

---

# Обзор протеома бактерии *Bacillus cereus* штамм А1

Азбукина Надежда<sup>1\*</sup>

<sup>1</sup>Факультет биоинженерии и биоинформатики Московский Государственный Университет им. Ломоносова

---

## 1 РЕЗЮМЕ

Данная работа посвящена исследованию протеома бактерии *Bacillus cereus* штамм А1 для выявления числа генов и белков, а также особенностей их распределения с помощью программы Excel. Результаты показали, что гены с вероятностью 83,26% распределены не случайно по прямой и обратной цепи. Также был найден интересный квазиоперон, содержащий 28 генов.

## 2 ВСТУПЛЕНИЕ

*Bacillus cereus* - грамположительные, спорообразующие с перитрихально расположенными жгутиками подвижные клетки. Встречаются повсеместно, могут вызывать у людей желудочные заболевания (диарею и др.), а также септицемию, эндокардит, поражения центральной нервной системы. Как правило, заболевания прекращаются без лечения, но зарегистрированы также и единичные смертельные случаи [1]. По типу питания *Bacillus cereus* является хемоорганогетеротрофом, по типу дыхания - факультативным анаэробом, способны к нитратредукции. Клетки крупные 3—4 мкм, эндоспores расположены центрально, не превышают размер клетки [2]. Штамм А1 был выделен из осадка в анаэробном ферментере, секвенирован 29 мая 2014 года. Размер генома составляет 5352307 п. н., количество генов - 5,648. Данный штамм гидролизует крахмал путем выделения амилазы в среду, а также сбраживает глюкозу и другие углеводороды до молекулярного водорода, ацетата и этанола, что делает его привлекательным объектом для исследований в области альтернативных источников энергии. Также авторы [3] предполагают наличие у данного штамма транскетолазного или трансальдолазного метаболического пути вместо фосфокетолазного, что вероятно позволяет этой бактерии метаболизировать пентозы более эффективно. Найдены гены, отвечающие за продукцию органических кислот (муравьиной, уксусной, молочной и масляной). Дальнейшее изучение метаболических путей штамма А1 может стать ключом к разгадке механизма продукции им молекулярного водорода и позволить увеличить выход этого продукта. Данная работа посвящена изучению особенностей распределения генов и белков в геноме бактерии, а также анализу имеющихся данных на предмет пересечения генов и составлению статистики белков по категориям достоверности их существования

## 3 МЕТОДЫ

---

При составлении данного была использована программа для работы с электронными таблицами Microsoft Excel 15.0 (Office 2016). В работе использовались функции ЧАСТОТА, ОКРУГЛЕНИЕ, графическое представление данных для построения гистограмм распределения длин белков и квазиоперонов, СЧЁТЕСЛИ для подсчета распределения числа различных видов генов по цепям и элементам генома.

Геном бактерии был скачан с базы данных Genome сайта NCBI (идентификатор CP015727.1). Данные о достоверности существования белков были получены с помощью базы сайта Uniprot.

## 4 РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЯ

### 4.1 Общая информация о распределениях числа генов по категориям

Геном данного штамма содержит 4 элемента: хромосома и 3 плазмиды (pBCA1, pBCA2, pBCA3). Распределение генов по этим элементам представлено в таблице 1.

Данный штамм содержит 5418 генов, кодирующих белки, 106 генов тРНК, 42 гена рРНК, 4 гена некодирующей РНК, 1 ген транспортно-матричной РНК и 77 псевдогенов. В дальнейшем в работе под словом “гены” понимаются все гены, за исключением псевдогенов. Примерное количество генов на 1 млн пар нуклеотидов составляет 1041.

### 4.2 Распределение длин белков

На рисунке 1а представлена общая гистограмма распределения длин белков, а на рисунке 1б – её масштабированный участок для диапазона длин 26-1016 аминокислотных остатков. Значение медианы для данного ряда составляет 246. Подавляющее число белков имеет длину 60-500 аминокислотных остатков. Самым большим белком является белок клеточной поверхности (ANE84937.1 cell surface protein) длиной в 5010 аминокислотных остатков, а самыми маленькими K<sup>+</sup> транспортирующая субъединица АТФазы и белок М стадии споруляции V (ANE86985.1 stage V sporulation protein M) оба длиной в 26 аминокислотных остатков (ANE84136.1 K<sup>+</sup>-transporting ATPase subunit F).

