

Level 6

Предисловие

Третья попытка выполнения этого задания

Причины пределок те же что и с 4 задачей

В этот раз постараюсь взять меньшую выборку, порядка 10 геномов, чтобы маленькие изменения который большинство были более явные а не превращались в погрешность

Выбор региона

Мой выбор пал на Челябинскую область, т.к. когда ещё выпадет возможность поисследовать родной регион

P.S. а ну и там всего 19 последовательностей достаточно разброшенных по времени, что даёт хорошую выборку

Геномы

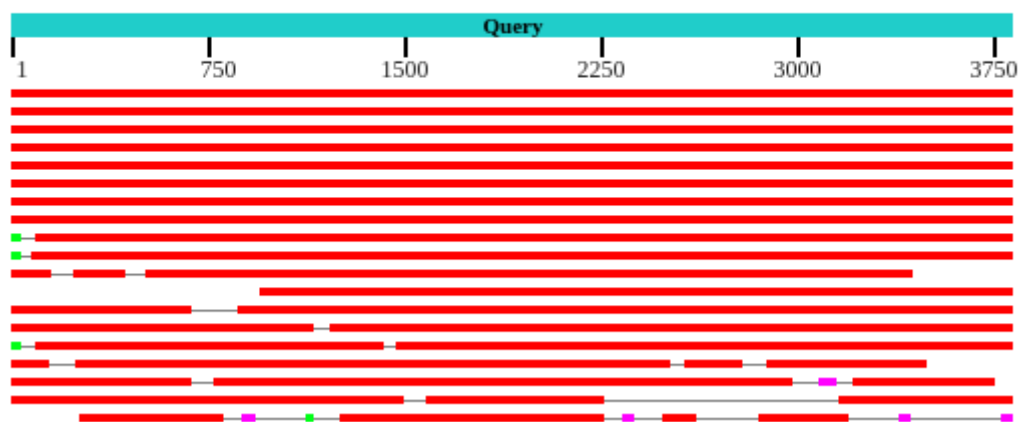
Геномы ковида - [gisaid](#)

Белок - [S-protein](#)

Выравнивание

Результат выравнивания от blastn расположен в файле [blastn](#)

