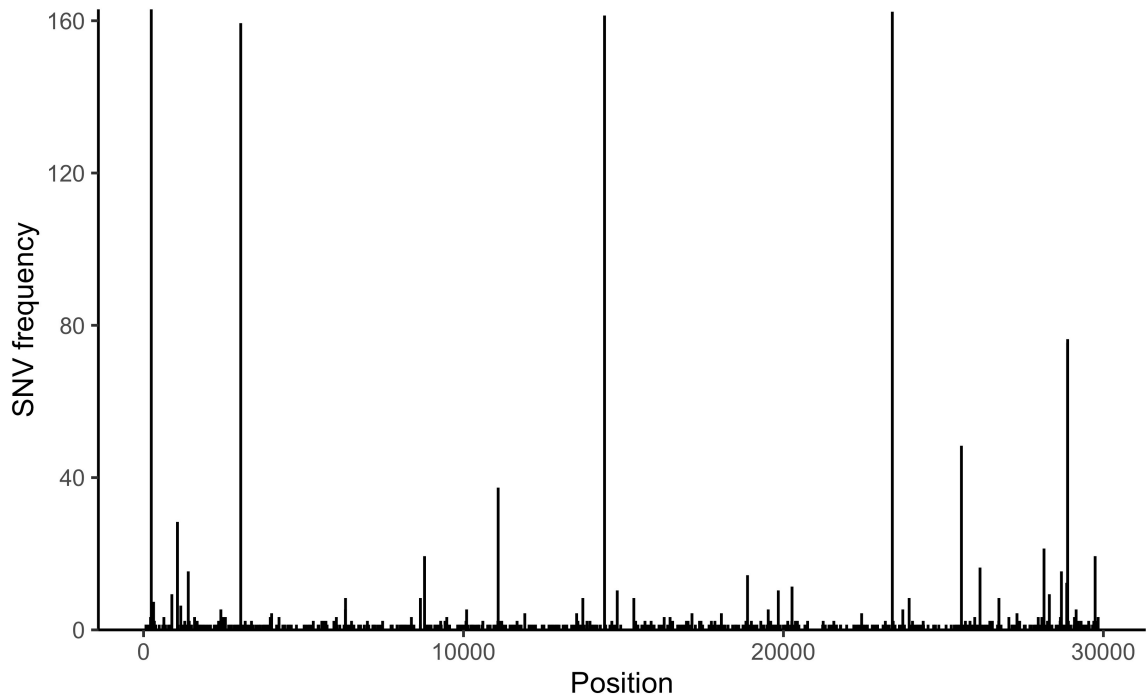
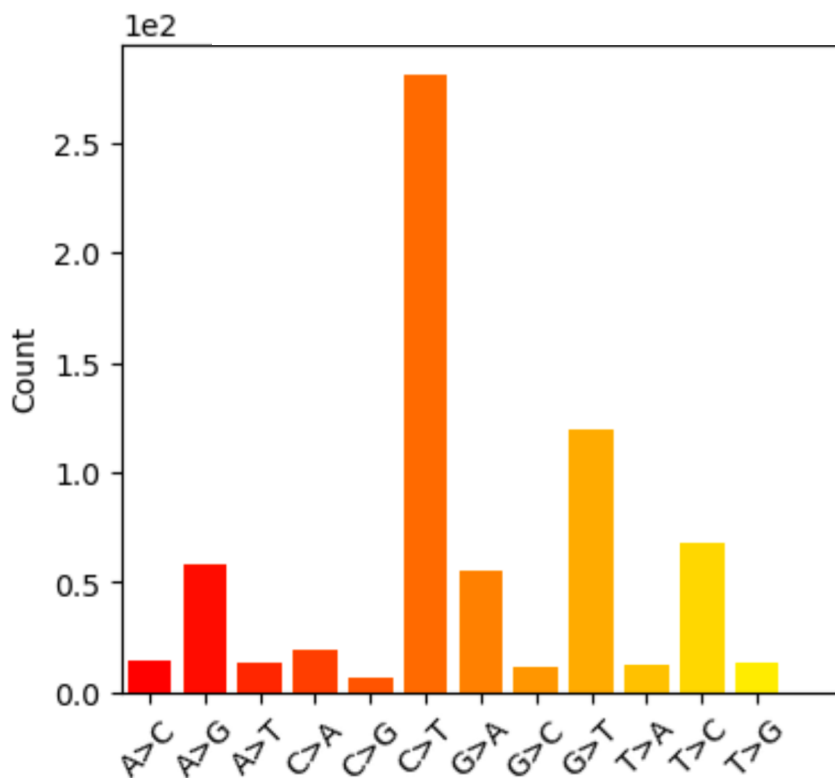


