# Examining SARS-CoV-2 variant data in the Spanish capital and the Balearic Islands

Alexandra Palacios

19 May, 2021

## Background and Overview

In December of 2019, the first few cases of the novel SARS-CoV-2 virus began circulating around Wuhan, China. In the following months, the virus made its way across country boundaries and was eventually declared a worldwide pandemic by the World Health Organization on March 11th, 2020. The very first SARS-CoV-2 case in Spain was reported on January 31st, 2020 in the Canary Islands (Ferrer, 2020). This first case from a German citizen was linked to a cluster of cases from Bavaria, Germany (Spiteri *et al.*, 2020). On February 9th, 2020 the first few cases were reported in the city of Palma on the island of Mallorca, in the Balearic Islands (Ferrer, 2020). Then, on February 24th, 2020 the first cases of the virus were reported in the Spanish mainland, and the first case in Madrid was reported the next day (Ferrer, 2020). Several weeks later, Spain, along with a couple other countries in western Europe, became the epicenter of the pandemic (Henríquez *et al.*, 2020). During this time, strict travel bans were set in place for international and national travel within Spain.

During this unique time, inhabitants of the Balearic Islands found themselves in a state of isolation. The Balearic Islands are a popular tourist spot in the Mediterranean Sea, but due to the travel restrictions, inhabitants of the islands not only found themselves isolated from the Spanish mainland, but also from island to island (Henríquez *et al.*, 2020); (Eguíluz *et al.*, 2020). Since then, travel restrictions for these islands have fluctuated in severity and as of May 2021, tourism to the islands is once again permitted for international travelers from certain countries. However, it is not known if this period isolation has had a lasting impact on the variants present in individuals on the islands and how much they may differ from those present in the Spanish mainland. Here, I analyzed SARS-CoV-2 variants in regions within the Balearic Islands of Ibiza and Mallorca, and compared them with variants found in the Spanish capital, Madrid. After downloading fastq data from 150 samples and configuring them into vcf files using a bash-pipeline adapted from Koyama *et al.* (2020) and parsing out SARS-CoV-2 variants in R, I found several variants from the island of Ibiza that were not found in Mallorca or Madrid. Additionally, the variants found in Mallorca and Madrid are also ones commonly found in Europe while those from Ibiza were not commonly found anywhere else. One variant that was common across all regions was the 23403 A $>$ G variant in the spike protein, which is also commonly found elsewhere in Europe.

## Methods

### Data Collection

I selected and downloaded a total of 150 SARS-CoV-2 samples from the NCBI Bioproject titled "WGS of SARS-CoV-2 circulating in Spain". The SRA Bioproject ID is PRJEB43166 and samples in this project were sequenced by the SeqCOVID-Spain consortium. This bioproject collected SARS-CoV-2 variant data from individuals of different ages, sex, and geographic regions within Spain. 100 of the samples I selected came from localities within two of the Balearic Islands in Spain (Ibiza and Mallorca) and were collected by Servicio de Microbiología, Hospital Universitario Son Espases. The other 50 samples came from the Spanish capital, Madrid, and were collected by Hospital General Universitario Gregorio Marañón. All samples were originally collected from individuals between January 01, 2021 and February 28, 2021.

## Variant Analysis

Using a bash pipeline created by Koyama *et al.* (2020) and modified by Naupaka Zimmerman, I downloaded all the raw Illumina fastq data selected from the SRA Bioproject, checked the data quality, trimmed unwanted sequence data, indexed and mapped sequences against the SARS-CoV-2 reference genome, sorted and processed reads, and configured reads to be processed in R as vcf files (Koyama *et al.*, 2020). Using the vcfR package in R (Knaus and Grünwald, 2017, 2016), I modified all vcf files to create a dataframe that included variant data along with genome annotations and metadata of the samples from the NCBI Bioproject. Finally, variants for each region and gene were parsed and plotted using the dplyr, ggplot2, and ggthemes packages in R (Wickham *et al.*, 2020; Wickham, 2016; Arnold, 2021). This entire pipeline was driven by a Makefile.

# Results

## Demographics of Sample Population

Of all 150 samples collected, 73 samples came from females while 77 came from males (**Figure 1**). The majority of samples came from individuals ages 20 to 50 years old, however; samples from individuals ages 1 to 20 and 50 to 100 years old were well represented in this sample population (**Figure 2**). There was also a relatively even representation of samples originating from Madrid, Ibiza, and Mallorca (**Figure 3**).

## Variant Analysis

For the entire sample population, the frequency of distinct SNPs was greatest for the spike protein gene, followed by the nucleocapsid gene, and the ORF3a gene (**Figure 4**). This trend was also found for the samples collected from Madrid and Mallorca, however; in Ibiza, the ORF8 gene had the same number of distinct SNPs as the ORF3a gene (**Figures 5, 6, 7**). In total, SNPs from the following known SARS-CoV-2 genes are represented in this sample population: E, M, N, ORF10, ORF3a, ORF6, ORF7a, ORF7b, ORF8, S (**Figure 4**). SNPs for the E gene were found in Madrid and Ibiza, SNPs for the ORF6 gene were only found in Mallorca, and SNPs for the ORF7b gene were found in Madrid and Mallorca (**Figures 5, 6, 7**).

The most common SNP for all three regions was the 23403 A > G mutation in the spike protein gene (**Figure 8; Tables 1, 2, 3**). This mutation translates to a D614G mutation in the protein sequence of the spike protein (Koyama *et al.*, 2020). Other variants in the spike protein gene that had at least two occurrences in all three regions are 21990 TTTATTA > TTTA, 22227 C > T, 23063 A > T, 23271 C > A, 23604 C > A, 23709 C > T, 24506 T > G, and 24914 G > C (**Figure 8; Tables 1, 2, 3**). Variants that had at least two occurrences in all regions for the nucleocapsid gene are 28280 G > C, 28281 A > T, 28282 T > A, 28881 G > A, 28882 G > A, 28883 G > C, 28932 C > T, 28977 C > T, and 29402 G > T (**Figure 9, Tables 4, 5, 6**). No common variants were present in all three regions for the ORF3a gene (**Figure 10**). Additionally, the count of unique SNPs in Madrid and Mallorca in the ORF3a gene were higher than the count of unique SNPs in Ibiza. However, the frequencies of each distinct SNP in the ORF3a gene did not exceed 5 counts for Madrid and Mallorca, while Ibiza had 3 distinct SNPs that exceeded 20 counts in this gene (**Figure 10; Tables 7, 8, 9**).

All regions had SNPs that weren't found in either of the other regions. However, only Ibiza had SNPs unique to its region that were present in high frequencies (greater than 15 occurrences). The following are all the SNPs unique to Ibiza that were detected at least 15 times in its population: 21855 C > T (S gene), 22909

T > C (S gene), 28651 C > T (N gene), 28747 G > T (N gene), 28869 C > T (N gene), 29422 G > T (N gene), 25505 A > G (ORF3a gene), 25906 G > C (ORF3a gene; **Figures 8, 9, 10; Tables 3, 6, 9**).

# Discussion

The 23403 A > G variant found most commonly in all three regions is also one of the most common spike protein variants found in Europe (Koyama *et al.*, 2020; Pachetti *et al.*, 2020). This D614G protein sequence change of the spike protein was also found to be one of the most dominant variants in Madrid during the first wave of the pandemic and has also found to be associated with higher death rates in Europe (Viedma *et al.*, 2021; Toyoshima *et al.*, 2020). Interestingly, the most common international ORF3a variant, 25563 G > T (Q57H protein sequence change), was found in a much lower frequency than other variants in Madrid and Ibiza, and it wasn't found at all in Mallorca (Koyama *et al.*, 2020; Bianchi *et al.*, 2021). A further investigation into the times that these variants first appeared and the times where the travel restrictions in Spain were most strictly enforced could explain why the 23403 A > G variant is present in this sample population and why the the 25563 G > T variant isn't as represented here.

