# First patient analysis using baseclick's "ClickTech Single Strain Mutation Mapping Kit for SARS-CoV-2"

**T. Frischmuth[1]*, A. Crisp[1], J. Sobotta[1], E. Schönegger[1], M. Muenchhoff[2]**

*[1] baseclick GmbH, Floriansbogen 2-4, 82061 Neuried, Germany*

*[2] Pettenkofer-Institute, Ludwig-Maximilians-University Munich (LMU), Pettenkoferstr. 9a, 80336 Munich, Germany*

***t.frischmuth@baseclick.eu**

## Introduction

The rapid spread of the newly emerging SARS-CoV-2 virus variant B.1.1.7 in Great Britain and other countries shows, changes in the genome of the SARS-CoV-2 virus can lead to new biological properties such as increased infectivity.

Until today more than 12,000 mutations within the genome of SARS-CoV-2 have been described (Korber *et al.*, 2020). Most of these mutations have been identified since February 2020 and so far did not change the biological properties of the virus. However, some of the almost 4,000 mutations that have been identified in the spike protein (S-Protein) have influenced the biological properties e.g. infectivity, disease progression and immunogenicity of the mutant SARS-CoV-2 virus. For example, in spring 2020 the emergence of the D614G mutation in the S-Protein led to an increase in the infectivity of this new SARS-CoV-2 mutant strain and has almost entirely replaced the Wuhan strain worldwide (Figure 1). The mutation A222V in the Spanish strain (20A.EU1), led again to an increase in infectivity and started to replace the D614G strain in some countries in Europe in the summer 2020. The mutation N501Y-del69-70 in B.1.1.7 in Great Britain, which was detected for the first time in November 2020, once again led to an increase in infectivity.

Therefore, one of the major questions in the COVID-19 pandemic is the variability of the infecting agent SARS-CoV-2. To monitor the spread and evolution of SARS-CoV-2 strains it is important to analyze a) the infecting virus transmitted from patient to patient (horizontal evolution) and b) virus strains present within a patient at the time of infection and changes during the infection (vertical evolution) (Figure 1).

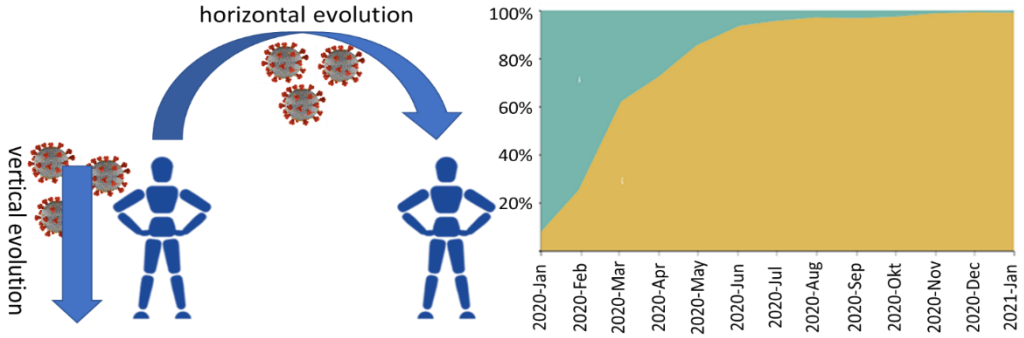


**Fig. 1:**
**Left:** Schematic drawing of possible routes how virus evolution is driven.
Horizontal - selection of a virus strain due to variation in the genetic background of the newly infected person.
Vertical – due to the missing proof reading of the viral replicase mutation can occur within a patient over time.
**Right:** Appearance of the spike protein mutant D614G in Europe and spread throughout the world (green = Wuhan Strain, yellow = D614G); Plot was produced with the software "Global genotypic frequency at position of SARS-CoV-2 Genome over time" provided by Nextstrain (https://nextstrain.org).

## Material and Methods

The newly developed SARS-CoV-2 sequencing kit from baseclick (**ClickTech Single Strain Mutation Mapping Kit for SARS-CoV-2**) was used to identify the infecting virus strain in a Munich COVID-19 patient by three overlapping PCR fragments (Figure 2).

