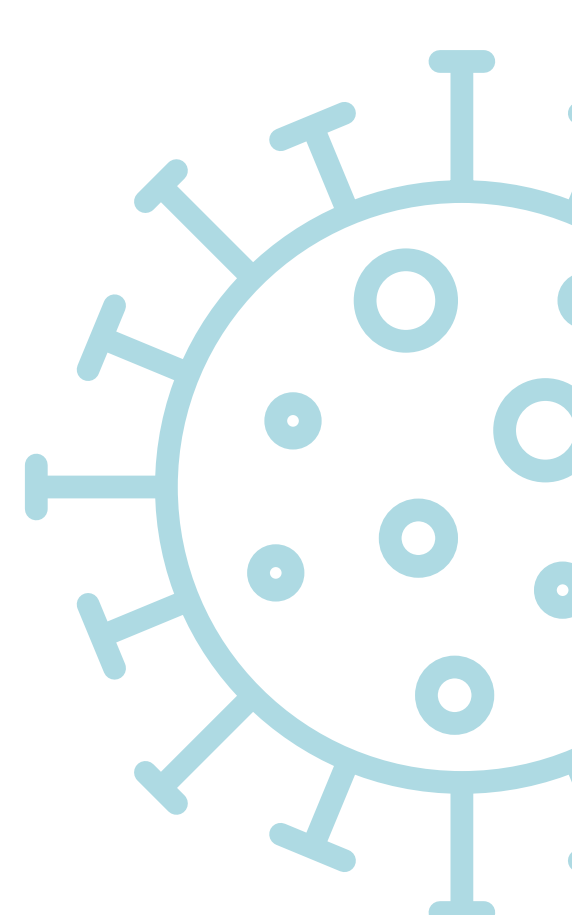


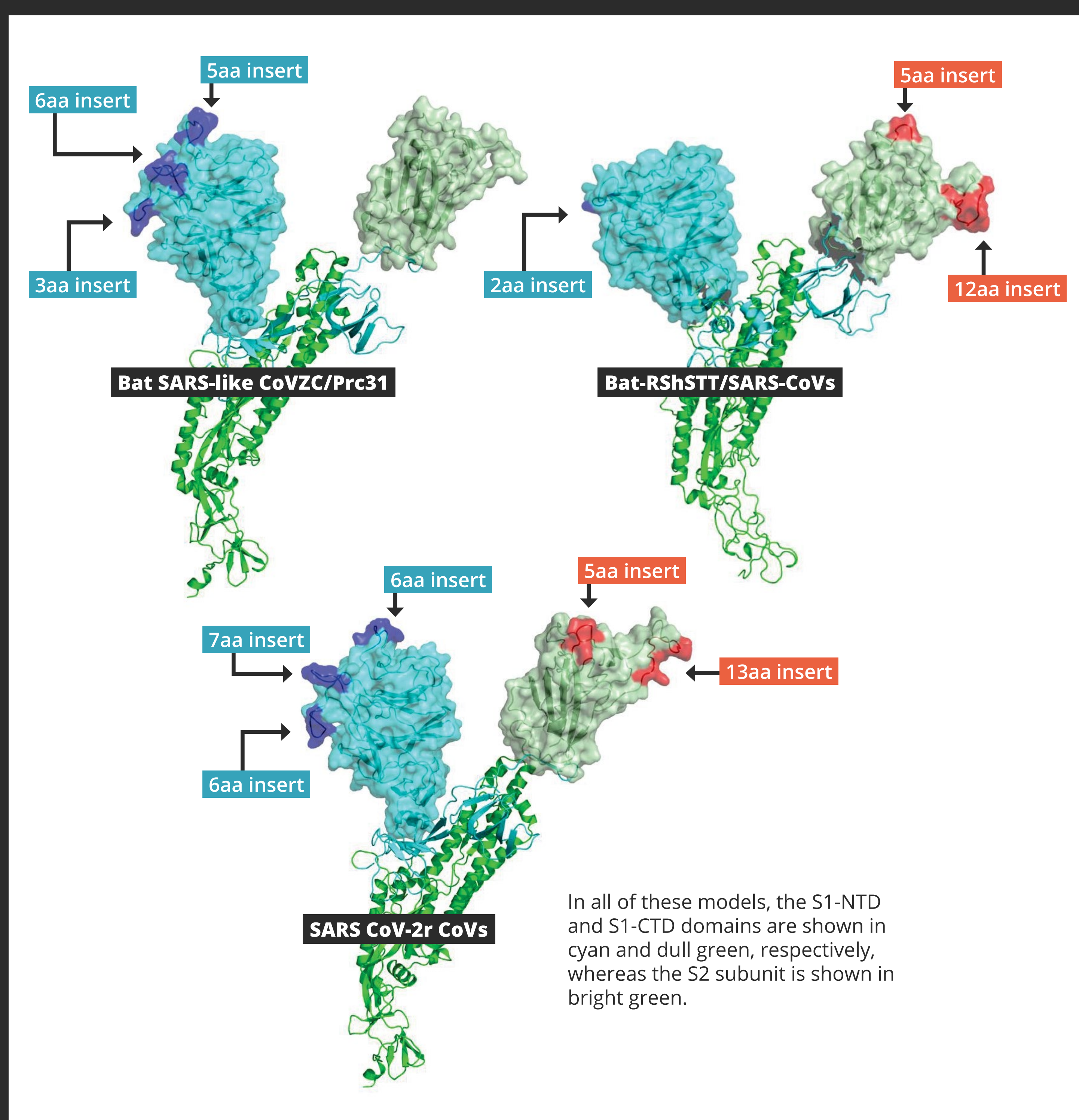
Conserved molecular signatures in the spike protein provide evidence indicating the origin of SARS-CoV-2 and a Pangolin-CoV (MP789) by recombination(s) between specific lineages of Sarbecoviruses

THIS STUDY

Both SARS-CoV-2 and SARS coronaviruses (CoVs) are members of the subgenus Sarbecovirus. **To understand the origin of SARS-CoV-2, sequences for the spike and nucleocapsid proteins from sarbecoviruses were analyzed.** Our aim was to identify molecular markers consisting of conserved inserts or deletions (termed CSIs) that are specific for either a particular clade of Sarbecovirus or are commonly shared by two or more clades of these viruses.



