

---

Биоинформатика

Обзор генома и протеома бактерии *Thermobifida sp. fusca* YX

Соболь Анастасия Павловна

Факультет биоинженерии и биоинформатики МГУ им. М. В. Ломоносова, ул. Ленинские Горы, 1с73,  
Москва

**Gmail: sobanas2000@gmail.com**

**20.12.2018**

---

## Резюме:

Цель данной работы - исследование протеома бактерии **Thermobifida sp. fusca** YX.

Для создания данной статьи был проанализирован её геном: получены количественные оценки для генов и белков с точки зрения их классификации по различным категориям, построена гистограмма распределения длин белков, рассмотрен вклад прямой и комплементарной цепей ДНК в кодирующие участки генома.

## 1. ВВЕДЕНИЕ:

**Thermobifida fusca** - умеренно теплолюбивая, палочковидная бактерия почвы. (Оптимальная температура размножения 55 градусов Целься)

Данная бактерия является грамположительной (бактерии, которые, в отличие от грамотрицательных бактерий, сохраняют окраску, не обесцвечиваются при промывке при использовании окраски микроорганизмов по методу Грама. Большинство Грам (+) бактерий имеют однослойную клеточную мембрану, без внешней мембраны.)

**Thermobifida fusca**- аэробная бактерия.

Видов этого рода отличает их способность формировать сгруппированные споры. Род **Thermobifida** включает 2 разновидности, **T. fusca** и **T. Элбу**. Из двух разновидностей, **T. fusca** наиболее теплолюбивая, её обнаруживают в местах с температурой до 75 °С.

Естественная среда обитания **Thermobifida** - органические материалы, подобно гниющему селу, компосту, удобрению или городским свалкам. Биологические и физиологические особенности этих бактерий соответственно приспособлены к условиям таких окружающих сред, а именно к высоким температурам и присутствию большого количества биополимеров естественного происхождения. Актиномицеты, к которым и относится **Thermobifida fusca**, хорошо приспособлены к этой окружающей среде, потому что они растут как совокупность гифов и способны проникать в клетки и разрушать нерастворимые полимеры типа лигноцеллюлозы.

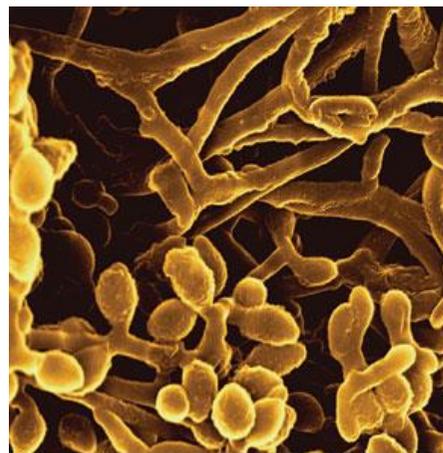


РИС.1 Изображение бактерии под микроскопом (рисунок получен с сайта: <http://www.abc.hu>)

Споры **Thermobifida**, как известно, вызывают аллергические заболевания дыхательных путей, названные «болезнями рабочего гриба» и «легким фермера». Эти болезни развиваются у сельскохозяйственных рабочих, которые по характеру их работы

вдыхают существенные количества спор актиномицетов от сена, компоста, и т.д.

**Thermobifida fusca.** Представляет интерес для человека ещё и потому, что способна разрушать(минерализовать) пластмассы.

Более того, как было сказано ранее, **Thermobifida** производит теплоустойчивые ферменты, вовлеченные в разрушение лигноцеллюлозы. Штаммы **Thermobifida** способны разрушать целлюлозу и лигноцеллюлозу, которые составляют большую часть сельскохозяйственных и городских отходов, и таким образом бактерии имеют высокое значение в установлении безопасных биотехнологических процессов переработки отходов.

Ключевые слова: бактерия, **Thermobifida sp. fusca** YX обзор, протеом.

<b>Domain</b>	<b>Bacteria</b>
<b>Kingdom</b>	<b>Terrabacteria group</b>
<b>Phylum</b>	<b>Actinobacteria</b>
<b>Class</b>	<b>Streptosporangiales</b>
<b>Order</b>	<b>Streptosporangiales</b>
<b>Family</b>	<b>Nocardiopsaceae</b>
<b>Genus</b>	<b>Thermobifida</b>
<b>Species</b>	<b>Thermobifida fusca</b>
<b>Strain</b>	<b>YX</b>

В таблице приведено таксономическое положение бактерии (Информация получена с <https://www.ncbi.nlm.nih.gov/> и изложена с помощью программы Excel)

## 2.МАТЕРИАЛЫ И МЕТОДЫ:

### 2.1 Используемые ресурсы и программы:

При поиске материалов для работы с геномом бактерии и для скачивания необходимых файлов использовался сайт NCBI. Для работы с данными (вычисления, составление таблиц) использовалась программа Microsoft Excel 2016. При работе с файлами был полезен Far Manager.