a



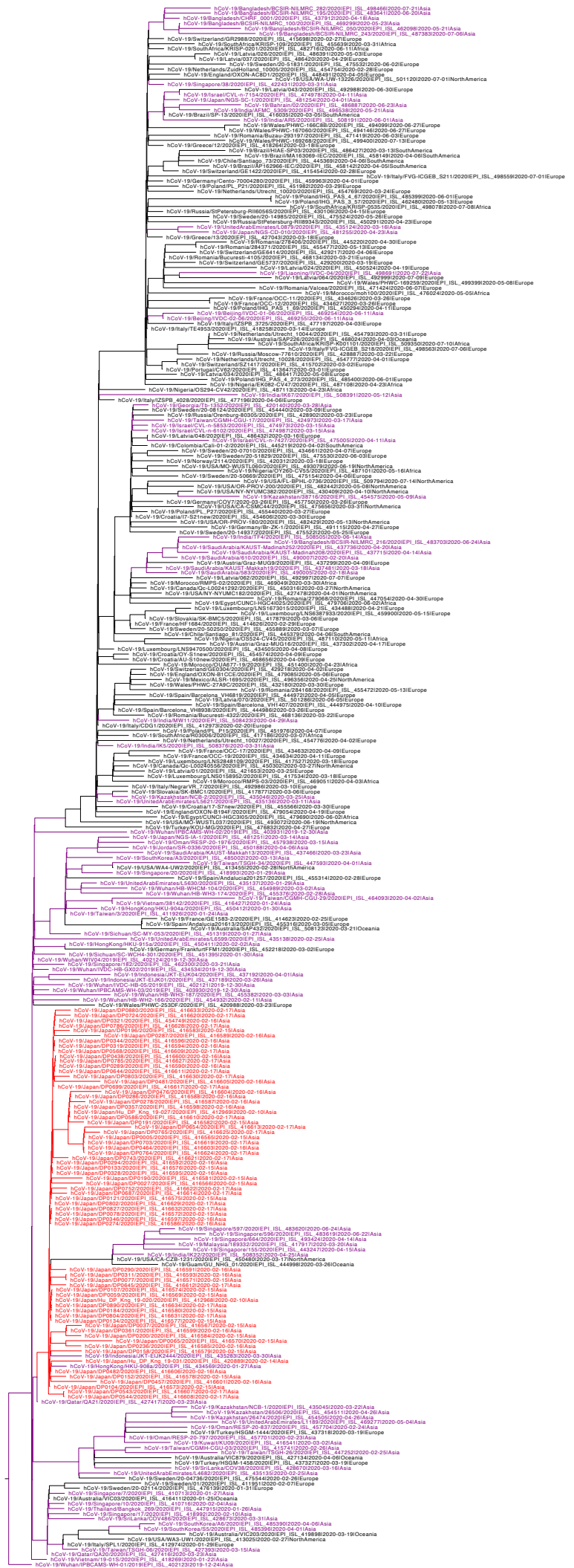
b

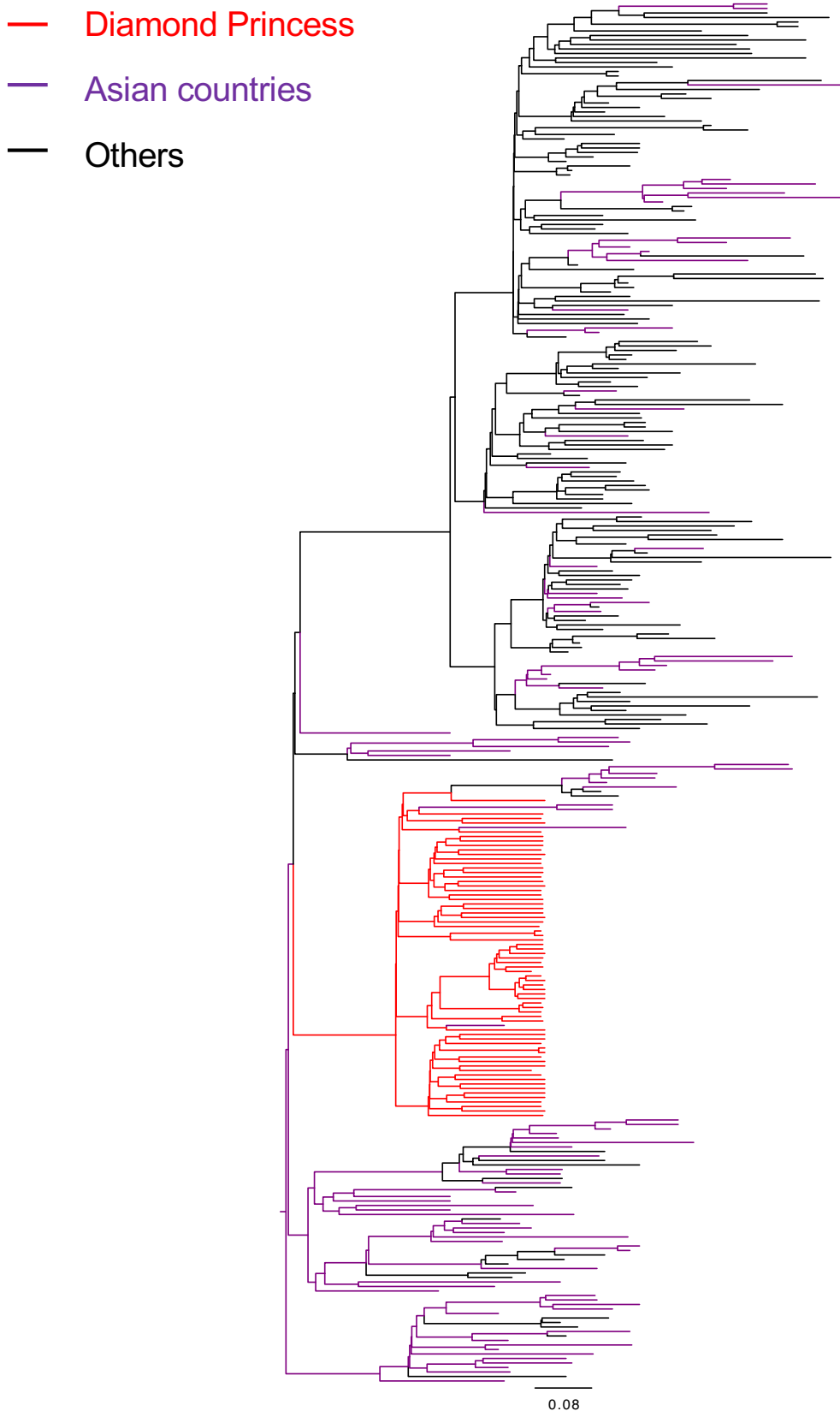


Supplementary Figure 1. Characteristics of single nucleotide variants in the reference dataset (n = 240). As error-rich regions, 50 nucleotides at the 5' and 3' terminals were masked. (a) Distribution of genetic variations in the reference sequences. Positions of genetic variations are presented according to NC_045512.2. (b) Detected nucleotide transition and transversion.

