Институт проблем передачи информации
им. А.А.Харкевича РАН

Биоинформатика, или молекулярная биология in silico

М.С.Гельфанд

15 января 2008

Расшифрован геном!

расшифрован геном человека - Поиск в Google - Mic...

File Edit View Favorites Tools Help

Back Search

Address http://www.google.ru/search?complete= Go Links

Google расшифрован геном человека Настройки

Веб Картинки Карты Новости Группы Почта Gmail

Войти

Google расшифрован геном человека

Поиск в Интернете Поиск

Веб Результаты **1** - **10** из примерно **73 200** для **расшифр**

Расшифрована структура **генома человека**. Статьи. Наука и техника
Расшифрована структура **генома человека**. Татьяна БАТЕНЕВА. С одной стороны, можно считать, что начался отсчет новой эпохи, сулящей победу над болезнями, ...
n-t.ru/tp/in/rs.htm - 13k -
Сохранено в кэше - Похожие страницы

Расшифрован геном человека
Фактически это означает завершение первого этапа грандиозного исследовательского проекта, известного как "**Геном Человека**". ...
www.ma-ma.ru/news/167 - 25k -
Сохранено в кэше - Похожие страницы

РИА Новости - Наука - **Расшифрован геном человека**
textLooks span.date { font-size:11px; color:#384787; padding-right: 10px} </style> <h5>**Расшифрован геном человека**</h5> <p><span ...
www.rian.ru/science/20030415/366929.html - 64k -
Сохранено в кэше - Похожие страницы

Диабет-Новости | ДОСТИЖЕНИЯ / ОТКРЫТИЯ
Специалисты утверждали, что на реализацию проекта '**Геном человека**' понадобится более 20 лет, но его **расшифровали** за 6. Уже в ближайшие 10 лет ученые ...
www.diabet-news.ru/info/otkr/otkr10.shtm - 37k -
Сохранено в кэше - Похожие страницы

Впервые **расшифрован геном** конкретного **человека**
В июне 2000 года совместно с руководителем проекта "**Геном человека**" Френсисом Коллинсом доктор Вентер объявил о первой реконструкции полного **генома человека** ...
www.bilogs.net/2007-09-04/vpervie_rasshifrovan_genom_konkretnogo_cheloveka.html - 11k -

Done Internet

genome deciphered - Поиск в Google - Microsoft ...

View Favorites Tools Help

Search Favorites

http://www.google.ru/search?complete=18 Go Links

human genome deciphered Настройки

ртинки Карты Новости Группы Почта Gmail ещ

Войти

ogle human genome deciphered

Поиск в Интернете Поиск ст

результаты **1** - **10** из примерно **373 000** для **human ge**

e.gov | 2000 Release: Two-Thirds **Human** equenced- [Перевести эту страницу]
s of **Human** DNA Script **Deciphered** by **Human** Project. Public Consortium To Complete "Working June. March 2000 ...
ome.gov/10002080 - 15k -
но в кэше - Похожие страницы

e Mania — **Deciphering** the **human** e: what does it
[Перевести эту страницу]
a gave extraordinary attention to the recent n of the draft sequence of the **human genome** by s of researchers in two leading ...
wersingenesis.org/docs2001/0309_**genome**.asp - 31k
ено в кэше - Похожие страницы

NewsHour: Breaking the Code -- December
- [Перевести эту страницу]
etic sequences would be comparatively easy to r with the aid .... If you printed out the whole **human** on pages-- and I don't know if ...
.org/newshour/bb/health/july-dec99/dna_12-2.html -
ранено в кэше - Похожие страницы

e **Genome Deciphered**; May Herald Breakthroughs- [Перевести эту страницу]
Genome Deciphered; May Herald Medical oughs ... Compared to the **human genome**, the nome is only about 1.5 percent different, ...
ionalgeographic.com/news/2007/04/070412-macaque-.html - 27k - Сохранено в кэше - Похожие страницы

cientists Start **Deciphering** Dog **enome**- [Перевести эту страницу]
an's best friend may soon have its **genome** ciphered. The National **Human Genome** Research stitute (NHGRI) in Bethesda, Maryland, recently added a dog to ...