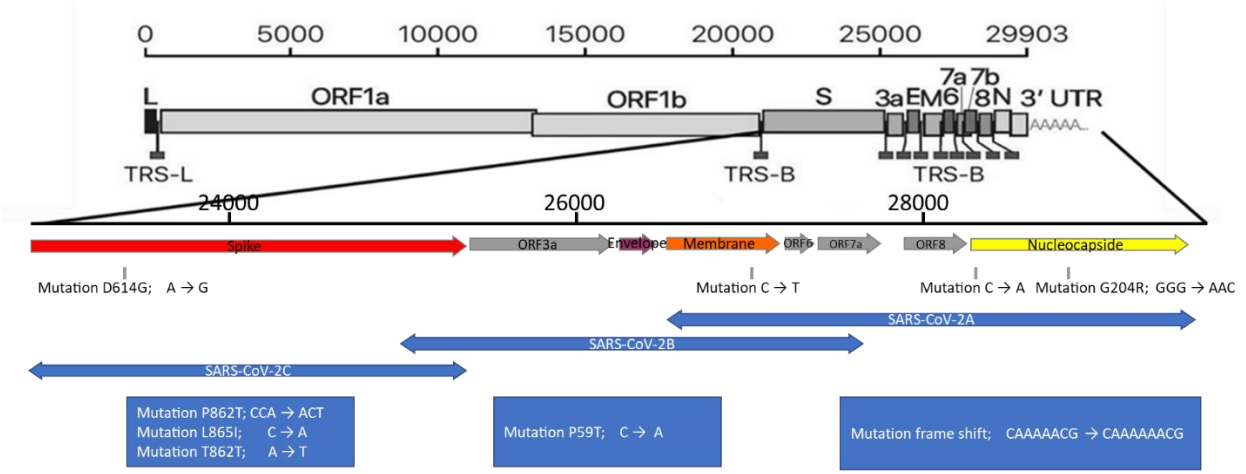


**Fig. 2: Schematic representation of the S-E-M-N SARS-CoV-2 genome segment**

**Top:** Schematic drawing of the SARS-CoV-2 genome. **Bottom:** Genomic positioning of the Spike Protein (red), ORF3a (grey), Envelope (purple), Membrane Protein (orange), ORF6 (grey), ORF 7a (grey), ORF8 (grey) und Nucleocapsid Proteins (yellow) – S-E-M-N segment - are shown. Overlapping PCR-fragments generated by the special ClickTech SARS-CoV-2 Kit are shown in blue. Position of major mutation detected in all analyzed genome sequences compared to the published sequence of the Wuhan strain are shown (D614G, C-T, C-A und G204R). Position of the five minor mutations (Table 1; 1-2%) are shown in blue.

The PCR products were sequenced on a Pacific Bioscience Sequel II machine. Sequence data for comparison were extracted from NCBI data base (www.ncbi.nlm.nih.gov/sars-cov-2/) and Nextstrain (https://nextstrain.org). The sequence data were analyzed with the software "Integrative Genomics Viewer". The intend of this first analysis using the SARS-CoV-2 sequencing kit was to get an insight into:

1. Accuracy of the kit by deep long read sequencing approach
2. Identity of the SARS-CoV-2 infecting strains in the patient (reference Wuhan strain)
3. Identify possible SARS-CoV-2 minor mutations in a patient – One strain in a patient or several?

## Results

### 1. Pacific Bioscience Sequel II Sequencing data summary

The produced PCR fragments were sequenced on Pacific Bioscience Sequel II, 8M SMRTcell, runtime 30h, analyzed either as single fragments or as mixture of all three fragments. After cleaning of raw sequence data (Pac Bio software), only reads with the correct primer orientation and an accuracy above 98% were used for further analysis. The usable read results of the sequencing are summarized in Table 1.

| Fragment Name | Number of usable reads |
|---|---|
| SARS-CoV-2A | 194,970 |
| SARS-CoV-2B | 359,485 |
| SARS-CoV-2C | 245,452 |
| SARS-CoV-2MIX | A:79,904; B:137,223; C:105,748 |

**Table 1: Summary of usable long reads of SARS-CoV-2 fragments**

The number of reads for each fragment was generally above 100,000. Even when fragments were mixed (to save costs per patient analysis), usable reads were above 50,000 reads and more than sufficient to identify minor mutations (Table 3).

The sequencing analysis of the COVID-19 patient virus isolate revealed that >98% of all sequenced molecules carried four mutations (Table 2).

**Table 2: Summary of identified SARS-CoV-2 major mutations**

| Position | Mutation | Consequence |
|---|---|---|
| 23,403 | A → G TYPE=snp | Aspartat (614) becomes Glycine (614) in the **spike protein (S).** |
| 27,046 | C → T TYPE=snp | No amino acid change in the **membrane protein (M) (silent mutation).** |
| 28,344 | C → A TYPE=snp | No amino acid change in the **nucleocapsid protein (N) (silent mutation).** |
| 28,881 | GGG → ACC TYPE= complex,snp | Glycine (204) becomes Arginine (204) in the **nucleocapsid protein (N).** |

**Position, mutation and consequence of the four major mutations on the viral genes involved.**

These four mutations in the SARS-CoV-2 genome have been described before (data from Nextstrain, https://nextstrain.org;). However, with the newly developed SARS-CoV-2 sequencing kit we show that all four mutations are in >98 % are on the same viral genome.