---

**Таблица 1.** Распределение генов по элементам генома

---

Элемент генома	Число генов
Хромосома	5332
pBCA1	226
pBCA2	78
pBCA3	12

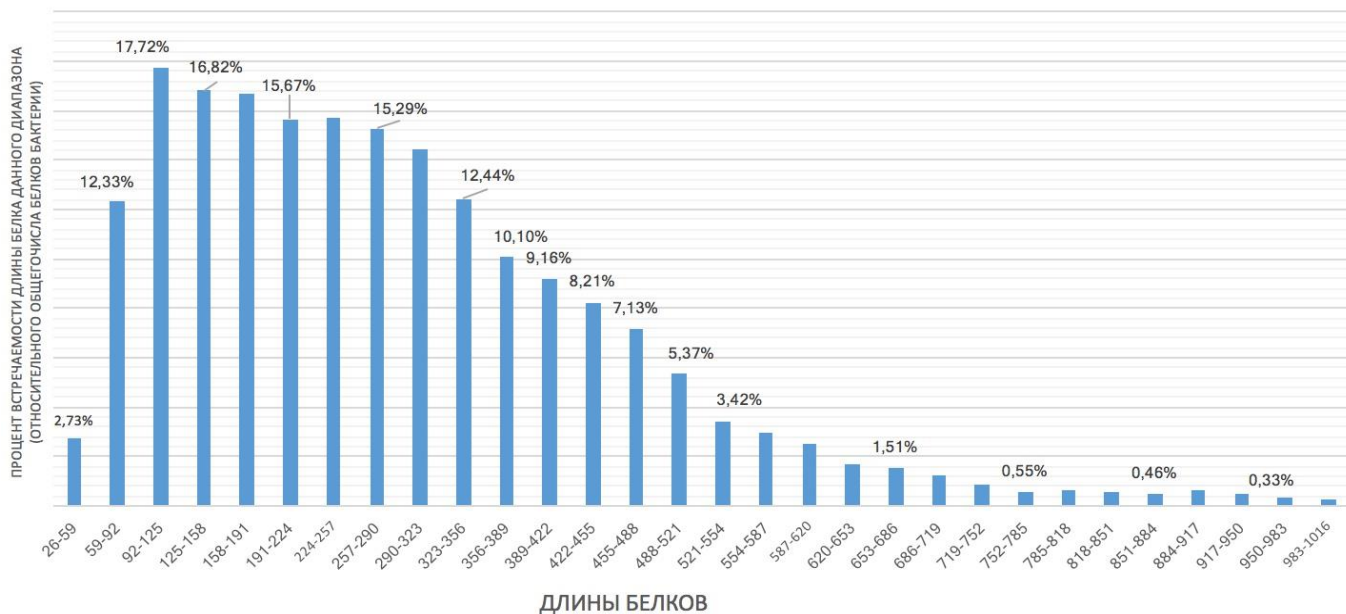
---



**Рисунок 1а.** Гистограмма распределения длин белков *Bacillus cereus* (штамм А1)

**Рисунок 1б.** Участок гистограммы распределения длин белков (26-1016 аминокислотных остатков) *Bacillus cereus* (штамм А1)

### Участок гистограммы распределения длин белков (26-1016 аминокислотных остатков) у *Bacillus Cereus* (штамм А1)



### 4.3 Распределение генов на прямой и обратной цепи

Распределение генов на прямой и обратной цепи представлено в таблице 2.

Дальше я проверила гипотезу о случайности распределения генов по цепочкам. Для этого была использована встроенная функция Excel БИНОМРАСП. В результате оказалось, что вероятность получить такое же отклонение или больше (если гены распределяются по цепям с вероятностью 0,5) равна 16,74%, то есть с вероятностью 83,26% гены распределены по прямой и обратной цепям не случайно.

