# Genetic Diversity and Conservation of E Gene of SARS-CoV-2 and its variants of concern with comparison with other human coronaviruses

**Benjamin M. Liu**

**Division of Pathology and Laboratory Medicine, Children's National Hospital, Washington, DC;**
**Departments of Pediatrics, Pathology, and Microbiology, Immunology & Tropical Medicine, George Washington University School of Medicine and Health Sciences, Washington, DC**

## 1. Abstract

**Introduction:** SARS-CoV-2 E gene PCRs have been widely used as the first-line assay with a higher sensitivity than those targeting N or RdRp gene. Given the currently available primers and probes were designed at the onset of the pandemic, it is unknown whether the SARS-CoV-2 VOCs have accumulated significant mutations that may affect E gene PCRs. In this study we aim to perform a comprehensive genetic analysis of SARS-CoV-2 E gene sequences to evaluate the impact of the emerging VOCs on E gene PCR performance.

**Methods:** 600 whole-genome sequences of 7 species of human coronaviruses (HCoVs) were retrieved from GenBank and GISAID, including Sarbecoviruses (SARS-CoV-2 variants B1.1.7, B1.351, P.1, B.1.617.2 and B.1.1.529, and SARS-CoV), Embecovirus (OC43, HKU1), Merbecovirus (MERS) and Alphacoronaviruses (229E, NL63). The E gene sequences were retrieved from full-length genomes of corresponding viruses and aligned by ClustalW multiple alignment. Phylogenetic, conservation and mutation analyses analysis of the enrolled sequences was performed.

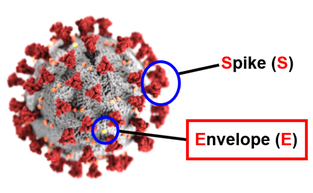
**Results:** E gene-based phylogenetic analysis yielded HCoVs typing results consistent with whole genome typing, suggesting E gene is a reliable locus for phylogenetic analysis and typing of HCoVs. Four pan-Sarbecovirus conserved E gene regions were identified with 100% conserved nucleotide similarity among SARS-CoV-2 and its VOCs, as well as SARS-CoV. These regions have appropriate G/C content which may be suitable for primer/probe design for E gene-based pan-Sarbecovirus screening assay. No significant E mutations were found in 137 retrieved SARS-CoV-2 and its VOCs. Interestingly, two novel variations, C26299U and T26354A, were identified in two of our SARS-CoV-2 strains. The latter variation occurred at the 3' end of the target region of the widely used Charité/Berlin (WHO) probe. This variant may lead to a potential failure of the first-line E gene PCR.

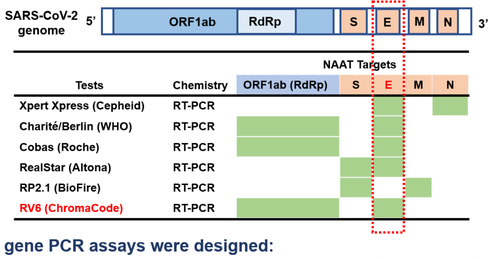
**Conclusions:** Our data shed light on the genetic diversity and conservation of E gene of SARS-CoV-2 and may be beneficial for future primer/probe design for novel first-line assay or SARS-CoV-2-specific E gene PCR. SARS-CoV-2 VOCs have not accumulated significant mutations in E gene so far. The impact of novel E gene variations C26299U and T26354A on molecular diagnostic testing warrants further investigation.

## 2. Rationale

### Will SARS-CoV-2 E gene PCRs continue to work?

- PCRs targeting E gene (E gene PCRs) that have been widely used as a first-line assay:
  - target pan-Sarbecovirus conserved regions in E gene (pan-Sarbecovirus detection)
  - have higher sensitivity than those targeting RdRp or N gene




- The currently available SARS-CoV-2 E gene PCR assays were designed:
  - at the onset of the pandemic
  - based on the reference sequence of the original "wild-type" viruses

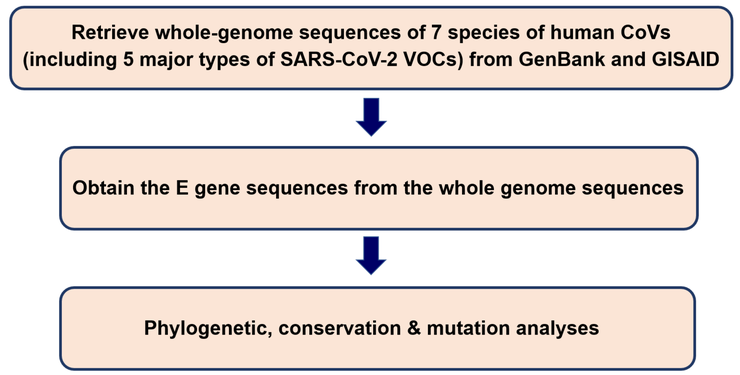
Peng et al. Nature 2020
Fung et al. EMI 2020
Tang YW et al. JCM 2020
Alanagreh et al. Pathogens 2020