Internet

расшифрован геном - Поиск в Google - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites

Address http://www.google.ru/search?complete=1&hl=ru&rlz=1T4GFRC_ruRU2 Go Links

Google расшифрован геном Настройки

Веб Картинки Карты Новости Группы Почта Gmail ещё Войти

Google расшифрован геном Поиск

Поиск в Интернете Поиск страниц на русском

Веб Результаты **1** - **10** из примерно **87 900** для **расшифрован геном**. (**0,27** секун

Расшифрован геном кошки - Медицинская библиотека сервера MedLinks.ru
Расшифрован геном кошки - Международная группа ученых, координируемая специалистами американского Национального института рака (National Cancer Institute), ...
www.medlinks.ru/article.php?sid=30736 - 46k -
Сохранено в кэше - Похожие страницы

Расшифрован геном винограда Pinot Noir | Газета.Ru: Новости
Расшифрован геном винограда Pinot Noir. Группа генетиков из Италии, США, Бельгии и Великобритании завершила предварительную расшифровку **генома** одного из ...
www.gazeta.ru/news/science/2007/12/19/n_1154765.shtml - 79k -
Сохранено в кэше - Похожие страницы

Элементы - новости науки: Полностью **расшифрован геном** собаки
Наряду с полной **расшифровкой генома** боксера были частично прочтены генетические коды еще нескольких пород собак, а также родственных видов — волка и койота. ...
elementy.ru/news/165035 - 30k - Сохранено в кэше - Похожие страницы

Коммерческая биотехнология | **Расшифрован геном** опоссума
Коммерческая биотехнология - сайт для всех, кто участвует в развитии российской биотехнологии: ученых, бизнесменов, производителей и продавцов оборудования ...
www.cbio.ru/modules/news/article.php?storyid=2784 - 36k -
Сохранено в кэше - Похожие страницы

Расшифрован геном устойчивого к антибиотикам штамма туберкулеза
Сегодня поступило сообщение, что исследователям из ЮАР удалось полностью секвенировать **геном** устойчивой к антибиотику формы туберкулеза. ...
rnd.cnews.ru/natur_science/news/line/index_science.shtml?2007/10/12/270287 - 27k -
Сохранено в кэше - Похожие страницы

ПОЛИТ.РУ \ НАУКА \ **Расшифрован геном** червя, вызывающего слоновью ...
Расшифрован геном червя, вызывающего слоновью болезнь. Как сообщает Reuters, международная группа ученых под руководством Элоди Гедин (Elodie

Internet

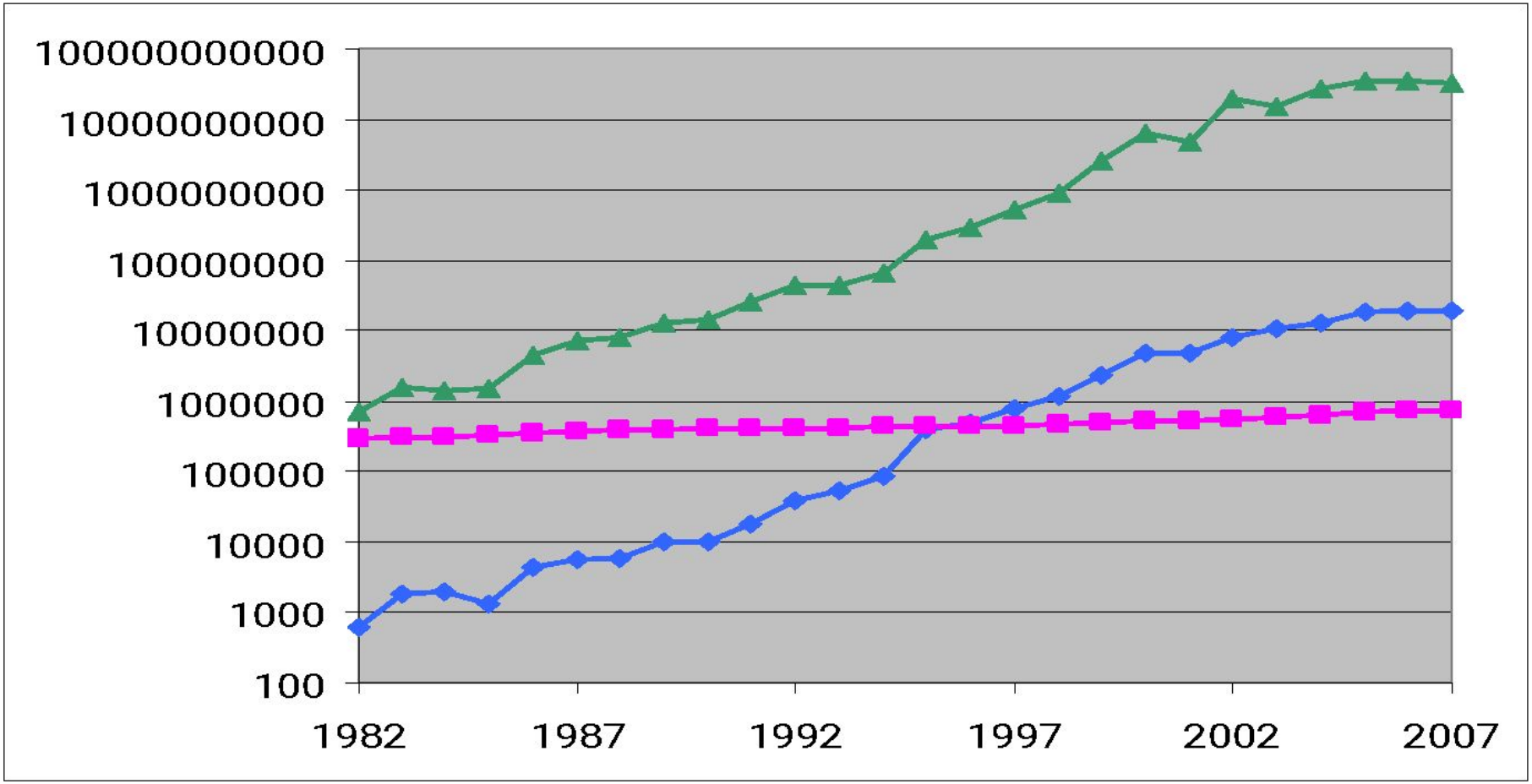
Перехватить зашифрованное сообщение – еще не значит его понять