Next we were looking for possible minor mutations in the virus isolate, using a threshold above 1% occurrence of the mutation in sequenced molecules. With this threshold sequencing arrows will not be considered. In total we identified five mutations fulfilling this criterion (Table 3).

**Table 3: Summary of identified SARS-CoV-2 minor mutations**

| Position | Mutation | Consequence |
|---|---|---|
| 29,051 | CAAAAACG → CAAAAAACG TYPE=ins | Frameshift starting from Arginine (262) of the nucleocapsid protein. Affects the next 158 AAs. **Probably not viable** |
| 26,697 | C → A TYPE=snp | Proline (59) becomes Threonine (59) of the **membrane protein (M).** |
| 24,146 | CCA → ACT,ACA TYPE=complex,snp | Proline (862) becomes Threonine (862) and/or Threonine (863) of the **spike protein (S).** |
| 24,155 | C → A TYPE=snp | Leucine (865) becomes Isoleucine (865) of the **spike protein (S).** |
| 24,160 | A → T TYPE=snp | No amino acid change in the **spike protein (S) (silent mutation).** |

**Position, mutation and consequence of the five minor mutations on the viral genes involved.**

## 2. Analysis of the four major and five minor mutations in the SARS-CoV-2 viral genome of the infected patient

Four mutations with occurrence of >98% in all long reads in the S-E-M-N genomic segment of the SARS-CoV-2 compared with the "Wuhan SARS-CoV-2" sequence were detected in the virus isolate form the infected COVID-19 patient (Table 2). These mutations have been described before and are

found in various parts of the world and appear and disappeared over time (data from Nextstrain, https://nextstrain.org; Figure 3).