Overall, the data demonstrate that the island of Ibiza has the most distinct SARS-CoV-2 SNP profile out of all three regions analyzed, particularly in the ORF3a and nucleocapsid genes. Given that travel was heavily restricted from the Spanish mainland to the Balearic Islands, it is surprising that samples from Mallorca did not have a similarly distinct SNP profile as its neighboring island. However, by the time all samples were collected, strict travel restrictions from the islands to the Spanish mainland had already been lifted to some extent. This could explain the why Madrid and Mallorca had similar SNP profiles. Ibiza also shared a lot of the same SNPs that were found in Madrid and Mallorca; however, the frequency of distinct SNPs that were only found on Ibiza suggest that individuals from Ibiza may have been more isolated from the Spanish mainland than Mallorca. The distinct SNPs found on Ibiza are also not commonly found elsewhere in Europe or any part of the world, which could suggest that these variants originated in Ibiza and not elsewhere (Koyama *et al.*, 2020; Pachetti *et al.*, 2020).

Another interesting find is that Mallorca and Madrid seem to have a higher relative frequency of distinct SNPs in the ORF3a gene than Ibiza. Despite having a lower variation in distinct SNPs, Ibiza had much higher occurrences of its most common SNPs for the ORF3a gene than Mallorca and Madrid had for their common SNPs in this gene. This can also suggest that the ORF3a variants found in Ibiza did in fact originate there.