Геном бактерии: несколько миллионов нуклеотидов

От 600 до 9 тысяч генов (примерно 90% генома кодирует белки)

На этом слайде – 0,1% генома *Escherichia coli*

Экспоненциальный рост объема данных



красный – статьи (PubMed)
синий – последовательности (GenBank)
зеленый – объем в нуклеотидах (GenBank)

из 18 миллионов ссылок, ~675 тыс. отвечают на "bioinformat* OR comput*"

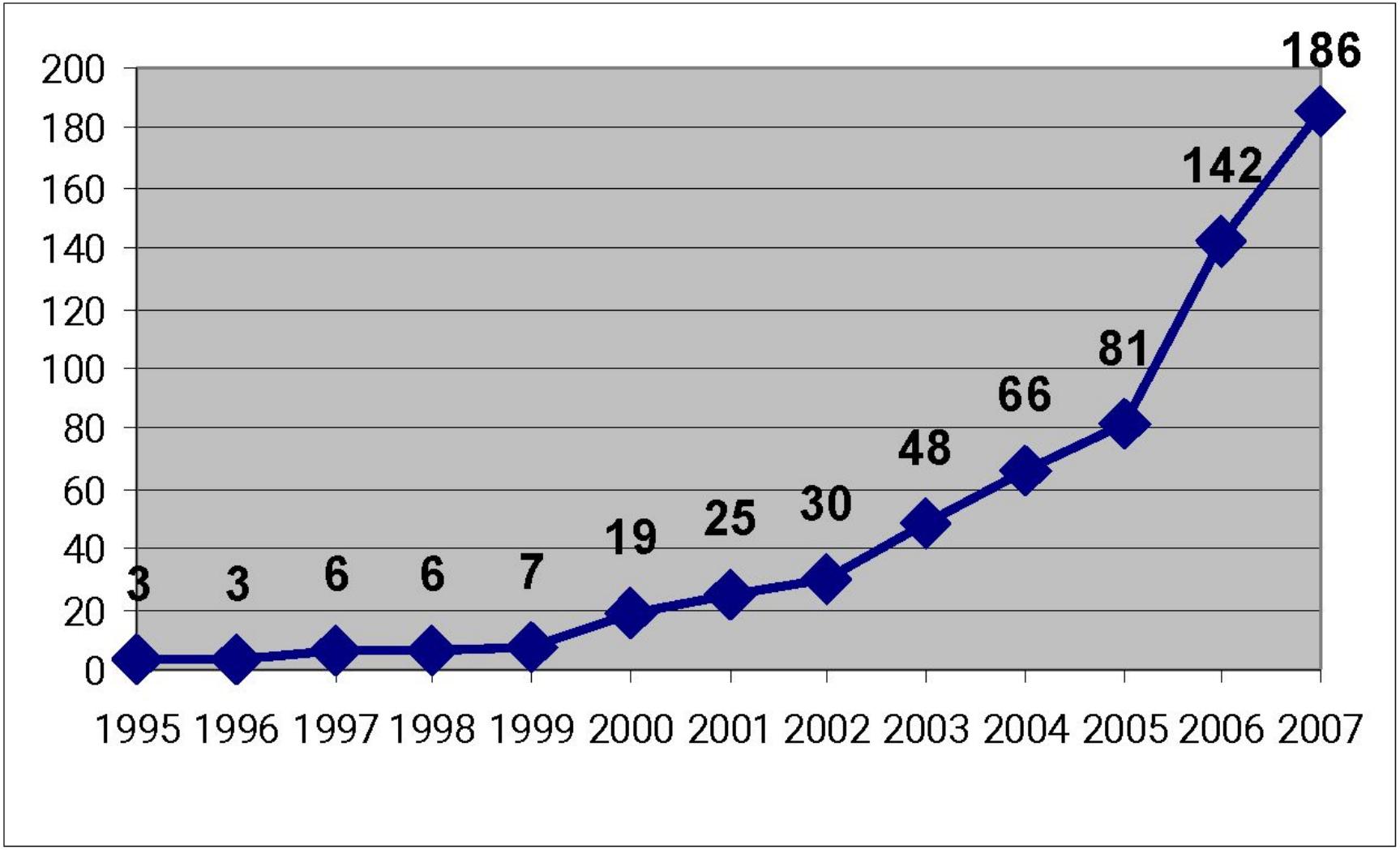
16 тыс. "bioinformat*"
65 тыс. "bioinformat* OR computat*"

