



# Тайны семейства гистидиновых триад

Анна Любанская, Мария Новикова, Яна Савченко,  
Ольга Бочкарева, Наталия Драненко, Вера Емельяненко, Айгуль Назибуллина, Александр Чистяков, Укрон Тиндрен

## Введение

Белки гистидиновой триады (Pht) – семейство белков внешней мембраны, играющих роль в массе процессов в клетке. У некоторых видов *Streptococcus* они также участвуют в патогенезе. В геномах может встретиться до четырех представителей этого семейства, однако их структура и функции остаются недоизученными. В рамках статьи мы нашли гены, кодирующие белки гистидиновой триады, во всех полных геномах стрептококков, доступных в базе GenBank, выровняли последовательности, проанализировали структуру оперонов, в которых они содержатся, и предсказали сайты связывания цинкового репрессора, ассоциированные с этими оперонами.

## Методы

RanASoTA: предварительный анализ; HMMer: поиск генов, кодирующих белки гистидиновой триады; zScan: предсказание сайтов связывания цинковых репрессоров; MUSCLE: выравнивания; iTOL: визуализация дерева.



## Обозначения:

- Большой длины
- Средней длины
- Малой длины

Список видов с количеством штаммов в каждой фазовой вариации

- S. pyogenes*
- S. sanguinis*
- S. mitis*
- S. galloyticus*
- S. canis*
- S. iniae*
- S. uberis*
- S. agalactiae*

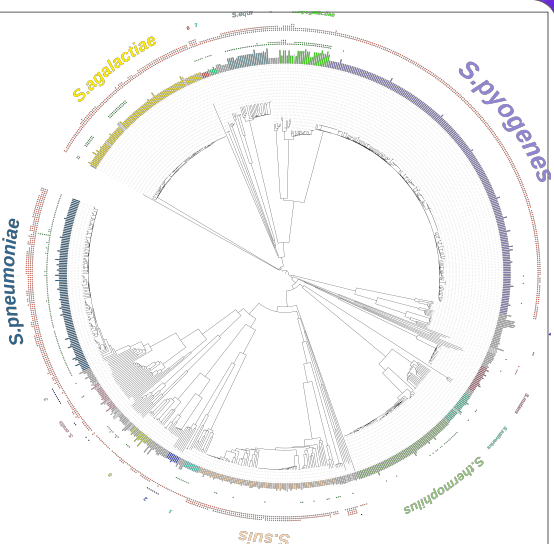
Филогенетическое дерево рода *Streptococcus*. Штаммы одного вида выделены одним цветом, при этом виды с количеством штаммов больше 17 обозначены текстом.

Зелеными цветом выделены гены, кодирующие белки гистидиновой триады (Pht), красными сайты связывания цинкового репрессора, не заполненные – не известны.

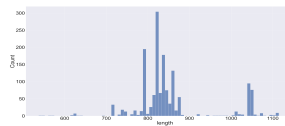
На дереве можно выделить два кластера: первый – это гены *pht1* от 10 видов, второй – гены *pht2* от 10 видов, у которых гены *pht1* от 10 видов не имеют *pht2* (исключениями).

Первый кластер в основном представлен представителями семейства Streptococcaceae в составе семейства Streptococcaceae и Streptococcaceae. Второй кластер представлен в основном представителями семейства Streptococcaceae в составе семейства Streptococcaceae и Streptococcaceae. В основном это представители рода *Streptococcus*. В первом кластере гены *pht1* и *pht2* имеют одинаковую длину, во втором – гены *pht1* и *pht2* имеют разную длину. В первом кластере гены *pht1* и *pht2* имеют одинаковую длину, во втором – гены *pht1* и *pht2* имеют разную длину. В первом кластере гены *pht1* и *pht2* имеют одинаковую длину, во втором – гены *pht1* и *pht2* имеют разную длину.

Второй кластер представлен видами *S. dysgalactiae*, *S. agalactiae* и *S. pyogenes* и др. В первом кластере гены *pht1* и *pht2* имеют одинаковую длину, во втором – гены *pht1* и *pht2* имеют разную длину. В первом кластере гены *pht1* и *pht2* имеют одинаковую длину, во втором – гены *pht1* и *pht2* имеют разную длину.

