

# Обзор генома *Sphingobium indicum* B90A

Никонов М.А.<sup>1</sup>

<sup>1</sup>ФГБОУ ВО «МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ М.В. ЛОМОНОСОВА»  
ФАКУЛЬТЕТ БИОИНЖЕНЕРИИ И БИОИНФОРМАТИКИ

## АННОТАЦИЯ

Данный обзор посвящен геному штамма B90A альфа-протеобактерии *Sphingobium indicum*. В данной работе был проведен анализ GC-состава, типов генов, k-меров, была построена диаграмма длин генов, определены координаты генов рибосомных белков и рРНК.

## 1. ВВЕДЕНИЕ

*Sphingobium indicum* - вид альфа-протеобактерий из семейства Sphingomonadaceae, живущий в почве. Отличительной чертой семейства является наличие гликофинголипидов на внешней мембране[1], чем и объяснено название семейства. Особенностью данного вида является способность метаболизировать гексахлороциклогексан (HCH)[2]. В данной работе был исследован геном штамма *Sphingobium indicum* B90A, который составляет примерно 3945 килобаз.

## 2. МАТЕРИАЛЫ И МЕТОДЫ

Данные о геноме были получены из открытой библиотеки NCBI[3]. Для анализа данных была использована программа LibreOffice Calc, а также применялись скрипты, написанные на Python. Визуализация полученной информации производилась также в LibreOffice.

## 3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Геном *Sphingobium indicum* B90A представлен одной кольцевой хромосомой и тремя плазмидами. ДНК данной бактерии в совокупности составляет 3945926 нуклеотидов точно. Доля гуанина и цитозина (GC) на хромосоме составляет 64.18%, на плазмидах: 64,10%, 65,9%, 62,5%. Из этого следует устойчивость ДНК *Sphingobium indicum* к денатурации.

	Белок		РНК				
	Ген	Псевдоген	тРНК	рРНК	тмРНК	7SL РНК	РНКазы Р
Кол-во	3670	76	52	6	1	1	1

Таблица генов *Sphingobium indicum* B90A по типам

### 3.1. Анализ k-меров

Наиболее часто встречающимися словами в данном геноме оказались CG, GC, GG, CC, наименее частые слова: ТА и ТТ для k = 2. При k = 3 наиболее встречающиеся слова: GCG, CGC и GGC, наиболее редкие: ТАА, TAG и GTA.

Диаграмма CB (Compositional Bias) k-меров для k = 2 по данным листа S1 таблицы Nikonov-supple-fin

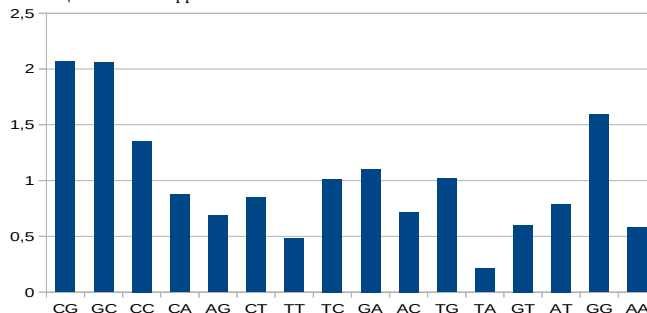
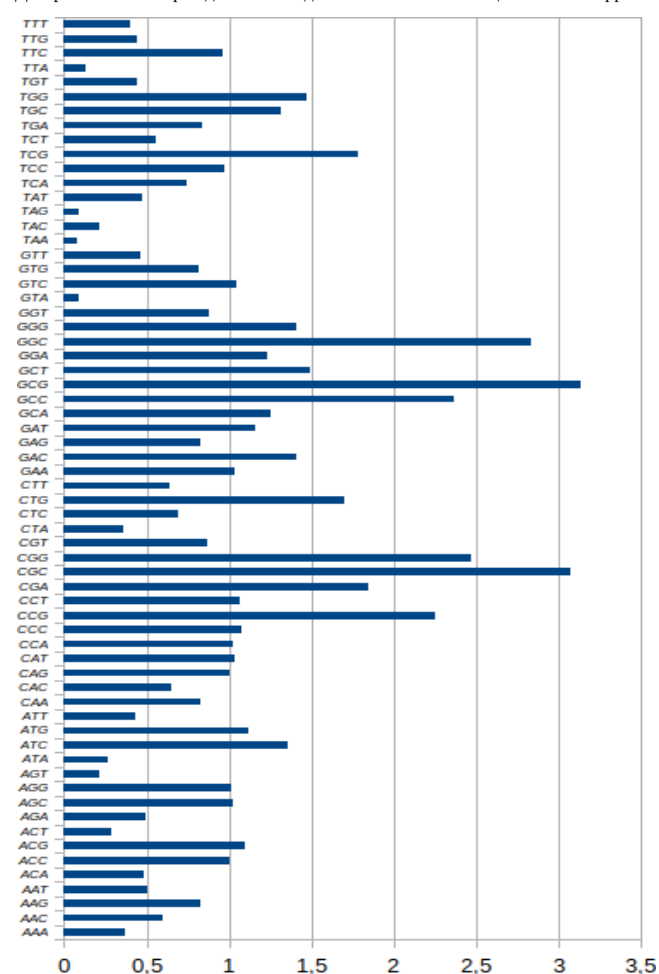