Distribution of the top 43 Blast Hits on 19 subject sequences



Если я правильно понял этот график выравнивание достаточно успешное

P.S. мне кажется не правильно

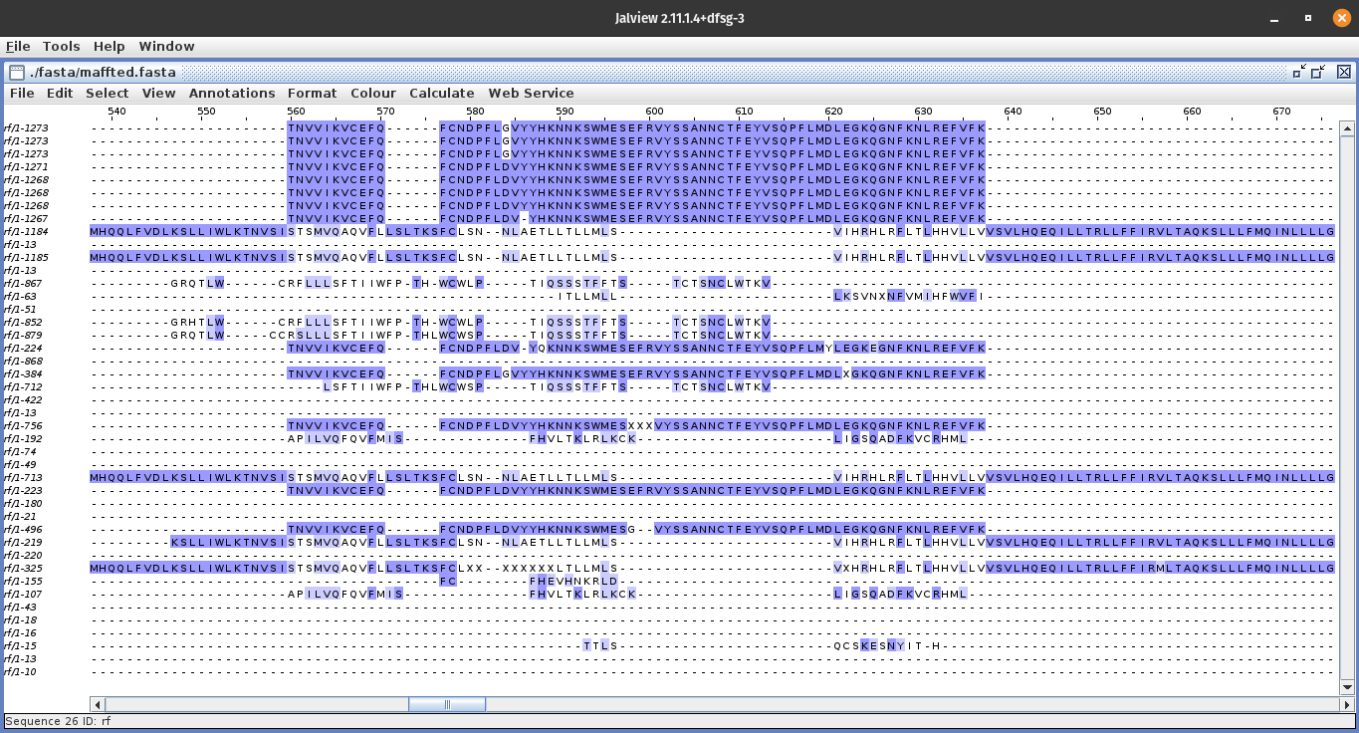
Трансляция

Отформатированный файл - [formded](#)

Перевод в белки - [translated](#)

MAFFT

Меньше чем за минуту, непривычный [результат](#)



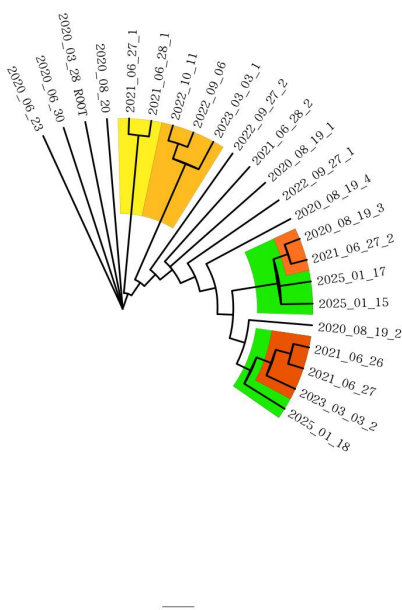
беглым взглядом можно выделить 3-4 группы, но группируются они точно не по дате, ощущение буд-то это три паралельно друг от друга развивающихся штампа

Также я позволил себе вольность, немного почистив получившуюся дату, убрав бессодержательные хвосты скопировав их в более содержательные геномы с той же датой, идея была в том чтобы похожее и неполноценное объединить в одно целое, чтобы различия между реально разными вещами были заметнее

хз, поможет это или нет

Mr.Bayers

В результате получаем достаточно структурированное дерево



Phylogenetic tree of SARS-CoV-2 sequences from the Hubei province, China, showing the evolution of the virus over time. The tree is rooted at the 2020_03_28 sequence. The sequences are color-coded: yellow for the first cluster (2021_06_27_1, 2021_06_28_1), orange for the second cluster (2022_10_11, 2022_09_06, 2023_03_03_1), green for the third cluster (2022_09_27_2, 2021_06_28_2, 2020_08_19_1), and red for the fourth cluster (2020_08_19_3, 2021_06_27_2, 2025_01_17, 2025_01_15, 2020_08_19_2, 2021_06_26, 2021_06_27, 2023_03_03_2, 2025_01_18). The tree shows a clear progression of the virus over time, with the most recent sequences (2025_01_17, 2025_01_15, 2025_01_18) being the most divergent from the root.

- Вся эта линия по всей видимости развивается из примерно одного штампа, так как те 4 генома что выявлены в 2020 году имеют почти 100% идентичность
- В июне 21 года была выявлена группа, но она по всей видимости не была достаточно успешной так что в 21 году её след и заканчивается
- В сентябре 22 года также отделилась небольшая группа, продержавшаяся ещё год, но не дожившая до наших дней
- В 20 году геном разделился на две группы обе оказались эволюционно успешными и имеют штампы этого года свежести

Таким образом в Челябинске проявлялась достаточно умеренная эволюционная активность и за 5 лет сформировалось только две эволюционно успешные различные группы