## Figures

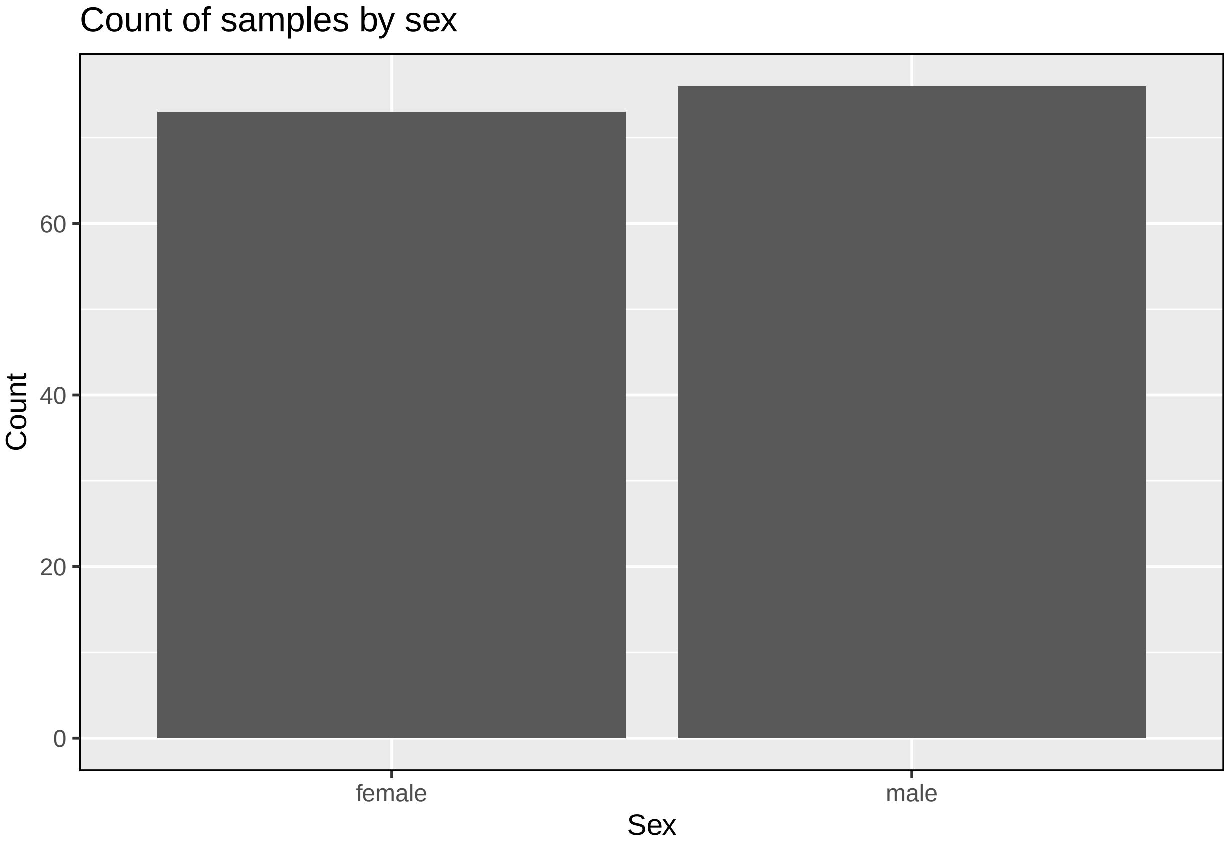

**Figure 1**: Count of individuals in the sample population by sex.

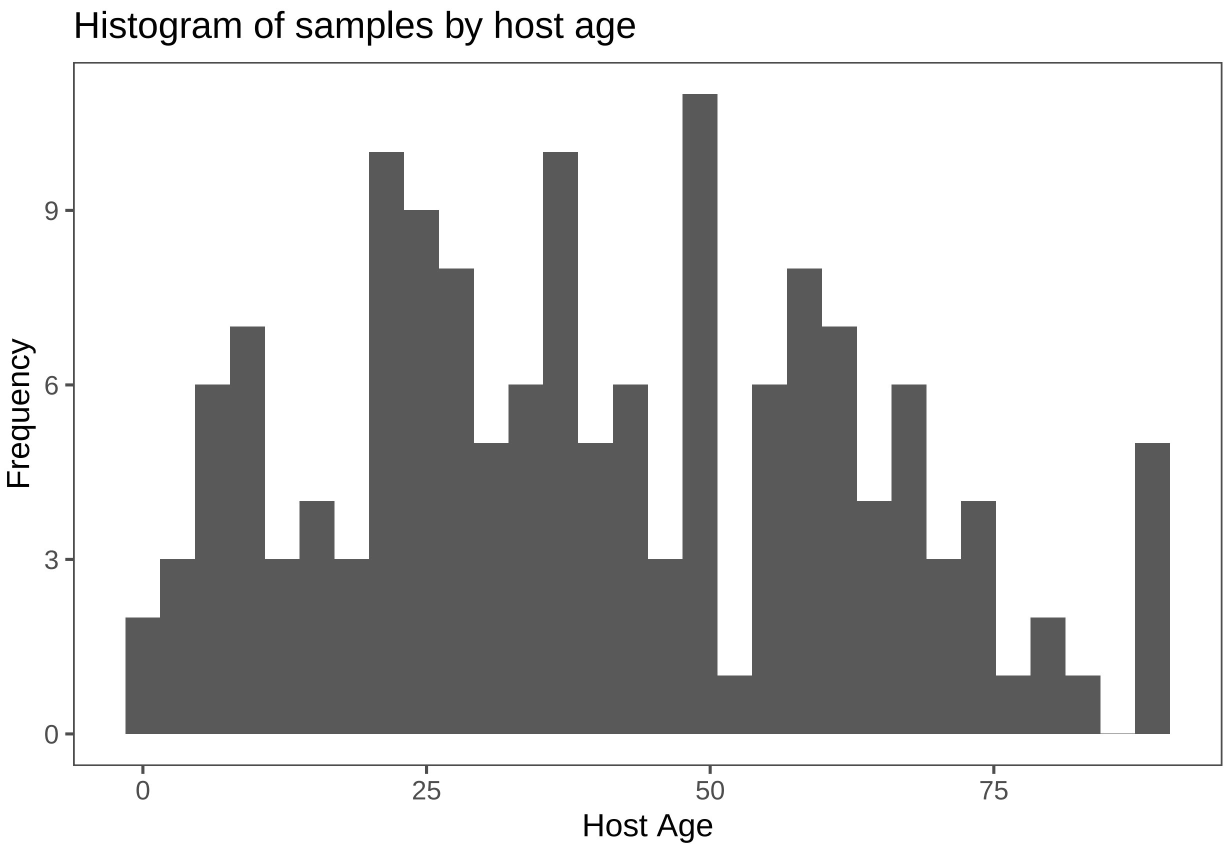

**Figure 2**: Distribution of individuals in the sample population by age.

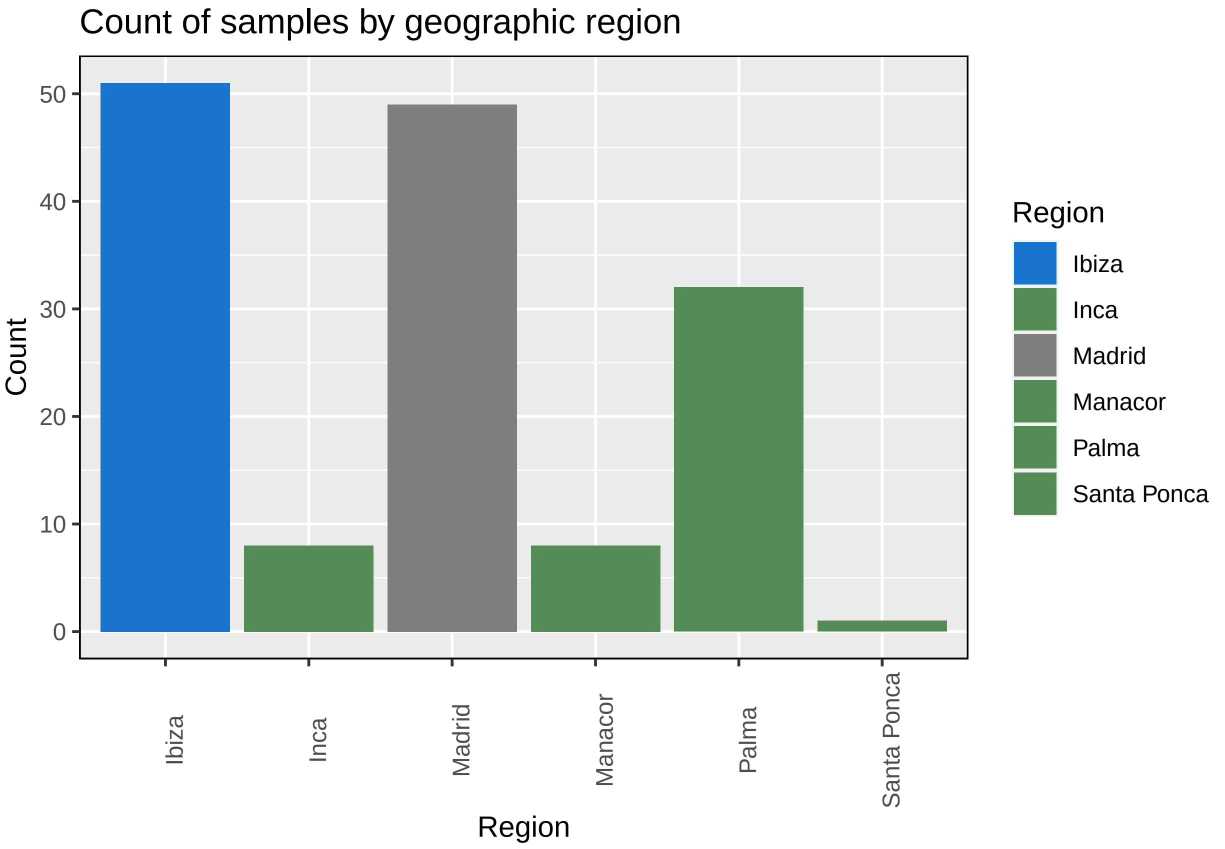

**Figure 3**: Counts of samples from each of the geographic regions represented in the sample population. All regions in green are localities within the island of Mallorca, regions in blue come from the island of Ibiza, and regions in gray come from the Spanish capital, Madrid.