Цель (локальная): аннотировать гены / белки in silico

- Что?
  - (биохимическая) функция
  - клеточная роль
- Когда?
  - Регуляция
    - Экспрессия
    - Время жизни (мРНК, белка)
- Где?
  - Локализация
    - Внутри/снаружи
    - Органеллы и компартменты
- Как?
  - Механизм
    - Специфичность, регуляция

Наиболее важные предсказания затем проверяются экспериментально

622 полных генома (прокариот)



Цель (глобальная)

Предсказать свойства организма путем (компьютерного) анализа его генома

(возможно, с использованием дополнительной информации: эпигенетика, белок-белковые взаимодействия и т.п.)

сейчас: метаболическая реконструкция, транспортные системы, ответ на стресс и т.д.

“Понять” эволюцию геномов/организмов

Сравнительная геномика

Базовые постулаты:

- Сходство => гомология (общность происхождения)
- Гомология => сходная функция

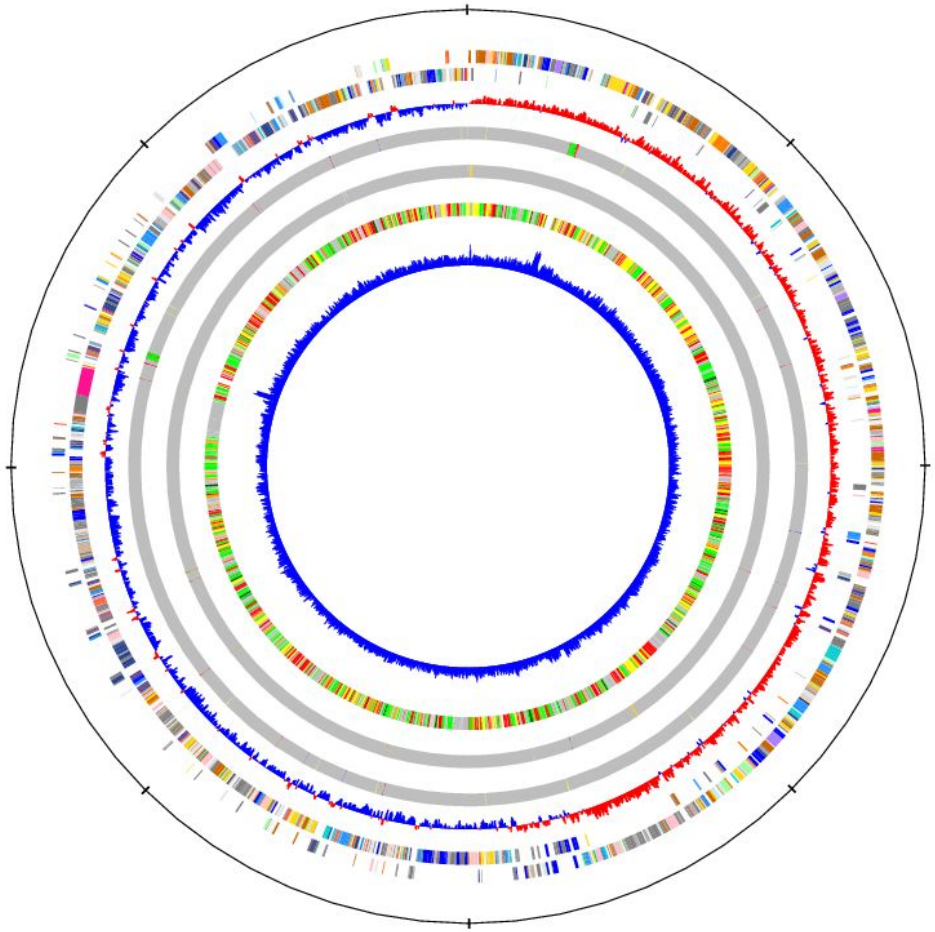
- Консервативно то, что важно
  - структурные и функциональные мотивы в белках
  - регуляторные сайты в ДНК

