

Bioinformatics HW2

Ershov Ivan

November 2021

Выбранный ген - ribosomal protein L6

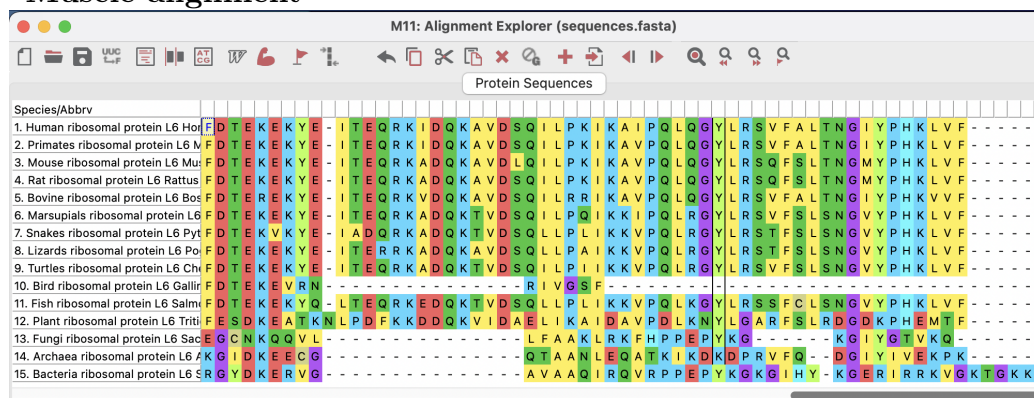
Задание 1. Найдем аминокислотные последовательности генов для различных видов и составим fasta-файл с последовательностями. Поставим в начало названия гена строку с названием организма, чтобы организм в будущем отображался на листьях дерева.

Вот ссылки на последовательности, которые я скачал:

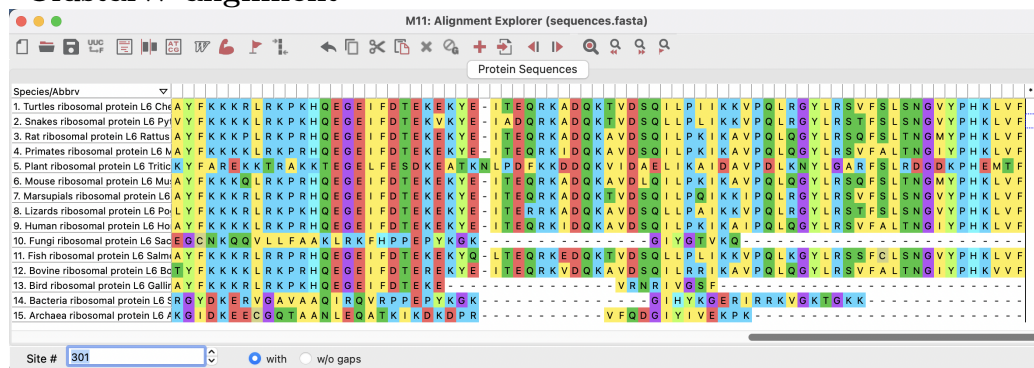
1. Человек - human, вид *Homo sapiens*
2. Обезьяна - primates, вид *Macaca mulatta*
3. Грузуны(мышь) - mouse, вид *Mus musculus*
4. Грузуны(крыса) - rat, вид *Rattus norvegicus*
5. Копытное - bovine, вид *Bos taurus*
6. Сумчатое - marsupials, вид *Vombatus ursinus*
7. Пресмыкающиеся(змея) – snakes, вид *Python bivittatus*
8. Пресмыкающиеся(ящерица) – lizards, вид *Pogona vitticeps*
9. Пресмыкающиеся(черепаха) – turtles, вид *Chelonia mydas*
10. Птица - bird, вид *Gallirallus okinawae*
11. Рыба - fish, вид *Salmo salar*
12. Растение - plant, вид *Triticum aestivum*
13. Пекарские дрожжи - fungi, вид *Saccharomyces cerevisiae*
14. Архея - archaea, вид *Archaeoglobus profundus* DSM 5631
15. Бактерия –bacteria, вид *Sulfobacillus acidophilus* DSM 10332

Задание 2. С помощью приложения Mega произведем множественное выравнивание последовательностей алгоритмами ClustalW и Muscle (представляю здесь скриншоты полученных выравниваний. К письму тоже приложу итоговые выравнивания файлами)

