

# Обзор особенностей генома бактерии *Aquifex aeolicus* VF5

Аксенова Ксения Олеговна

Московский Государственный Университет им. М. В. Ломоносова,  
факультет биоинженерии и биоинформатики, Москва, Россия

## Аннотация

*Aquifex aeolicus* VF5 - гипертермофильная хемоавтотрофная бактерия с суперкоротким геномом, всего 1,5 млн нуклеотидов. Обитает она на дне океана, около подводных гейзеров. Для добычи энергии окисляет водород и соединения серы. В работе были рассмотрены GC% состав ее генома и представлено сравнение частоты встречаемости различных старт-кодонов.

## Введение

Таксономическая принадлежность *Aquifex aeolicus* VF5 [1]:

Домен: *Bacteria*

Тип: *Aquificae*,

Отряд: *Aquificales*

Семейство: *Aquificaceae*

Род: *Aquifex*

Вид: *Aquifex aeolicus* VF5.

Род состоит из двух видов, описанных в 1992 году: типового вида *Aquifex pyrophilus* (штамм Kol5a; DSM 6858) и *Aquifex aeolicu*, который ещё не был официально опубликован

*A. aeolicus* — грамотрицательная, палочковидная, подвижная, жгутиковая, не образующая спор бактерия размером примерно  $0,5 \times 2-6$  мкм [1]. Это гипертермофильная бактерия, и, поскольку она оптимально растет при температуре  $85\text{ }^{\circ}\text{C}$  (и до  $95\text{ }^{\circ}\text{C}$ ), она является одной из самых экстремальных гипертермофильных бактерий. Будучи облигатным хемолитоавтотрофом, он не использует для жизнедеятельности свет или органические молекулы, а питается исключительно газами и неорганическими питательными веществами [1, 4]. *A. aeolicus* — это бактерия, живущая в гремучем газе, микроаэрофильный дыхательный микроорганизм, получающий энергию из водорода, но также способный окислять восстановленные неорганические соединения серы (элементарную серу и тиосульфат) [1, 3].

Это один из самых древних и глубоко ветвящихся представителей бактерий, что делает его важным объектом для изучения эволюции жизни. Его геном показывает признаки обширного латерального переноса генов, в том числе от архей и других термофильных бактерий, что усложняет его филогенетическую классификацию [3].

Бактерия *Aquifex aeolicus* является модельным организмом для глубоко укоренённого типа *Aquificae* [1]. Этот «производитель воды» — микроаэрофильный организм, окисляющий  $\text{H}_2$ , процветает в чрезвычайно жарких морских средах. Кроме  $\text{H}_2$ , *A. Aeolicus* окисляет соединения серы, обычно встречающиеся в вулканической среде

[1,2]. *A. aeolicus* обладает сверхстабильными белками [2] и полностью секвенированным геномом [1].

*A. aeolicus* стала первой гипертермофильной бактерией, геном которой был секвенирован [1]. Геном (кольцевая хромосома) довольно мал для свободноживущей бактерии: его размер составляет всего 1 551 335 пар оснований, что примерно в три раза меньше размера генома *Escherichia coli* [2]. Содержание гуанина и цитозина в геноме составляет 43 %, что является относительно низким показателем, хотя для 16S рРНК (65 %) этот показатель очень высок, как и для других гипертермофилов [1]. *A. aeolicus* также содержит плазмиду длиной 39 456 пар оснований, в которой находится 63 гена, большинство из которых кодируют белки с неизвестной функцией. На хромосоме идентифицирован только 1731 ген (47 из них являются псевдогенами), из которых 1632 кодируют белки, а 52 — РНК [1].

### Материалы и Методы

Таблица особенностей генома бактерии *Aquifex aeolicus* VF5 была получена с использованием языка Bash. С сервера NCBI были скачаны файлы [https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/008/625/GCF\\_000008625.1\\_ASM862v1](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/008/625/GCF_000008625.1_ASM862v1) с полной последовательностью генома в формате fasta, последовательностью всех белок-кодирующих участков в формате fasta, таблицей локальных особенностей генома (1, 2).

