

# Bioinformatics HW3

Ershov Ivan

December 2021

## **Задание 1.**

Я выбрал 9 геномов коронавируса людей из следующих стран:

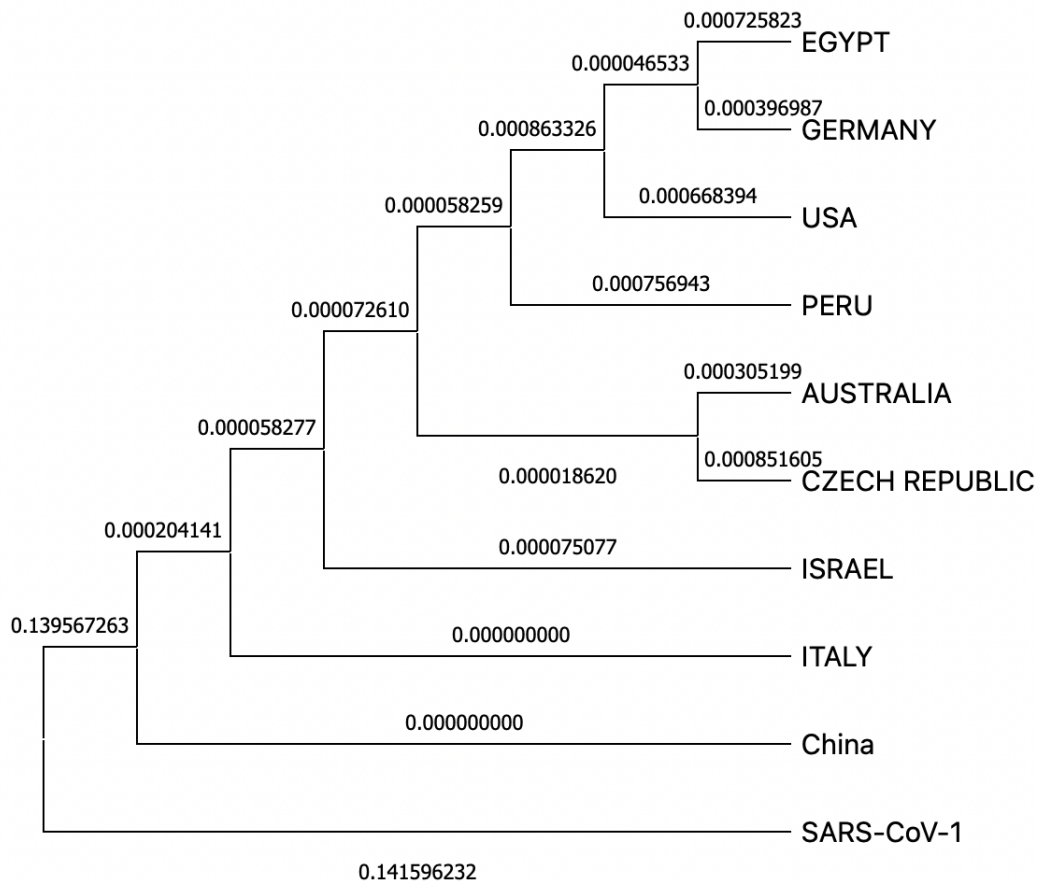
1. Australia
2. Chine
3. Czech Republic
4. Egypt
5. Germany
6. Israel
7. Italy
8. Peru
9. USA

А также взял геном коронавируса SARS-CoV-1.

Теперь в программе MEGA выравниваем эти последовательности при помощи алгоритма Muscle. После этого построим филогенетические деревья методом расстояний (Neighbor-Joining и UPGMA), методом максимального правдоподобия (Maximum Likelihood) и методом максимальной бережливости (Maximum Parsimony). Будем строить без бутстрэпа, так как он имеет смысл только при большом количестве мутаций.

