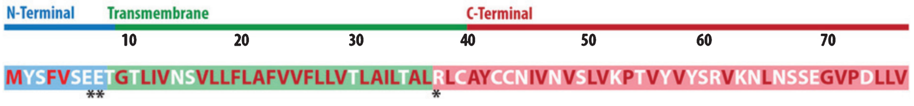
The M protein is the most abundant structural protein in the viral envelope. It interacts with all the major structural proteins in the CoV assembly process. The necessary interaction of S and M proteins maintains the S protein in the endoplasmic reticulum/Golgi complex for integration into new virions. The combination of the M and E protein consists to the structure of the viral envelope.

The E protein is the smallest of the structural proteins (containing 76–109 amino acids, in the range of 8.4 to 12 kDa in size), but it is the most important for viral replication [15]. During replication, the E protein is upregulated on the infected cell endoplasmic reticulum, and only a small quantity is incorporated into the new virion envelope. Studies have shown that recombinant CoVs that lack the E protein exhibit reduced viral titters and impaired viral maturation [15]. The CoV E protein has a short, hydrophilic amino terminus comprised of 7–12 amino acids. This is followed by a large hydrophobic transmembrane domain consists of 25 amino acids, and ending with an extended, hydrophilic carboxyl terminus that constitutes the majority of the protein. The transmembrane domain has two neutral amino acids: valine and leucine, that account for the E protein's hydrophobicity [26] (Fig. 2). The SARS-CoV E protein has a PDZ binding motif located in the last four amino acids of the C terminus. On co-immunoprecipitation and pull-down assays, the SARS-CoV E PDZ domain binds to PALS1 in mammalian cells, a tight junction-associated protein crucial for the establishment and maintenance of epithelial polarity in mammals [25]. Such functions and interactions exhibit multiple ways in which the E protein critically mediates SARS-CoV pathogenesis [25].

Analyses of 103 genetic populations of SARS-CoV-2 genomes showed that these viruses developed into two dominant strains, called L and S. The S type is the original strain and was less aggressive and less prevalent than L (30% versus 70% of cases, respectively). The L type strain was already more frequent in the early stages of the Wuhan outbreak, but decreased in frequency after early January 2020 [31]. Later this year, further mutations were found, defining additional strains called A, B, and C [32]. Over time It is normal to collect random genomic mutations, which depend on age. Genomes develop variable mutations, constituting markers of disease spread. By building a phylogeny, it is possible to gather information about the epidemiological phenomena occurring to such pathogens, such as spread, their timeline of appearance and epidemic growth rate [33]. There are websites dedicated to tracking strains and updated daily, as the numbers of strains have increased [33]. Nextstrain.org identifies 5 major clades that correspond to GISAID nomenclature (in parenthesis): 19A (L) & (V), 19B (S), 20A (G), 20B (GR), 20C (GH). Different methods of classification and nomenclature are in place [34]. The GISAID initiative has classified the major SARS-CoV-2 clades and named them according to marker mutations in 6 major phylogenetic groups starting from the initial split of S and L, to the later division of L into V and G and later of G into GH and GR. So far 6 major clades have been identified based on 9 marker variants [34]:.

- **S**: C8782T, T28144C includes NS8-L84S
- **L**: C241, C3037, A23403, C8782, G11083, G25563, G26144, T28144, G28882 (WIV04-reference sequence)
- **V**: G11083T, G26144T NSP6-L37F+NS3-G251
- **G**: C241T, C3037T, A23403G includes S-D614G
- **GH**: C241T, C3037T, A23403G, G25563T includes S-D614G+NS3-Q57H
- **GR**: C241T, C3037T, A23403G, G28882A includes S-D614G+N-G204R [34].

Based on the data available from GISAID the Geographical worldwide distribution of COVID-19 clades (GISAID) is illustrated in Fig. 3.


N-Terminal
Transmembrane
C-Terminal
10
20
30
40
50
60
70
MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPTVYVYSRVKNLNSSEGVPDLLV
**
*

Amino Acid Sequence and Domains of the SARS-CoV E Protein. The SARS-CoV E protein consists of three domains, i.e. The Amino (N)-Terminal domain, the transmembrane domain (TMD), and the Carboxy (C)-Terminal domain. Amino acid properties are indicated: Hydrophobic (red), Hydrophilic (white) polar, charged (asterisks)

Fig. 2. Illustration SARS-CoV Protein E. Adapted from [15]. *Illustrated by Dr. Joe Bolanos.*