Функции, использованные для работы с данными в Excel.	Функция
Средняя длина белка	СРЗНАЧ
Стандартное отклонение	СТАНДОТКЛОН
Медианная длина белка	МЕДИАНА
Минимальная длина белка	МИН
Максимальная длина белка	МАКС
Число белков в определенном диапазоне	СЧЁТЕСЛИМН

## 2.2 Подсчёт числа генов белков и РНК:

Для подсчёта числа генов белков по всем категориям использовались фильтры “CDS” и “with\_protein”. По полученным данным было найдено количество различных генов. Для генов транспортных белков был установлен фильтр “transport”, для рибосомальных – “ribosomal”, для гипотетических - “hypothetical”; оставшиеся были посчитаны вычитанием из общего числа вышеперечисленных групп. Для подсчёта числа генов РНК были проделаны те же действия, но были изменены фильтры. Были произведены такие же расчёты, что и для поиска генов белков. Все результаты находятся в пункте 3.1 (см. Таблица 1, 2)

## 2.3 Распределение длин белков:

Для подсчёта количества длин белков на заданных промежутках (с шагом в 200 а.о.) сначала была составлена таблица всех длин, а после была проанализирована частота встречаемости длин белков на данных промежутках. В результате была составлена гистограмма длин белков (см. Рисунок 2; см сопровождающую Таблицу 3)

## 2.4 Подсчёт числа генов белков, РНК и псевдогенов на прямой и обратной цепях ДНК:

Для подсчёта числа генов белков РНК и псевдогенов на прямой и обратной цепях ДНК были использованы фильтры “+” - если ген белка находился на прямой цепи, и “-” - если на комплементарной цепи. (см. Таблица 4)

## 3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЯ:

### 3.1 Распределение генов по категориям:

В Таблицах 1, 2 представлены результаты исследования распределения двух основных классов генов, а именно, РНК и белков по категориям.

Всего генов	3131
-------------	------

Как видно из полученных данных, количество генов белков, отнесенных к категории «остальные», превосходит количество распределенных по именованным категориям, и составляет 70,17% от общего числа. Скорее всего, это связано с тем, что данная категория по сути является «сборной» некоторого количества подкатегорий, значимостью которых по отдельности в данной работе было решено пренебречь. Можно заметить, что значительную часть представляют гены, кодирующие гипотетические белки. Это может быть связано с тем, что данный штамм еще недостаточно изучен.

БЕЛКИ		3064
Категория	Число генов	В % от общего числа генов
Транспортные	203	6,48%
Рибосомальные	74	2,36%
Гипотетические	590	18,84%
Остальные	2197	70,17%

Таблица 1. Гены белков. Распределение по категориям

По распределению генов, кодируемых РНК-геномом, можно сказать, что бóльшую часть из них составляют гены тРНК. Интересно заметить, что в отличие от генов белков, к категории «остальные» был отнесен всего лишь 1 ген.

РНК		67
Категория	Число генов	В % от общего числа генов
рРНК	12	0,38%
тРНК	52	1,66%
Остальные	3	0,10%

Таблица 2. Гены РНК. Распределение по категориям

Стоит отметить, что более 97% генов кодируют белки, тогда как на РНК приходится около 2%.

### 3.2 Статистика длин белков:



РИС.2

На Рисунке 2 четко видно, что большинство белков состоят из 800-1200 аминокислотных остатков, тогда как всего несколько белков имеют длину больше 3000 аминокислотных остатков. Причём далее встречаемость генов уменьшается.

Показатель	Значение
Минимальная длина	71
Максимальная длина	10890
Средняя длина	978
Медиана	846
Стандартное отклонение	672

Таблица 3. Статистика длин белков из протеома бактерии *Thermobifida sp. fusca* YX

Можно заметить, что разница между минимальной и максимальной длинами очень велика, а также, что медиана и средняя длина отличаются друг от друга.

Размер одного из белков, обладающий максимальной длиной, сильно отличается от остальных. Думаю, большой размер белка может быть связан со сложностью структуры, которую он кодирует.

### 3.3 Распределение кодирующих последовательностей по цепям ДНК:

Результат подсчета генов на разных цепях ДНК представлен в Таблице 4.

Хорошо видно, что распределение кодирующих последовательностей по цепям ДНК случайное и количество генов прямой цепи примерно равно генам обратной цепи.

Цепь	Гены белков	Псевдогены	Гены РНК
Прямая	1520	6	27
Обратная	1544	11	40

Таблица 4. Расположение генов по цепям ДНК

### 4.СОПРОВОДИТЕЛЬНЫЕ МАТЕРИАЛЫ:

Все действия с таблицей и вычисления находятся в [файле](#)

### 5.БЛАГОДАРНОСТИ:

*Хотелось бы поблагодарить преподавательский состав факультета биоинженерии и биоинформатики МГУ им. М. В. Ломоносова за отзывчивость, помощь в освоении сложного материала и полезные подсказки.*

### 6.СПИСОК ЛИТЕРАТУРЫ:

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1899369/>
- <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=269800>
- [https://ru.wikipedia.org/wiki/%D0%93%D1%80%D0%B0%D0%BC%D0%BF%D0%BE%D0%BB%D0%BE%D0%B6%D0%B8%D1%82%D0%B5%D0%BB%D1%8C%D0%BD%D1%8B%D0%B5\\_%D0%B1%D0%B0%D0%BA%D1%82%D0%B5%D1%80%D0%B8%D0%B8](https://ru.wikipedia.org/wiki/%D0%93%D1%80%D0%B0%D0%BC%D0%BF%D0%BE%D0%BB%D0%BE%D0%B6%D0%B8%D1%82%D0%B5%D0%BB%D1%8C%D0%BD%D1%8B%D0%B5_%D0%B1%D0%B0%D0%BA%D1%82%D0%B5%D1%80%D0%B8%D0%B8)
- [https://www.genome.jp/kegg-bin/show\\_organism?org=tfu](https://www.genome.jp/kegg-bin/show_organism?org=tfu)
- <https://www.naik.hu/>