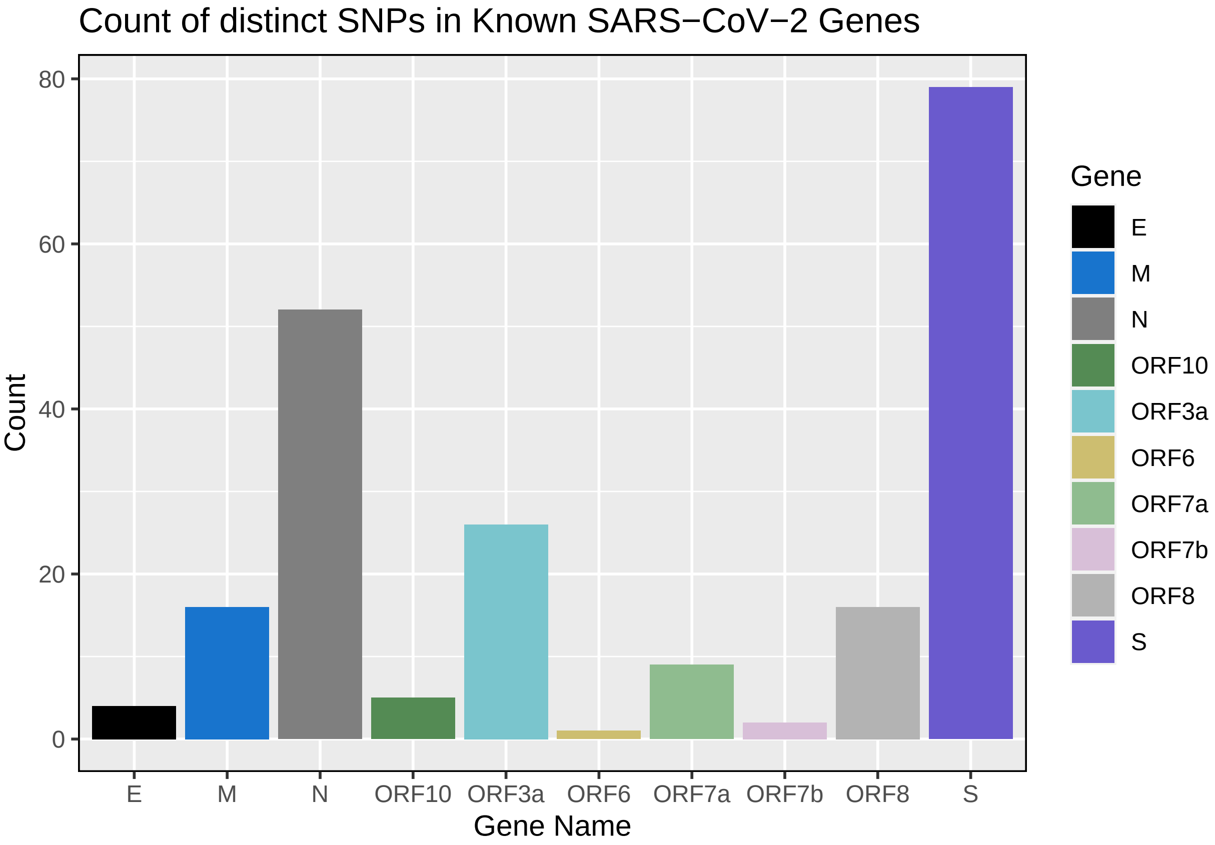

**Figure 4**: Frequencies of distinct SNPs of SARS-CoV-2 genes for the whole sample population.

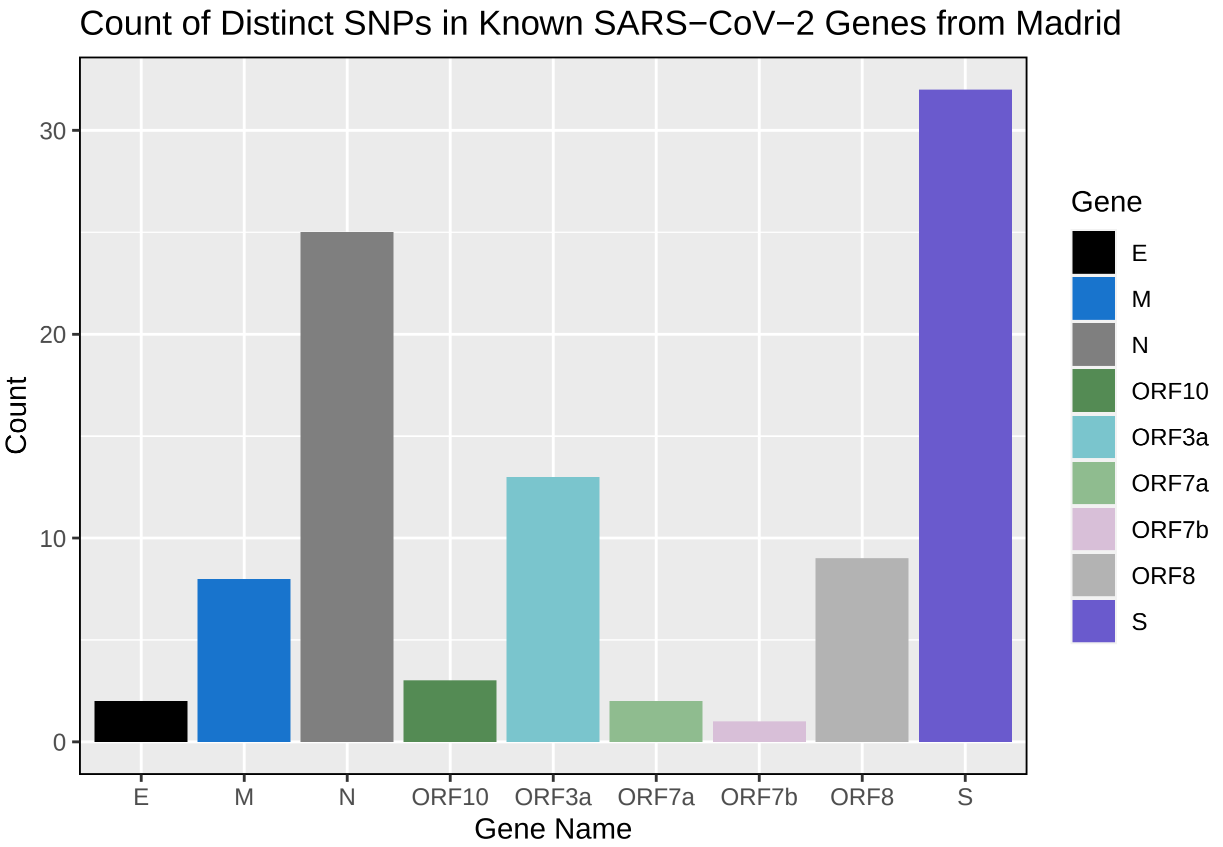

**Figure 5**: Frequencies of distinct SNPs of SARS-CoV-2 genes for samples collected from Madrid, Spain.

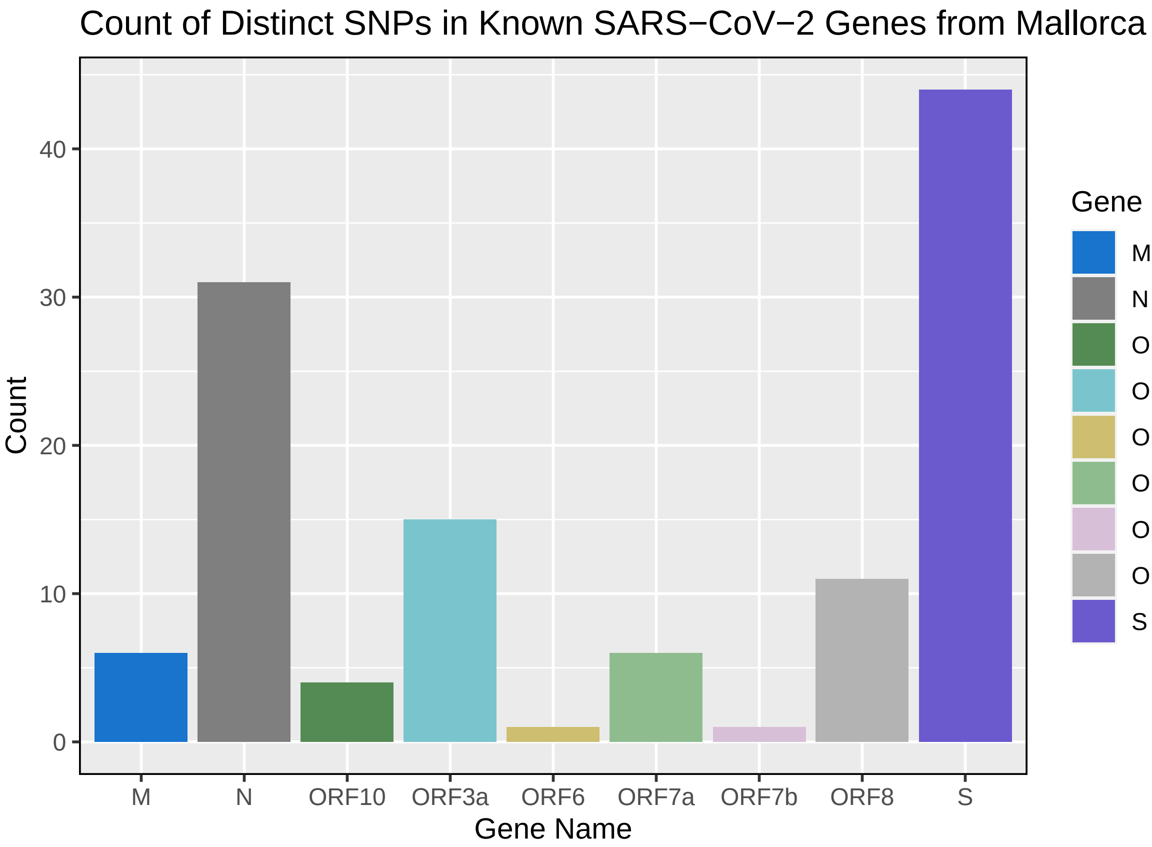

**Figure 6**: Frequencies of distinct SNPs of SARS-CoV-2 genes for samples collected from all localities within the island of Mallorca.

