **Potential common ancestor of SARS-CoV-2 and BatCoV RatG13**
- Whole genome identity from PangoCoV to SARS-CoV-2: **91.02 %**
  - **Second closest known relative to SARS-CoV-2**
- Highly similar spike protein and nucleocapsid protein
  - **Potential functional similarities**

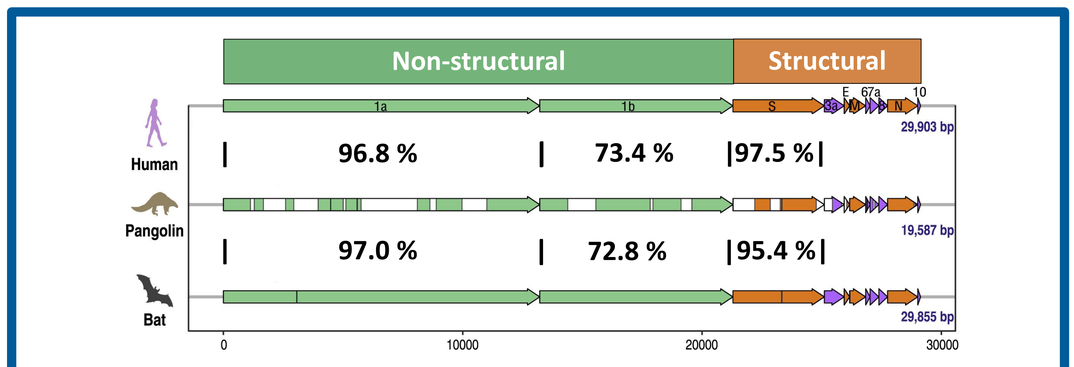

Figure 3: Comparison of common genome organization similarity among SARS-CoV-2, Pangolin-CoV, and BatCov RaTG13.

Comparison of the genomic structure

Whole genome identity:

- PangoCoV/SARS-CoV-2: **91.02 %**
- PangoCoV/BatCoV RaTG13: **90.55 %**
- BatCoV RaTG13/SARS-CoV-2: **96.20 %**

- Some drafted PangoCoV ORF's showed higher similarity to SARS-CoV-2 than to BatCoV RaTG13

→ orf1b, orf7a, orf10, **spike protein**

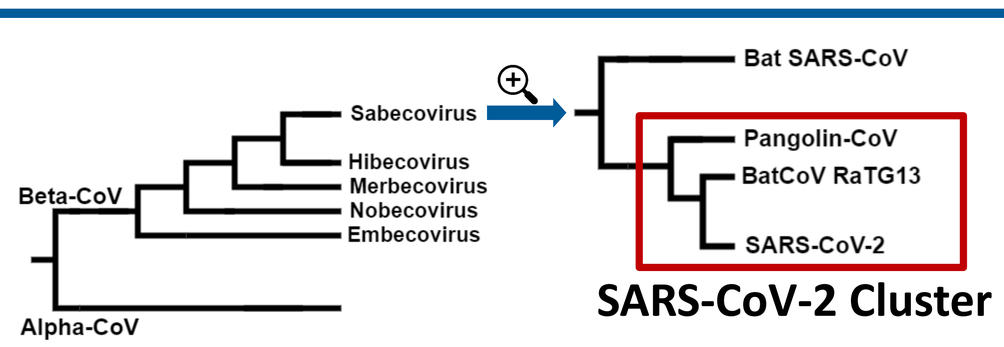

Figure 4: Schematic representation of the phylogenetic relationships of CoVs based on the whole genome.

Structure Of The Coronavirus

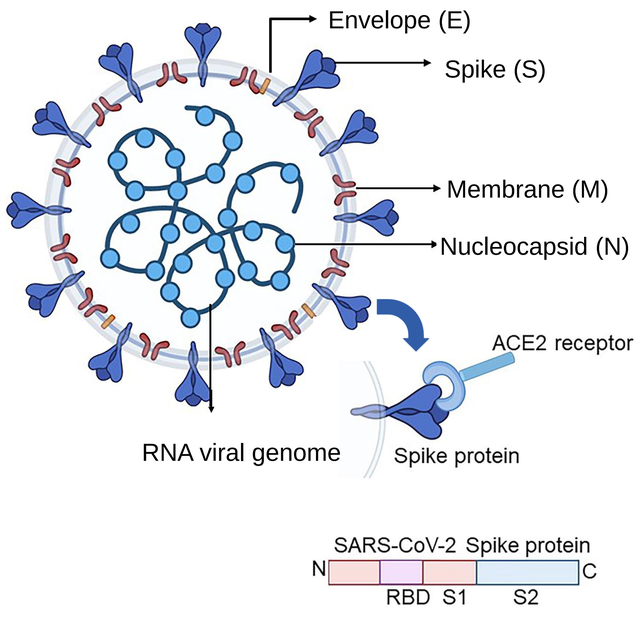

Figure 5: Schematic representation of the structure of SARS-CoV-2 and the organization of its spike protein [4].

Spike protein:

- Enables cell entry into receptor-expressing cells
  - RBD binding to ACE2 receptors in human cells
  - **Insights into the pathogenic potential**

Priming-phase:

- (S) protein gets cleaved into S1/S2-subunits
- **Cleavage motif necessary**

Nucleocapsid protein:

- Highly immunogenic phosphoprotein
  - Very conserved & used for diagnostic assays

Envelope protein:

- Support protein taking part in multiple viral mechanisms

Membrane protein:

- Support protein taking part in the assembly of newly generated virions

Cleavage Motif

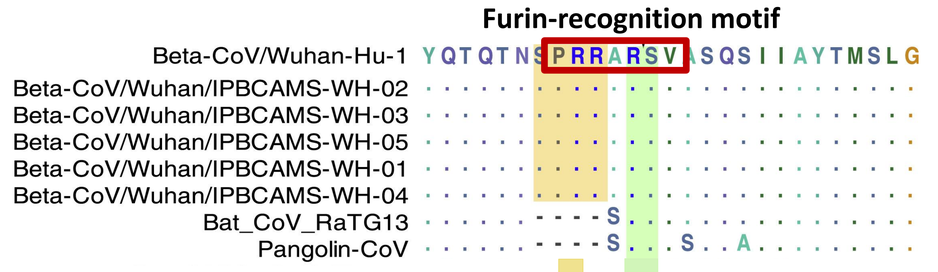

Figure 6: CoV (S) protein S1/S2-cleavage site and cleavage motif.

Insertions

- Insertions in SARS-CoV-2 variants form a furin-recognition cleavage motif
- **No insertions in BatCoV RaTG13 and Pangolin-CoV**

S1/S2-cleavage motif

- **PRRARSV**-sequence works as cleavage motif for host cell proteases

Conclusion

- First report on a potential close relative to SARS-CoV-2 inherited by pangolins called PangoCoV
- Further research on the pathologic potential of PangoCoV is necessary
- To be confirmed if pangolins are a probable candidate related to the COVID-19 outbreak
- Suggestion to classify the pangolin species as a natural reservoir for SARS-CoV-like CoV's