

Заведите файл Hxxxx_Protocol.docx или Hxxxx_Protocol.txt или Hxxxx_Protocol.odt. Здесь Hxxxx – ваша фамилия латинскими буквами, например Ivanov (чтобы у нас не перепутались файлы). Помимо ответов на вопросы, заносите в протокол имена создаваемых файлов и другую информацию о сделанных упражнениях.

Обязательные задания по банкам данных

I. Содержимое записей

1. По адресу <https://www.ebi.ac.uk/ena/browser/api/embl/X00218.1> можно увидеть одну из записей банка EMBL. Занесите в протокол ответы на следующие вопросы:
 - а) Из какого организма взята ДНК, последовательность которой описывается?
 - б) Авторы записи.
 - в) Номер доступа (AC) записи.
 - г) Каковы координаты последовательности лейциновой тРНК в этой записи?
2. Пользуясь текстовым редактором, создайте файл, в котором в Fasta-формате находилась бы последовательность лейциновой тРНК из записи EMBL предыдущего задания.
3. По адресу <https://www.ebi.ac.uk/ena/browser/api/embl/J02459.1> находится запись EMBL, описывающая геном фага λ. Найдите в этой записи информацию о закодированных в геноме компонентах капсида (capsid component) фага. Занесите в протокол:
 - а) координаты генов компонентов капсида;
 - б) номера доступа записей банка Swiss-Prot, описывающих закодированные этими генами белки;
 - в) на сколько статей ссылается запись о геноме? В какие годы выходили эти статьи (самая ранняя — самая поздняя);
 - г) (* — необязательно) какая интересная (на ваш взгляд) информация содержится в поле CC данной записи?Сохраните **аминокислотную** последовательность одного из компонентов капсида в отдельном файле в Fasta-формате.
4. По адресу <https://www.uniprot.org/uniprot/P03305.txt> находится одна из записей банка UniprotKB. Занесите в протокол ответы на следующие вопросы:
 - а) Из какого организма взят белок, последовательность которого описана?
 - б) Что это за белок?
 - в) Каковы идентификатор и номер доступа записи?
 - г) Каковы номера доступа записей банка EMBL, в которых описан ген этого белка?
 - д) Каким (зрелым) белкам принадлежат остатки 500, 1000, 2000 этой последовательности?
 - е) Последовательность каких зрелых белков слегка варьируется у разных изолятов данного штамма?

II. Поиск в банках данных

1. С помощью поиска на сайте NCBI найдите в RefSeq запись, описывающую геном или сегмент генома какого-либо вируса.

Выбор вируса – за вами. Лучше выбирать не очень большие записи – менее 20 000 п.н., в которых описаны гены нескольких белков (а не одного полипротеина).

2. Сохраните выбранный геном в двух форматах – GenBank (т.е. полную запись) и fasta.

Названия файлов — Xxxxxx_AAAAAAA_genome.gb, Xxxxxx_AAAAAAA_genome.fasta, где Xxxxxx — ваша фамилия, а AAAAAAA – номер доступа (AC, Accession) записи RefSeq с геномом вируса, например, AF030154 или NC_044935.

3. Кратко опишите в протоколе содержимое записи (название вируса, таксономия, число пар нуклеотидов, число описанных в записи генов и закодированных белков).

4. Скопируйте из Uniprot или RefSeq файлы с полной записью и с последовательностью в fasta-формате одного из белков вируса. Опишите в протоколе всю понятную вам информацию из записи.

Названия файлов — Xxxxxx_AAAAA.uniprot или Xxxxxx_AAAA.refseq и Xxxxxx_AAAAA.fasta ; AAAAA – код доступа записи Uniprot или RefSeq. Выбор белка за вами. Выбирайте белок, про который хоть что-то известно (есть название, отличное от “unknown protein”).

5. Найдите и сохраните в файле в формате fasta последовательности из Uniprot 10–15 белков, имеющих такое же название, что и выбранный вами белок (см. 4.) и принадлежащий вирусам из того же семейства, что и выбранный вами вирус (см. 2.). Уровень родства и название белка можно, при необходимости, разумно варьировать.

Имя файла — Xxxxxx_TTTTTT_prot.fasta, где TTTTTT — краткое название семейства.

Когда закончите, вышлите протокол и файлы на sas@belozersky.msu.ru

Важно: протокол должен быть читаем без обращения к тексту задания. То есть нельзя писать:

1. а) Свинья б) Иванов в) X01234 и т.д.

Можно, например, так:

1. Запись EMBL: а) ДНК взята из свиньи. б) Автор записи — Иванов. в) номер доступа — X01234 и т.д.