

# Краткий обзор генома археи *Thermococcus litoralis* DSM 5473

Иван Старовойтов

Представленная работа содержит в себе результаты исследования протеома и генома археи *Thermococcus litoralis* DSM 5473. В ходе работы были рассмотрены особенности распределения белков по их длинам; закономерности распределения генов, кодирующих белки и РНК, учитывая направленность цепочки ДНК, на которой они располагаются.

## 1 Введение

*Thermococcus* – род экстремофилов в семействе *Thermococcaceae*. Все представители рода являются Археями, имеют термофильные-гипертермофильные черты. Чаще всего встречается коккоидная форма, в диаметре от 0,6 до 2 мкм. Некоторые виды неподвижны, другие же используют жгутик в качестве основного способа передвижения. В основном виды анаэробы, также большинство термофилов являются барофилами, живя на глубине от 200 до 1000 футов. Наиболее благоприятным значениями pH является 5.6 - 7.9. Как оказалось, NaCl не нужен для данных организмов в качестве субстрата, как было показано в исследовании представителей *Thermococcus* в горячих водных системах Новой Зеландии, однако они нуждаются в низкой концентрации ионов лития для собственного роста. Представители рода были найдены во многих системах гидротермальных источников, включая территории от морей Японии до побережья Калифорнии. Вид *Thermococcus litoralis* является органотрофом, не мобилен, проживает при температуре 55-100 °C. Интересен тем, что производит альтернативную ДНК-полимеразу в отличии от классической Taq полимеразы.

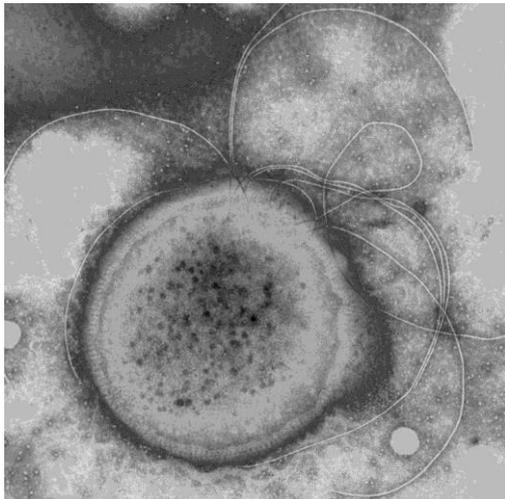


Рис. 1. *Thermococcus* с жгутиками

Геном археи содержит 2570 генов, которые кодируют 2511 белков. Также встречаются последовательности, кодирующие 10 псевдогенов, 4 рРНК, 45 тРНК.

## 2 Методы

Данные о протеоме археи были получены с сайта-NCBI, а для анализа была использована программа Microsoft Office Excel 2013. После скачивания архива GCA\_000246985.3\_ASM24698v3\_feature\_table.txt.gz с протеомом, данные были импортированы и представлены в виде таблицы.

При работе для каждой задачи были составлены отдельные таблицы с нужной информацией, которая с помощью фильтра данных была извлечена из основной таблицы. На протяжении работы для подсчета элементов были использованы функции: «СЧЁТЕСЛИМН()», «СЧЁТЕСЛИ()». Помимо этого использовались функции: «МАКС()», «МИН()», «СРЗНАЧ()», «МЕДИАНА()».

### РЕЗУЛЬТАТЫ

## 3 АНАЛИЗ ПРОТЕОМА АРХЕИ

В геноме *Thermococcus litoralis* DSM 5473 насчитывается 2570 генов, что можно увидеть в таблице 1.

Таблица 1. Разновидности последовательностей в геноме изучаемой археи

|        | Белок | Псевдоген | рРНК | тРНК |
|--------|-------|-----------|------|------|
| Кол-во | 2511  | 10        | 4    | 45   |

Также была построена таблица, отражающая основные статистические свойства протеома изучаемой археи

Таблица 2. Основные статистические свойства протеома *Thermococcus litoralis* DSM 5473.

| MAX  | MIN | AVERAGE  | STAND    | MED |
|------|-----|----------|----------|-----|
| 5745 | 90  | 796,6177 | 553,3165 | 684 |

Распределение длин белков отражено на гистограмме (диаграмма 1).

