
Мини-обзор генома *Prevotella intermedia*

Ахтямова Карина

Московский государственный университет имени М. В. Ломоносова

Факультет Биотехнологии и биоинформатики

РЕЗЮМЕ

В данной работе разобраны некоторые особенности бактерии *Prevotella intermedia*. Проанализирован состав генома двух последовательностей хромосом находящихся в этой бактерии, также был проанализирован внезапно малый GC-состав, как для безболезненной бактерии. Изучены длины белков находящихся в ней и проведено исследование свойств особо длинных белков.

Важно изучить особенности генома и протеома безболезненной бактерии, вызывающей пародонтит, с которым сталкивалось до 90% процентов взрослого населения по заключению ВОЗ.

КЛЮЧЕВЫЕ СЛОВА

Prevotella intermedia, GC-состав, длины белков, NZ_CP024727.1, NZ_CP024728.1

ВВЕДЕНИЕ

Prevotella intermedia — это грамотрицательная облигатная анаэробная патогенная бактерия.

Является одним из бактериальных патогенов, которые, как предполагается, вызывают пародонтит, и часто выделяются из поддесневого налета у пациентов, страдающих острым некротизирующим гингивитом, гингивитом при беременности и хроническим пародонтитом, встречается у 55% пациентов с хроническим пародонтитом, это меньше других сопутствующих бактерий, но гипер локализация (повышенная приспособляемость бактерии к тому или иному организму) встречается у 7,7% пациентов. Следовательно, бактерия является значимым прогностическим признаком тяжести поражения пародонта. Для *P. intermedia*, нет различий по полу по встречаемости и степени поражения при равной обсемененности.

Положение бактерии в современной систематике:

Царство Bacteria

Тип Bacteriota

Класс Bacteroidia

Отряд Bacteroidales

Семейство Bacteroidaceae

Род *Prevotella*

Вид *Prevotella intermedia*

Геном бактерии представлен двумя хромосомами: NZ_CP024727.1, с количеством генов равным 2011558 и NZ_CP024728.1, с количеством генов равным 753182, также важно отметить количество других аспектов бактерии - 2133 CDS, 48 tРНК, 12 rРНК (CDS-кодирующая последовательность (coding sequence))

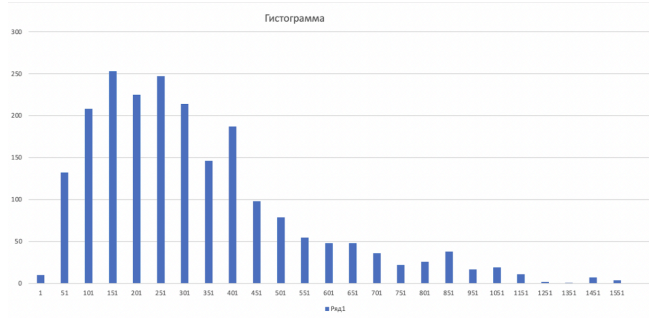
МЕТОДЫ

1. Для анализа состава генома хромосом, а также определения имени и GC-состава каждой, были использованы программы на языке программирования Python [1]
2. Для анализа длины белков, и создания гистограммы, были использованы функции СЧЕТЕСЛИМН, СУММ, а также возможность создания гистограммы программы Google Sheets. [2]
3. Для нахождения имен и GENEID особо длинных белков была использована функция ВПР программы Google Sheets. [2]
4. Для подсчета степени случайности были использованы функции СЧЕТЕСЛИМН, BINOM.DIST, МИН программы Google Sheets. [2]

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

1. Были проанализированы составы хромосом входящие в бактерию. Важно отметить, что составы аденина и тимина примерно равны друг другу, однако разность значений увеличивается с ростом количества нуклеотидов кодирующих хромосомы (Для сравнения разность составов аденина и тимина в хромосомах NZ_CP024727.1 и NZ_CP024728.1 соответственно равны 11514 и 494 соответственно, важно отметить что для второй хромосомы количество тимина больше чем для аденина, однако точной информации почему это происходит найти не удалось). GC-состав для хромосом NZ_CP024727.1 и NZ_CP024728.1 соответственно равен 42,86% и 44,85% соответственно. Удивительно что состав настолько мал, как для безболезненной бактерии вызывающей одного из самых часто встречающихся болезней стоматологического спектра, ведь данный состав характеризует устойчивость бактерии. Однако это может объясняться тем что *Prevotella intermedia* редко существует в одиночку. Обычно она стабилизируется сопутствующими бактериями, которые так или иначе воздействуют на иммунную систему (а также на состав слюны о чем свидетельствуют некоторые исследования, однако их достоверность находится под вопросом, но является интересной информацией)

2)



(Рис 1.) Гистограмма с шагом 50 и 100

На представленной гистограмме видно что пиков 2 с длиной 101-151 и 201-251, однако не это меня заинтересовало, а то что после объективного снижения практически до нулевых значений количества белков после длины 1351, был внезапный всплеск длинных белков длиной 1351- 1551. Также самый длинный белок с длиной 2536 был совершенно внезапен, так как второй по длине составлял всего 1891. Было интересно проанализировать более точные данные для длин белков, для чего была создана дополнительная гистограмма с большим шагом представленная ниже:



(Рис 1.) Гистограмма с шагом 10

На ней лучше заметны пики, и особо длинные белки бактерии.