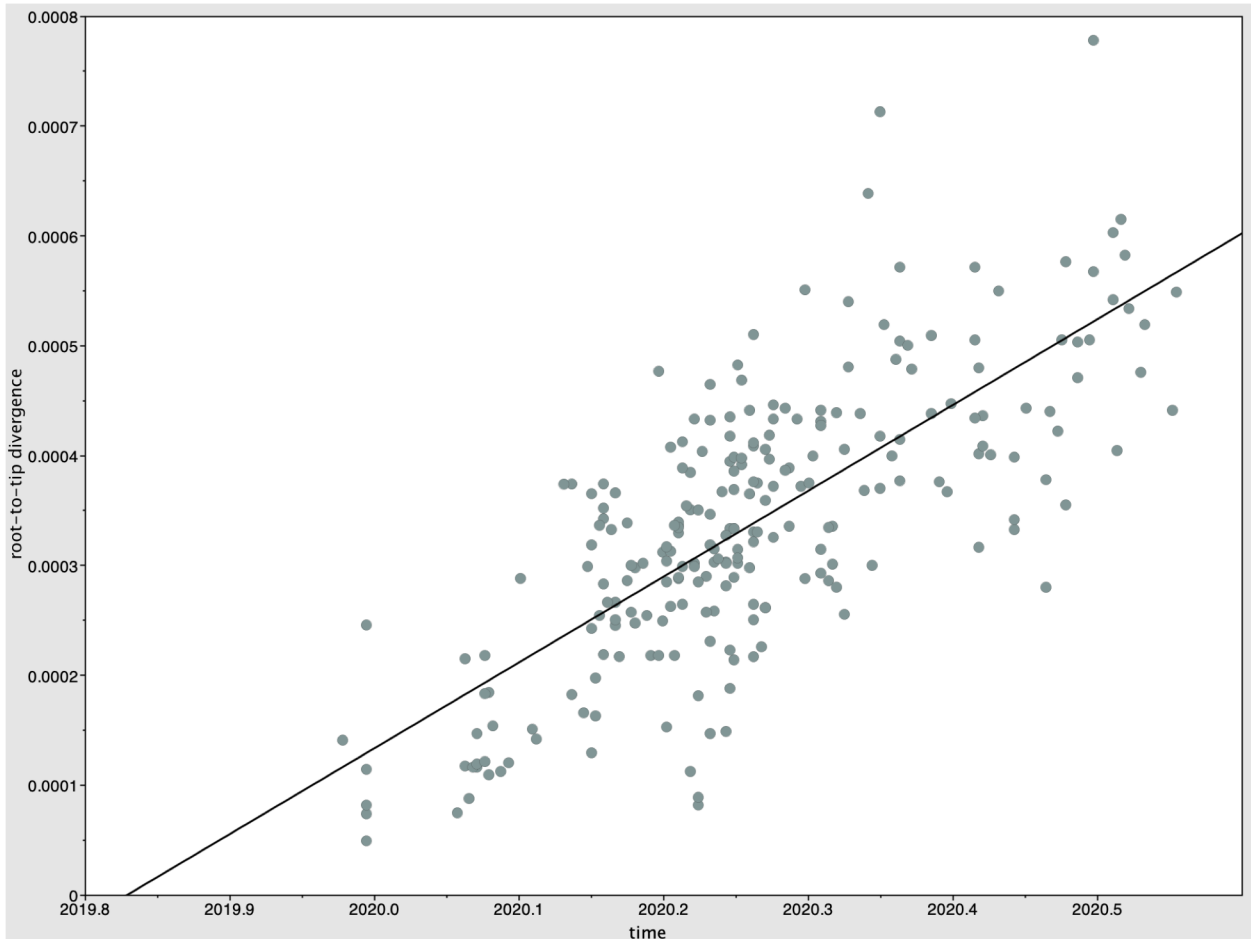
- Diamond Princess
- Asian countries
- Others

Supplementary Figure 2. Maximum likelihood tree based on SARS-CoV-2 whole-genome sequences of 240 global reference strains and 67 isolates sampled from individuals on board the *Diamond Princess*. Branches and tip names of sequences from the *Diamond Princess* and those from Asian countries are colored red and purple, respectively. The tree was rooted using the earliest sample in the dataset (IPBCAMS-WH-01, GenBank entry: MT019529.1). Branch length reflects units of nucleotide substitutions per site according to the provided scale.





Supplementary Figure 3. Maximum clade credibility tree based on SARS-CoV-2 whole-genome sequences of 240 global reference strains and 67 isolates sampled from individuals on board the *Diamond Princess*. Branches of sequences from the *Diamond Princess* and those from Asian countries are colored red and purple, respectively. The branch length reflects units of year according to the scale provided.



Supplementary Figure 4. Root-to-tip regression analysis of the maximum likelihood (ML) phylogenetic trees. ML phylogenetic trees were constructed using the SARS-CoV-2 whole-genome sequences of 240 reference sequences using RAxML (v.8.2.9) with 1000 replicates. Root-to-tip regression analysis was performed using Tempest (v. 1.5.3). The estimated mean evolutionary rate was 7.82×10^{-4} substitutions per site per year ($R^2 = 0.556$), and the estimated time to the most common ancestor was 2019.83 (November 30, 2019).