

Обзор генома бактерии *Brevundimonas naejangsanensis* Strain B1

Бушмакин И. М.

Факультет Биоинженерии и Биоинформатики, МГУ
ilbumi@kodomo.fbb.msu.ru

Brevundimonas naejangsanensis Strain B1 – недавно выделенный, факультативно анаэробный штамм бактерий, который производит водород с крайне высокой эффективностью. Изучение этой бактерии, может привести ясность в механизмы переработки биомассы в водород.

Ключевые слова: *Brevundimonas naejangsanensis*; геном; протеом.

1. Введение

Вид *B. naejangsanensis* был описан в 2009. Это грамотрицательная бактерия, палочка, длиной до 4 мкм. Передвигаются посредством единственного полярного жгутика. Формирует гладкие, блестящие, серо-желтые колонии диаметром 1-2 мм после двухдневной инкубации при 30 °C¹. Последовательность генома штамма B1 была установлена в 2014². Авторы исследования опубликовали аннотацию генома, который содержит 2789 кодирующих последовательностей, из них 51 кодируют РНК. Также они определили 351 подсистему в геноме и вычислили долю G+C пар (67.5%). С момента выхода статьи данные о геноме были уточнены, добавлена информация о кодирующих последовательностях. Хотя существует много статей, описывающих использование этой бактерии в переработке растительного сырья³, они не касаются ее генома.

2. Материалы и методы

Таблица кодирующих последовательностей (сборка GCA_000635915.2) был взята с FTP сервера NCBI^a. Данные обрабатывались в Microsoft Excel 2016 и LibreOffice Calc 5.4.2.2, гистограмма в статье построена пакетом pgfplots в ЛАТ_EX. Для расчета вероятности по биномиальному распределению был написан скрипт на Python 2.7^b. Классификация белков по доказательствам их существования проводилось по данным Uniprot.

^aftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/635/915/GCA_000635915.2_ASM63591v2/GCA_000635915.2_ASM63591v2_feature_table.txt.gz

^b<https://github.com/SnicsIB/phenal/blob/master/bincalc.py>

3. Результаты

3.1. Общая информация

Геном *B. naejangsanensis* Strain B1 содержит 2859 генов (2772 кодирующих белок последовательностей, 28 псевдогенов и 59 РНК). Распределение продуктов по классам представлено в Табл. 1. Длина генома $2,97 * 10^6$ п.н., средняя плотность генов - 962,13 генов на 10^6 п.н.

Таблица 1. Классы продуктов и их количество.

Класс	Кол-во
Предсказанные гипотетические белки	533
Транспортные белки	51
Рибосомальные белки	55
Другие белки	2133
rRNA	6
tRNA	49
ncRNA	3
tmRNA	1

3.2. Белки

Как видно из Рис. 1, преобладают белки длиной 100-300 аминокислотных остатков. Самым длинным белком является НАД-глутамат дегидрогеназа, играющий важную роль в метаболизме азотсодержащих соединений. Второй по длине – белок, содержащий домен DUF490. Среди самых длинных белков также синтаза энтеробактина, субъединицы РНК и ДНК полимераз и несколько гипотетических предсказанных белков. Статистическая информация о белках приведена в Табл. 2.

Таблица 2. Статистика белков.

	Значение
Минимальная длина	39
Максимальная длина	1620
Средняя длина	316,275974
σ	209,7136351
Медиана длины	266