Для создания гистограммы распределения длин белков был использован Google Sheets. С помощью функции СЧЁТЕСЛИМН в исходной таблице генома было подсчитано количество белков,

длины которых (рассчитанные как длина нуклеотидной последовательности, деленная на 3) попадают в заданные интервалы. На основе полученных данных о количестве белков в каждом интервале и самих интервалов была построена гистограмма (1).

Гистограмма процентного содержания гуанина и цитозина (GC-состава) по кодирующим последовательностям (CDS) также построена в Google Sheets. Для этого столбец таблицы с показателями GC% был скопирован на новый лист, после чего на его основе была создана гистограмма (1).

Для выполнения собственного задания я написала код на Python, он находится в сопроводительных материалах (4), а также воспользовалась Google Sheets, для создания диаграммы, отображающей отношение количеств белков с разными старт-кодонами (3).

## **Результаты**

### **1. Распределение длин белков *Aquifex aeolicus* VF5.**

На основе CDS бактерии были найдены длины (в аминокислотах) всех продуктов трансляции у данного вида. Затем построена гистограмма, которая отражает соответствие количества белков каждому интервалу возможной длины.

## Гистограмма длин белков

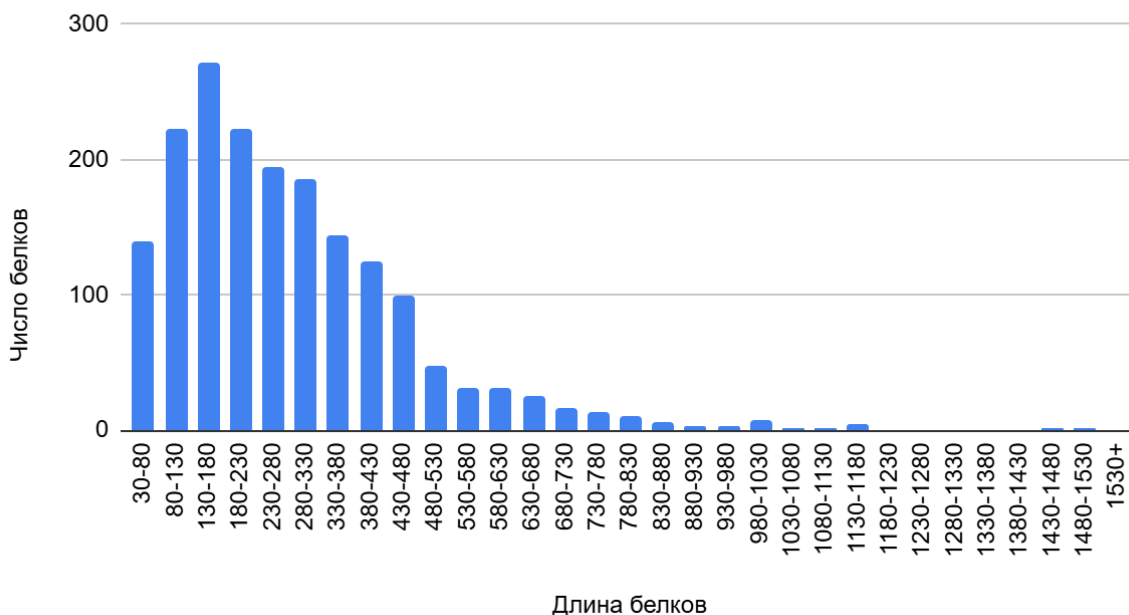


Рис. 1. Гистограмма длин белков в АМК.