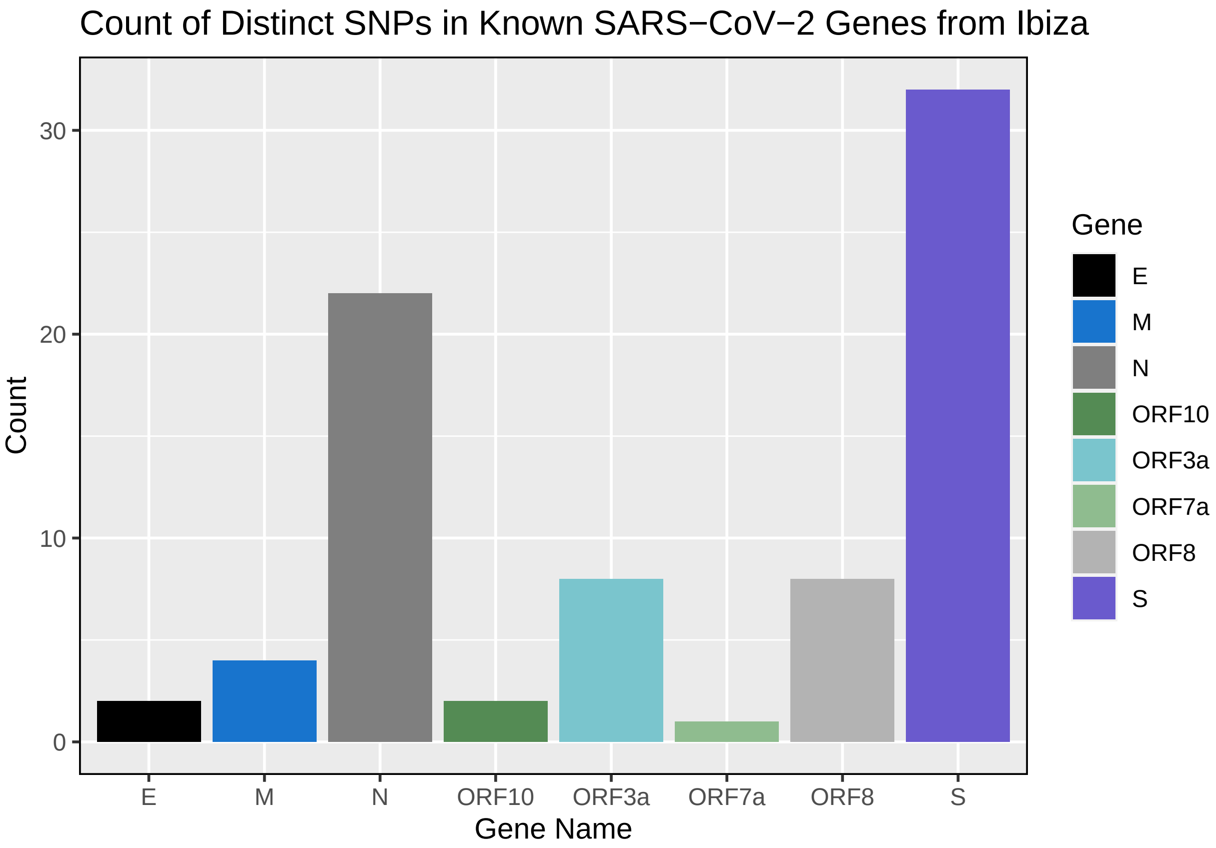

**Figure 7**: Frequencies of distinct SNPs of SARS-CoV-2 genes for samples collected from all localities within the island of Ibiza.

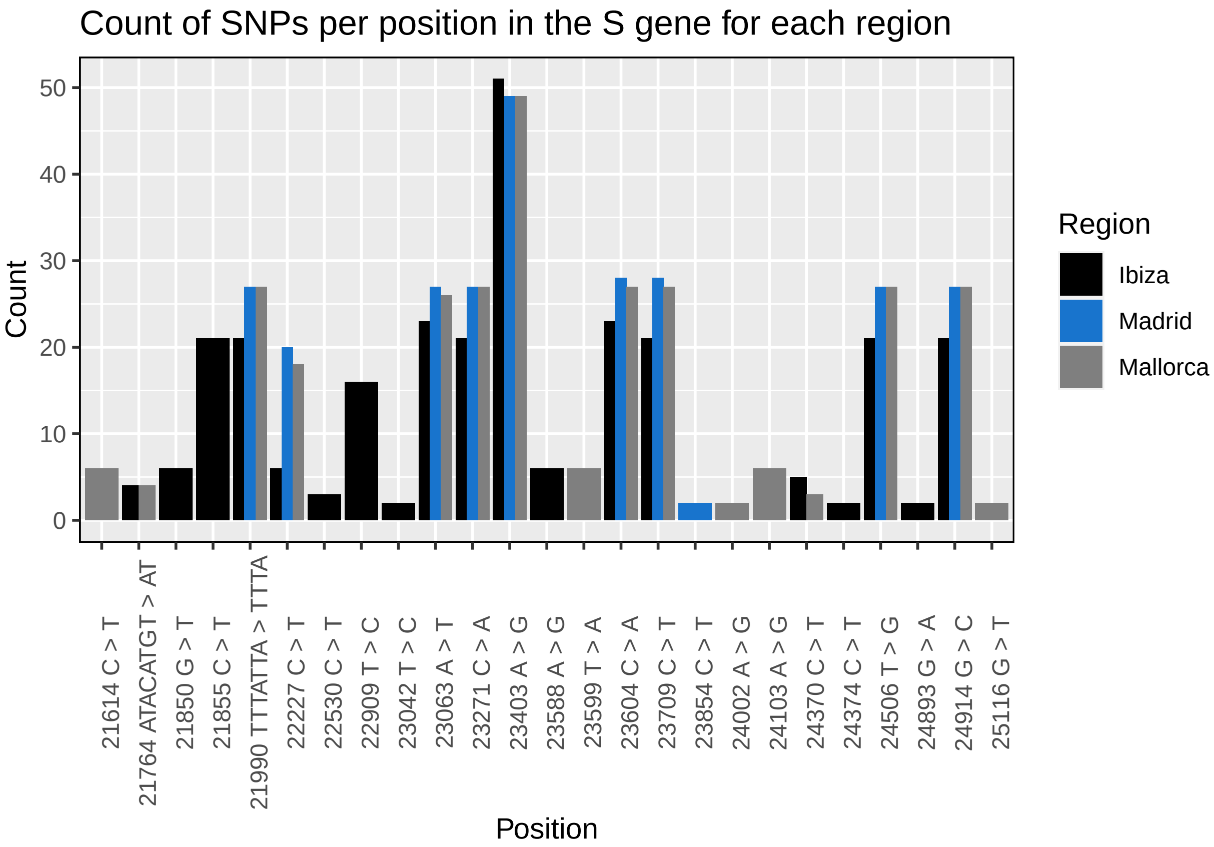

**Figure 8**: Frequencies of the most common SNPs within the spike protein gene for each region. All SNPs represented here have frequencies of greater than 10 occurrences.

