length to the corresponding proteins in SARS-CoV.

Of the four structural genes, SARS-CoV-2 shares more

than 90% amino acid identity with SARS-CoV except

for the S gene, which diverges''”’. The replicase gene

covers two thirds of the 5’ genome, and encodes a large

polyprotein (pplab),which is proteolytically cleaved into

16 non-structural proteins that are involved in transcrip-

tion and virus replication. Most of these SARS-CoV-2

non-structural proteins have greater than 85% amino

acid sequence identity with SARS-CoV”.

The phylogenetic analysis for the whole genome

shows that SARS-CoV-2 is clustered with SARS-CoV

and SARS-related coronaviruses (SARSr-CoVs) found

in bats, placing it in the subgenus Sarbecovirus of the

genus Betacoronavirus. Within this clade, SARS-CoV-2

is grouped in a distinct lineage together with four horse-

shoe bat coronavirus isolates (RaTG13, RmYN02, ZC45

and ZXC21) as well as novel coronaviruses recently iden-

tified in pangolins, which group parallel to SARS-CoV