



# nCoV Genetics

February 3, 2020

## Key findings for public health

- A PCR diagnostic test can differentiate 2019-nCoV from other coronaviruses.
- The outbreak was initiated from either a single introduction into humans or very few animal-to-human transmission events.
- nCoV-2019 and SARS-CoV use the same cellular receptor, ACE2, which could be used as a starting point for creating therapeutics for nCoV-2019.

## Background

Coronaviruses, including the pneumonia-causing novel coronavirus currently known as nCoV-2019, are enveloped, nonsegmented, positive-sense RNA viruses. Coronavirus genomes have some of the largest genomes among RNA viruses, with approximately 25–32 kilobases.<sup>1</sup> The typical CoV genome includes a 5'-cap, 5'-untranslated region (UTR), open reading frames, a 3'-UTR, and 3'-poly(A) tail. The first two thirds of the genome typically codes for nonstructural proteins from 2 open reading frames that form the replicase complex. The last third of the genome encodes primarily structural proteins.<sup>2</sup> There are 4 conserved structural proteins across CoVs: the spike (S) protein, membrane (M) protein, envelope (E) protein, and nucleocapsid (N) protein.<sup>1</sup> The S protein is responsible for binding to host cell receptors and viral entry to host cells. The M, E, and N proteins are part of the nucleocapsid of viral particles.

## nCoV genetics

In a recently published paper,<sup>3</sup> viral sequences collected from the earliest patients recognized in the ongoing nCoV-2019 outbreak were assessed and compared to known viral sequences. Sequence analysis of 11 samples found that nCoV-2019 is in the same species as SARS-CoV; the 2 viruses are 94.6% similar in amino acid sequence (80% nucleotide sequence similarity) across the genome.<sup>3</sup> However, other studies do not consider nCoV-2019 to be the same species as SARS-CoV, as it differs from SARS-CoV by more than 10% in the replicase genes.<sup>4</sup>

Further analysis demonstrated that nCoV-2019 was less than 75% homologous to nearly all strains of SARS-CoV in the spike protein. A single isolate of a bat coronavirus, named BatCoV RaTG13, shared 96.7% sequence homology with nCoV-2019, suggesting nCoV-2019 originated in bats and shares a common ancestor with SARS-CoV.<sup>3</sup> Other teams found nCoV-2019 had over 85% sequence homology with bat SARS-like CoVs.<sup>4,5</sup> nCoV-2019's S protein is most closely related to bat coronaviruses.<sup>6,7</sup>

In addition to the typical coronavirus structural proteins and replicase genes, nCoV-2019 has several currently unidentified nonstructural open reading frames in its genome.<sup>4</sup> nCoV-2019 can be differentiated from other coronaviruses, including SARS-CoV, using PCR primers specific to a highly variable region of the spike protein, meaning a PCR diagnostic test can differentiate this virus from other coronaviruses.<sup>3</sup>

Phylogenetic analysis of 30 publicly available nCoV-2019 samples concluded that emergence of nCoV-2019 into the human population likely occurred in mid-November 2019.<sup>8</sup> The sequences have limited variability in consensus sequences, suggesting the outbreak was initiated from either a single introduction into humans or from a very few animal-to-human transmission events.<sup>9</sup> The mutation rate has been estimated in various groups, ranging from about  $1.05 \times 10^{-3}$  to  $1.26 \times 10^{-3}$  substitutions per site per year, which is similar to some estimates of MERS-CoV mutation rates.<sup>8,10-12</sup> As more viral genomes are made publicly available, scientists will better be able to track viral evolution and mutation rates, so the exact estimates will vary.

