

# Краткий обзор протеома бактерии *Synechococcus elongatus* PCC 7942

Ульяна Пиунова\*

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

## РЕЗЮМЕ

Данная работа посвящена анализу протеома и генома бактерии *Synechococcus elongatus* PCC 7942, направленному на исследование характера распределения белков по длинам, распределения генов по прямой и обратной цепям ДНК, их пересечения и объединения в квазиопероны.

## 1 ВВЕДЕНИЕ

*Synechococcus elongatus* — это одноклеточные, палочковидные бактерии (рис. 1, фото L.A. Sherman, D.M. Sherman, Purdue University) Могут существовать отдельно, объединяться в пары, соединяться линейно или в небольшие кластеры. Данные цианобактерии способны плавать или скользить в воде, несмотря на отсутствие жгутиков или ресничек. Движение не зависит от световой стимуляции.

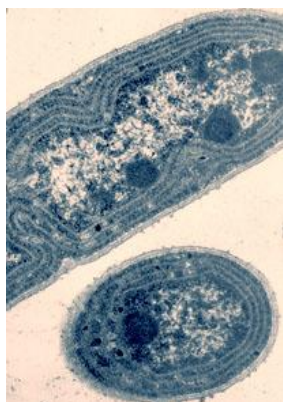


Рис. 1. Микрофотография [1]

## 2 МЕТОДЫ

Данные о протеоме были взяты с сервера NCBI ([ftp://ftp.ncbi.nlm.nih.gov/genomes/archive/old\\_refseq/Bacteria/](ftp://ftp.ncbi.nlm.nih.gov/genomes/archive/old_refseq/Bacteria/)). Использованы файлы NC\_007595.ptt, NC\_007604.ptt и NC\_007604.rnt [2].

Все результаты были получены при помощи программы Microsoft Office Excel 2013, путем использования различных встроенных функций. При помощи СЧЁТЕСЛИМН была составлена гистограмма длин белков. Расчет числа генов на + (прямой) и - (обратной) цепях ДНК производился с помощью функции СЧЁТЕСЛИ. Для определения количества квазиоперонов и пересечений генов использовались логические функции. Для статистического анализа данных использованы функции СРЗНАЧ и МЕДИАНА.

## 3 РЕЗУЛЬТАТЫ

Геном *Synechococcus elongatus* PCC 7942 представлен одной кольцевой хромосомой и двумя плазмидами.

Круговая хромосома длиной 2695903 пар нуклеотидов (содержание GC равно 55,5%).

Хромосома бактерии содержит гены, кодирующие 2612 белков и 53 РНК. Одна из плазмид бактерии в настоящее время секвенируется в лаборатории доктора Голдена в Техасском университете A&M (Texas A&M University) в сотрудничестве с Объединенным институтом геномов (DOE Joint Genome Institute) [3]. Другая несет информацию о 50 белках.

### 3.1 Распределение белков по длинам

Изучение протеома показало, что самый длинный белок, информация о котором находится в хромосоме *Synechococcus elongatus* PCC 7942, имеет длину 1807 аминокислот, самый короткий – 29. В плазмиде – 999 и 41 соответственно. Большинство белков, закодированных в хромосоме бактерии, имеют длину от 101 до 400 аминокислот. В плазмиде – от 101 до 200.

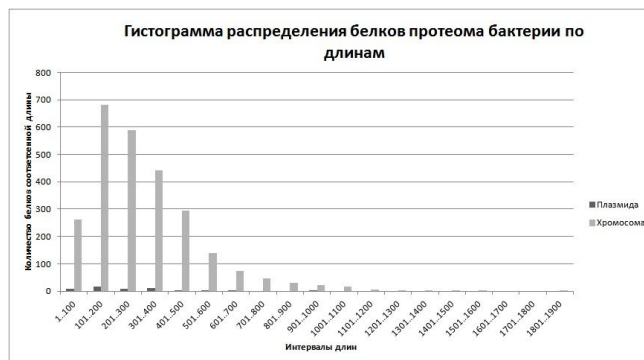


Рис. 2. Распределение белков по длинам

### 3.2 Распределение белков по цепям ДНК

Статистика распределения генов по цепям представлена в таблице 1. На прямой цепи ДНК в хромосоме бактерии – 1277 генов белков, на обратной – 1335. На прямой цепи – 32 гена РНК и 21 – на обратной.

В плазмиде 28 генов белков находится на прямой цепи и 22 – на обратной.