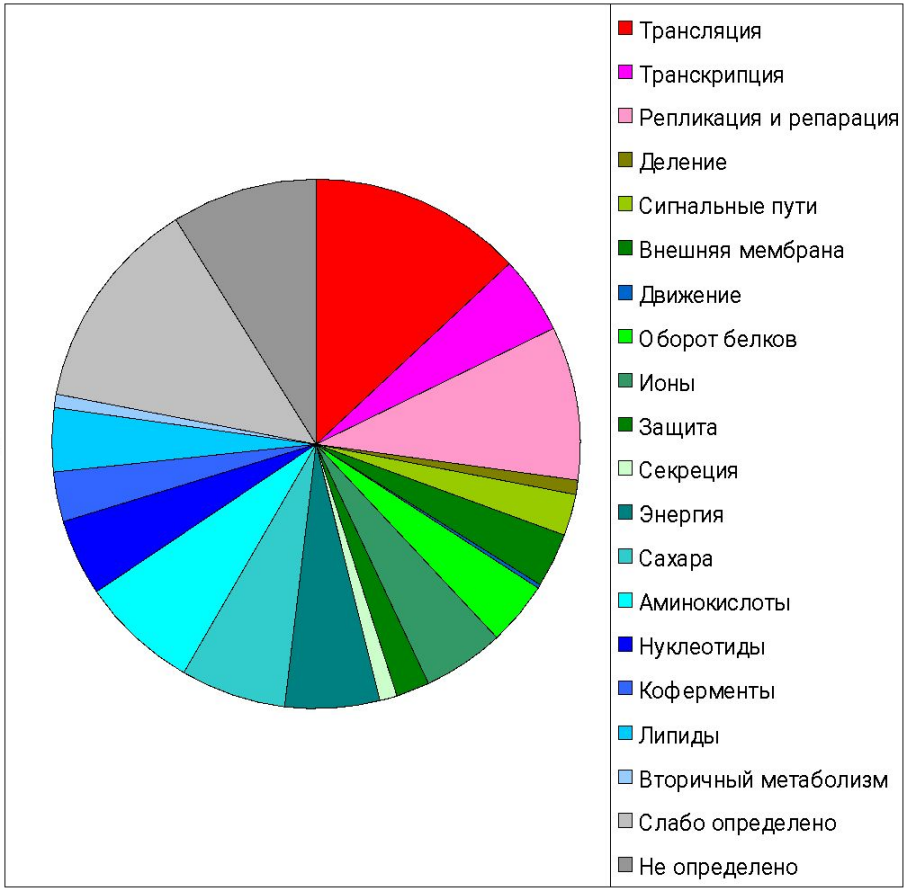
Математические и алгоритмические проблемы

- Формализация понятия сходства
  - теория вероятностей: значимость наблюдаемого уровня сходства
  - вычислительная геометрия: сходство структур
- Алгоритмические проблемы:
  - быстрый поиск сходных последовательностей
    - большой объем базы данных (растет быстрее, чем быстродействие процессоров)
  - множественное выравнивание
    - оптимальный алгоритм имеет полиномиальное время работы, но степень равна числу последовательностей
  - построение эволюционных деревьев
    - баланс между биологическими соображениями и вычислительными возможностями
- Идентификация функциональных и регуляторных мотивов в последовательностях
  - теория предсказания образов: нейронные сети, поддерживающие вектора и т.п.

