

Emergence of a new virus prompts immediate comparison of the genetic sequence to sequences of other known viruses. This can provide initial indications of important characteristics, such as transmission route, transmissibility, and virulence, if the new virus is a close match to other viruses for which these properties are known. These comparative analyses also can elucidate the evolutionary history of the new virus, and suggest the non-human species the virus emerged from or passed through on its way to humans.

In the case of SARS-CoV-2, the evolutionary history of the virus remains unclear, despite computational investigations by many skilled laboratories. The closest relative of the virus was reported by Chinese researchers in bats, which host thousands of coronaviruses, including some capable of infecting humans (such as viruses related to the viruses that cause Ebola, SARS, and MERS). However, it is now well-established that the genome sequences in key parts of SARS-CoV-2 differs substantially from this related bat virus. In particular, the SARS-CoV-2 spike protein, which mediates attachment to host cells, contains a furin cleavage site between the S1 and S2 domains, encoded by a 12-base pair insertion. This site is not found in the related bat virus or in any SARS-related coronaviruses. The presence of this site is expected to increase infectivity in the human respiratory tract, and thereby promote more efficient spread than other SARS-related coronaviruses.

SARS-CoV-2 is an RNA virus. While viruses with RNA genomes accumulate mutations at a higher rate than DNA viruses, these mutations are generally single-nucleotide changes, additions, or deletions. Larger insertions, such as the 12-base pair insertion that encodes the furin cleavage site in SARS-CoV-2, generally are not seen to arise all at once.

Larger genomic segments like the 12-base pair insertion can arise naturally through recombination, when two viral genomes infecting the same cell exchange genetic material (as occurs, for example, when a novel influenza strain arises through exchange of genetic material between two influenza strains, which may derive from different animal species). This insertion could have occurred during co-infection of a non-human host by two different coronaviruses—one related to the bat coronavirus that is the closest known relative of SARS-CoV-2, and another coronavirus (not yet identified) with unique characteristics found in the SARS-CoV-2 spike protein.