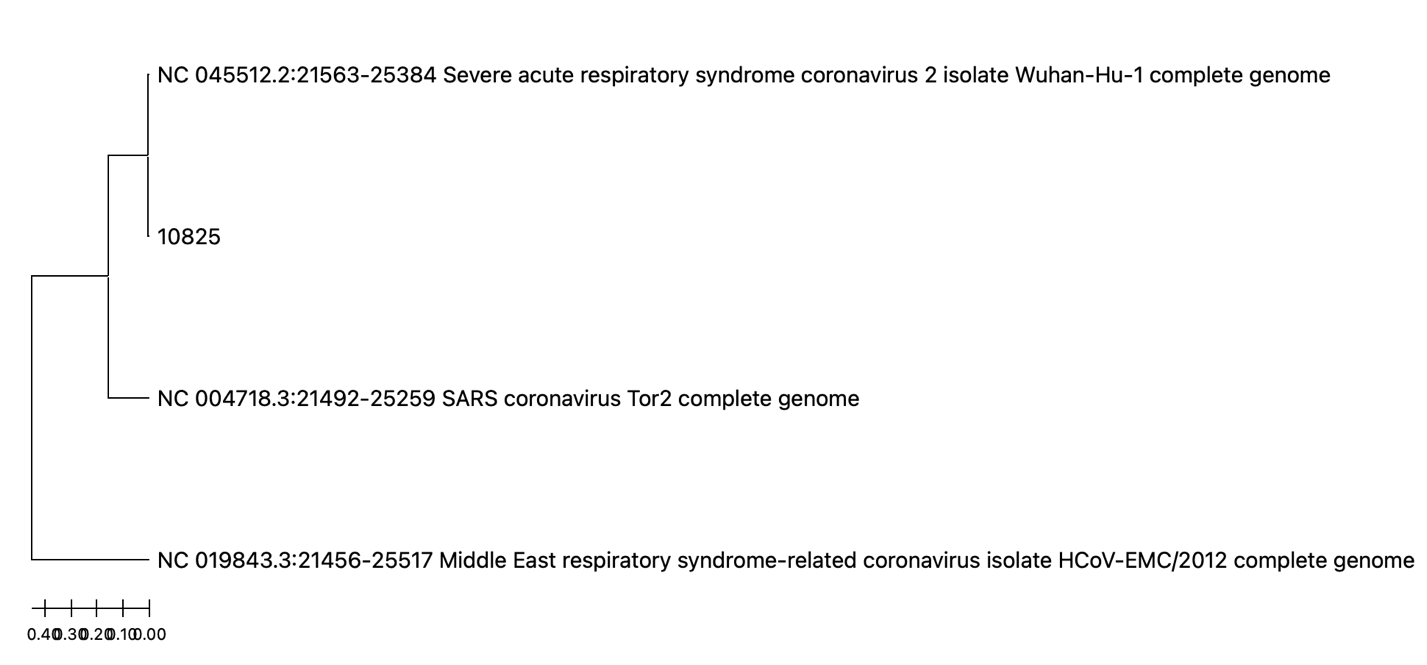
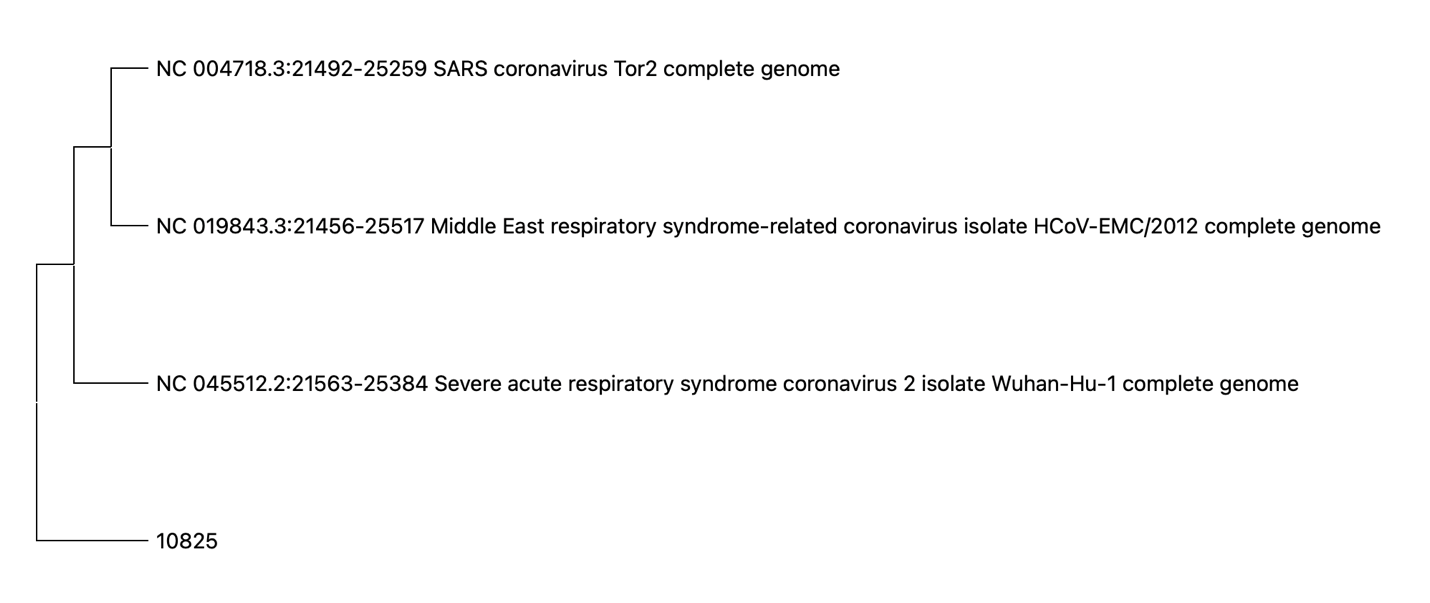
Задание 7

