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 (Hu et al., 2021). SARS-CoV-2 was initially detected in Wuhan, China, and has now spread to all corners of globe (Huang et al., 2020). Approximately 770,000,000 people have contracted the virus and over 7,000,000 people have died due to complications caused by the virus . This figure is most likely even higher as a result of governments’ 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). The economic burden has been far-reaching, estimated losses are up to $16 trillion worldwide and unemployment levels have soared (source). Nationwide lockdowns, that existed at the height of the pandemic, have not only had an economic effect, but, have also impacted education and literacy rates. It has been predicted that by 2030 less than 40% of pupils in the UK will achieve a pass in GCSE English and Mathematics (source). SARS-CoV-2’s effects will continue to resonate, even as though the virus becomes less prevalent.

**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**

Figure 1, Number of sequences per day

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

**Mutation analysis**

What I did?

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

**Interpret the results**

1. **Correlation between geographical distribution and specific mutations**
2. **Mutation hotspots? Natural selection + selective pressure**

**Link back to previous studies**

**Implications for public health**

1. **Vaccine design**
2. **Impact of mutations on diagnostics and therapeutic intervention**

**Limitations and future directions**

1. **Limitations of the dataset**
2. **Limitations of techniques used**
3. **Future research**

Conclusion

**Summary of the main findings**

**Importance of studying SARS-CoV-2**

Methods and Materials

**Data sources**

**R packages used**

**Statistical analysis**

**Any additional stuff used**

**Link to GitHub**

Hu, B. et al. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nature reviews. Microbiology*, 19 (3), pp.141–154.

Huang, C. et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*, 395 (10223), pp.497–506.