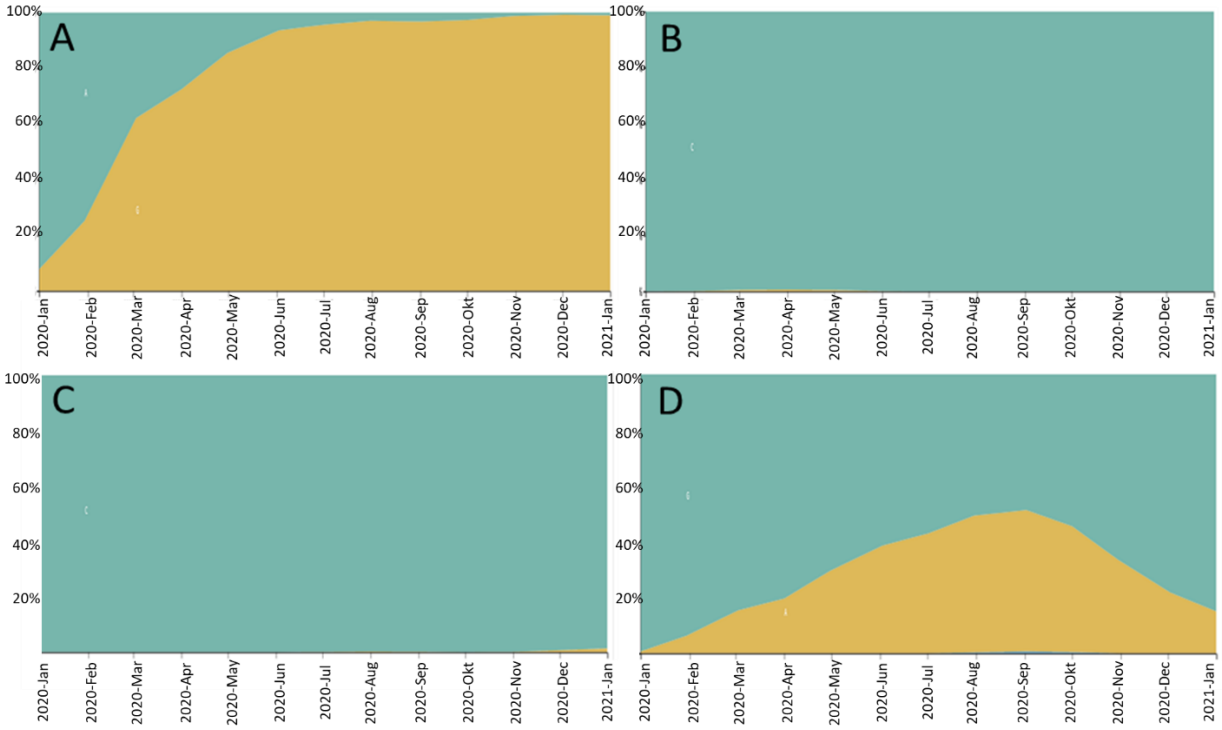


**Fig. 3: Statistical analysis of the appearance and distribution of SARS-CoV-2 mutants in the world population**

**A.** Mutation D614G in the spike protein at position 23,403 was first identified in European in January/February 2020 and replaced in most part of the world the Wuhan strain. **B.** The silent mutation in the membrane protein at position 27,046, first identified in Europe in April 2020 and disappeared in July 2020. **C.** The silent mutation in the nucleocapsid protein at position 28,344 was first identified in Europe in July 2020 and since December started to be more frequently detected in analyzed virus isolates. **D.** Mutation G204R in the nucleocapsid protein at position 28,881, appeared in January 2020 had the maximum in September and starts to disappear since then.

Plots were produced with the software “Global genotypic frequency at position of SARS-CoV-2 Genome over time” provided by Nextstrain (https://nextstrain.org).

Plots were produced with the software “Global genotypic frequency at position of SARS-CoV-2 Genome over time” provided by Nextstrain (https://nextstrain.org).The COVID-19 patient from Munich was infected with a virus strain containing four major mutations, the D614G spike protein mutation which is associated with a much higher infectivity rate then the Wuhan strain (Korber *et. al.*, 2020, Toyoshima *et. al.*, 2020). Mutation 2 was only present so far in few sequenced SARS-CoV-2 isolates from patients infected between April and July 2020. This corresponds with time frame of the isolation of the SARS-CoV-2 virus from the analyzed patient. The silent N-Protein mutations 3 only occurred as late as July 2020 and is now starting to become more frequently present in patient virus isolates since December 2020. The N-Protein mutation G204R (mutation 4) was dominant in September 2020 but is now disappearing (Figure 3). All four mutations distinct from the Wuhan strain have been accumulated in the infecting SARS-CoV-2 isolate in this patient indicating that either patients can be infected with SARS-CoV-2 variants at the same time and by recombination combine to new viruses or that certain mutations increase the viability of the SARS-CoV-2 strain.

The five newly minor detected mutations in the patient sample were not described yet. Three mutants alter the amino acid sequence of the M-Protein and S-Protein, one is a silent mutation not changing

the amino acid sequence and one mutant in the N-Protein has such a strong impact on the amino acid sequence, that we think this mutant is only viable with a helper virus providing the correct N-Protein.

Taken together the results of this first analysis of a COVID-19 patient virus isolate indicates that probably transmission of minority virus variant populations might occur and secondly that during the course of an infection mutations in the viral genome accumulate in a patient.

## Summary

The newly developed long read SARS-CoV-2 sequencing kit "ClickTech Single Strain Mutation Mapping Kit for SARS-CoV-2" from baseclick can be used for detailed analysis of the S-E-M-N genomic region of the SARS-CoV-2 and is able to connect mutations to a single virus genome. Due to the deep sequencing with the NGS long read Pac Bio Sequel II technology, not only major mutations in the viral genome are detectable but also minor mutations in a virus isolate in a patient can be detected. This is particular important to evaluate the plasticity of the SARS-CoV-2 genome in respect of:

- ***Determine appearance of new SARS-CoV-2 mutants***
- ***Evade recognition by the immune system even in recovered infected patient – problem reinfection?***
- ***Resistance to drugs?***
- ***Resistance to vaccines?***
- ***Receptor recognition variations?***
- ***etc. …***

## Literature

**Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus**
B. Korber, W.M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E.E. Giorgi, T. Bhattacharya, B. Foley, K.M. Hastie, M.D. Parker, D.G. Partridge, C.M. Evans, T.M. Freeman, T.I. de Silva, C. McDanal, L.G. Perez, H. Tang, A. Moon-Walker, S.P. Whelan, C.C. LaBranche, E.O. Saphire, D.C. Montefiori (2020) ***Cell:*** 182, 812–827

**SARS-CoV-2 genomic variations associated with mortality rate of COVID-19**
Y. Toyoshima, K. Nemoto, S. Matsumoto, Y. Nakamura, K. Kiyotani (2020) ***Journal of Human Genetics:*** 65, 1075–1082