polymorphism at nucleotide position 28,144, which results in amino acid substitution of Ser for Lys at residue 84 of the ORF8 protein. Those variants with this mutation make up a single subclade labelled as 'clade S'[33,34]. Currently, however, the available sequence data are not sufficient to interpret the early global transmission history of the virus, and travel patterns, founder effects and public health measures also strongly influence the spread of particular lineages, irrespective of potential biological differences between different virus variants.

Animal host and spillover

Bats are important natural hosts of alphacoronaviruses and betacoronaviruses. The closest relative to SARS-CoV-2 known to date is a bat coronavirus detected in *Rhinolophus affinis* from Yunnan province, China, named 'RaTG13', whose full-length genome sequence is 96.2% identical to that of SARS-CoV-2 (REF.[11]). This bat virus shares more than 90% sequence identity with SARS-CoV-2 in all ORFs throughout the genome, including the highly variable S and ORF8 (REF.[11]). Phylogenetic analysis confirms that SARS-CoV-2 closely clusters with RaTG13 (FIG. 2). The high genetic similarity between SARS-CoV-2 and RaTG13 supports the hypothesis that SARS-CoV-2 likely originated from bats[35]. Another closely related coronavirus has been identified in a *Rhinolophus malayanus* bat sampled in Yunnan. This novel bat virus, denoted 'RmYN02',