

ОБЗОР ПРОТЕОМА БАКТЕРИИ *DICKEYA DADANTII* 3937

АВТОР

Зайцева Елизавета, Факультет Биоинженерии и Биоинформатики МГУ им. М. В. Ломоносова

В рамках этой работы было проведено общее исследование генома бактерии *Dickeya dadantii* 3937. Определено распределение генов белков и РНК по категориям. Построена гистограмма длин белков из протеома исследуемой бактерии. Определена закономерность распределения генов белков и РНК на прямой и обратной цепях ДНК. Посчитано число «квазиоперонов» в геноме бактерии.

Ключевые слова: *Dickeya dadantii*; геном; бактерия.

1. Введение

Систематическое положение: царство *Bacteria*, нетаксономическая группа *Gracilicutes*. *Dickeya dadantii* - факультативный анаэроб, паразиты культурных растений. Так, часто становятся пораженными клубни, корни, мясистые листья, проводящие элементы стеблей картофеля, сахарного тростника, сорго, банана, злаков и декоративных растений. Также в литературе встречается видовое название *Erwinia chrysanthemi*. [1]

В данной работе приведены результаты общего исследования бактерии *Dickeya dadantii* наиболее распространенного штамма 3937.

2. Материалы и методы

2.1. Используемые ресурсы и программы

Использованы данные с сайта NCBI: GCA_000147055.1_ASM14705v1_feature_table.txt.gz [2] Данные обрабатывались в Microsoft Office Excel: построение таблиц, диаграммы, использование функций логического сравнения ЕСЛИ, счета при множественном условии СЧЕТЕСЛИМН, статистических функций СРЗНАЧ, МИН, МАКС, СТАНДОТКЛОНА, МЕДИАНА. Также использовался фильтр для выбора CDS, gene, RNA. Для проверки гипотезы о случайном распределении генов по цепям была написана программа на Python 3.6.

2.2. Используемые понятия

Гены, про которых в столбце “class” было написано “protein_coding”, “pseudogene” или какая-либо “RNA”, считались соответственно кодирующими белок,

некодирующими или кодирующими какой-либо тип РНК. “Квазиоперонами” считались гены на одной цепи ДНК, расстояние между которыми не превышает 100 пар нуклеотидов.

3. Результаты и обсуждение

Штамм содержит 4692 гена, из которых, как может быть видно в Таб.1, 4549 кодируют белок, 122 кодируют различные типы РНК, а 22 являются некодирующими. Из всех генов, что видно из Таб.2, 2315 генов расположены на прямой цепи ДНК, а 2377 – на обратной. После обработки полученных с помощью программы, написанной на Python 3.6, данных и сопоставления их с данными об исследуемой бактерии *Dickeya dadantii* 3937 было выдвинуто предположение о случайности расположения генов на цепях ДНК. В Таб.3 продемонстрировано, что вероятность случайного распределения генов по цепям ДНК составляет более 60%. Число “квазиоперонов” – 2626.

Таблица 1. Распределение генов бактерии *Dickeya dadantii* 3937 по категориям.

Тип гена	Число генов
Protein coding	4549
tRNA	75
Pseudogene	22
rRNA	22
ncRNA	13
Antisense RNA	3
RNase P RNA	1
tmRNA	1
SRP RNA	1

Таблица 2. Количество генов на прямой и на обратной цепях ДНК.

Цепь	Protein coding	Pseudogene	RNA	Суммарно по цепи
+	2254	10	51	2315
-	2295	12	70	2377
Суммарно по категориям	4549	22	121	4692

Таблица 3. Определение случайности расположения генов на прямой и обратной цепях.

Гены на прямой цепи	2315
Гены на обратной цепи	2377
Суммарно на двух цепях	4692
Отклонение от среднего ожидаемого	31
Вероятность большого отклонения	62,5%

В рамках работы было исследовано распределение белков штамма бактерии относительно их длин. Оказалось, что около 93% белков имеет длину в диапазоне 65 - 665 аминокислотных остатков, что видно из приведенной диаграммы. Длины всех белков находятся в диапазоне 32 - 4215 аминокислотных остатков. Другие статистические данные приведены в Таб.4.

Таблица 4. Статистические данные по белкам бактерии *Dickeya dadantii* 3937

минимальная длина	32
максимальная длина	4215
стандартное отклонение	231,9
среднее арифметическое	311,5
медиана	274

Диаграмма 1. Число белков в определенном диапазоне длин.



Были предприняты попытки исследовать функции нескольких “псевдооперонов”. Так, среди “псевдооперонов” были найдены: оперон из 6 генов, основной функцией которого является транспорт рибозы, оперон из 5 генов, отвечающий за транспорт мальтозы, оперон из 5 генов, регулирующий преобразование рибозы в фруктозу, оперон из 5 генов, отвечающий за синтез и деградацию целлюлозы. Данные об этих оперонах приведены в Таб.5.

Таблица 5. Исследованные опероны
 бактерии *Dickeya dadantii* 3937

Координата начала оперона	Координата конца оперона	Число генов в опероне	Белки, закодированные в опероне
6696	12538	6	predicted cytoplasmic sugar-binding protein; fused D-ribose transporter subunits of ABC superfamily: ATP-binding components; D-ribose transporter subunit; membrane component of ABC superfamily; ribokinase; periplasmic-binding component of an ABC superfamily D-ribose transporter; Ribose operon repressor
70225	75608	5	Maltose/maltodextrin ABC transporter, substrate binding periplasmic protein MalE; Sugar ABC transporter, permease protein; Sugar ABC transporter, permease protein; Maltose/maltodextrin transport ATP-binding protein MalK; Metal-dependent hydrolase
92176	100077	6	D-ribulokinase; C-compound and carbohydrate metabolism; Dihydrodipicolinate synthase; Putative cytoplasmic protein; Transporter, MFS superfamily; Fructose repressor FruR, LacI family
122078	132067	5	Cellulose synthase (UDP-forming); Cyclic di-GMP binding protein precursor; Cellulose synthase operon protein C precursor; hypothetical protein; endoglucanase Y

В заключение обсуждения можно сказать, что необходимы дальнейшие исследования генома *Dickeya dadantii* 3937, так как функции многих генов все еще непонятны.

4. Сопроводительные материалы

- <https://kodomo.fbb.msu.ru/~elizaveta-biloba/term1/pr13.xlsx>

5. Благодарности

Выражаю благодарность преподавательскому составу Факультета Биоинженерии и Биоинформатики МГУ им. М. В. Ломоносова за помощь при подготовке данной работы и обучение навыкам работы на языке программирования Python 3.6 и работы с базами данных.

6. Список источников

1. <https://ru.scribd.com/document/228652254/Erwinia-chrysanthemii> - Data Sheets on Quarantine Pests *Erwinia chrysanthemii*
2. ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/147/055/GCA_000147055.1_ASM14705v1/GCA_000147055.1_ASM14705v1_feature_table.txt.gz - особенности генома *Dickeya dadantii* 3937
3. http://medbiol.ru/medbiol/canc_2010/000045e3.htm - ABC-Белки (ABC-транспортеры, суперсемейство транспортных белков ABC)