

Local Hack Day 2020

Кейс: Биоинформатический анализ генома коронавируса SARS-CoV-2

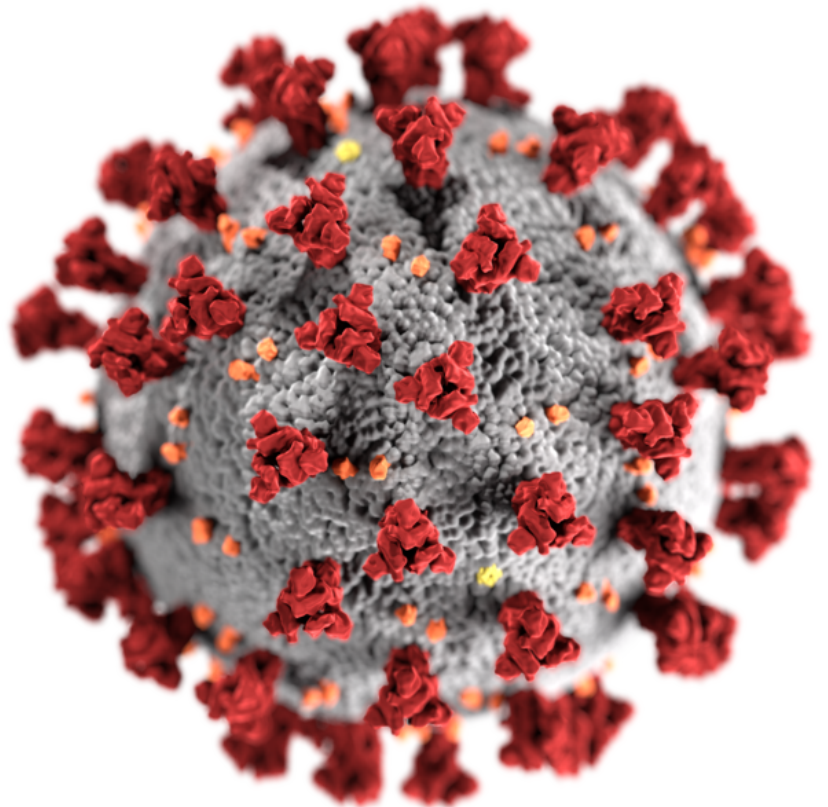
Команда: **Команд04ка**

Кириллова Дарья

Раснюк Никита

Афонин Данил

Фильчак Полина



Сравнительный анализ SARS-COV-2 и других вирусов

Геном SARS-CoV-2 : https://www.ncbi.nlm.nih.gov/nuccore/NC_045512

Было получено 30 последовательностей по 1000 нк (кроме seq30)

Далее с помощью сервиса BLAST они были выровнены, получены следующие филогенетические деревья (см. [phylogenetic_trees.zip](#))

Поделенные последовательности находятся в файле Sequences SARS-CoV-2.pdf

SARS-CoV, по данным BLAST, имеет 98% перекрытия и 80% сходства в пределах перекрытия с SARS-CoV-2. При этом на филогенетических деревьях, построенных по данным BLAST по референсным базам, SARS-CoV, вместе с одним из коронавирусов летучих мышей, принадлежит к ветви, сестринской для SARS-CoV-2. Однако, на деревьях, построенных с учетом данных нереференсных баз, сестринским для SARS-CoV-2 оказывается другой вирус летучих мышей, отсутствующий в референсных базах. Таким образом, SARS-CoV является близкородственным для SARS-CoV-2, но не является для него предковым.

MERS-CoV, по данным BLAST, имеет 30% перекрытия и 67% сходства в пределах перекрытия с SARS-CoV-2. Можно сделать вывод, что эти вирусы являются весьма дальними родственниками.

Популяционное исследование SARS-COV-2

Были отобраны последовательности из разных стран и разных дат из доступной базы данных NCBI (16 последовательностей находятся в папке «последовательности_коровы» в формате FASTA)

Для их выравнивания была использована программа CodonCode Aligner.

Внимание: выравнивание находится в файле alignment.pdf, а также в файле alignment.ccap (доступна для открытия в программе)

В таблице файла mutations.pdf вы можете наблюдать все данные

В той же программе было построено филогенетическое дерево (contig1.nwk или)

Также был произведен расчет скорости мутации вируса. (за основу были взяты 12 последовательностей, так как 3 из 16 не имели четкой даты, и одна представлена ранее референсной /но такую последовательность использовать нельзя, так как она не входит в базу данных RefSeq, а значит является недостаточно верной/)

Наиболее часто происходят мутации типа неконсервативный миссенс - 34,72(2)% ,т.е замена кислоты приводит к сильным изменениям свойств молекулы. Наиболее заменяющая аминокислота в таких мутациях - Гистидин (His) с вероятностью 28%, она является положительно-заряженной при нейтральном pH. Если анализировать другие мутации, но наиболее заменяющая в консервативных миссенс мутациях будет Лейцин(Leu) - 43,48%, в сайлент мутациях также Лейцин(Leu) - ~ 25,00%.

Нон-сенс мутации обнаружены не были.

Всего вместе с ИнДел и всеми возможными однонуклеотидными заменами обнаружено 95 мутаций.

Наиболее мутированная вариация вируса - MT292570, просеквенированная 10 марта 2020 года в Испании.

В среднем у всех фрагментов длина схожая, и поправку на длину можно не учитывать, но за счет делеции мы можем видеть что количество нуклеотидов разнится до 200нк максимум.

Ссылаясь на рекомендации Всемирной Организации Здравоохранения, метод ПЦР-РТ на 15.03.2020 признается эффективным для диагностики коронавирусной инфекции. Стоит заметить, что последующие мутации в области праймеров могут снизить эффективность анализа. На данный момент ВОЗ экстренно одобрила метод изотермной ПЦР. В клинической практике диагноз COVID-19 не ставится только по результатам ПЦР, но и по КТ грудной клетки или флюорографии. С учетом наблюдаемых мутаций в вирусах комплементарность сохраняется, но это не отрицает факт того, что новые варианты мутаций могут ее нарушить.

Primer	Sequence (5' to 3')
F3	TCCAGATGAGGATGAAGAAGA
B3	AGTCTGAACAACTGGTGTAAAG
FIP(F1c+F2)	AGAGCAGCAGAAAGTGGCACAGGTGATTGTGAAGAAGAAGAG
BIP(B1c+B2)	TCAACCTGAAGAAGAGCAAGAAGTGTGCTCCTCACTGCC
LoopF	CTCATATTGAGTTGATGGCTCA
LoopB	ACAAACTGTTGGTCAACAAGAC

Праймера на
изотермную ПЦР

Table 2. RT-LAMP primers alignment with other coronaviruses

Virus	GenBank ID	% Mismatch
COVID-19	MN908947	0
Bat SARS-like CoV 2015	MG772933.1	27.22
Bat SARS-like CoV 2017	MG772934.1	30.77
Bat SARS CoV RM1/2004	KY417144.1	40.83
SARS CoV ZS-C	AY395003.1	42.60
Civet SARS CoV SZ16/2003	AY304488.1	39.64
SARS CoV	NC_004718.3	40.24
SARS CoV MA15	FJ882957.1	40.24
Middle East Respiratory CoV	NC_019843.3	52.07
Belacoronavirus England 1	NC_038294.1	54.44
Murine hepatitis virus	NC_001846.1	53.25
Human Coronavirus 229E	NC_002645.1	51.48
Human Coronavirus NL63	NC_005831.2	54.44
Human Coronavirus HKU1	NC_006577.2	53.25
Human Coronavirus OC43	NC_006213.1	49.11