Заведите файл Xxxxx\_Protocol.docx или Xxxxx\_Protocol.txt. Здесь Xxxxx – 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.

# Обязательные задания по банкам данных

# Содержимое записей

# 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:

# Из какого организма взята ДНК, последовательность которой описывается?

# Авторы записи.

# Из какого класса данных данная запись?

# В каком таксономическом разделе EMBL находится запись? Для чего предназначен этот раздел (какие последовательности содержит)?

# Каков ее номер доступа (AC)?

# Каковы координаты последовательности лейциновой тРНК в этой записи?

# Пользуясь текстовым редактором, создайте файл, в котором в Fasta-формате находилась бы последовательность лейциновой тРНК из записи EMBL предыдущего задания.

* 1. 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 координаты генов, а также номера доступа соответствующих записей банка Swiss-Prot. Сохраните последовательность одного из белков в отдельном файле в Fasta-формате.
  2. По адресу <https://www.uniprot.org/uniprot/P03305.txt>находится одна из записей банка UniprotKB. Занесите в протокол ответы на следующие вопросы:
     1. Какому организму принадлежит белок, последовательность которого описана?
     2. Что это за белок?
     3. Каковы идентификатор и номер доступа записи?
     4. Каковы номера доступа записей банка EMBL, в которых описан ген этого белка?
     5. Каким (зрелым) белкам принадлежат остатки 500, 1000, 2000 этой последовательности?
     6. Последовательность какого зрелого белка слегка варьируется у разных изолятов данного штамма?
     7. (\*) В какой публикации описана структура генома? Открыт ли к ней бесплатный доступ? Скопируйте в протокол аннотацию статьи.

1. **Поиск в банках данных**
   1. С помощью поиска на сайте ENA найдите в разделе STD банка EMBL 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 – EMBL (полную запись) и fasta.   
      *Названия файлов — Xxxxx\_AAAAAAA\_genome.embl, Xxxxx\_AAAAAAA\_genome.fasta, где Xxxxx — ваша фамилия, а AAAAAAA – код доступа записи EMBL с геномом вируса, например, JX453331.*
   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 файлы с полной записью и с последовательностью в fasta-формате одного из белков вируса. Опишите в протоколе всю понятную вам информацию из записи Uniprot.   
      *Названия файлов — Xxxxx\_AAAAA.uniprot, Xxxxx\_AAAAA.fasta ; AAAAA – код доступа записи Uniprot. 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 — Xxxxx\_TTTTTTT\_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) STD *и т.д.  
Можно, например, так:*1. Запись EMBL. а) ДНК взята из свиньи. б) Автор записи — Иванов. в) Запись принадлежит классу STD *и.т.д.*