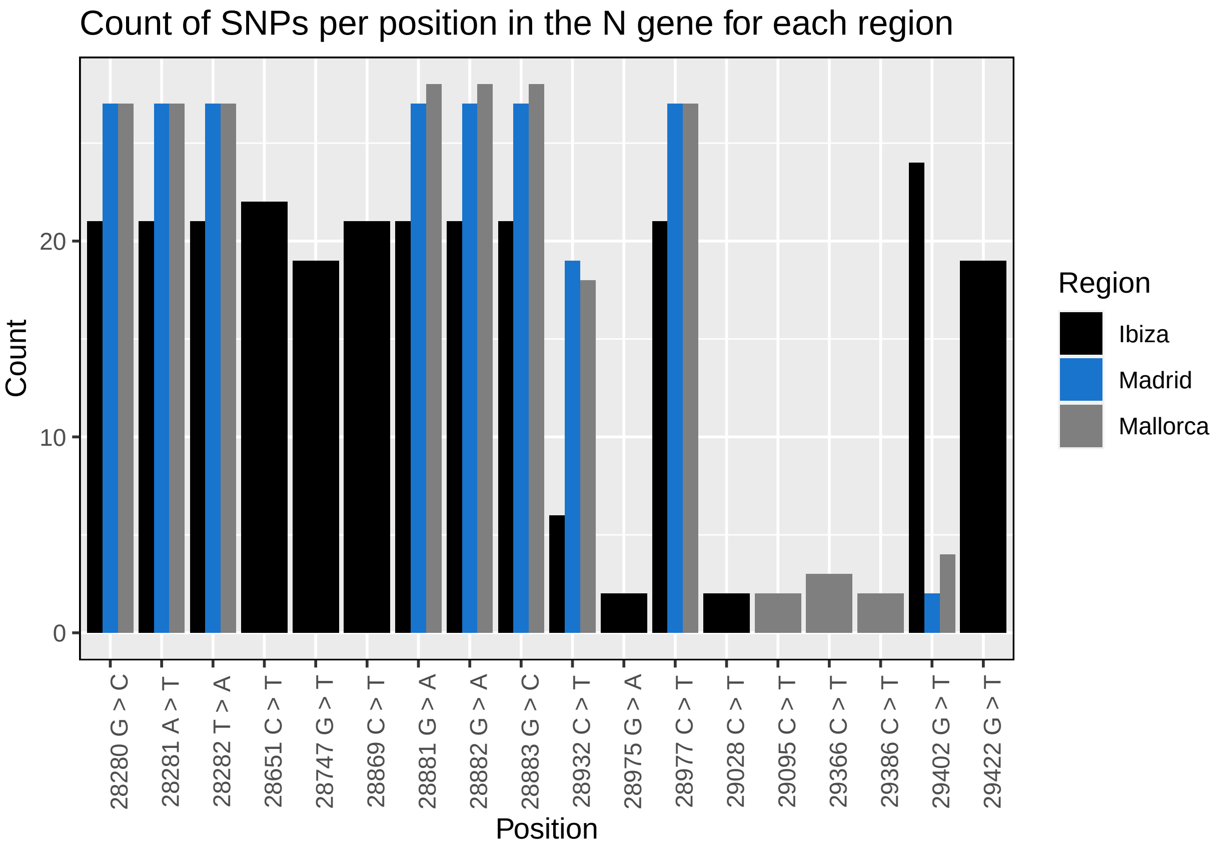

**Figure 9**: Frequencies of the most common SNPs within the nucleocapsid gene for each region. All SNPs represented here have frequencies of greater than 10 occurrences.

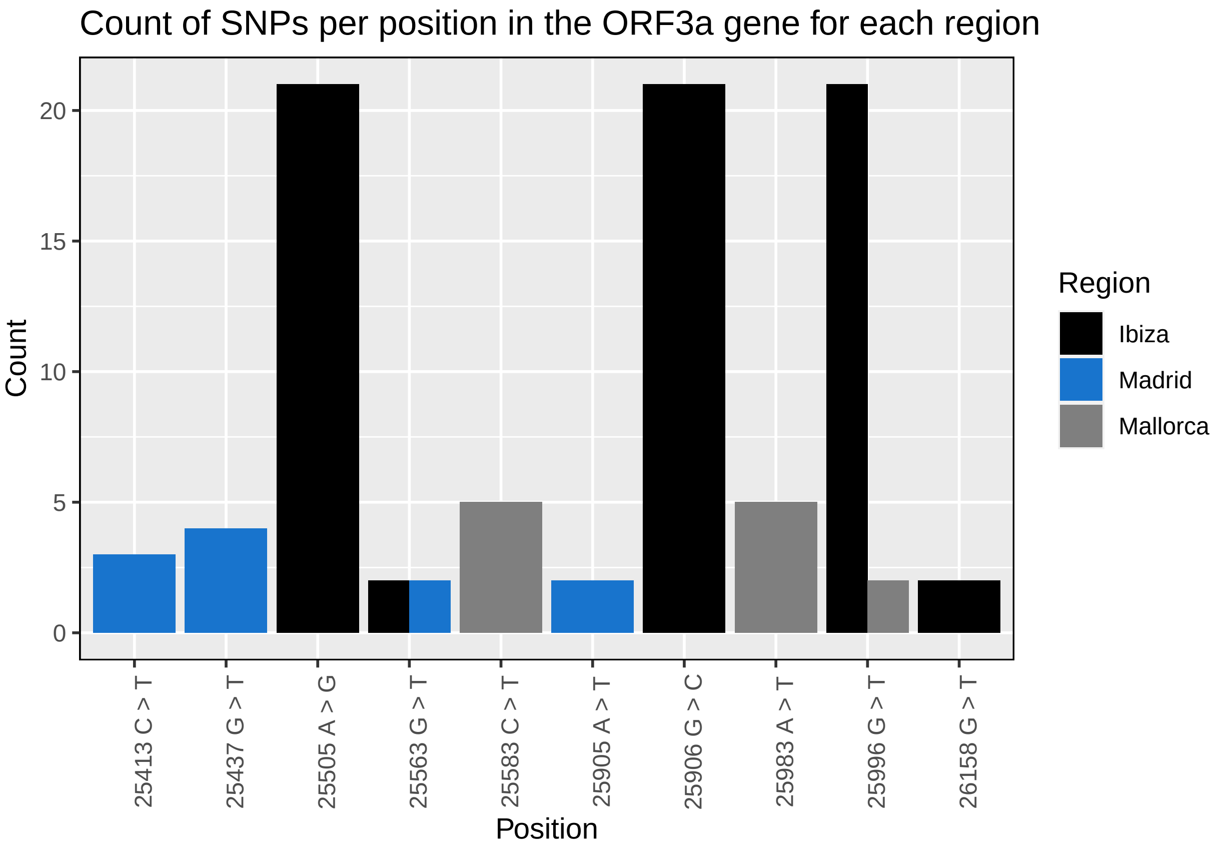

**Figure 10**: Frequencies of the most common SNPs within the ORF3a gene for each region. All SNPs represented here have frequencies of greater than 10 occurrences.

## Tables

**Table 1**: Frequencies of all SNPs found within the spike protein gene from samples collected in Madrid

| Reference | Alternate | Position | Gene | Region | Count |
|---|---|---|---|---|---|
| A | G | 23403 | S | Madrid | 49 |
| A | T | 23063 | S | Madrid | 27 |
| ATACATGT | AT | 21764 | S | Madrid | 1 |
| C | A | 23271 | S | Madrid | 27 |
| C | A | 23604 | S | Madrid | 28 |
| C | T | 21614 | S | Madrid | 1 |
| C | T | 22227 | S | Madrid | 20 |
| C | T | 22432 | S | Madrid | 1 |
| C | T | 23277 | S | Madrid | 1 |
| C | T | 23638 | S | Madrid | 1 |
| C | T | 23709 | S | Madrid | 28 |
| C | T | 23854 | S | Madrid | 2 |
| C | T | 24174 | S | Madrid | 1 |
| C | T | 24334 | S | Madrid | 1 |
| C | T | 24381 | S | Madrid | 1 |
| C | T | 24844 | S | Madrid | 1 |
| C | T | 25000 | S | Madrid | 1 |
| C | T | 25006 | S | Madrid | 1 |

| Reference | Alternate | Position | Gene | Region | Count |
|---|---|---|---|---|---|
| G | A | 21898 | S | Madrid | 1 |
| G | A | 22992 | S | Madrid | 1 |
| G | C | 21974 | S | Madrid | 1 |
| G | C | 24914 | S | Madrid | 27 |
| G | T | 23587 | S | Madrid | 1 |
| G | T | 24764 | S | Madrid | 1 |
| G | T | 25244 | S | Madrid | 1 |
| T | C | 21579 | S | Madrid | 1 |
| T | C | 23042 | S | Madrid | 1 |
| T | C | 23677 | S | Madrid | 1 |
| T | C | 24721 | S | Madrid | 1 |
| T | C | 24835 | S | Madrid | 1 |
| T | G | 24506 | S | Madrid | 27 |
| TTTATTA | TTTA | 21990 | S | Madrid | 27 |