**Have the SARS-CoV-2 VOCs accumulated significant mutations that may affect E gene PCRs?**

## 3. Aim of the study

**Study genetic conservation and diversity of E gene sequences of SARS-CoV-2 and its VOCs with comparison with other human CoVs**

## 4. Study design




## 5. Methods and Results

### 5.1. E gene-based phylogenetic analysis yielded CoV typing results consistent with whole genome typing

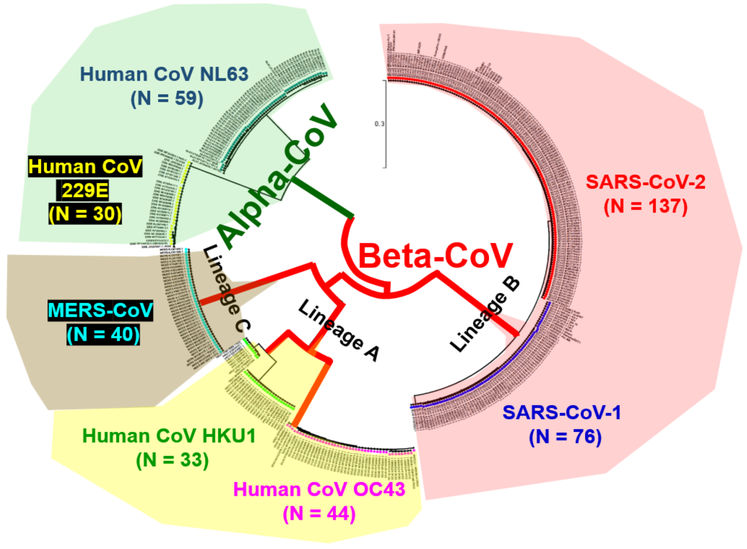



**Fig. 5.1 E gene-based phylogenetic analysis yielded CoV typing results consistent with whole genome typing. A total of 419 whole genome sequences of human CoVs were retrieved. The classification of these CoVs were made based on whole genome typing. When the E gene sequences of these human CoVs were used to perform this phylogenetic analysis, it yielded tying results consistent with the whole genome typing.**

### 5.2. Alignment of E gene sequences of all retrieved human CoVs

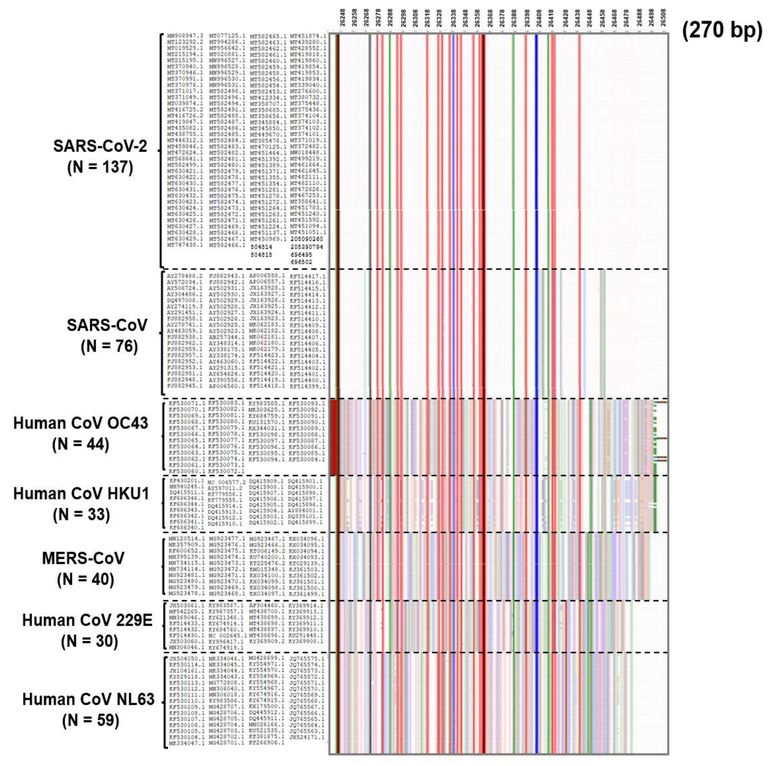



**Fig. 5.2 Alignment of E gene sequences of all 419 retrieved human CoVs. To examine genetic conservation and diversity of E gene, we performed Clustal W alignment of E gene sequences of all 419 retrieved human CoVs. This is a snapshot of the multi-sequence alignment, with accession numbers listed aside.**

### 5.3. Point mutation analysis for SARS-CoV-2

- **No significant E mutations were found in 137 retrieved SARS-CoV-2 and its VOCs**
- **C26299U was identified in one SARS-CoV-2 strain MT630422 (not a VOC)**

## 6. Conclusions

- **Our data shed light on the genetic diversity and conservation of E gene of SARS-CoV-2 and may be beneficial for future primer/probe design for novel first-line assay or SARS-CoV-2-specific E gene PCR.**
- **SARS-CoV-2 VOCs have not accumulated significant mutations in E gene so far. The impact of novel E gene variations C26299U and T26354A on molecular diagnostic testing warrants further investigation.**

Correspondence: Benjamin Liu (Email: bliu1@childrensnational.org)