Selection analysis of the genome suggests that 2 genes in the nCoV-2019, the S and N genes, are under episodic selection as the virus is transmitted between humans.<sup>7</sup> This is normal for emerging viruses and means that part of the genome are undergoing positive selection.<sup>13,14</sup> Mutations and adaptation in the S and N genes could affect virus stability and pathogenicity.<sup>7</sup> As more genomes are made publicly available, analysis of the genome sequence diversity across samples has revealed the highest diversity occurring in the structural genes, especially the S protein, ORF3a, and ORF8.<sup>15</sup>

The nCoV-2019 nucleocapsid protein is not cross-reactive with any other human coronavirus except SARS-CoV. Serological testing of patient samples showed strong IgG antibodies specific for the virus 20 days after disease onset in 5 of 7 patients. Patient sera was able to neutralize nCoV-2019 in Vero E6 cells. The virus could also be neutralized by horse anti-SARS-CoV serum dilutions, further indicating the cross-reactivity between SARS-CoV and nCoV-2019. nCoV-2019 was able to enter HeLa cells expressing the ACE2 receptor, but not cells lacking this receptor. This suggests that ACE2 is the cell receptor used by nCoV-2019 to enter cells, which is the same receptor used by SARS-CoV. Cells lacking other receptors used by other coronaviruses were still infected by nCoV-2019, suggesting ACE2 is the primary cell receptor for nCoV-2019.<sup>3</sup>

1. Masters PS. Coronavirus genomic RNA packaging. *Virology* 2019;537:198-207. doi:10.1016/j.virol.2019.08.031
2. Lo C-Y, Tsai T-L, Lin C-N, Lin C-H, Wu H-Y. Interaction of coronavirus nucleocapsid protein with the 5' - and 3' -ends of the coronavirus genome is involved in genome circularization and negative-strand RNA synthesis. *FEBS J* 2019;286(16):3222-3239. doi:10.1111/febs.14863
3. Zhou P, Yang X-L, Wang X-G, et al. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. *Microbiology* 2020. doi:10.1101/2020.01.22.914952
4. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020. doi:10.1056/NEJMoa2001017
5. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 2020. doi:10.1016/S0140-6736(20)30251-8
6. Robertson D. nCoV's relationship to bat coronaviruses & recombination signals (no snakes). *Virological* January 22, 2020. <http://virological.org/t/ncovs-relationship-to-bat-coronaviruses-recombination-signals-no-snakes/331>. Accessed January 24, 2020.
7. Benvenuto D, Giovanetti M, Ciccozzi A, Spoto S, Angeletti S, Ciccozzi M. The 2019-new Coronavirus epidemic: evidence for virus evolution. *bioRxiv* January 2020. doi:https://doi.org/10.1101/2020.01.24.915157
8. Rambaut A. Phylodynamic analysis of nCoV-2019 genomes- 27-Jan-2020. *Virological* January 27, 2020. <http://virological.org/cl-novel-2019-coronavirus/331/latest>. Accessed January 28, 2020.
9. Bedford T, Richard N, Hadfield J, Hodcroft E, Muller N, Llcisin M. Narrative: Genomic analysis of nCoV spread. Situation report 2020-01-23. nextstrain. <https://nextstrain.org/narratives/ncov/sit-rep/2020-01-23?n=1>. Published January 23, 2020. Accessed January 24, 2020.
10. Cotten M, Watson SJ, Zumla AI, et al. Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus. *mBio* 2014;5(1). doi:10.1128/mBio.01062-13
11. Dudas G, Carvalho LM, Rambaut A, Bedford T. MERS-CoV spillover at the camel-human interface. *eLife* 2018;7:e31257. doi:10.7554/eLife.31257
12. Duchene S, Featherstone L, Haritopoulou-Sinanidou M. Temporal signal and the evolutionary rate of 2019 n-CoV using 47 genomes collected by Feb 01 2020. *Virological* January 29, 2020. <http://virological.org/t/ncovs-relationship-to-bat-coronaviruses-recombination-signals-no-snakes-no-evidence-the-2019-ncov-lineage-is-recombinant/331/22>. Baric RS, Yount B, Hensley L, Peel SA, Chen W. Episodic evolution mediates interspecies transfer of a murine coronavirus. *J Virol* 1997;71(3):1946-1955. doi:10.1128/JVI.71.3.1946-1955.1997
13. Sironi M, Cagliani R, Forni D, Clerici M. Evolutionary insights into host-pathogen interactions from mammalian sequence data. *Nat Rev Genet* 2015;16(4):224-236. doi:10.1038/nrg3905
14. Nextstrain / ncov (2/3/2020). nextstrain. [https://nextstrain.org/ncov?m=num\\_date](https://nextstrain.org/ncov?m=num_date). Accessed February 3, 2020. February 3, 2020.