# Задания по темам «выравнивание» и «поиск»

*Xxxxx в заданиях и подсказках – ваша фамилия, написанная латинскими буквами.*

МНОЖЕСТВЕННОЕ ВЫРАВНИВАНИЕ. РЕКОНСТРУКЦИЯ ФИЛОГЕНИИ.

**1. Постройте выравнивание последовательностей белков, отобранных вами на предыдущем занятии (задание II.5) и сохраните филогенетическое дерево.** В протоколе укажите

a) участок («от позиции … до позиции …») “правильного” выравнивания (на котором в каждой колонке все аминокислотные остатки гомологичны); и участок сомнительного выравнивания (на котором нельзя утверждать, что в колонках стоят только гомологичные аминокислотные остатки)

b) на филогенетическом дереве две самых близких последовательности и две самых далеких

Результат — **протокол**, **файл с выравниванием** в формате проекта JalView (Xxxxx\_alignment.jvp), **файл с филогенетическим деревом** в формате Newick (Xxxxx.tre) и **графический файл с изображением дерева** (Xxxxx.png).

**Указания.**

- используйте программу Jalview;

- откройте файл с последовательностями в формате fasta;

- выровняйте последовательности одной из программ (muscle – часто используемая; mafft – быстрая, Tcoffee – хорошая, но медленная, любая другая) с умолчательными параметрами.

Программы доступны в меню Web Services => Alignment

Выравнивание появляется в новом окне  
- раскрасьте выравнивание по методу Clustalx и подберите порог “by conservation” так, чтобы лучше было видно сходство и различия между последовательностями;

- проверьте выравнивание! В нем могут оказаться последовательности негомологичных белков (что видно по малому числу совпадений) или короткие фрагменты белков. Удалите “плохие” последовательности из выравнивания;

– постройте филогенетическое дерево (Calculate → Calculate Tree → Average Distance Using % Identity).

– постройте филогенетическое дерево меодом BioNJ (Neighbour Joining) с помощью веб-сервиса

<http://phylogeny.lirmm.fr/phylo_cgi/index.cgi>

В окошко положите выравнивание в формате фаста. Чтобы сохранить выравнивание в Jalview выделите все (Esc → Ctrl A), правую кнопку мыши нажмите на именах, Selection → Output to textbox.

– сохраните дерево в формате Newick (имя файла XXXXXXX.tre где XXXXXXX – ваша фамилия латинскими буквами)

– сохраните «Проект JalView» в виде файла XXXXXXX\_homologs.jvp (в самом большом окне File → Save Project → находите свою папку, в нижнем меню выбираете JalView, в верхнем окошке пишете имя файла).

– Откройте дерево (файл с расширением tre) программой Mega. Укорените в среднюю точку. Придайте ему какую-нибудь красивую форму и сохраните изображение.

СРАВНЕНИЕ ГЕНОМОВ ВИРУСОВ С ПОМОЩЬЮ КАРТЫ ЛОКАЛЬНОГО СХОДСТВА

**2. Найдите и опишите в протоколе две пары вирусов, геномы которых находятся в следующих отношениях:**

a) **(первая пара)** гомологичны почти по всей длине: сумма длин гомологичных участков больше половины каждого из геномов;

b) **(вторая пара)** имеются сравнительно короткие гомологичные участки: суммарная длина гомологичных участков меньше половины каждого из геномов.

Протокол должен включать

- **семейство**, **название каждого вируса** из пары и **его хозяина**;

- **идентификаторы записей с геномом**;

- **длины геномов** (в п.н.);

- **суммарный процент гомологичных участков** относительно длины одного из геномов;

- **карту локального сходства** геномов;

- **характеристику сходства геномов** – свободный текст (о длинах гомологичных участков, их достоверности – E-value, и типичном проценте совпадающих букв в выравниваниях гомологичных участков).

**Рекомендации**

- откройте страницу геномов: NCBI Genome (<http://www.ncbi.nlm.nih.gov/genome/>) => Viruses => Browse viral genomes by family

- в другой вкладке откройте страницу “BLAST двух последовательностей”, доступную со страницы BLAST на сайте NCBI <http://blast.ncbi.nlm.nih.gov/Blast.cgi> → Nucleotide BLAST → поставьте галочку в чекбоксе «Align two or more sequences».  
Чтобы оперативно проверять сходство выбранных геномов:  
 в окошках можно указать идентификаторы записей (типа NC\_123..) вместо самих последовательностей;

в меню “Optimize for” выбирайте “ Somewhat similar sequences (blastn)”;

дождавшись результата, смотрите на карту локального сходства (DotMatrix).

