

Обзор протеома археи *Methanopyrus kandleri* AV19

Желтова А.В.

Факультет биоинженерии и биоинформатики МГУ им. М.В.Ломоносова

РЕЗЮМЕ

Данная работа посвящена исследованию протеома археи *Methanopyrus kandleri*, штамм AV19. С помощью программы Microsoft Office Excel 2010 было уставлено количество генов белков и генов РНК, расположенных на прямой цепи и обратной цепи. Также была построена гистограмма распределения длин белков.

1 ВВЕДЕНИЕ

M. kandleri относится к роду Метаногены. Впервые данная архея была обнаружена на стене одного из черных курильщиков в Калифорнийском заливе. Её геном состоит из кольцевой хромосомы, длина последовательности которой составляет 1694969 п.н. В данной работе была изучена совокупность всех белков данного организма – протеом. Всего протеом *M. kandleri* содержит гены, кодирующие 1687 белков, и 38 генов РНК.

2 МАТЕРИАЛЫ И МЕТОДЫ

Данные полного секвенирования генома археи *Methanopyrus kandleri* AV19, были получены из базы данных National Centre of Biotechnology Information. Идентификатор записи в базе данных RefSeq на сайте NCBI, который содержит информацию о полном геноме археи: NC_003551.

Работа была выполнена в программе Microsoft Office Excel 2010. Данные из файлов NC_003551.ptt и NC_003551.rnt были импортированы в данную программу, с помощью которой осуществлялась их дальнейшая обработка. При подсчете количества белков, длина которых находится в указанном диапазоне, и количества генов белков и генов РНК, расположенных на «+» и «-» - цепях была использована функция СЧЁТЕСЛИМН. С помощью функции БИНОМ.РАСП была проверена гипотеза о том, что гены распределены по цепочкам случайно с вероятностью 0.5. Для установления распределения длин белков с помощью Microsoft Excel была построена столбчатая гистограмма.

3 РЕЗУЛЬТАТЫ

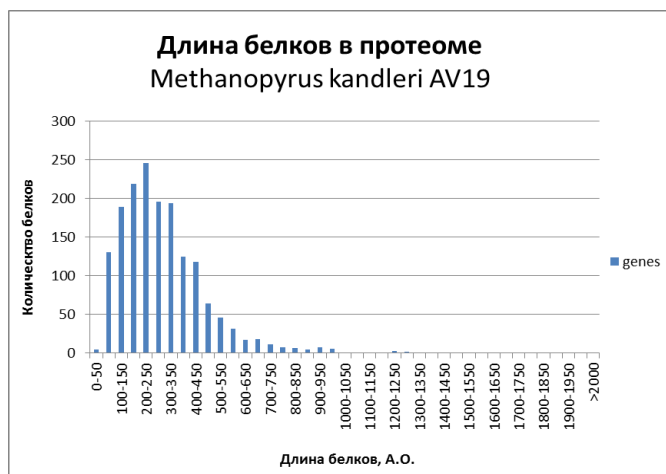
Число генов, расположенных на прямой и обратных цепях практически одинаково. Полученные величины p -value подтверждает гипотезу о том, что гены распределены по цепочкам случайно с вероятностью 0.5.

	гены белков	гены РНК
Прямая цепь	843	25
Обратная цепь	844	21
Общее количество	1687	46

p-value	1	0,99992174
---------	---	------------

Таблица 1. Количество генов белков и РНК на прямой и обратной цепях, p -value для данных величин.

На гистограмме 1 представлено распределение длин белков. Наибольшее число белков имеет длину от 200 до 250 а.о.



Гистограмма 1. Распределение длин генов из протеома *Methanopyrus kandleri* AV19. По вертикали расположено количество белков данной длины, а по горизонтали – диапазоны длин белков в а.о.

4 ОБСУЖДЕНИЕ

Число генов, расположенных на прямой и обратной цепях почти одинаково. Наибольшее число генов имеют длину от 100 до 400 а.о. 246 генов белков имеют длину 200-250 а.о. Для бактерий данное значение можно считать стандартной средней длиной. Длина только одного белка превышает 2000 а.о. Гипотезу о том, что гены распределены по цепочкам случайно с вероятностью 0.5, подтвердилась.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

[proteome.xlsx](#)- файл с гистограммой распределения длин белков и таблицей распределения генов белков и генов РНК по цепям.

БЛАГОДАРНОСТИ

Автор выражает благодарность А.В. Алексеевскому и Ивану Русинову.