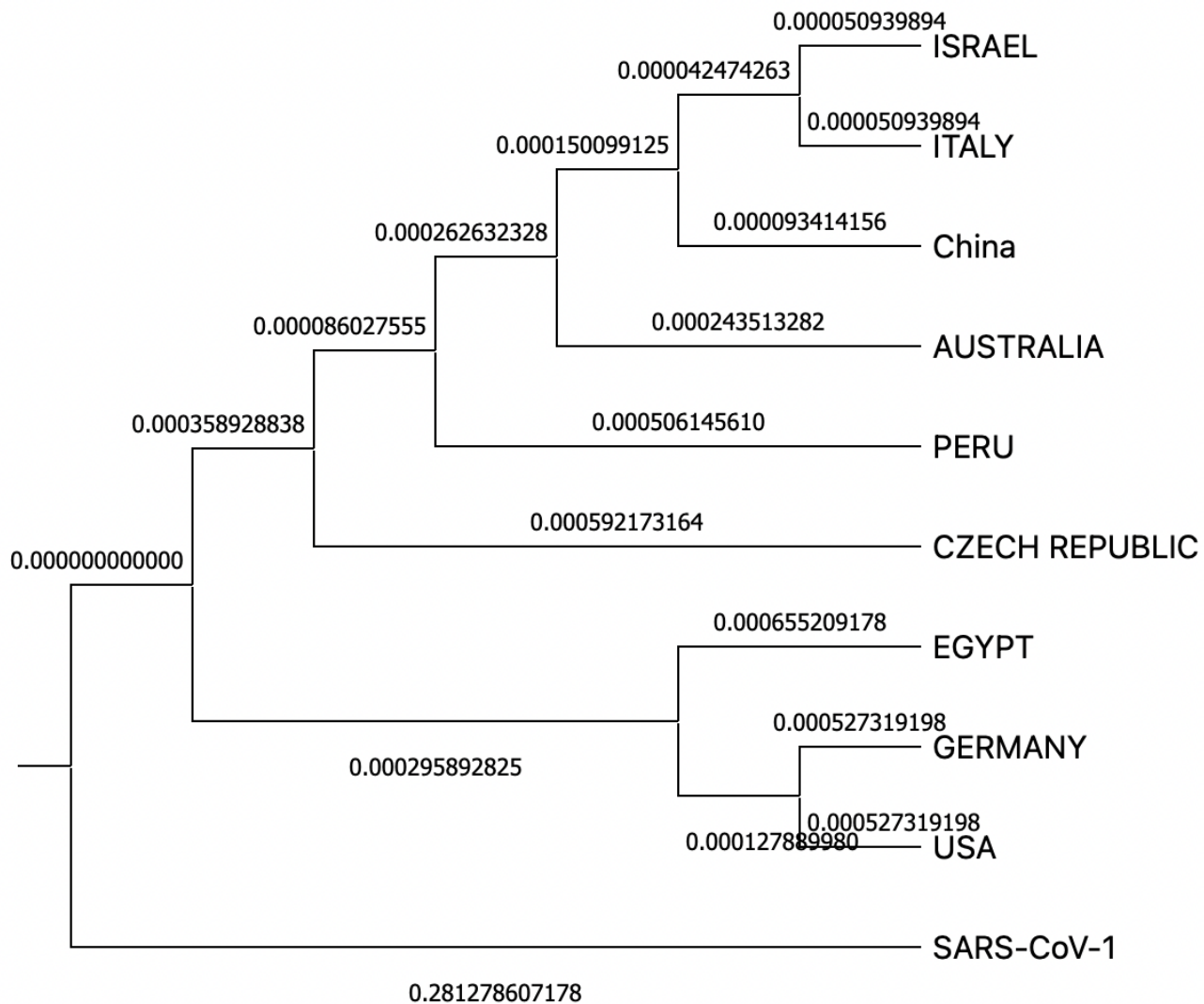
Будем считать суммы длин ветвей от SARS-CoV-1 до других геномов. Геномы коронавируса, которые расположены в дереве "дальше всего" от SARS-CoV-1, имеют больше всего мутаций относительно SARS-CoV-1, а значит, человек, из которого взяли образец генома данного коронавируса, был заражен позже всех. Аналогично: геном, который ближе к SARS-CoV-1, имеет меньше всего мутаций, а значит вирус заразил человека раньше остальных.