Исходя из данных рисунка 1 видно, самый высокий столбец имеет длину от 130 до 180 АМК. После 180 АМК можно наблюдать плавный спад до 480, потом резкий в диапазоне 480-530. белков длиной больше, чем 830 АМК, уже практически нет. Минимальная длина белка - 32 АМК, максимальная - 1574 АМК

## 2. GC состав генов

Исходя из данных CDS была построена гистограмма процентного содержания GC в генах бактерии. Она помогает наглядно увидеть распределение GC в генах и определить его средний показатель

## Гистограмма процентного содержания GC

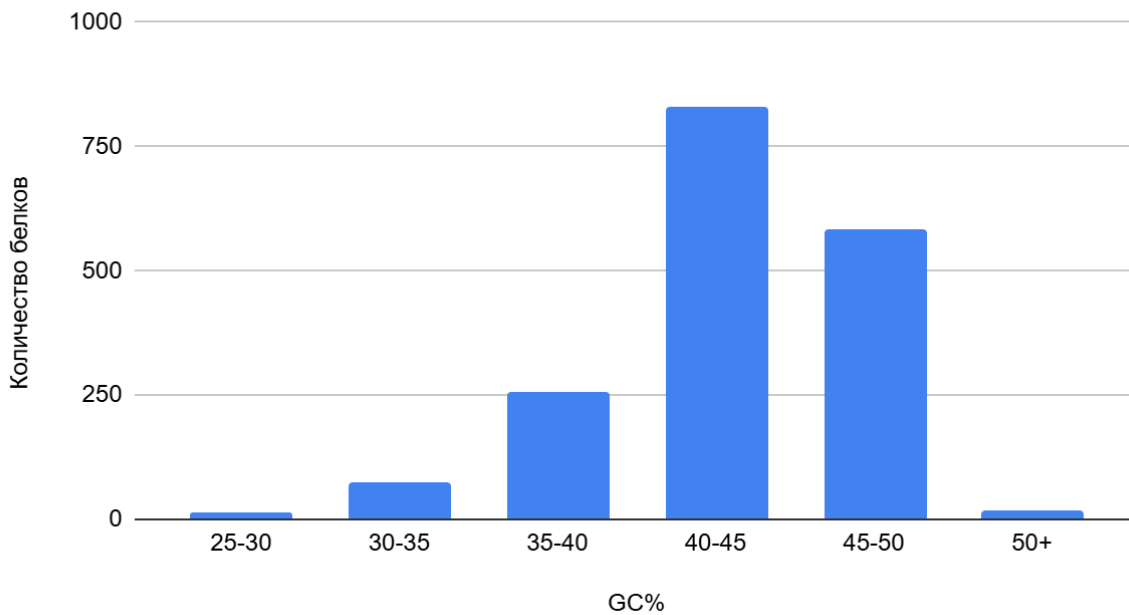


Рис 2. гистограмма процентного содержания GC.

Из данных рисунка 2 видно, что у бактерии преобладают гены с содержанием GC 40-45%. Это необычно, поскольку данная бактерия - гипертермофил. У таких организмов содержание GC, как правило, значительно выше, поскольку связи GC обладают повышенной устойчивостью. Однако данные результаты подтверждают данные о среднем уровне GC в районе 43% [1].

### 3. Распределение белков с различными старт-кодонами

С помощью гистограммы визуализируем количество белков, начинающихся с различных старт-кодонов.