Первый российский бактериальный геном - *Acholeplasma laidlawii*

Секвенирование: ИФХМ МЗ РФ, аннотация: ИППИ РАН

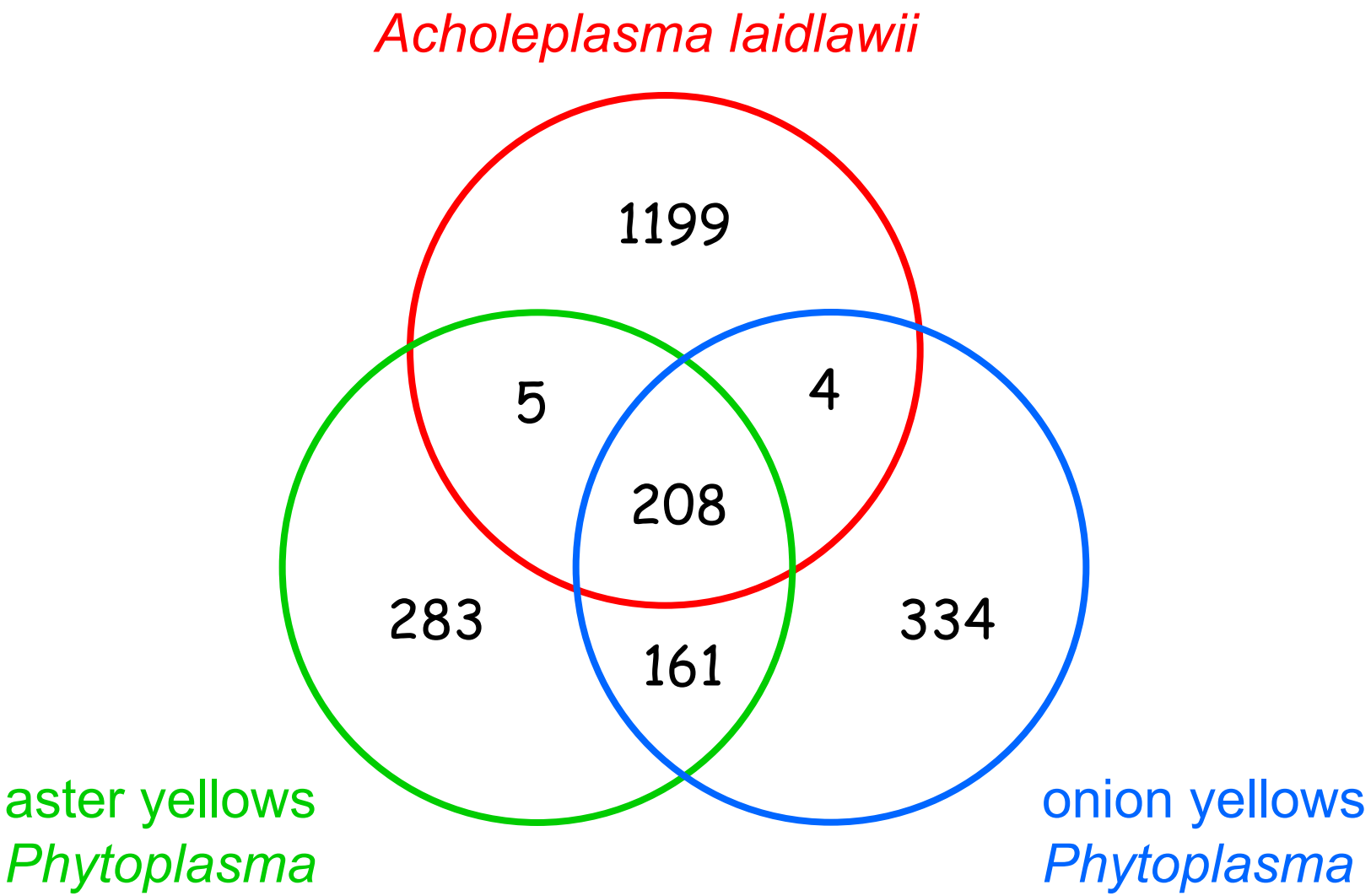




~1,5 Mb; ~1400 генов.