### Neighbor-Joining



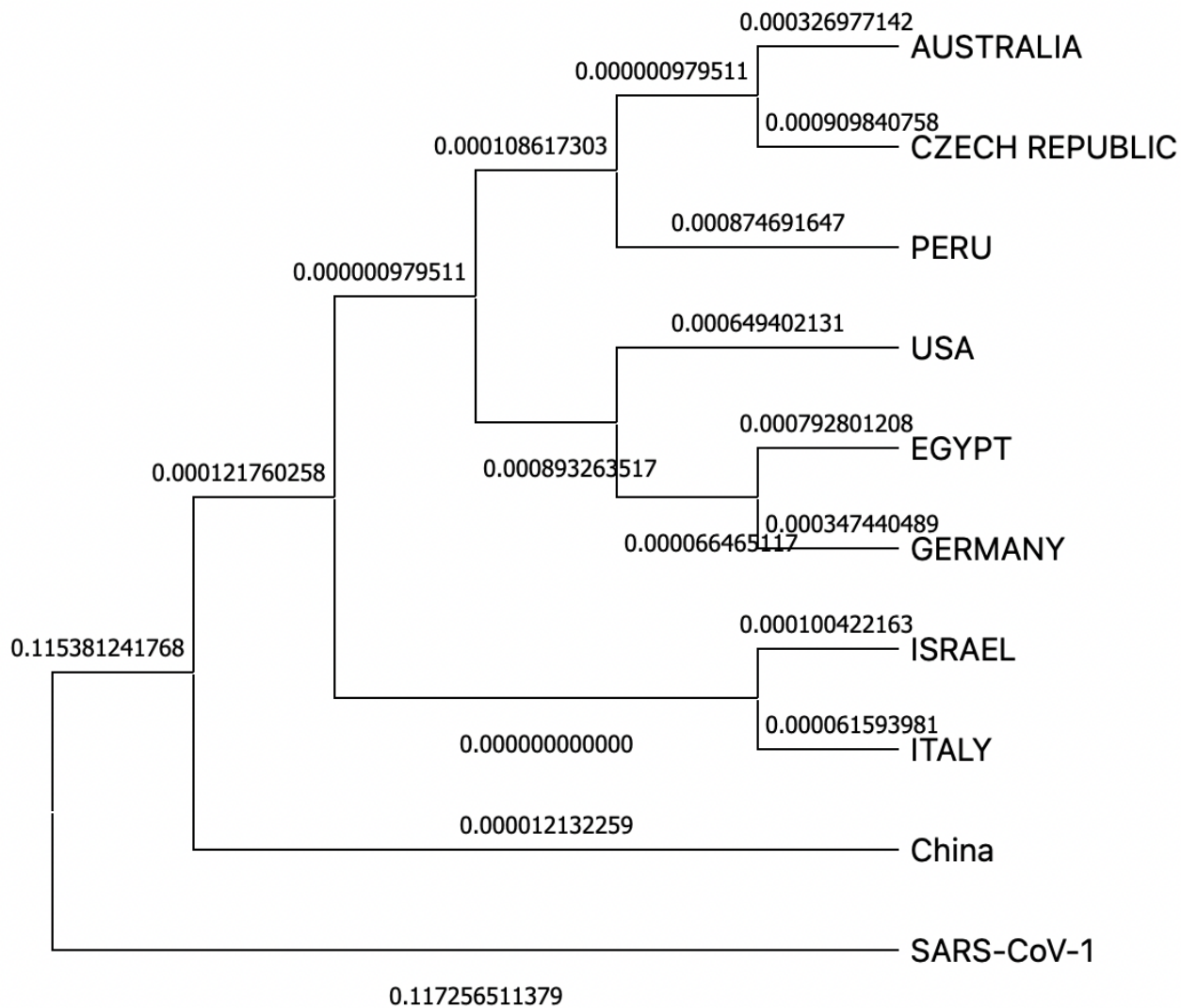
Заметим, что в данном случае самый ближайший геном к SARS-CoV-1 - это China, а самый дальний - Germany и Egypt

## UPGMA



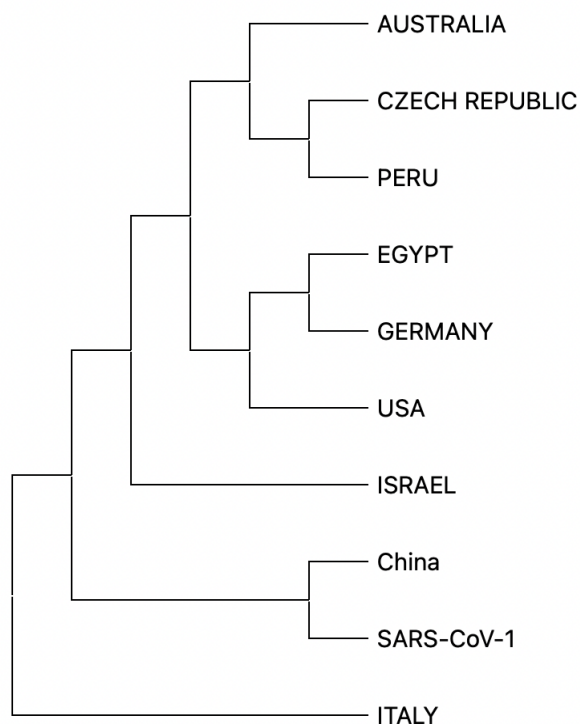
Заметим, что в данном случае самый ближайший геном к SARS-CoV-1 - это China, а самый дальний - Germany, USA и Egypt