### Количество белков, кодируемых определенным старт-кодоном

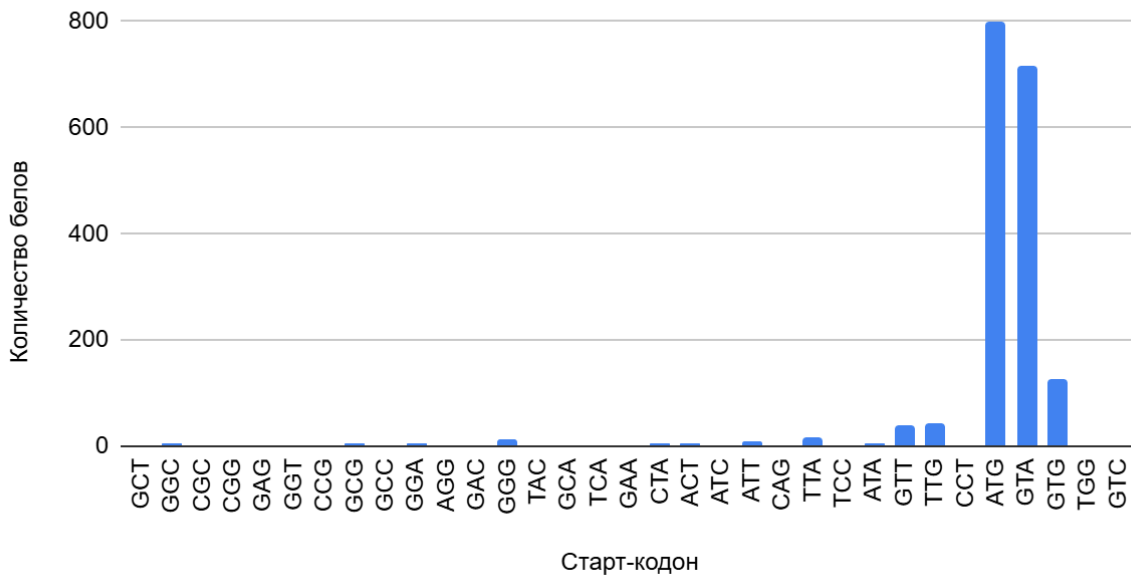


Рис. 3. Количество белков, кодируемых определенным старт-кодоном.

Из данных рисунка 3 видно, что большая часть белков начинаются со старт-кодонов ATG и GTA, немного меньше с GTG. Также заметно, что белков, начинающихся с других старт-кодонов практически нет. эти данные подтверждаются диаграммой (рис 4).