### 4.4 Анализ числа и распределения квазиоперонов

Квазиоперон- это участок ДНК, содержащий гены, предположительно транслирующиеся с одной мРНК (так как расстояние между ними невелико). Квазиоперон может состоять и из одного гена. Если принять, что максимальное расстояние (шаг) между генами в квазиопероне составляет 50

пар нуклеотидов, то обнаружилось 4009 квазиоперонов, при шаге в 100 п.н. – 3440, при шаге в 200 п.н. -2553. При увеличении длины шага количество квазиоперонов уменьшается, что довольно логично, ведь большее количество генов оказываются сцепленными. На рисунке 2 представлена гистограмма распределения квазиоперонов по числу генов для шага в 50 нуклеотидов. Очевидно, что их подавляющее число состоит из 1 гена (2327 штук), 880 состоит из 2-3 генов. Для длины в 4-12 генов есть небольшой разброс, а опероны, состоящими из большего числа генов достаточно редки и представлены единичными случаями. Самым большим оказался квазиоперон из 28 генов, состоящие из различных белков 30S и 50S субъединиц рибосом, а также белка, переносащего другие белки через мембрану [4], аденилат киназы, метионин аминопептидазы I, фактора инициации транскрипции IF-1. Можно предположить, что данный оперон служит для синтеза рибосомы. Подробную информацию нём смотри в приложении (Лист «Самый большой оперон»).

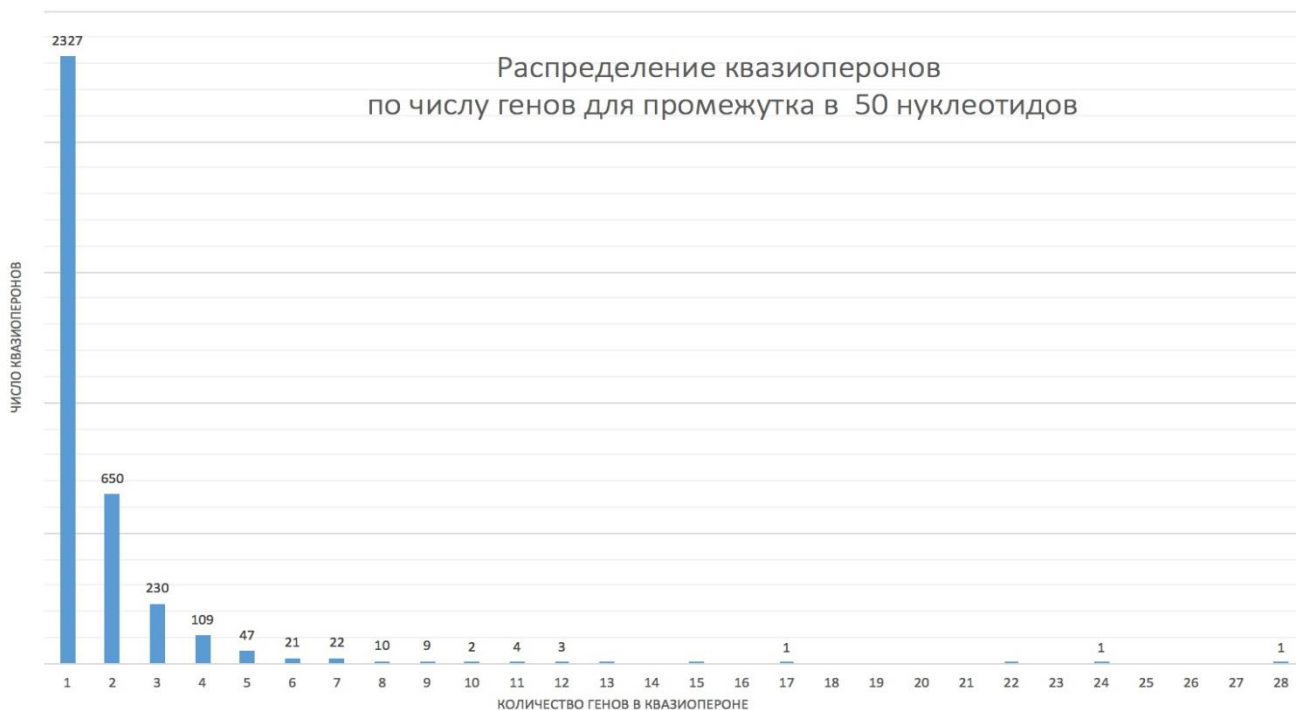


Рисунок 2. Распределение квазиоперонов по числу генов для шага в 50 нуклеотидов

**Таблица 2.** Число генов белков и РНК на прямой и комплементарной цепи ДНК

Категория генов	Прямая цепь	Обратная цепь	Сумма
Число генов РНК	118	35	153
Число генов белков	2631	2787	5418
Сумма	2749	2822	5571