## Maximum Likelihood



Заметим, что в данном случае самый ближайший геном к SARS-CoV-1 - это China, а самый дальний - Germany и Egypt

### Maximum Parsimony



Здесь мы не можем считать длины ветвей и искать самые ближние и самые дальние геномы к SARS-CoV-1, но по дереву видно, что оно соответствует тем результатам, которые были получены, анализируя предыдущие деревья. А именно, что самый ближайший геном к SARS-CoV-1 - это China, а самый дальний - Germany, Egypt

**Вывод:** человек, зараженный в Китае, переболел коронавирусом раньше всех, а человек из Германии и человек из Египта - позже всех. В целом такой результат не удивителен, так как я взял геном коронавируса из жителя Ухани (город, в котором началась эпидемия).

**Задание 2.** Теперь сравним последовательности геномов коронавируса из людей Китая и из Германии и найдем 5 мутаций. Возьмем координаты этих мутаций и определим, в каком гене они произошли.

Я нашел следующие 5 мутаций:

Species/Abbrv

\*

\*

\*

\*

\*

\*

1. China

G

A

A

A

A

C

T

2. GERMANY

G

A

A

A

A

T

T

Site # 745

1)

Species/Abbrv

\*

\*

\*

\*

\*

1. China

A

G

A

A

A

2. GERMANY

A

T

A

A

A

Site # 1048

2)

Species/Abbrv

\*

\*

\*

\*

\*

1. China

T

T

T

C

T

2. GERMANY

T

T

T

T

T

Site # 3037

3)

Species/Abbrv

\*

\*

\*

\*

\*

\*

\*

\*

1. China

A

A

A

A

A

G

C

T

2. GERMANY

A

A

A

A

A

G

T

C

Site # 4181

4)

Species/Abbrv

\*

\*

\*

\*

\*

\*

\*

\*

\*

1. China

T

C

T

A

A

A

C

C

A

2. GERMANY

T

C

T

A

A

A

C

T

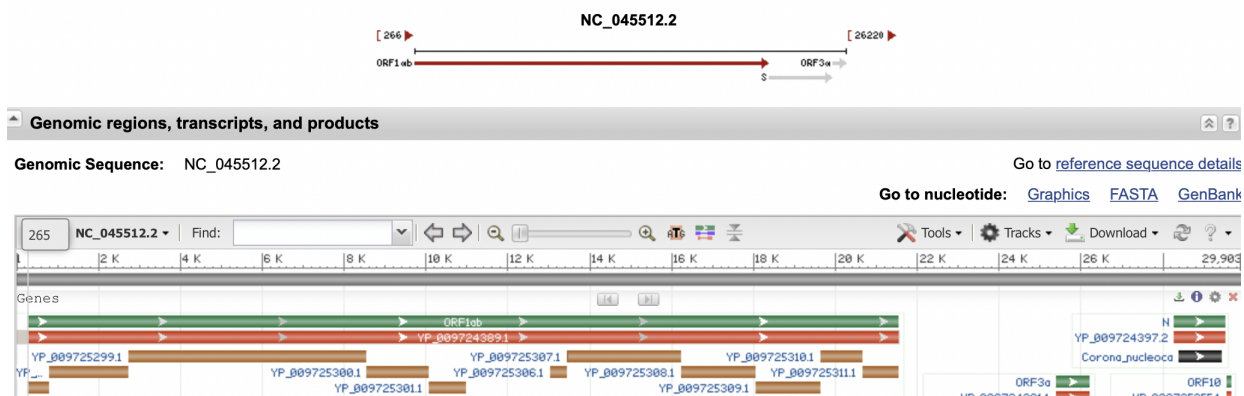
A

Site # 6402

5)

Теперь посмотрим на разметку генов:

Sequence: NC\_045512.2 (266..21555)



Ясно, что найденные мной мутации произошли в гене ORF1ab.