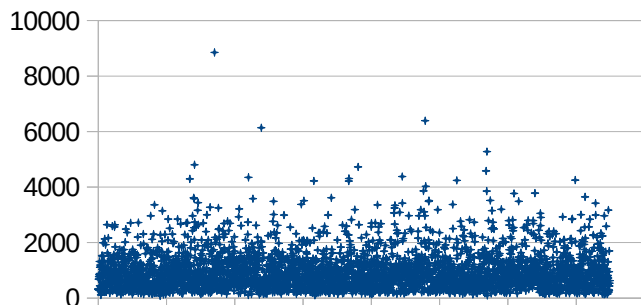
Диаграмма CB k-меров для k = 3 по данным листа S2 таблицы Nikonov-suppl-fin



### 3.2. Анализ протеома

Геном *Sphingobium indicum* B90A в совокупности содержит 3670 белок-кодирующих генов. Причем 1823 из них лежат на (-)-цепи, что свидетельствует о случайном распределении генов белков по цепям с вероятностью более 90%

Длины генов в нуклеотидах по данным листа S3 таблицы Nikonov-suppl-fin



Среднее значение длины белок-кодирующего гена — 952,5 нуклеотида. Самым длинным геном оказался Tandem 95 repeat protein, составляющий 8851 нуклеотид. Функция белка, который кодирует данный ген, неизвестна, но он содержит повторяющийся VCBS-домен, характерный для родов *Vibrio*, *Colwellia*, *Bradyrhizobium* и *Shewanella*[4]. Предполагается, что данный домен играет некоторую роль в адгезии[4].

### 3.3. Расположение генов рибосомных белков и рРНК

Ниже представлена таблица координат рибосомных белков и РНК *Sphingobium indicum* B90A. Видна закономерность расположения генов тРНК: все они располагаются близко друг от друга за исключением копий двух генов рРНК 5S и 23S малой субъединицы рибосомы, находящихся на одной из плазмид (PSRL1).

Таблица координат генов рибосомных белков и рРНК по данным листа S4 таблицы Nikonov-suppl-fin

Название	Цепь	Начало	Конец
50S ribosomal protein L7/L12	+	39243	40136
50S ribosomal protein L10	-	40355	40735
50S ribosomal protein L28	-	50927	51148
50S ribosomal protein L27	-	241805	242395
50S ribosomal protein L21	-	242669	242938
50S ribosomal protein L1	-	245812	247542
50S ribosomal protein L11	-	247796	248494
30S ribosomal protein S2	+	584876	586699
50S ribosomal protein L31	+	596274	597053
30S ribosomal protein S4	+	923385	925292
30S ribosomal protein S1	+	948572	948943
ribosomal protein S18-alanine N-acetyltransferase	-	1058237	1059394
50S ribosomal protein L25/general stress protein Ctc	-	1129376	1129951
30S ribosomal protein S15	-	1348845	1351820
30S ribosomal protein S6	-	1380284	1381204
30S ribosomal protein S18	-	1381274	1381855
50S ribosomal protein L9	-	1381855	1382748
30S ribosomal protein S12 methylthiotransferase RimO	-	1636863	1637798

