

# Описание протеома бактерии *Sulfurihydrogenibium azorense* Az-Fu1

Головачев Ярослав Алексеевич

Биологический факультет МГУ им. М. В. Ломоносова

## РЕЗЮМЕ

В данной работе изучается протеом бактерии *Sulfurihydrogenibium azorense* Az-Fu1, распределение длин белков и расположение генов белков и РНК на различных цепях ДНК. Данная работа важна для изучения жизнедеятельности бактерии.

## ВВЕДЕНИЕ

Протеом – совокупность всех белков организма. Его изучение позволяет выяснить механизмы жизнедеятельности, сигнальные пути организма, что в будущем позволит использовать эти механизмы с помощью биоинженерии в различных областях человеческой жизни. С развитием биоинформатических методов изучение протеома становится значительно более лёгким, что позволяет увеличить эффективность исследований.

## МАТЕРИАЛЫ И МЕТОДЫ

Из банка данных NCBI были взяты данные полного секвенирования генома бактерии *Sulfurihydrogenibium azorense* Az-Fu1 (идентификатор записи NCBI – NC\_012438). В работе использовались программы Far manager, Microsoft Excel 2010, Microsoft Excel 2003. Из файлов NC\_012438.ptt и NC\_012438.rnt были взяты данные по положению закодированных белков и РНК в геноме и импортированы в файл Excel на листы proteins и rnas соответственно. Для установления расположения генов на цепях ДНК («+» или «-») была создана таблица на странице strands, где для каждого гена указана цепь, на которой он закодирован, и вид гена (CDS или RNA). Для определения же числа генов, закодированных на цепях, на странице distribution файла Excel была создана сводная таблица, на которой можно увидеть, сколько генов белков и РНК закодированы на той или иной цепи. Затем были обособлены длины белков (страница protein\_length файла) в порядке возрастания, были подсчитаны средняя длина белка и мода, а также установлены максимальный и минимальный по длине ген. Была построена диаграмма длин белков, в которой количество белков определённой длины (укладывающейся в заданный промежуток) ставилась в зависимость от длины самого промежутка.

## РЕЗУЛЬТАТЫ

В ходе изучения распределения генов белков и РНК по цепям ДНК были получены такие данные:

Цепь	Белок	РНК	Общий итог
-	821	25	846
+	902	20	922
Общий итог	1723	45	1768

Также была построена диаграмма распределения белков по их длинам:

Гистограмма распределения длин белков



На гистограмме можно видеть, что распределение длин белков асимметрично, максимальное количество белков имеют длину от 100 до 200 а.о., минимальное – >1500 а.о.

Средняя длина белка – 299,2687 а.о., медианное значение длины белка – 262 а.о., максимальное – 1605 а.о., минимальное – 29 а.о.

## ОБСУЖДЕНИЕ

Тот факт, что гены белков и РНК примерно в равном количестве распределены по «+» и «-» - цепям ДНК, не должно вызывать удивления. Каждая цепь ДНК может считываться от 5' к 3'-концу, соответственно на каждой из них могут располагаться гены.

Полученные результаты длин белков также ожидаемы: не наблюдается белков со слишком большой длиной (>1700 а.о.), так как такие длинные белковые цепи весьма неудобны в работе, белки с длинной цепью редко бывают необходимы, гораздо выгоднее в данном случае – сложный белок из нескольких субъединиц.

## БЛАГОДАРНОСТИ

Хочется принести благодарности преподавателям: Андрею Владимировичу Алексеевскому, Борису Эдуардовичу Нагаеву и Ивану Сергеевичу Русинову за помощь в написании практикума и объяснение непонятных моментов.

## СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Файл Excel – [kodomofbb.msu.ru/~polietilenglikol/term1/S.azorense\\_Az-Fu1\\_proteome.xlsx](http://kodomofbb.msu.ru/~polietilenglikol/term1/S.azorense_Az-Fu1_proteome.xlsx)