

<http://онлайн-читать.рф/%D0%BC%D0%B0%D1%80%D0%BA-%D1%82%D0%B2%D0%B5%D0%BD-%D0%BF%D1%80%D0%B8%D0%BA%D0%BB%D1%8E%D1%87%D0%B5%D0%BD%D0%B8%D1%8F-%D1%82%D0%BE%D0%BC%D0%B0-%D1%81%D0%BE%D0%B9%D0%B5%D1%80%D0%B0/2>

**Белки китайского коронавируса.**

На досуге я начал разбираться с белками китайского коронавируса из любопытства. Временно, он называется 2019-nCoV. Те три белка, на которые я смотрел, дали результаты, не совпадающие с публикациями в журналах и в интернете. Другие авторы сравнивали геномы, а я люблю белки.

Так и быть, могу разрешить некоторым студентам взять один белок и определить, на последовательности белков каких вирусов похожа его последовательность. Можно работать маленькими группами из 2х – 3х человек.

За разрешение надо: (1) Обязательно поблагодарить меня через форму (2) Написать отчёт, в конце отчёта должно быть указано кто что делал. (3) Считать меня соавтором отчёта

О нужных инструментах, рассказано будет на занятиях в конце семестра. Для меня слишком долго ждать. Поэтому минимальные советы см ниже. Подсказывать по существу не буду.

На вопросы допущенных групп, заданные через форму, отвечать буду. Но не гарантирую ответ в тот же день.

**Состав группы.**

Определяете самостоятельно. Записываетесь в форму. Ожидаю, что группы состоят из студентов одной учебной группы. Среди вас, возможно, есть продвинутые студенты: те, которые запускали программу BLAST или строили выравнивание. Одного такого я знаю. Так вот**,** в группу не должны входить два или более продвинутых студента. Как известно, фантазия лучше работает у наивных и детей.

**Что и как можно делать.**

* Выбрать белок, запомнить его идентификатор (вроде YP\_009725306.1; P – от protein), последовательность в формате fasta.
* Найти белки с похожей последовательностью у других коронавирусов.
* Отобрать похожие из нескольких разных видов вирусов.
* Описать различия между последовательностями белков.
* Посмотреть литературу и интернет.
* Написать отчёт

**Выбор белка.**

Выбираете самостоятельно. Координируете выбор так, чтобы внутри одной учебной группы выбирались разные белки. Можете договориться о координаторах.

* Список геномов 2019-nCoV в GenBank <https://www.ncbi.nlm.nih.gov/genbank/2019-ncov-seqs/>
  + В записях с описанием последовательности используется название «Wuhan seafood market pneumonia virus»
  + Интерактивный рисунок “Reference Genome” поможет выбрать ген белка. При наведении курсора на ген появляется меню. Можно получить последовательность белка (Fasta view); последовательность с описанием белка (GenBank View); запустить поиск похожих белков (BLAST Protein) в базе данных
  + Годятся зрелые белки – светло коричневые полоски, или красные полоски, под которыми нет светло коричневых полосок
  + Из “Nucleotide Sequences” выберите полный геном (complete) из колонки RefSeq. Содержится последовательность генома с описанием его.
  + “PubMed Articles” – статьи по теме

**Найти белки с похожей последовательностью у других коронавирусов**

* BLAST protein (по правой кнопке мыши)
  + Organism надо указать Coronavirus, меня интересуют только находки белков коронавирусов  
    Обнаружил, что из Coronavirus’ов иногда исключаются (а иногда нет!?) Betacoronavirus’ы, к которым относят 2019-nCoV. Поэтому запустите BLAST и по Betacoronavirus’ам, сравните и объедините результаты, если надо. Предполагаю, что в базе данных поступают в связи с большим интересом к 2019-nCoV
  + По желанию можно менять Algorithm parameters
    - Max target sequences - максимальное число выдаваемых находок
    - Expect threshold - чем меньше, тем меньше ненадежных находок. Можно поставить, например, 0.1
    - Word size – если поменять 6 на 3, то надежных находок будет больше, но работать будет дольше
  + BLAST

**Запасной вариант поиска**

* Если нет ни одной надежной находки, кроме белков 2019-nCoV то на странице BLAST выберите TBLASTN. Все остальное то же. TBLASTN сравнивает последовательность с нуклеотидной последовательностью геномов, транслируя ее в аминокислотную последовательность.

**Отобрать похожие из нескольких разных видов вирусов**

* Находки BLAST упорядочены по надежности. Надежность меряется параметром E-value, чем меньше E, тем надежнее.
* Похожесть найденной последовательности меряется двумя параметрами Identity и Coverage, см. словарик
* Среди находок в списке будут находки с указанием, что они из одного и того же вируса. Причина такая. Например, геном 2019-nCoV секвенирован из образцов от многих десятков больных. И все эти геномы лежат в базе данных.  
  Для одного вируса выбирайте одну находку, наиболее похожую на вашу последовательность.
* Ориентировочно отбирайте около 5 последовательностей. Не менее трех, не более десяти.

**Описать различия между белками.**

* В списке находок BLAST отметьте галочкой те, которые вы выбрали.
* Сохраните все последовательности в одном файле формата fasta. Download - та, что над табличкой.
* Добавьте в файл последовательность, по которой велся поиск.
* Дайте последовательностям имена по именам вирусов
* Выровняйте последовательности на сайте Uniprot.org, далее align и в окошко поместите содержимое fasta-файла с последовательностями.
* Опишите различия как сумеете. Если вопросы – спрашивайте.

**Словарик.**

* aa – amino acid, аминокислотный остаток белка.
* Polyprotein – большой белок, который разрезается на много частей, называемых зрелыми пептидами
* Mature peptide – белок, который получается из полипротеины при протеолизе, т.е. разрезании на части. Зрелый белок имеет функцю.
* Gene – ген, в данном контексте, последовательность нуклеотидов кодирующая белок или РНК.
* CDS – CoDing Sequence, кодирующая последовательность, в данном случае – то же, что ген.
* Host - хозяин вируса
* Alignment – выравнивание. В вашей последовательности и в находке похожими могут быть только участки, а не последовательности целиком. Аминокислотные остатки участков схожести сопоставляются. Такое сопоставление называется выравниванием. Сопоставленные аминокислотные остатки стоят друг под другом. Выравнивание можно увидеть щелкнув по названию находки.
* Coverage – покрытие. Coverage равно проценту длины участка похожести в вашем белке от полной длины вашего белка
* Identity –Identity равно проценту сопоставленных одинаковый аминокислотных остатков (букв) от длины выравнивания.
* Positives (синоним Symilarity) – процент сопоставленных родственных по свойствам аминокислотных остатков (включая совпадающие) от длины выравнивания. Посмотрев на выравнивания можете увидеть какие пары аминокислотных остатков родственны.

**Как и когда сдавать отчёт.**

* Отчет оформляете как учили.
* Лучше всего – как word-файл.
* Сообщаете через форму веб-адрес отчета. Пока по почте. Форму сделаю позже.
* Я проверю состояние дел в понедельник 17го февраля.
* Окончательно проверю отчеты проверю до 12 апреля



Взгляд с необычной точки зрения