50S ribosomal protein L19	+	1823676	1824713
30S ribosomal protein S16	-	1826214	1826981
50S ribosomal protein L11 methyltransferase	-	1844999	1846042
30S ribosomal protein S20	-	1990517	1991032
50S ribosomal protein L34	+	2044199	2044816
30S ribosomal protein S21	-	2148835	2149980
50S ribosomal protein L15	+	2164370	2165287
50S ribosomal protein L30	+	2165476	2165766
30S ribosomal protein S5	+	2165763	2166074
50S ribosomal protein L18	+	2166089	2166982
50S ribosomal protein L6	+	2166979	2167653
30S ribosomal protein S8	+	2167650	2168267
30S ribosomal protein S14	-	2168371	2169240
50S ribosomal protein L5	-	2169242	2170642
50S ribosomal protein L24	-	2170685	2171311
50S ribosomal protein L14	-	2171308	2171718
30S ribosomal protein S17	-	2171791	2172621
50S ribosomal protein L29	-	2172658	2173233
50S ribosomal protein L16	-	2173361	2174281
30S ribosomal protein S3	-	2174399	2176066
50S ribosomal protein L22	-	2176085	2177098
30S ribosomal protein S19	+	2177389	2177970
50S ribosomal protein L2	+	2177974	2178711
50S ribosomal protein L23	+	2178708	2179109
50S ribosomal protein L4	+	2179141	2180034
50S ribosomal protein L3	+	2180058	2181155
30S ribosomal protein S10	-	2181472	2182032
30S ribosomal protein S7	+	2183466	2185337
30S ribosomal protein S12	+	2185384	2186631
30S ribosomal protein S9	+	2187111	2187575
50S ribosomal protein L13	+	2187581	2188975
50S ribosomal protein L32	+	2721800	2722306
50S ribosomal protein L36	-	2816552	2817379
30S ribosomal protein S13	-	2881988	2882482
30S ribosomal protein S11	-	2882472	2882687
50S ribosomal protein L17	+	2883435	2883908
30S ribosomal protein S14	-	3070861	3071475
50S ribosomal protein L33	-	3151721	3152635
5S ribosomal RNA	-	3196033	3196947
23S ribosomal RNA	-	3196947	3197657
16S ribosomal RNA	-	3199708	3200415
50S ribosomal protein L20	+	3331826	3332884
50S ribosomal protein L35	+	3332881	3333405
		PSRL1	
5S ribosomal RNA	+	3649427	3650098
		PSRL1	
16S ribosomal RNA	+	3651903	3652793

## **ЗАКЛЮЧЕНИЕ**

Согласно результатам исследования геном *Sphingobium indicum* B90A составляет 3945926 пар нуклеотидов точно. Доля GC в ДНК составляет 64.18%, что может свидетельствовать о ее устойчивости к денатурации. Согласно результату анализа k-меров наиболее частыми словами оказались CG, GC, GG, CC (для k=2), GCG, CGC и GGC (k=3), наименее частыми - ТА и ТТ (k=2), ТАА, ТАГ и GTA (K=3). Среди 3670 белок-кодирующих генов закономерности расположения по цепям не наблюдалось. Средняя длина белок-кодирующего гена составляет 952,5 нуклеотидов. Самым длинным геном (8851 пар нуклеотидов) оказался ген Tandem 95 repeat protein, кодирующий белок с неизвестной функцией с VCBS-доменом, возможно, участвующий в адгезии. Закономерностей в расположении рибосомных белков не наблюдалось, но гены рРНК сконцентрированы в одной области кроме копий двух генов рРНК малой субъединицы, находящихся на плазмиде.

## **БИБЛИОГРАФИЯ**

- [1] Garrity GM, Brenner DJ, Krieg NR et al. (2005). "Bergey's Manual of Systematic Bacteriology. Two The Proteobacteria, Part C: The Alpha-, Beta-, Delta-, and Epsilonproteobacteria". New York, New York: Springer.
- [2] Pal, R. (2005). "Hexachlorocyclohexane-degrading bacterial strains *Sphingomonas paucimobilis* B90A, UT26 and Sp+, having similar *lin* genes, represent three distinct species, *Sphingobium indicum* sp. nov., *Sphingobium japonicum* sp. nov. and *Sphingobium francense* sp. nov., and reclassification of [*Sphingomonas*] *chungbukensis* as *Sphingobium chungbukense* comb. nov". International Journal of Systematic and Evolutionary Microbiology.
- [3] Директория с данными NCBI о геноме *Sphingobium indicum* B90A  
[ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/264/945/GCF\\_000264945.2\\_ASM26494v2](ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/264/945/GCF_000264945.2_ASM26494v2)
- [4] Данные NCBI по домену VCBS  
<https://www.ncbi.nlm.nih.gov/Structure/cdd/TIGR01965>

## **СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ**

1. Таблица Nikonov-suppl-fin.xlsx  
[kodo.fbb.msu.ru/~w0lfgang/term1/Nikonov\\_suppl\\_fin.xlsx](https://kodo.fbb.msu.ru/~w0lfgang/term1/Nikonov_suppl_fin.xlsx)
2. Скрипты Python  
<https://kodo.fbb.msu.ru/~w0lfgang/term1/scripts>