Muscle alignment



ClustalW alignment

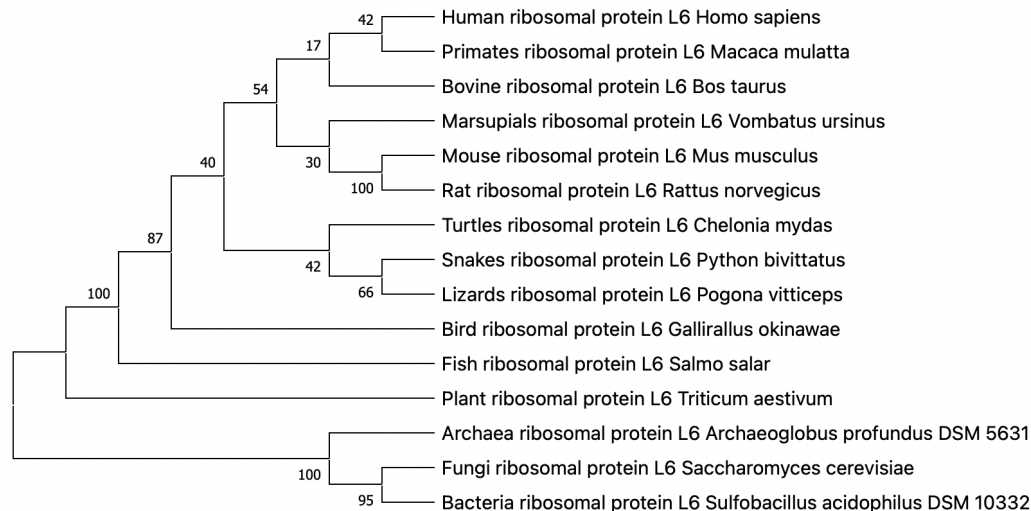


Полученные выравнивания отличаются по крайней мере расположением пропусков (это явно видно на скриншотах)

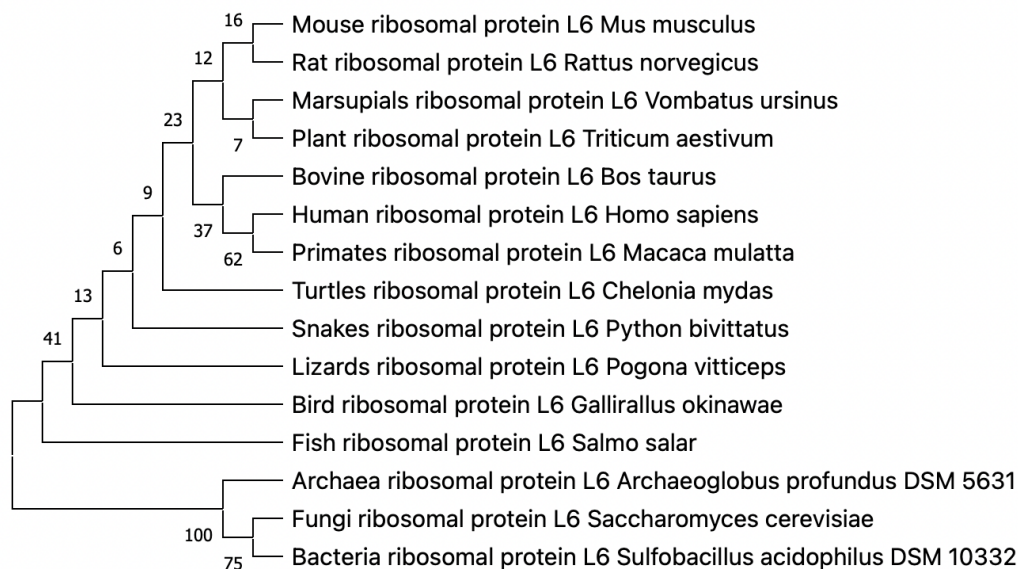
Задание 3.