3) Получив данные выше было принято решение попытаться связать малый состав GC, а также особо длинные белки бактерии. Составив таблицу связывающих длину белка и его имя было выявлено то, что самым длинным белком был белок клеточной оболочки бактерии. Что означает высокомолекулярность клеточной оболочки, особенно это видно при сравнении общей длины белков в бактерии. Высокомолекулярность оболочки означает ее устойчивость следовательно можно пренебречь устойчивостью в геноме. Следовательно малый GC-состав объясняется не только сопутствующими бактериями наблюдаемых при пародонтите в микрофлоре, но и сложным составом клеточной оболочки защищающей бактерию от ферментов слюны и иммунитета человека (или животного).

name	GeneID	product_length
cell surface protein SprA	57246799	2536
DUF4465 domain-containing protein	57245829	1891
alpha-2-macroglobulin	57246098	1879
translocation/assembly module TamB domain-containing protein	57247194	1703
choice-of-anchor J domain-containing protein	57246680	1514
N-6 DNA methylase	57246809	1497
translocation/assembly module TamB domain-containing protein	57247035	1491
choice-of-anchor J domain-containing protein	57246917	1481
choice-of-anchor J domain-containing protein	57247158	1479
cobaltochelatase subunit CobN	57246896	1469
DNA-directed RNA polymerase subunit beta'	57245277	1454
type II CRISPR RNA-guided endonuclease Cas9	57245793	1380
DNA-directed RNA polymerase subunit beta	57245278	1269
isoleucine-tRNA ligase	57246908	1253
phosphoribosylformylglycinamide synthase	57245777	1250
DNA polymerase III subunit alpha	57246036	1238
hypothetical protein	57245721	1236
choice-of-anchor J domain-containing protein	57245717	1235
hypothetical protein	57245641	1199
pyruvate:ferredoxin (flavodoxin) oxidoreductase	57246399	1188
hypothetical protein	57246028	1183
type IX secretion system sortase PorU	57245826	1181
DUF5112 domain-containing protein	57246317	1172
DEAD/DEAH box helicase family protein	57247376	1160
transcription-repair coupling factor	57246009	1156
type VI-B CRISPR-associated RNA-guided ribonuclease Cas13b	57246985	1148

(Рис 3.) Таблица имен и длин белков

4) Было проанализировано количество белков закодированных на двух цепях наибольшей хромосомы в бактерии, было выявлена статистическая значимость равная 0,0039, это означает что вероятность случайного распределения очень мала, а также различия важны для статистики.

+	838	Статистическая значимость	0,0039
-	723		

(Рис 4.) Таблица для оценки статистической значимости



(Рис 5.) Микробные колонии Prevotella intermedia на исходном уровне в тестовой группе. (b) Микробные колонии Aggregatibacter actinomycetemcomitans на исходном уровне тестовой группы. (c) Микробные колонии Porphyromonas gingivalis на исходном уровне тестовой группы a b c

На рисунке рассматривается различия внешнего вида заметных человеческому глазу для культур основанных на анализе слюны людей больных пародонтитом.

ЗАКЛЮЧЕНИЕ

Изучение бактерий вызывающих, те или иные заболевания очень важны для продвижения медицины. Изучение протеома и генома является одними из основных факторов при поиске и разработке лекарственных препаратов и методов лечения. Для бактерии Prevotella intermedia была выявлена закономерность длин белков и GC-состава генома. Также была проанализирована статистическая значимость распределения ДНК.

СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ

ВСЕ СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ НАХОДЯТСЯ В ПАПКЕ НА GOOGLE DRIVE:
[HTTPS://DRIVE.GOOGLE.COM/DRIVE/FOLDERS/1r6JdMJ3yJGE0RRLBNHprI5qOMQxAlQaL?usp=sharing](https://drive.google.com/drive/folders/1r6JdMJ3yJGE0RRLBNHprI5qOMQxAlQaL?usp=sharing)

СПИСОК ЛИТЕРАТУРЫ

1. Michal Potempa, Jan Potempa, Tomasz Kantyka, Ky-Anh Nguyen, Katarzyna Wawrzonek, Surya P. Manandhar, Katarzyna Popadiak, Kristian Riesbeck, Sigrun Eick, Anna M. Blom, Interpain A, Cysteine Proteinase from *Prevotella intermedia*, Inhibits Complement by Degrading Complement Factor C3
<https://doi.org/10.1371/journal.ppat.1000316>
2. Aena Jain Pundir, Anju Vishwanath, Siddharth Pundir, Mutyala Swati One-stage Full Mouth Disinfection Using 20% Propolis Hydroalcoholic Solution: A Clinico-microbiologic Study DOI:10.4103/ccd.ccd_544_17
3. Anna M. Blom, Teresia Hallstrom, Kristian Riebeck Complement evasion strategies of pathogens—Acquisition of inhibitors and beyond
<https://doi.org/10.1016/j.molimm.2009.04.025>