

Обзор протеома бактерии *Hiprea maritima* DSM 10411

Шошинова М.С.

Факультет биоинженерии и биоинформатики МГУ имени М.В. Ломоносова

ABSTRACT

В статье представлен обзор протеома бактерии *Hiprea Maritima*, штамм DSM 10411, проведённый в целях ознакомления с методикой написания статей. Итогами исследования стали следующие выводы:

- I. Геном бактерии размером 1.7Мб кодирует 1677 белков и 57 РНК, среди которых преобладает тРНК.
- II. Гены расположены как на прямой цепи, так и на обратной, причём их количественное соотношение для обоих направлений примерно одинаково.
- III. Большая часть белков имеет длину до 400 аминокислотных остатков
- IV. Существуют участки, кодирующие белки неопределённого назначения

1 INTRODUCTION

Программа курса биоинформатики предполагает написание научных статей в общепринятом формате, и данная работа является ознакомительной. Обзор не претендует на глубину или полноту, поскольку основывается на небольшом количестве исходных данных, предоставленных [1] и [2]. В качестве учебного объекта была получена бактерия *Hiprea Maritima*, штамм DSM 10411. Эта бактерия относится к хемолитотрофным и существует за счёт энергии, получаемой при окислении неорганических соединений, в частности, серы.

На момент написания статьи занимает следующее положение в систематике: Bacteria → Proteobacteria → Deltaproteobacteria → Desulfurellales → Desulfurellaceae → *Hiprea* → *Hiprea Maritima*. Целью изучения протеома этой бактерии является получение представлений о методах сбора и анализа информации и способах представления данных. Поэтому анализ протеома проводится по критериям длин белков и числа генов без рассмотрения их роли в жизненном цикле бактерии.

2 METHODS

Информация о геноме и протеоме бактерии получена из [1] с помощью плагина NetBox файлового менеджера FAR3.0 build 4040, просмотрена в редакторе EmEditor ver. 14.5.4 и интерпретирована с помощью электронных таблиц Excel 2007 пакета Microsoft Office (12.0.4518.1014, русифицированная версия). Данные и результаты обзора представлены в про-

граммах Word 2007 пакета Microsoft Office (12.0.4518.1014, русифицированная версия) с последующей конвертацией в формат pdf.

При анализе данных внимание уделялось распределению кодирующих последовательностей в цепи ДНК, а также соотношениям длин белков и РНК и их количеству. Полученные данные оценивались с помощью функций Excel.

3 RESULTS

Длина генома *Hiprea Maritima* составляет 1694430 пар оснований, с помощью которых бактерия может производить 1677 белков и 57 РНК.

В ходе исследования было выявлено, что гены, кодирующие аминокислотные последовательности белков *Hiprea Maritima*, расположены как на прямой, так и на обратной цепи ДНК, причём на обеих цепях их количество примерно одинаково.

Максимальная длина белка составляет 1584, минимальная – 35 аминокислотных остатков. Распределение белков по длинам максимально в области до 400 аминокислотных остатков. Гистограмма распределения приведена на рисунке 1. Число рРНК в 8 раз меньше числа тРНК.

347 генов кодируют белки, обозначенные как «hypothetical protein» (см. таблицу 1). В описании 3 генов РНК стоят прочерки. Для остальных белков и РНК определены тип и функция, приведённые в таблице Excel.

4 DISCUSSION

В описании протеома попадались строки «hypothetical protein», а среди РНК некоторые были отмечены прочерком. Исходя из этого был сделан вывод, что не для всех синтезируемых соединений определена их роль в жизненном цикле бактерии.

Отношение числа нуклеотидов, входящих в гены, к общему их числу близко к 1/2, что может отражать как эффективную эволюцию в отношении «лишних» участков генома, так и сравнительно недавнее происхождение вида. Однако при таком поверхностном рассмотрении, которое даже не включало в себя анализ типов производимых белков, степень их необходимости, частоту востребования в жизненном цикле бактерии и

объёмы синтеза в единицу времени, невозможно даже примерно судить о такого рода закономерностях.

5 SUPPLEMENTARY MATERIALS

Таблица 1. Соотношения длин белков и РНК

Proteins		RNA's	
Proteins summary:	1677	RNA summary:	57
Maximum protein length:	1584	number of tRNA:	48
Minimum protein length:	35	number of rRNA:	6
Median of protein length:	278	not defined:	3
Number of "hypothetical":	347	RNA length sum:	14661
Protein length sum (in nb):	1571605	of all DNA chains length, %	0,432623
of all DNA chains length, %	46,37562		
Strand:	830	Strand:	28
Reverse:	847	Reverse:	29

Соотношение белков и РНК, кодируемых на прямой и обратной цепи, близко к 50%. Среди РНК есть неопределённые участки, помеченные прочерком в исходных данных и отображённые в таблице как not defined

Рисунок 1. Распределение длин белков по интервалам



Excel-файл с исходными и итоговыми данными доступен по адресу:

<http://kodomofbb.msu.ru/~arma/term1/Excel/HipMa.xlsx>

ACKNOWLEDGEMENTS

Ведущую роль в освоении программных средств сыграли упражнения и презентации, составленные преподавателями факультета биотехнологии и биоинформатики, а также множество энтузиастов, размещавших в интернете справочные материалы и примеры использования, за что им выражается огромная благодарность!

REFERENCES

1. <ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/> - источник информации о геноме и протеоме *Hippea Maritima*
2. <http://kodomofbb.msu.ru/> - руководство по составлению данной статьи и справочные материалы
3. <https://ru.wikipedia.org/wiki/%CF%F0%E0%E2%E5%E1%E0%EA%F2%E5%F0%E8%E8> - статья Википедии о протеобактериях
4. <http://kodomofbb.msu.ru/~arma/term1/Excel/HipMa.pdf> - адрес настоящей мини-статьи