Установлены функции ~80% генов; проведена метаболическая реконструкция

Сравнение с родственными геномами



Сравнительная геномика - 2

- Не обязательно последовательности:
  - структура белка и РНК
  - расположение генов на хромосоме (ко-локализация)
  - ко-регуляция и ко-экспрессия генов
  - филогенетические образцы (совместное появление в геномах)
- Предсказав структурные особенности белка, можно определить его функциональный класс
- Изучение геномного контекста позволяет отнести ген (белок) к функциональной подсистеме

- Задача: формализация этих подходов
  - Полногеномные сравнения
  - Статистическая значимость
  - Распознавание образов и экспертные системы

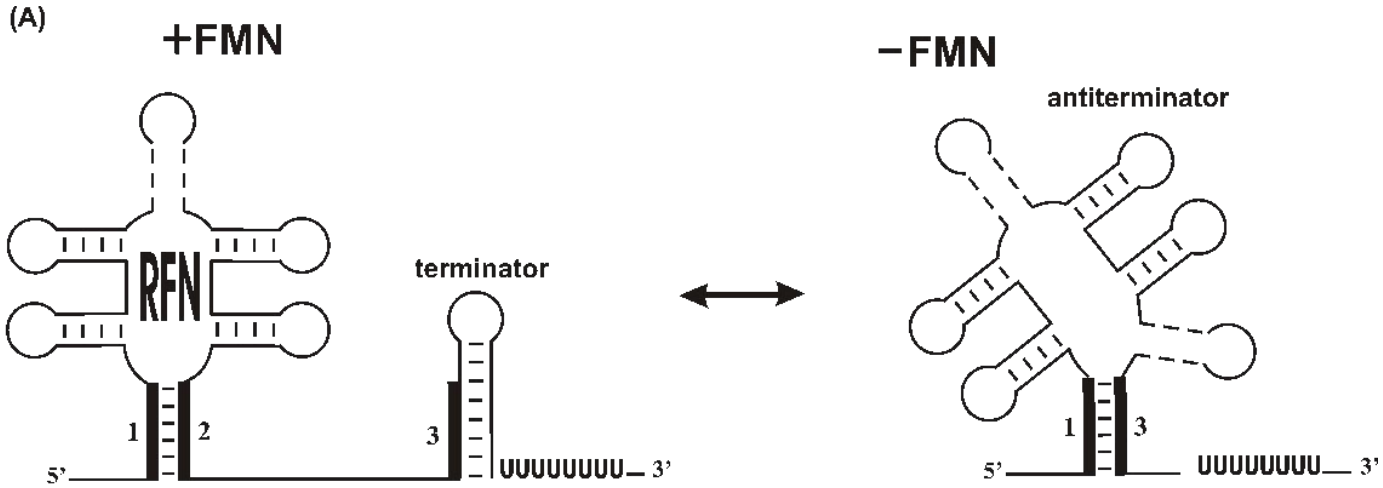
РНК-переключатели: от биоинформатического анализа к экспериментальной проверке

- Новый универсальный механизм регуляции экспрессии генов за счет формирования альтернативных структур РНК и прямого связывания малых молекул

- Структуры и механизм предсказаны биоинформатически и затем подтверждены экспериментально

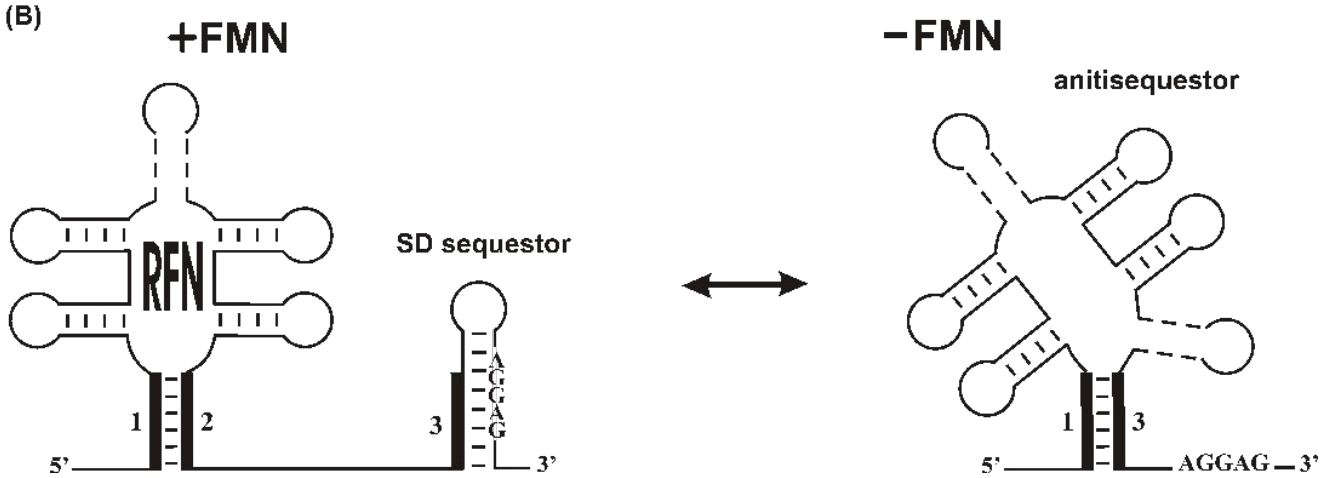
Регуляция экспрессии генов за счет формирования альтернативных структур РНК

