Analysis of SARS-CoV-2 spike protein mutations in the UK

Abstract

Introduction

**Background: COVID-19 pandemic**

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the strain of coronavirus responsible for the ongoing COVID-19 pandemic (source). SARS-CoV-2 was initially detected in Wuhan, China, and has now spread to all corners of globe (source). Approximately 770,000,000 people have contracted the virus and over 7,000,000 people have died due to complication caused by the virus (source). This figure is most likely even higher as a result of government's inability to report cases, as well as, corruption with countries attempting to hide the true values. An estimated value of the number of deaths sits at around 18,000,000 – 32,000,000 (source).

**Background: SARS-CoV-2 structural and molecular information**

**Background: Spike glycoprotein**

**Aims of the investigation:**

1. **Investigate the clustering and distribution of SARS-CoV-2 spike protein mutations in the UK**
2. **Identify and characterise common spike mutations + non vs syn mutations**
3. **Analyse their potential effects on viral infectivity**

**Methods overview**

1. **Summarise the dataset used**
2. **Mention of the clustering techniques used, R-based analysis, various packages used and tools used for visualisation**

**Summary of the main findings**

Results

**Clustering and the distribution of mutations**

1. **Geographic distribution of spike protein mutations**
2. **Map visualising the distribution in the UK**

**Mutation analysis**

1. **Summary of the most common mutations**
2. **Non vs syn mutations**
3. **Frequency of mutations**

**Evolutionary Insights**

1. **Phylogenetic tree highlighting relationships among clusters**

**Impact of mutations**

1. **Link mutations to known functional effects**

Discussion

Conclusion