**1. SARS-Cov Spike protein**

|  |  |
| --- | --- |
| Страна | Название последовательности |
| **China** | >NC\_045512.2:21563-25384 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome |
| **Canda, Toronto** | >NC\_004718.3:21492-25259 SARS coronavirus Tor2, complete genome |
| **Saudi Arabia** | >NC\_019843.3:21456-25517 Middle East respiratory syndrome-related coronavirus isolate HCoV-EMC/2012, complete genome |
| **?, из семинара** | 10825 |

UPGMA tree:



Maximum Likelihood Tree:

Сравнение методов:

* Дерево UPGMA использует предположение о равномерной скорости эволюции, что упрощает реальный процесс. В данном случае кластеризация, скорее всего, отражает географическое или временное распределение штаммов.
* Maximum Likelihood алгоритм ищет такое филогенетическое дерево, которое **наиболее вероятно** объясняет, как эволюционировали последовательности. Эволюционная модель дерева используется по умолчанию, т.к. сами мы ее не задавали. Модель выбирает дерево, где вероятность перехода из одного сиквенса в другой наиболее вероятна. Такой результат более надежен, чем у UPGMA.

Результаты:

Я бы ожидала, что последовательности из Китая и Саудовской Аравии будут кластеризоваться вместе, поскольку они ближе географически, но оказалось, что Китаю ближе вариант Spike-белка из Канады. Это может объясняться тем, что, наверное, по работе люди из Китая и Канады могут летать чаще друг к другу, чем в Саудовскую Аравию. Но это немного натягивание совы на глобус)

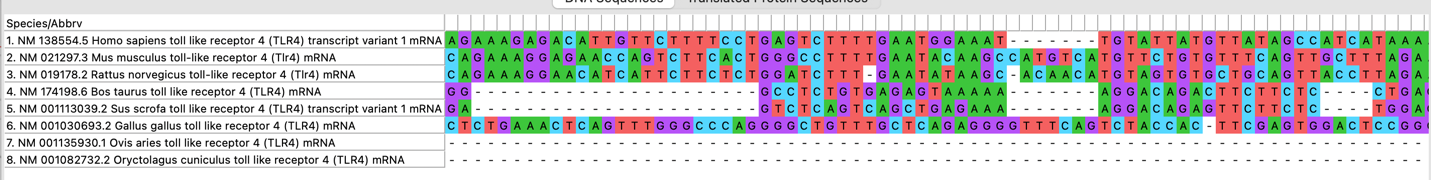
**2. TLR-4 orthologs**

Белок, кодируемый этим геном, является членом семейства Toll-подобных рецепторов (TLR), которое играет фундаментальную роль в распознавании патогенов и активации врожденного иммунитета. TLR высококонсервативны и имеют структурные и функциональные сходства. Они распознают ассоциированные с патогенами молекулярные паттерны (PAMP) и опосредуют выработку цитокинов, необходимых для развития иммунного ответа.