**Table 2**: Frequencies of all SNPs found within the spike protein gene from samples collected in Mallorca

| Reference | Alternate | Position | Gene | Region | Count |
|---|---|---|---|---|---|
| A | G | 22169 | S | Mallorca | 1 |
| A | G | 23403 | S | Mallorca | 49 |
| A | G | 24002 | S | Mallorca | 2 |
| A | G | 24103 | S | Mallorca | 6 |
| A | G | 24454 | S | Mallorca | 1 |
| A | T | 21631 | S | Mallorca | 1 |
| A | T | 23063 | S | Mallorca | 26 |
| A | T | 23541 | S | Mallorca | 1 |
| A | T | 24774 | S | Mallorca | 1 |
| ATACATGT | AT | 21764 | S | Mallorca | 4 |
| C | A | 23271 | S | Mallorca | 27 |
| C | A | 23604 | S | Mallorca | 27 |
| C | T | 21614 | S | Mallorca | 6 |
| C | T | 21762 | S | Mallorca | 1 |
| C | T | 21846 | S | Mallorca | 1 |
| C | T | 21855 | S | Mallorca | 1 |
| C | T | 21859 | S | Mallorca | 1 |
| C | T | 22227 | S | Mallorca | 18 |
| C | T | 22432 | S | Mallorca | 1 |
| C | T | 22858 | S | Mallorca | 1 |
| C | T | 23613 | S | Mallorca | 1 |
| C | T | 23709 | S | Mallorca | 27 |
| C | T | 24370 | S | Mallorca | 3 |
| G | A | 22302 | S | Mallorca | 1 |
| G | C | 21770 | S | Mallorca | 1 |
| G | C | 23915 | S | Mallorca | 1 |
| G | C | 24914 | S | Mallorca | 27 |
| G | T | 21724 | S | Mallorca | 1 |
| G | T | 21786 | S | Mallorca | 1 |
| G | T | 21898 | S | Mallorca | 1 |
| G | T | 22205 | S | Mallorca | 1 |
| G | T | 23224 | S | Mallorca | 1 |
| G | T | 23593 | S | Mallorca | 1 |

| Reference | Alternate | Position | Gene | Region | Count |
|---|---|---|---|---|---|
| G | T | 25049 | S | Mallorca | 1 |
| G | T | 25088 | S | Mallorca | 1 |
| G | T | 25116 | S | Mallorca | 2 |
| G | T | 25273 | S | Mallorca | 1 |
| T | A | 23599 | S | Mallorca | 6 |
| T | C | 21771 | S | Mallorca | 1 |
| T | C | 22828 | S | Mallorca | 1 |
| T | C | 24847 | S | Mallorca | 1 |
| T | G | 24307 | S | Mallorca | 1 |
| T | G | 24506 | S | Mallorca | 27 |
| TTTATTA | TTTA | 21990 | S | Mallorca | 27 |

**Table 3**: Frequencies of all SNPs found within the spike protein gene from samples collected in Ibiza

| Reference | Alternate | Position | Gene | Region | Count |
|---|---|---|---|---|---|
| A | C | 22005 | S | Ibiza | 1 |
| A | G | 23403 | S | Ibiza | 51 |
| A | G | 23588 | S | Ibiza | 6 |
| A | T | 23063 | S | Ibiza | 23 |
| A | T | 24774 | S | Ibiza | 1 |
| ATACATGT | AT | 21764 | S | Ibiza | 4 |
| C | A | 23271 | S | Ibiza | 21 |
| C | A | 23604 | S | Ibiza | 23 |
| C | T | 21614 | S | Ibiza | 1 |
| C | T | 21846 | S | Ibiza | 1 |
| C | T | 21855 | S | Ibiza | 21 |
| C | T | 22227 | S | Ibiza | 6 |
| C | T | 22530 | S | Ibiza | 3 |
| C | T | 23625 | S | Ibiza | 1 |
| C | T | 23709 | S | Ibiza | 21 |
| C | T | 24054 | S | Ibiza | 1 |
| C | T | 24370 | S | Ibiza | 5 |
| C | T | 24374 | S | Ibiza | 2 |
| C | T | 24418 | S | Ibiza | 1 |
| C | T | 24642 | S | Ibiza | 1 |
| G | A | 23867 | S | Ibiza | 1 |
| G | A | 24893 | S | Ibiza | 2 |
| G | C | 24914 | S | Ibiza | 21 |
| G | T | 21850 | S | Ibiza | 6 |
| G | T | 22346 | S | Ibiza | 1 |
| G | T | 25314 | S | Ibiza | 1 |
| T | C | 21628 | S | Ibiza | 1 |
| T | C | 22909 | S | Ibiza | 16 |
| T | C | 23042 | S | Ibiza | 2 |
| T | C | 24152 | S | Ibiza | 1 |
| T | G | 24506 | S | Ibiza | 21 |
| TTTATTA | TTTA | 21990 | S | Ibiza | 21 |

**Table 4**: Frequencies of all SNPs found within the nucleocapsid gene from samples collected in Madrid

| Reference | Alternate | Position | Gene | Region | Count |
|---|---|---|---|---|---|
| A | T | 28281 | N | Madrid | 27 |
| C | A | 28751 | N | Madrid | 1 |
| C | T | 28453 | N | Madrid | 1 |
| C | T | 28677 | N | Madrid | 1 |
| C | T | 28744 | N | Madrid | 1 |
| C | T | 28887 | N | Madrid | 1 |
| C | T | 28928 | N | Madrid | 1 |
| C | T | 28932 | N | Madrid | 19 |
| C | T | 28977 | N | Madrid | 27 |
| C | T | 29253 | N | Madrid | 1 |
| C | T | 29272 | N | Madrid | 1 |
| C | T | 29386 | N | Madrid | 1 |
| C | T | 29421 | N | Madrid | 1 |
| G | A | 28881 | N | Madrid | 27 |
| G | A | 28882 | N | Madrid | 27 |
| G | A | 29399 | N | Madrid | 1 |
| G | C | 28280 | N | Madrid | 27 |
| G | C | 28883 | N | Madrid | 27 |
| G | C | 28975 | N | Madrid | 1 |
| G | T | 28307 | N | Madrid | 1 |
| G | T | 28690 | N | Madrid | 1 |
| G | T | 29402 | N | Madrid | 2 |
| G | T | 29513 | N | Madrid | 1 |
| T | A | 28282 | N | Madrid | 27 |
| T | C | 28297 | N | Madrid | 1 |