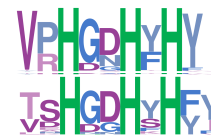


## Длины белков гистидиновой триады



Найденные белки гистидиновой триады очень варьировались по длине. Чаще всего была построена гистограмма длин последовательностей белков. Из нее видно, что большая часть белков имеет длину 500-1150 аминокислот. Остальные белки разделились на группы для <math>< 500</math> и <math>> 1150</math> аминокислот.

## Количество гистидиновых триад



Лого первой и второй гистидиновой триады основной выборки

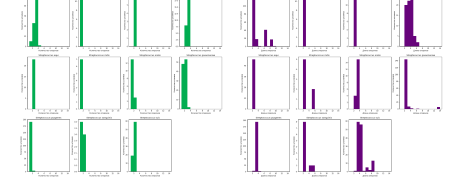
Первые две гистидиновые триады белков основной выборки наиболее консервативны. Было замечено, что количество гистидиновых триад в белке коррелирует с его длиной. Так, в белках большой длины может встретиться более 6 гистидиновых триад, а в аминокислотных последовательностях основной выборки обычно около 3-4. Также встречались случаи, когда в коротких белках отсутствовали гистидиновые триады. У бактерий вида *S. thermophilus* все найденные белки относятся к группе коротких, а в них отсутствуют гистидиновые триады, а их гены не имеют сайта связывания цинкового репрессора. Учитывая тот факт, что эти организмы не являются патогенами, мы предполагаем, что наличие этих белков в данном случае – артефакт.

## Выравнивание белков гистидиновой триады вида Streptococcus agalactiae

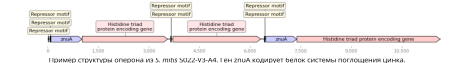
Были построены выравнивания белков гистидиновой триады всех групп для каждого вида *Streptococcus*. При анализе выравниваний выяснилось, что часть белков малой длины на самом деле являются фрагментами последовательностей белков средней длины, что можно увидеть на примере *S. agalactiae*. Также это можно хорошо наблюдать на филогенетическом дереве. Рамки считывания данных пар коротких белков расположены в геномах последовательно.

На выравнивании выше представлены 4 последовательности: белок средней длины одного из штаммов *S. agalactiae*, 2 белка малой длины другого штамма данного вида, у которого отсутствует данная последовательность. Нижняя последовательность – транскрипционная нуклеотидная последовательность от начала рамки считывания первого короткого белка до стоп-кодона второго.

Из литературы известно, что некоторые гены, кодирующие белки гистидиновой триады, регулируются цинковыми репрессорами. Мы выполнили поиск потенциальных сайтов связывания цинковых репрессоров и обнаружили, что большинство этих генов находятся под контролем цинковых репрессоров. Кроме того, зачастую гены, кодирующие белки гистидиновой триады, лежат в оперонах.



Гистограмма числа оперонов под контролем цинковых репрессоров для оперонов вида *S. pyogenes* (слева) и *S. agalactiae* (справа).



Пример структуры оперона из *S. mitis* Z422V9-A4. 1-ый *pht* кодирует белок системы поглощения цинка.

В рамках вида количество оперонов, кодирующих белки гистидиновой триады, достаточно стабильно и в среднем составляет около двух единиц. Длина оперона, в отличие от их количества, более вариативна. Применительно то, что во многих родах опероны обычно одной длины и включают в себя только гены белков гистидиновой триады и гены импорта цинка, но есть и исключения. У *S. mitis* опероны, включающие гены белков гистидиновой триады, могут содержать либо 2, либо 5 генов, но состоят они всегда только из *pht1* и *pht2* генов. *S. agalactiae*, *S. dysgalactiae* и *S. suis* обладают большим разнообразием длин оперонов. *S. pneumoniae* включает опероны, содержащие и 13 генов, но мы предполагаем, что это связано с наличием большого количества обломков псевдогенов.