- выбирайте вирусы с нефрагментированным геномом длины меньше 20 000 п.н. (соблюдение этих условий не обязательно, но упрощает анализ);

- для a) выбирайте вирусы из одного семейства или даже рода;

- для с) выбирайте вирусы из разных родов или семейств; используйте свои знания из вирусологии!

Наверное, придется перебрать несколько пар вирусов пока найдете подходящие!

ВЫРАВНИВАНИЕ ГОМОЛОГИЧНЫХ ГЕНОМОВ

**3. Постройте выравнивание полных геномов из задания 2а и укажите координаты самого длинного участка выравнивания с гомологичными последовательностями (номера позиций от – до).**   
  
Результат. **Протокол с координатами** участка выравнивания. **Файл с выравниванием** в формате проекта JalView (XXXXXXX\_2genomes.jvp)

**Указания.**

- Сохраните обе последовательности в одном файле в формате fasta. Найдите каждую из последовательностей в GenBank (если не закрыли окно – щелканьем по идентификатору). Сохраните в fasta формате (send => to file, формат fasta); вторую добавьте в тот же файл.

- Откройте файл программой Jalview (File => Input from alignment)  
- раскрасьте по нуклеотидам ( color => nucleotides) и

- установите раскраску только при 100% совпадении букв в колонке (color => Above identity threshold)

- выровняйте последовательности (Web service => alignment => Muscle (default) или любая др. программа выравнивания, которая не откажется работать)

- раскрасьте так же выровненные последовательности

- найдите достаточно длинные (более 20 столбцов) участки, на которых нет гомологичности; это либо участки с делециями/вставками, либо такие, на которых идентичных колонок меньше половины; занесите в протокол координаты участка с гомологичными последовательностями; не могу предвидеть, как будет выглядеть ваше выравнивание, поэтому проявляйте здравый смысл по мере необходимости))) Или спрашивайте преподавателей.

- (\*) для интереса найдите рамку считывания какого-нибудь гена по чередованию: две консервативные позиции, одна неконсервативная и т.д. (ясно ли, в чем дело?)

- coхраните результат как «проект JalView» в виде файла XXXXXXX\_2genomes.jvp (в самом большом окне File → Save Project → находите свою папку, в нижнем меню выбираете JalView, в верхнем окошке пишете имя файла).

ВЫРАВНИВАНИЕ КОДИРУЮЩИХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ: ВЫРАВНИВАТЬ ГЕНЫ ИЛИ БЕЛКИ?

**4. Для пары энтеровирусов сравните процент гомологичных участков, определяемых выравниванием нуклеотидных последовательностей геномов (программа blastn) и формальных трансляций этих геномов в шести рамках (программа tblastx)**

**Указания.**

- используйте ту же страницу “BLAST двух последовательностей”, запомните процент гомологичных участков и сохраните карту в протоколе;

- вернитесь (Edit and resubmit) и в о второй раз выберите в верхнем меню программу tblastx вместо blastn

- оформите протокол (две карты, два процента “покрытия” и комментарий)

АННОТАЦИЯ ПО СХОДСТВУ

**5. Предскажите функцию одного из белков, предсказанных автоматической програмимой в *de novo* секвенированном геноме паразита водорослей *Amoeboaphelidium protococcarum*.**

**Указания.**

- используйте поиск blastp в базе данных Refseq Proteins;

- выберите лучшую находку с хорошо аннотированным белком;

- опишите результат в протоколе (E-value, coverage %, identity %, функция, таксономия выбранной находки).

**6. Найдите два белка с разной доменной архитектурой, но с одинаковым доменом. [Постройте карту локального сходства их последовательностей - дополнительное, т.е. не обязательное задание]**

В протоколе сохраните **доменные архитектуры белков в виде картинок** с подписями названий доменов.

Два варианта действий. (1) Выберите «любимый» белок. Посмотрите его доменную архитектуру, введя код доступа в окошко БД Pfam (<http://pfam.xfam.org/>). Выберите один домен, щелкнув по нему перейдете на страницу домена, на ней есть ссылка на доменные архитектуры, включающие домен, на выравнивание последовательностей домена из разных организмов и много всего полезного. (2) На странице Pfam найдите страницу домена поиском по ключевым словам или через Browse. Посмотрите доменные архитектуры с этим доменом