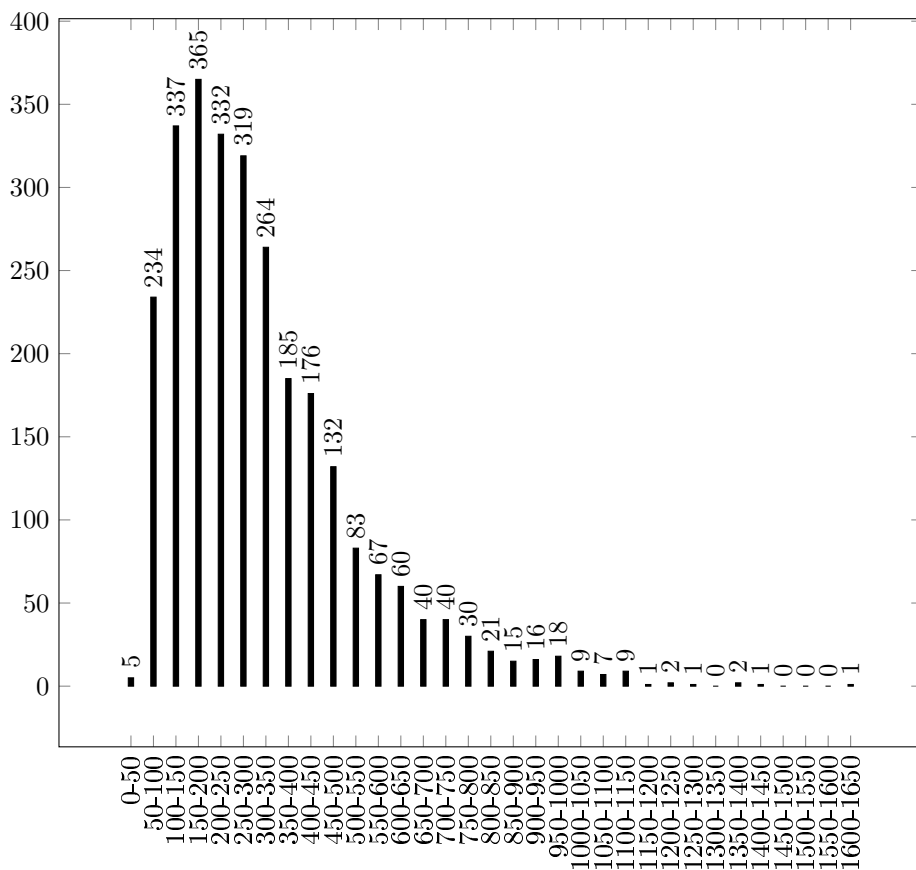


Рис. 1. Распределение длин белков

3.3. Гены

3.3.1. Распределение по цепям

Распределение генов, представленное в Табл. 3, по цепям соответствует случайному. Если гипотеза о случайности цепи верна, вероятность получить такое же или менее равномерное распределение равна 60,81% в случае белков, 29,76% для РНК, 34,49% для псевдогенов. Для всех кодирующих последовательностей эта вероятность составляет 82,24%.

3.3.2. Перекрывание генов

Из 2772 генов 714 пересекаются. Существуют пересечения как генов на одной цепи, так и на противоположных.

Таблица 3. Распределение генов по цепям.

Цепь	Белки	РНК	Псевдогены
+	1400	25	11
-	1372	34	17

3.4. Существование белков

В базе данных Uniprot на момент написания статьи существуют записи о 2708 белках из 2772. Бактерия еще не изучена достаточно. Как видно из Табл. 4, большинство белков имеют статус предсказанных или предполагаемых по гомологии. Всего для двух белков представлены доказательства на уровне транскрипции. Это АТФ-зависимая протеаза La и шаперон Hsp70 (белок теплового шока).

Таблица 4. Категории существования

Категория	Кол-во
Predicted	1666
Inferred from homolog	1040
Evidence at transcript level	2
Total	2708

4. Сопроводительные материалы

Данные, использовавшиеся в работе, приведены в таблице формата ODT (LibreOffice Calc): <http://www.kodomo.fbb.msu.ru/~ilbumi/src/pr13.xlsx>

Литература

1. Kang SJ, Choi NS, Choi JH, Lee JS, Yoon JH, Song JJ, Brevundimonas naejangsanensis sp. nov., a proteolytic bacterium isolated from soil, and reclassification of mycoplasma bullata into the genus brevundimonas as brevundimonas bullata comb. nov., *International journal of systematic and evolutionary microbiology* **59**(12):3155–3160, 2009.
2. Su H, Zhang T, Bao M, Jiang Y, Wang Y, Tan T, Genome sequence of a promising hydrogen-producing facultative anaerobic bacterium, brevundimonas naejangsanensis strain b1, *Genome announcements* **2**(3):e00542–14, 2014.
3. Wang S, Zhang T, Su H, Enhanced hydrogen production from corn starch wastewater as nitrogen source by mixed cultures, *Renewable Energy* **96**:1135–1141, 2016.