

## ОБЗОР ПРОТЕОМА АРХЕИ *DESULFUROCOCCLUS KAMCHATKENSIS* 1221N

БАЙКУЗИНА ПОЛИНА

*Факультет биоинженерии и биоинформатики Московский государственный университет имени  
Ломоносова, Москва, Россия  
polina.bai09@gmail.com*

Данная работа посвящена исследованию протеома археи *Desulfurococcus kamchatkensis* 1221n. Работа выполнялась в рамках курса биоинформатики. Анализ проводился с помощью программы Microsoft Excel 2010. Исследовалось распределение длин белков и расположение генов на разных цепях ДНК. В результате работы было установлено белки какой длины преобладают протеоме археи *Desulfurococcus kamchatkensis* 1221n.

*Основные понятия:* Протеом; Длина белка; Распределение генов.

### 1. Введение

*Desulfurococcus kamchatkensis* - анаэробная органотрофная гипертермофильная архея, найденная в наземных горячих источниках. Ее геном состоит из одной кольцевой хромосомы без внехромосомных элементов. Всего в геноме 1538 генов, из них 1471 кодируют белки и 67 – тРНК и рРНК.

Цель данной работы – исследование протеома археи *Desulfurococcus kamchatkensis* 1221n и овладение методами анализа данных в программе Microsoft Excel 2010.

### 2. Материалы и методы

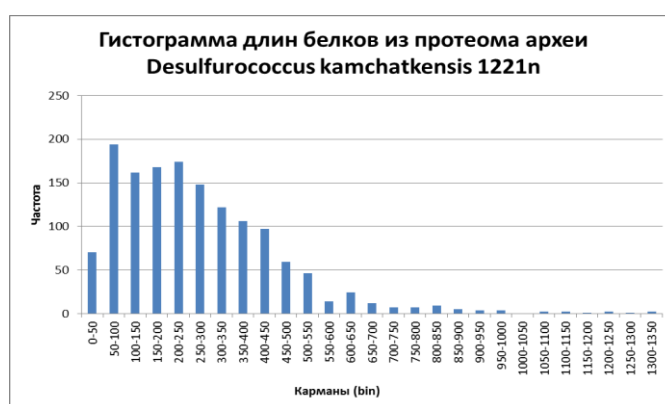
Исходные данные о протеоме *Desulfurococcus kamchatkensis* 1221n были получены на сайте <ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/>. Для обработки данных использована программа для работы с электронными таблицами – Microsoft Excel 2010. Были использованы следующие функции:

- «СЧЁТЕСЛИ» (COUNTIF) – для подсчета количества генов на прямой и обратной цепях ДНК;
- «СЧЁТЕСЛИМН» (COUNTIFS) – для подсчета количества белков, длина которых находится в указанном диапазоне;
- «МИН» (MIN) и «МАКС» (MAX) – для определения минимальной и максимальной длины генов соответственно;
- построение диаграммы и гистограммы по набору значений.

### 3. Результаты

#### 3.1. Распределение длин белков

Распределение длин белков из протеома археи *Desulfurococcus kamchatkensis* 1221n представлено в виде гистограммы (рис.1). В ходе работы было выяснено, что большинство белков имеют длину от 50 до 100 аминокислотных остатков, что составляет около 13% от общего числа белков (294 белка). По рисунку видно, что большинство белков имеет длину от 50 до 450 а.о. Самый длинный белок протеома содержит 1334 а.о., самый короткий – 37 (табл.1).



**Рис.1.** Распределение длин белков в протеоме археи *Desulfurococcus kamchatkensis* 1221n. По горизонтали указаны диапазоны длин белков в аминокислотных остатках (а.о.), по вертикали – количество белков с такой длиной.

| Наибольший и наименьший пептид протеома |  |      |
|---|--|------|
| Максимальная длина белка                |  | 1334 |
| Минимальная длина белка                 |  | 37   |

**Табл.1.** Размер наибольшего и наименьшего пептида

#### 3.2. Распределение генов РНК и белков на прямой и комплементарной цепях ДНК

В ходе работы было установлено, что на прямой цепи расположено 606 генов белков, что составляет примерно 41,2 % от общего числа генов белков. На комплементарной цепи – 865 (58,8%). Генов РНК на прямой цепи 21 (31,3% от всех генов РНК), на комплементарной – 46 (68,7%) (табл.2). Была проверена гипотеза о

том, что гены распределены по цепочкам случайно с вероятностью 0,5. Гены распределены по цепям неравномерно (рис.2 и рис.3).

|                      | Число генов белков | Чило генов РНК |
|----------------------|--------------------|----------------|
| <b>Прямая цепь</b>   | 606                | 21             |
| <b>Обратная цепь</b> | 865                | 46             |
| <b>Всего</b>         | 1471               | 67             |

Табл.2. Соотношение генов белков и РНК на прямой и комплементарной цепочках ДНК.

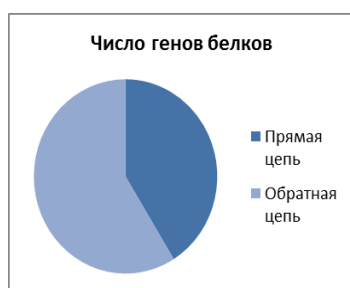


Рис.2. Соотношение генов белков на прямой и обратной цепи



Рис.3. Соотношение генов РНК на прямой и обратной цепи

#### 4. Обсуждение

Около 80% белков имеют длину в диапазоне от 50 до 450 а.о., что является нормой для прокариотических организмов. Гипотеза о том, что гены распределены по цепочкам случайно с вероятностью 0,5, не подтвердилась. Большинство генов расположено на обратной цепи. Возможно, такое распределение не случайно. С другой стороны, это может быть обусловлено тем, что выборка имела небольшой размер.

#### 5. Сопроводительные материалы

[Baikuzina\\_pr15.xlsx](#) – файл с гистограммой и таблицей распределения генов по цепям.

#### Благодарности

Выражаю благодарность Андрею Алексеевскому и Ивану Русинову за обучение основам работы в программе Excel.

#### Литература

1. <ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/>
2. <http://jb.asm.org/content/191/7/2371.full>