Построим деревья для Muscle alignment

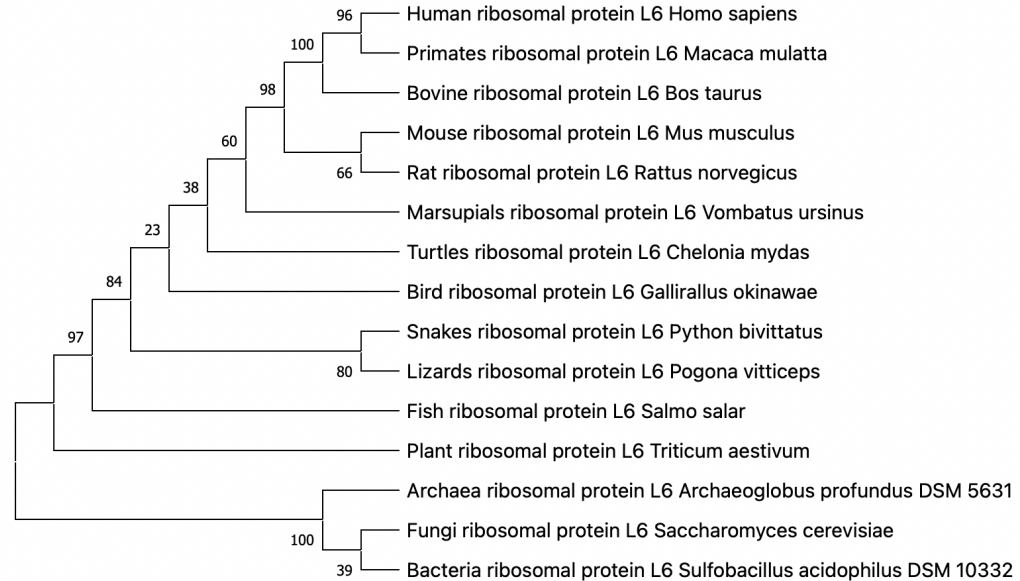
- методом расстояний (UPGMA) Bootstrap 100