#### 4.5 Пересечения генов

Для подсчета количества перекрывающихся генов была использована встроенная функция СЧЁТЕСЛИМН. Результаты представлены в таблице 3. Как видно из таблицы, больше всего перекрываний образуется без сдвига рамки считывания и на прямой цепи. Перекрывания со сдвигом рамки считывания на 2 нуклеотида очень мало – всего 5 из 469. Также при выполнении этой задачи была обнаружена следующая ошибка в базе данных : при секвенировании хромосомы и плазмиды рVCA1 начало отсчета участка генома попало в середину гена, поэтому его номер нуклеотида конца гена получился очень большой. Например 1 ген хромосомы, кодирующий фруктозо-6-фосфат альдозу, начинается с 1 нуклеотида и кончается 5352307. Это, конечно же, не так. Удостовериться в этом мы можем, посмотрев на соответствующую колонку длины продукта, имеющую значение 215 аминокислотных остатков, или на колонку, показывающую длину интервала (648 п. н.).

#### 4.6 Кодирующие последовательности, длина которых не кратна 3

С помощью функций ОСТАТ и СЧЁТЕСЛИМН были найдены последовательности, длина которых не кратна 3. Таких генов не обнаружилось, это значит, что в них нет сдвига рамки считывания.

#### 4.7 Статистика белков по категориям достоверности их существования

К сожалению, в базе Uniprot не оказалось данных относительно белков данного штамма, поэтому анализ достоверности существования белков я сделала по данным для общего вида *Bacillus cereus*. Из 5418 белков штамма A1 2218 совпали с таковыми у *Bacillus cereus*.

Из них существование 11 доказано на уровне продуктов трансляции, 7 на уровне транскриптов, 689 по данным гомологии (у родственных видов существуют гены-ортологи) и о существовании 1511 нет этих доказательств (predicted). Подробнее о категориях достоверности существования белков можно узнать на сайте Uniprot [5].

## 5 СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

[http://kodomofbb.msu.ru/~azbukinanadezda/term1/Azbukina\\_supplementary.xlsx](http://kodomofbb.msu.ru/~azbukinanadezda/term1/Azbukina_supplementary.xlsx)

## 6 БЛАГОДАРНОСТИ

Автор выражает благодарность доценту Факультета биоинженерии и биоинформатики МГУ им. Ломоносова Алексеевскому А. В. за помощь в поиске данных в базе Uniprot.

## 7 СПИСОК ЛИТЕРАТУРЫ

- [1] <http://mibio.ru/contents.php?id=93>
- [2] [https://ru.m.wikipedia.org/wiki/Bacillus\\_cereus#](https://ru.m.wikipedia.org/wiki/Bacillus_cereus#)
- [3] Zhang T. и др. Genome Sequence of *Bacillus cereus* Strain A1, an Efficient Starch-Utilizing Producer of Hydrogen // *Genome Announc.* 2014. Т. 2. № 3. e00494-14.
- [4] <https://www.wikigenes.org/e/gene/e/3991223.html>
- [5] [http://www.uniprot.org/help/protein\\_existence](http://www.uniprot.org/help/protein_existence)

**Таблица 3.** Перекрывание генов на прямой и обратной цепи

Вид цепочки	Со сдвигом рамки считывания на 1 нуклеотид	Со сдвигом рамки считывания на 2 нуклеотида	Без сдвига рамки считывания	Сумма
На "+" цепочке	107	1	130	238
На "-" цепочке	88	0	129	217
цепочка с "+"	6	4	4	14
цепочкой				
Общее число	201	5	263	469