**Table 5**: Frequencies of all SNPs found within the nucleocapsid gene from samples collected in Mallorca

| Reference | Alternate | Position | Gene | Region | Count |
|---|---|---|---|---|---|
| A | G | 28336 | N | Mallorca | 1 |
| A | T | 28281 | N | Mallorca | 27 |
| A | T | 28390 | N | Mallorca | 1 |
| C | A | 28830 | N | Mallorca | 1 |
| C | T | 28651 | N | Mallorca | 1 |
| C | T | 28657 | N | Mallorca | 1 |
| C | T | 28708 | N | Mallorca | 1 |
| C | T | 28869 | N | Mallorca | 1 |
| C | T | 28887 | N | Mallorca | 1 |
| C | T | 28932 | N | Mallorca | 18 |
| C | T | 28977 | N | Mallorca | 27 |
| C | T | 29095 | N | Mallorca | 2 |
| C | T | 29171 | N | Mallorca | 1 |
| C | T | 29200 | N | Mallorca | 1 |
| C | T | 29218 | N | Mallorca | 1 |
| C | T | 29366 | N | Mallorca | 3 |
| C | T | 29386 | N | Mallorca | 2 |
| C | T | 29409 | N | Mallorca | 1 |
| C | T | 29445 | N | Mallorca | 1 |
| G | A | 28396 | N | Mallorca | 1 |
| G | A | 28808 | N | Mallorca | 1 |
| G | A | 28881 | N | Mallorca | 28 |

| Reference | Alternate | Position | Gene | Region | Count |
|---|---|---|---|---|---|
| G | A | 28882 | N | Mallorca | 28 |
| G | C | 28280 | N | Mallorca | 27 |
| G | C | 28883 | N | Mallorca | 28 |
| G | T | 28747 | N | Mallorca | 1 |
| G | T | 29260 | N | Mallorca | 1 |
| G | T | 29402 | N | Mallorca | 4 |
| G | T | 29422 | N | Mallorca | 1 |
| T | A | 28282 | N | Mallorca | 27 |
| T | G | 28393 | N | Mallorca | 1 |

**Table 6**: Frequencies of all SNPs found within the nucleocapsid gene from samples collected in Ibiza

| Reference | Alternate | Position | Gene | Region | Count |
|---|---|---|---|---|---|
| A | G | 28742 | N | Ibiza | 1 |
| A | T | 28281 | N | Ibiza | 21 |
| A | T | 28295 | N | Ibiza | 1 |
| C | A | 29367 | N | Ibiza | 1 |
| C | T | 28651 | N | Ibiza | 22 |
| C | T | 28869 | N | Ibiza | 21 |
| C | T | 28932 | N | Ibiza | 6 |
| C | T | 28977 | N | Ibiza | 21 |
| C | T | 29028 | N | Ibiza | 2 |
| C | T | 29466 | N | Ibiza | 1 |
| G | A | 28881 | N | Ibiza | 21 |
| G | A | 28882 | N | Ibiza | 21 |
| G | A | 28975 | N | Ibiza | 2 |
| G | A | 29260 | N | Ibiza | 1 |
| G | C | 28280 | N | Ibiza | 21 |
| G | C | 28514 | N | Ibiza | 1 |
| G | C | 28883 | N | Ibiza | 21 |
| G | T | 28747 | N | Ibiza | 19 |
| G | T | 29402 | N | Ibiza | 24 |
| G | T | 29422 | N | Ibiza | 19 |
| T | A | 28282 | N | Ibiza | 21 |
| T | C | 28642 | N | Ibiza | 1 |

**Table 7**: Frequencies of all SNPs found within the ORF3a gene from samples collected in Madrid

| Reference | Alternate | Position | Gene | Region | Count |
|---|---|---|---|---|---|
| A | T | 25905 | ORF3a | Madrid | 2 |
| C | T | 25413 | ORF3a | Madrid | 3 |
| C | T | 25452 | ORF3a | Madrid | 1 |
| C | T | 25463 | ORF3a | Madrid | 1 |
| C | T | 25549 | ORF3a | Madrid | 1 |
| C | T | 25710 | ORF3a | Madrid | 1 |
| C | T | 25904 | ORF3a | Madrid | 1 |
| C | T | 26060 | ORF3a | Madrid | 1 |
| G | T | 25437 | ORF3a | Madrid | 4 |
| G | T | 25563 | ORF3a | Madrid | 2 |
| G | T | 25726 | ORF3a | Madrid | 1 |

| Reference | Alternate | Position | Gene | Region | Count |
|---|---|---|---|---|---|
| G | T | 25785 | ORF3a | Madrid | 1 |
| T | A | 25551 | ORF3a | Madrid | 1 |

**Table 8**: Frequencies of all SNPs found within the ORF3a gene from samples collected in Mallorca

| Reference | Alternate | Position | Gene | Region | Count |
|---|---|---|---|---|---|
| A | G | 25505 | ORF3a | Mallorca | 1 |
| A | T | 25905 | ORF3a | Mallorca | 1 |
| A | T | 25983 | ORF3a | Mallorca | 5 |
| C | A | 25693 | ORF3a | Mallorca | 1 |
| C | A | 26029 | ORF3a | Mallorca | 1 |
| C | A | 26060 | ORF3a | Mallorca | 1 |
| C | T | 25583 | ORF3a | Mallorca | 5 |
| C | T | 25665 | ORF3a | Mallorca | 1 |
| C | T | 25854 | ORF3a | Mallorca | 1 |
| G | C | 25906 | ORF3a | Mallorca | 1 |
| G | T | 25437 | ORF3a | Mallorca | 1 |
| G | T | 25440 | ORF3a | Mallorca | 1 |
| G | T | 25455 | ORF3a | Mallorca | 1 |
| G | T | 25563 | ORF3a | Mallorca | 1 |
| G | T | 25996 | ORF3a | Mallorca | 2 |

**Table 9**: Frequencies of all SNPs found within the ORF3a gene from samples collected in Ibiza

| Reference | Alternate | Position | Gene | Region | Count |
|---|---|---|---|---|---|
| A | G | 25505 | ORF3a | Ibiza | 21 |
| C | A | 26060 | ORF3a | Ibiza | 1 |
| C | T | 25931 | ORF3a | Ibiza | 1 |
| G | C | 25906 | ORF3a | Ibiza | 21 |
| G | T | 25455 | ORF3a | Ibiza | 1 |
| G | T | 25563 | ORF3a | Ibiza | 2 |
| G | T | 25996 | ORF3a | Ibiza | 21 |
| G | T | 26158 | ORF3a | Ibiza | 2 |

## Sources Cited

Arnold,J.B. (2021) Ggthemes: Extra themes, scales and geoms for 'ggplot2'.

Bianchi,M. *et al.* (2021) SARS-cov-2 orf3a: Mutability and function. *International journal of biological macromolecules*, **170**, 820–826.

Eguíluz,V.M. *et al.* (2020) Risk of secondary infection waves of covid-19 in an insular region: The case of the balearic islands, spain. *Frontiers in medicine*, **7**, 905.

Ferrer,R. (2020) COVID-19 pandemic: The greatest challenge in the history of critical care. *Medicina intensiva.*

Henríquez,J. *et al.* (2020) The first months of the covid-19 pandemic in spain. *Health Policy and Technology*, **9**, 560–574.

Knaus,B.J. and Grünwald,N.J. (2016) VcfR: An r package to manipulate and visualize VCF format data. *BioRxiv*.

Knaus,B.J. and Grünwald,N.J. (2017) VCFR: A package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, **17**, 44–53.

Koyama,T. *et al.* (2020) Variant analysis of sars-cov-2 genomes. *Bulletin of the World Health Organization*, **98**, 495.

Pachetti,M. *et al.* (2020) Emerging sars-cov-2 mutation hot spots include a novel rna-dependent-rna polymerase variant. *Journal of translational medicine*, **18**, 1–9.

Spiteri,G. *et al.* (2020) First cases of coronavirus disease 2019 (covid-19) in the who european region, 24 january to 21 february 2020. *Eurosurveillance*, **25**, 2000178.

Toyoshima,Y. *et al.* (2020) SARS-cov-2 genomic variations associated with mortality rate of covid-19. *Journal of human genetics*, **65**, 1075–1082.

Viedma,E. *et al.* (2021) Genomic epidemiology of sars-cov-2 in madrid, spain, during the first wave of the pandemic: Fast spread and early dominance by d614g variants. *Microorganisms*, **9**, 454.

Wickham,H. (2016) Ggplot2: Elegant graphics for data analysis Springer-Verlag New York.

Wickham,H. *et al.* (2020) Dplyr: A grammar of data manipulation.