## Частота старт-кодонов

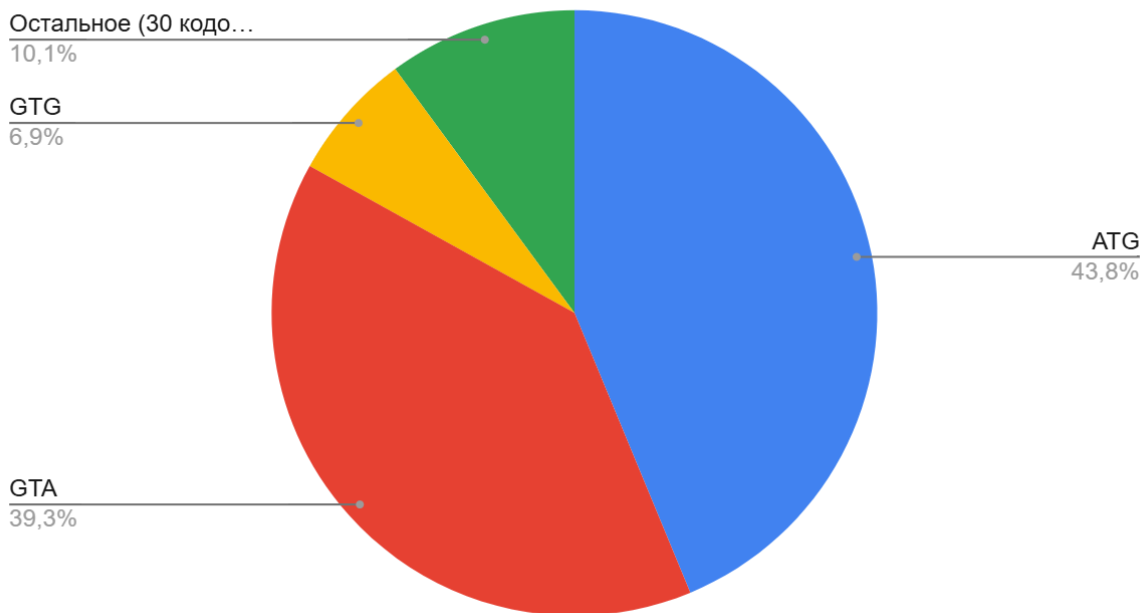


рис.4 Диаграмма частоты встречаемости старт-кодонов

## 4. Средняя длина белка в зависимости от старт-кодона

### Длина белка в зависимости от старт-кодона

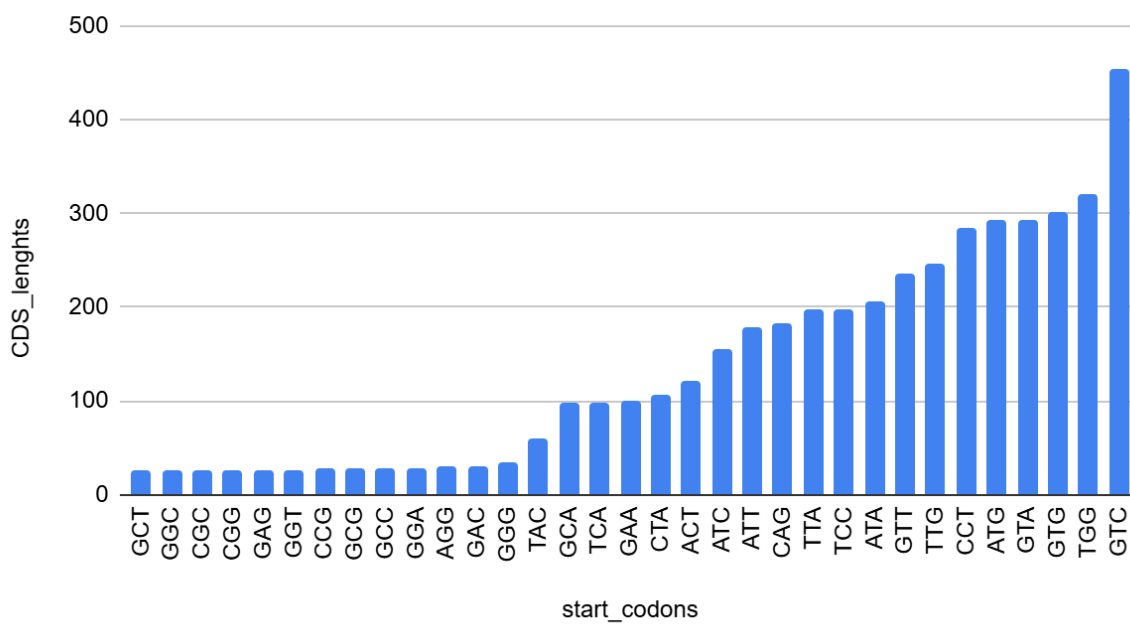






Рис 5. Столбчатая диаграмма зависимости длины кодирующей последовательности от старт-кодона.

Исходя из данных гистограммы мы видим, что самые длинные последовательности у белков, начинающихся с GTC, старт-кодона GCT, GGC, CGC, CGG, GGT, CCG, GCG, GCC, AGG, GAC, GGG дают начало действительно коротким белкам, длиной около 25 АМК. Средняя длина белков самого распространенного кодона, АТG, около 300 АМК

### Список литературы

- [1]. Guiral M., Giudici-Orticoni M. T. Microbe Profile: Aquifex aeolicus: an extreme heat-loving bacterium that feeds on gases and inorganic chemicals //Microbiology. – 2021. – Т. 167. – №. 1. – С. 001010.
- [2].Guiral M. et al. The hyperthermophilic bacterium Aquifex aeolicus: from respiratory pathways to extremely resistant enzymes and biotechnological applications //Advances in Microbial Physiology. – 2012. – Т. 61. – С. 125-194.
- [3].Eveleigh R. J. M. et al. Being Aquifex aeolicus: untangling a hyperthermophile’s checkered past //Genome biology and evolution. – 2013. – Т. 5. – №. 12. – С. 2478-2497.
- [4]. Prioretti L. et al. Carbon fixation in the chemolithoautotrophic bacterium Aquifex aeolicus involves two low-potential ferredoxins as partners of the PFOR and OGOR enzymes //Life. – 2023. – Т. 13. – №. 3. – С. 627.

### Сопроводительные материалы.

1.  CDS from genome of Aquifex aeolicus VF5
2.  Genomic features of Aquifex aeolicus VF5
3.  основа графиков
4. Папка с кодом и файлами генома бактерии:  Архив