	Хромосома		Плазмида
	CDS	RNA	CDS
+	1277	32	28
-	1335	21	22

Таблица 1. Распределение генов по цепям

\*ulyana@fbb.msu.ru

### 3.3 Статистика квазиоперонов

Квазиоперон – это группа генов (но могут состоять и из одного гена), расположенных на одной цепи, таких, что расстояние между ними – не больше какой-то заданной "пороговой" величины.

Было проведено исследование квазиоперонов в геноме *Synechococcus elongatus* PCC 7942. Так, из анализа плазмиды видно, что при заданном расстоянии в 100 пар нуклеотидов число квазиоперонов составляет 66% от общего числа генов, при 200 п. н. – 56%, при 50 п. н. – 74%.

В хромосоме при заданном расстоянии 100 п. н. число квазиоперонов генов белков составляет 56,34%, при 200 п. н. – 44,87%, при 50 п. н. – 69,45%. При длине 100 п. н. – 90,57% квазиоперонов генов РНК, при 200 п. н. – 86,79%, при 50 п. н. – 94,34%. Из таблицы 2 видно, что количество квазиоперонов на прямой и обратной цепях примерно одинаково при одном заданном расстоянии, кроме генов РНК (в хромосоме), где в среднем прямая цепь содержит на 20% квазиоперон больше, чем обратная.

Количество квазиоперонов в зависимости от порогового расстояния						
	Плаزمида		Хромосома			
	CDS		CDS		RNA	
	+	-	+	-	+	-
100	18	15	737	734	29	19
200	14	14	595	576	28	18
50	18	19	904	910	30	20

**Таблица 2.** Количество квазиоперонов на прямой и обратной цепях в зависимости от порогового расстояния

### 3.4 Пересечения генов

После проведения исследований, в хромосоме из числа генов белков пересекающимися оказались 206, что составляет 8% от общего числа генов. Средняя длина пересечений – 7,13 п. н., медиана длин – 2. В данном случае медиана совпадает с наиболее распространенной длиной: количество таких пересечений составляет 61,65% от их общего числа в хромосоме. Среди генов РНК пересечений нет.

В плазмиде бактерии нашлось 8 пересекающихся генов, что составляет 8% от общего числа генов. Средняя длина пересечений составляет 18,25 пар нуклеотидов, медиана – 16 п. н. Наиболее характерна длина пересечения, равная 2 п. н. (50%).

Всего в геноме 7,88% генов от общего числа являются пересекающимися.

## ОБСУЖДЕНИЕ

Касательно распределения генов по прямой и обратной цепям ДНК можно заметить, что на обеих цепях их примерно поровну: в хромосоме 1309 и 1356 соответственно, в плазмиде – 28 и 22.

При анализе данных было обнаружено, что с увеличением "порогового" расстояния количество квазиоперонов уменьшается, с уменьшением, наоборот, увеличивается. Такая закономерность наблюдается, поскольку, увеличивая заданное расстояние, мы получаем меньшее количество квазиоперонов, состоящих из одного гена, и наоборот. Исключение составляет число квазиоперонов на прямой цепи плазмиды при заданных расстояниях 100 и 50 п. н. (число квазиоперонов одинаково). Это объясняется тем, что не оказалось расстояний в интервале от 50 до 100 п. н. между соседними генами, то есть гены, составившие квазиопероны, одни и те же в обоих случаях. По результатам исследования пересекающихся генов можно сделать предположение, что для генома *Synechococcus elongatus* PCC 7942 характерны пересечения генов длиной 2 пары нуклеотидов – 61,21% от числа всех пересечений.

## 4 СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

Ссылка на xlsx-файл с вычислениями и результатами:

[http://kodom.fbb.msu.ru/~ulyana/term1/Piunova\\_supplementary.xlsx](http://kodom.fbb.msu.ru/~ulyana/term1/Piunova_supplementary.xlsx)

### ССЫЛКИ

1. <http://genome.jgi.doe.gov/synel/synel.home.html>
2. [ftp://ftp.ncbi.nlm.nih.gov/genomes/archive/old\\_refseq/Bacteria/Synechococcus\\_elongatus\\_PCC\\_7942\\_uid58045/](ftp://ftp.ncbi.nlm.nih.gov/genomes/archive/old_refseq/Bacteria/Synechococcus_elongatus_PCC_7942_uid58045/)
3. [https://microbewiki.kenyon.edu/index.php/Synechococcus\\_elongatus/](https://microbewiki.kenyon.edu/index.php/Synechococcus_elongatus/)