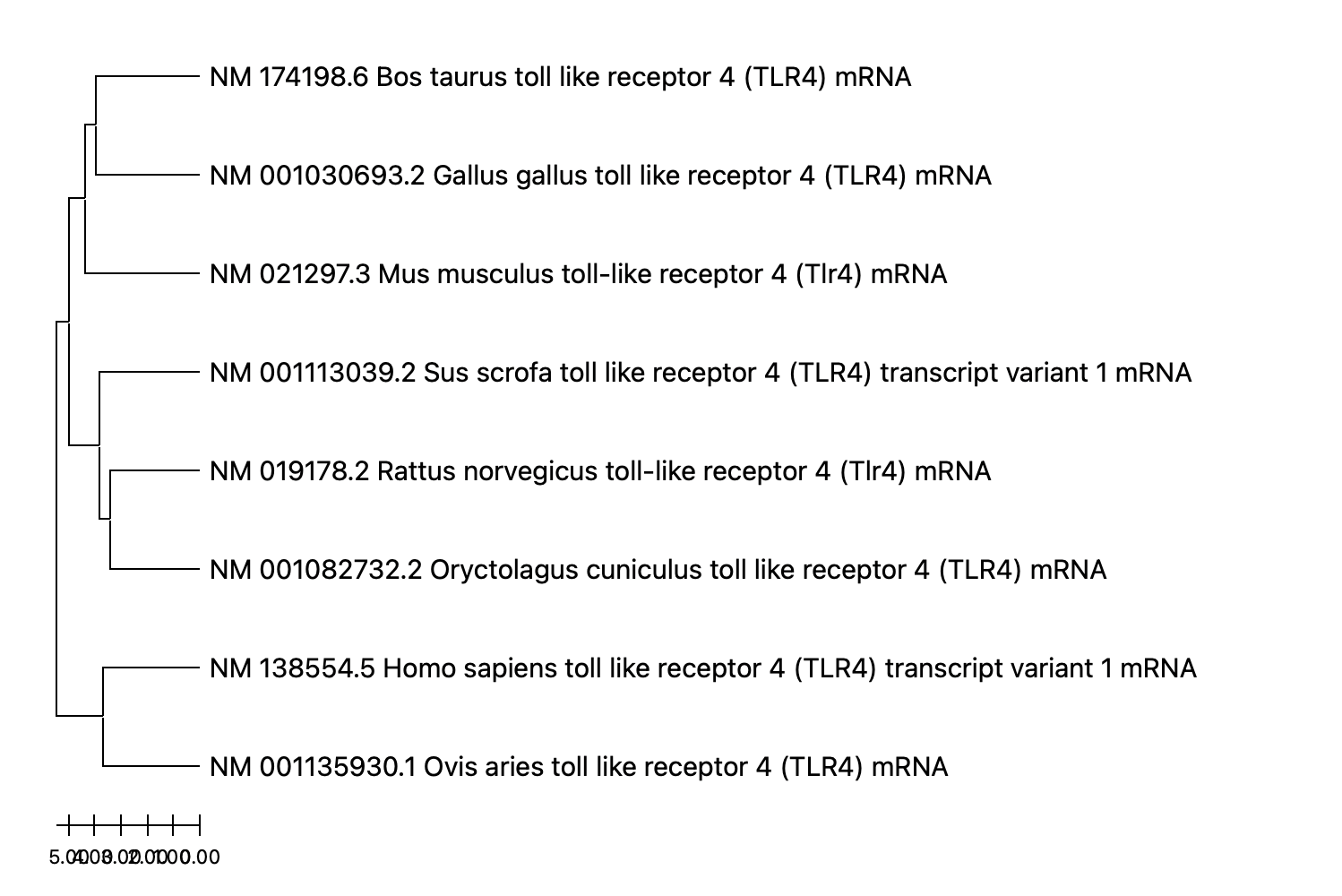
Взяты последовательности:

1. Homo sapiens
2. Mus musculus
3. Rattus norvegicus
4. Bos taurus
5. Sus scrofa
6. Gallus gallus
7. Ovis aries
8. Oryctolagus cuniculus

Привожу кусок выравнивания, где видно, что последовательности достаточно сильно отличаются, некоторые значительно короче.



UPGMA tree:



Maximum Likelihood не построилось, оно висело пару часов и никуда не продвинулось. По дереву UPGMA достаточно неожиданные результаты. Почему-то домовая мышь и крыса не в родственном кластере, но крыса и кролик – в одном, что подтверждает их родство как грызунов. Очень непонятно, почему все парнокопытные *Sus scrofa* (кабан), и в особенности [*Bos taurus* и *Ovis aries* (овца домашняя)](плюс похожие условия проживания!) оказались в трех разных кластерах.

Попробуем объяснить такую кластеризацию. Поскольку TLR – рецептор, который узнает молекулярные паттерны, ассоциированные с патогенами, например бактериальный жгутик или липополисахарид грамотрицательных бактерий, то они очень высококонсервативны. TLR-4 критически важен для врождённого иммунитета, поэтому его последовательности у разных видов могут быть очень похожи. Соответственно, даже небольшие различия (например, единичные замены) UPGMA интерпретирует как значимые, создавая ложные кластеры.