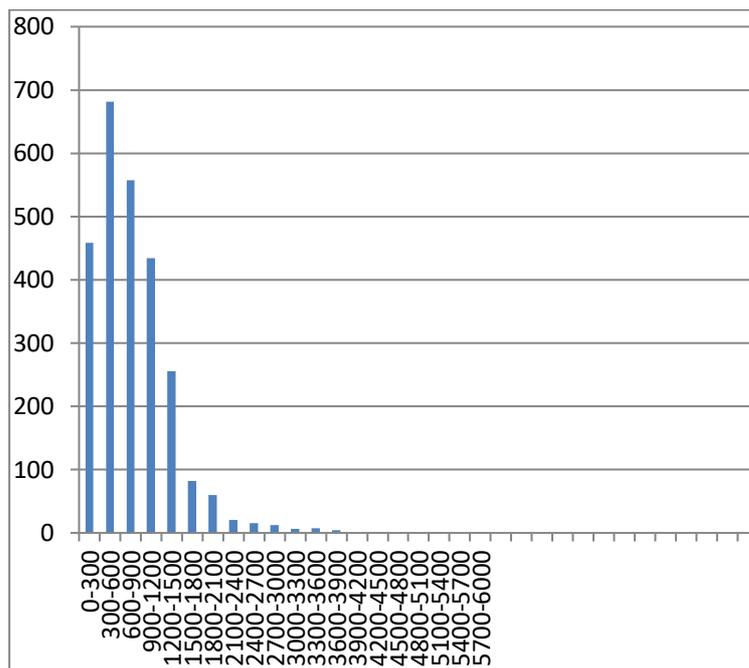


Диаграмма 1. Распределение белков археи по длине

По полученным данным можно заметить, что основное количество белков имеют длину до 1500 аминокислотных остатков. Несложно заметить отсутствие белков с длиной более 6000 аминокислотных остатков.

Так же можно отметить довольно большую разницу между средними длинами и их медианами, что говорит о том, что содержание генов, кодирующих продуктов с маленькой длиной больше.

#### 4 АНАЛИЗ ГЕНОМА АРХЕИ

После анализа расположения генов по разным цепям ДНК была составлена таблица 2.

Таблица 2. Распределение генов бактерии по цепочкам ДНК в различных частях генома.

| Цепь ДНК | Гены Белков | Псевдогены | Гены РНК |
|----------|-------------|------------|----------|
| Прямая   | 1285        | 2          | 22       |
| Обратная | 1225        | 8          | 27       |

#### 5 ВЫВОДЫ

При выполнении этой работы была проведена работа с данными генома *Thermococcus litoralis* DSM 5473 с помощью программы Excel 2013.

Были определены количество и виды последовательностей в геноме, посчитаны некоторые статистические величины для описания некоторых свойств генома, а также наглядно продемонстрировано различие в количестве белок-кодирующих последовательностей разной длины.

МЕСТО РАСПОЛОЖЕНИЯ ДОКУМЕНТА С РАСЧЕТНЫМИ ДАННЫМИ И ВСПОМОГАТЕЛЬНЫМИ ТАБЛИЦАМИ:

<http://kodomo.fbb.msu.ru/~13thlucifer/term1/excel.html>

Благодарности.

Благодарю Косарецкого Егора и Ивана Подъякова, студентов 1 курса ФББ МГУ, за оказанную мне помощь в проведении расчётов. Также следует благодарить мою кошку, заставлявшую отвлекаться меня от монитора хотя бы на несколько минут.

Список литературы:

<https://en.wikipedia.org/wiki/Thermococcus>

[https://en.wikipedia.org/wiki/Thermococcus\\_litoralis](https://en.wikipedia.org/wiki/Thermococcus_litoralis)