- методом расстояний (NJ) Bootstrap 100

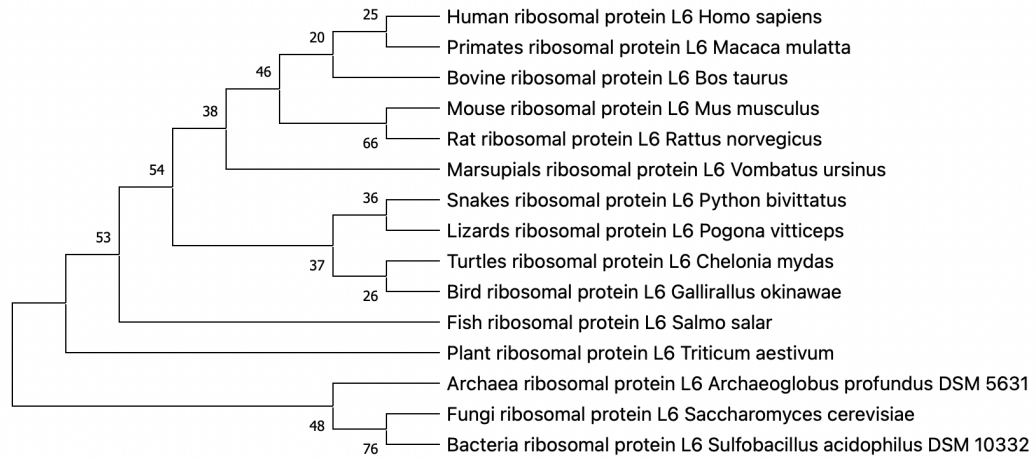


- методом максимального правдоподобия Bootstrap 100

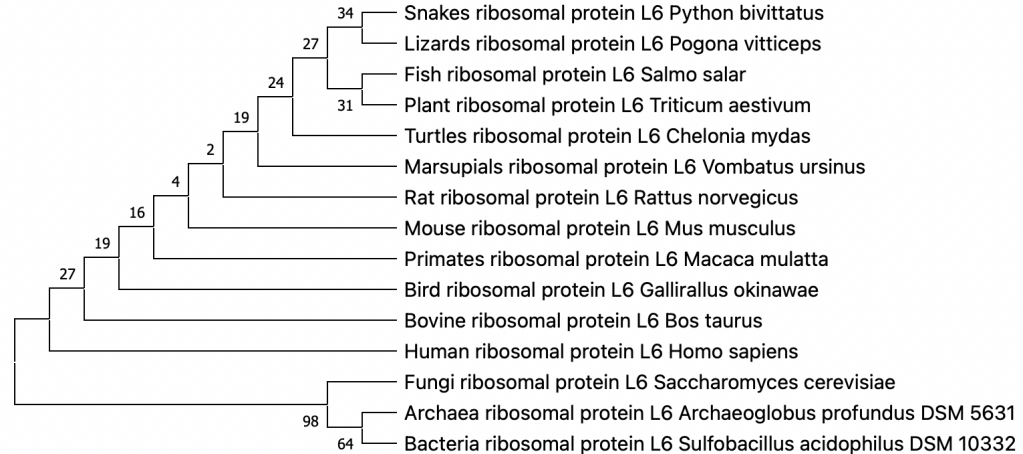


Построим деревья для ClustalW alignment

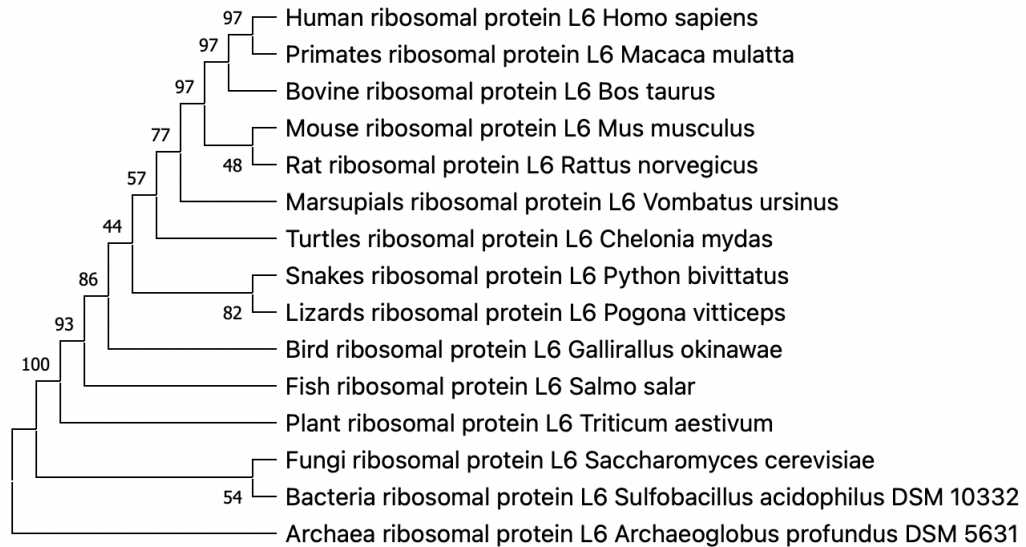
- методом расстояний (UPGMA) Bootstrap 100



- методом расстояний (NJ) Bootstrap 100



- методом максимального правдоподобия Bootstrap 100



Вывод:

Какой алгоритм выравнивания лучше сработал - ClustalW или Muscle?

Сложно сравнивать два этих алгоритма, но, мне кажется, метод Muscle отработал лучше. Дерево, построенное методом расстояний (NJ) с ClustalW, обладает маленьким количеством различных разветлений, разные виды по большей части соединены последовательно, а некоторые виды, которые, казалось бы должны быть вместе, такие как ящерица, змея и черепаха, расположились даже не близко. В это же время в Muscle тот же метод объединил по крайней мере мышь и крысу, а уже упомянутые змеи, ящерицы и черепахи расположены близко.

Одинаковая ли получилась топология деревьев при построении разными методами?

В целом, можно заметить, что у деревьев построенных одним методом, но для последовательностей выровненных по-разному, имеют сходства в узлах и ветвях. Некоторые деревья (например построенные методом расстояний UPGMA) очень похожи (хотя опять же, не идентичны), а некоторые скорее непохожи (как описанные в предыдущем пункте деревья, построенные при помощи NJ) и имеют мало сходств.

Одинаковые ли получились бутстрэп-значения?

Ответ в большинстве случаев нет, для этого достаточно взглянуть на построенные деревья и удостовериться в этом. Однако почти все бутстрэп-значения деревьев, построенные одним методом, но разными выравниваниями, различаются в ± 20 .

Совпадают ли деревья, построенные по одному гену с принятыми деревьями видов?

Есть принятая таксономия, как произошли виды. Она составлена на основе, как и обычной биологии, так и молекулярной биологии. На сайте NCBI можно построить дерево таксономии по выбранным видам организмов. Я так и сделал и объединил выбранные мной виды в дерево (дерево на следующей странице)