STRUCTURAL LOCATIONS OF THE CSIS SPECIFIC FOR SOME SARBECOVIRUSES IN THE SPIKE PROTEIN

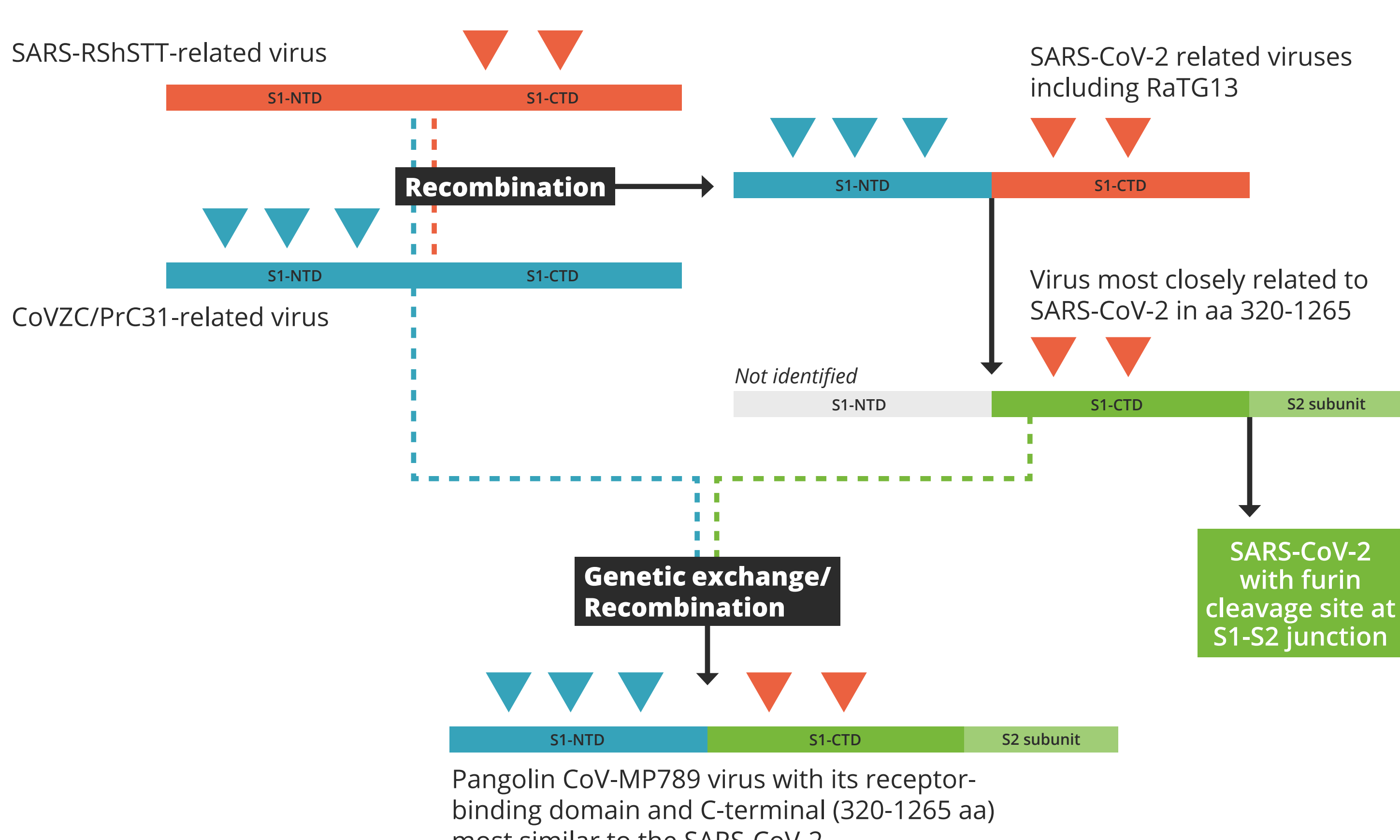


RESULTS

The distribution patterns of the identified signatures (CSIs) and sequence similarity studies reveal that the **SARS-CoV-2 (Covid-19)-related viruses have originated from the recombination of a Bat-CoVZC/Prc31-related virus and a SARS (or SARS-related) virus.**

Our results also show that **recombination between a virus very closely related to SARS-CoV-2 (>70% of the sequence for this virus known) and Bat-CoVZC/Prc31 virus has led to the formation of Pangolin-CoV (MP789).**

SEQUENCE CHARACTERISTICS OF THE SPIKE PROTEIN S1-REGION



CONCLUSIONS

Several other CSIs reported here are specific for other clusters of sarbecoviruses. Structural mapping studies show that the identified CSIs form distinct loops/patches on the surface of the spike protein. **It is hypothesized that these novel loops/patches on the spike protein should play important roles in the biology/pathology of SARS-CoV-2 virus.**

The CSIs specific for different clades of sarbecoviruses including the SARS-CoV-2r cluster provide **novel means for identification of these viruses and other potential applications.**