- **Transcription attenuation**



Структура предсказана на основе сравнительного анализа выравненных последовательностей

- **Translation attenuation**

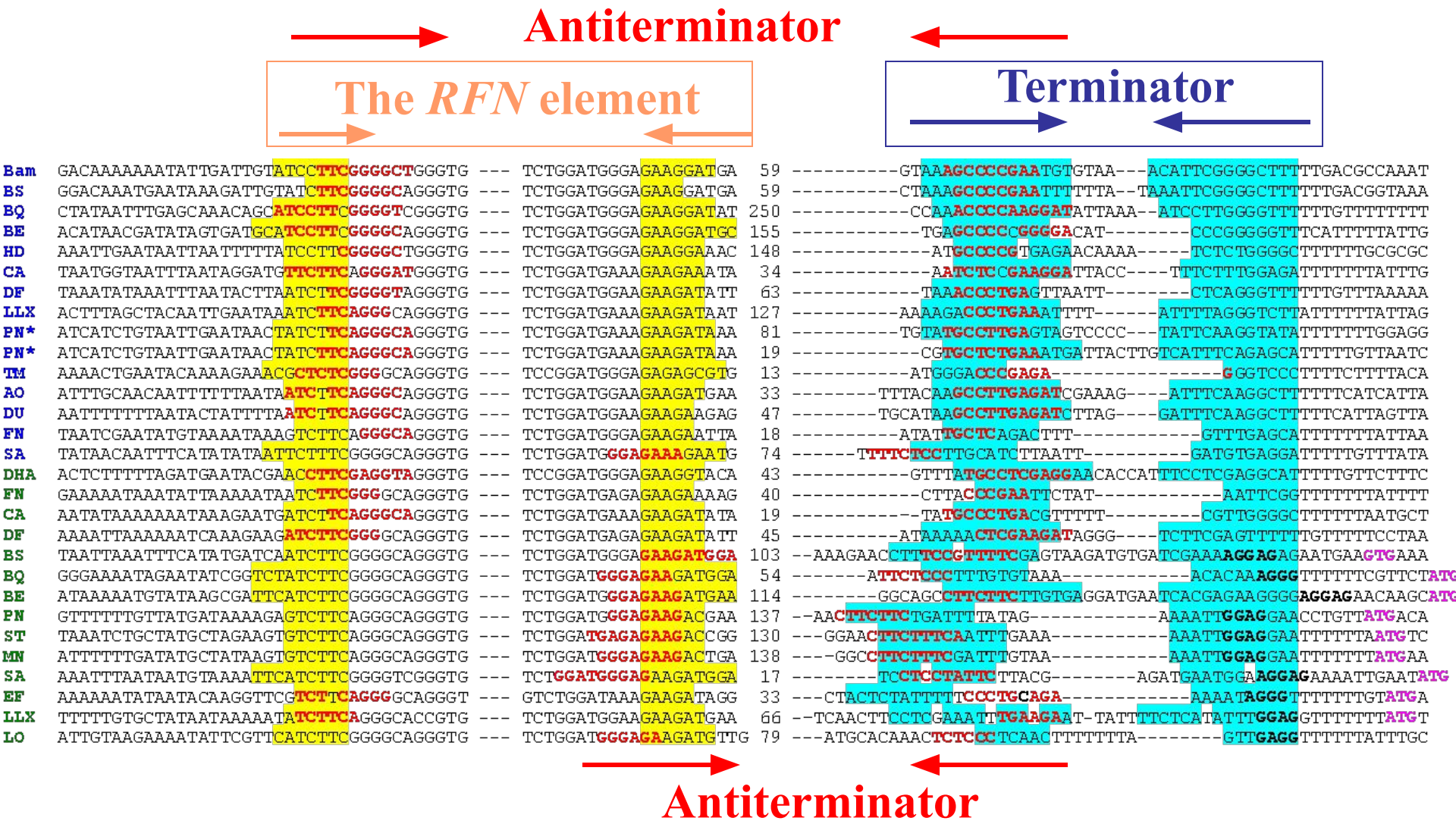


Механизм предсказан на основе литературных данных и анализа структурных особенностей

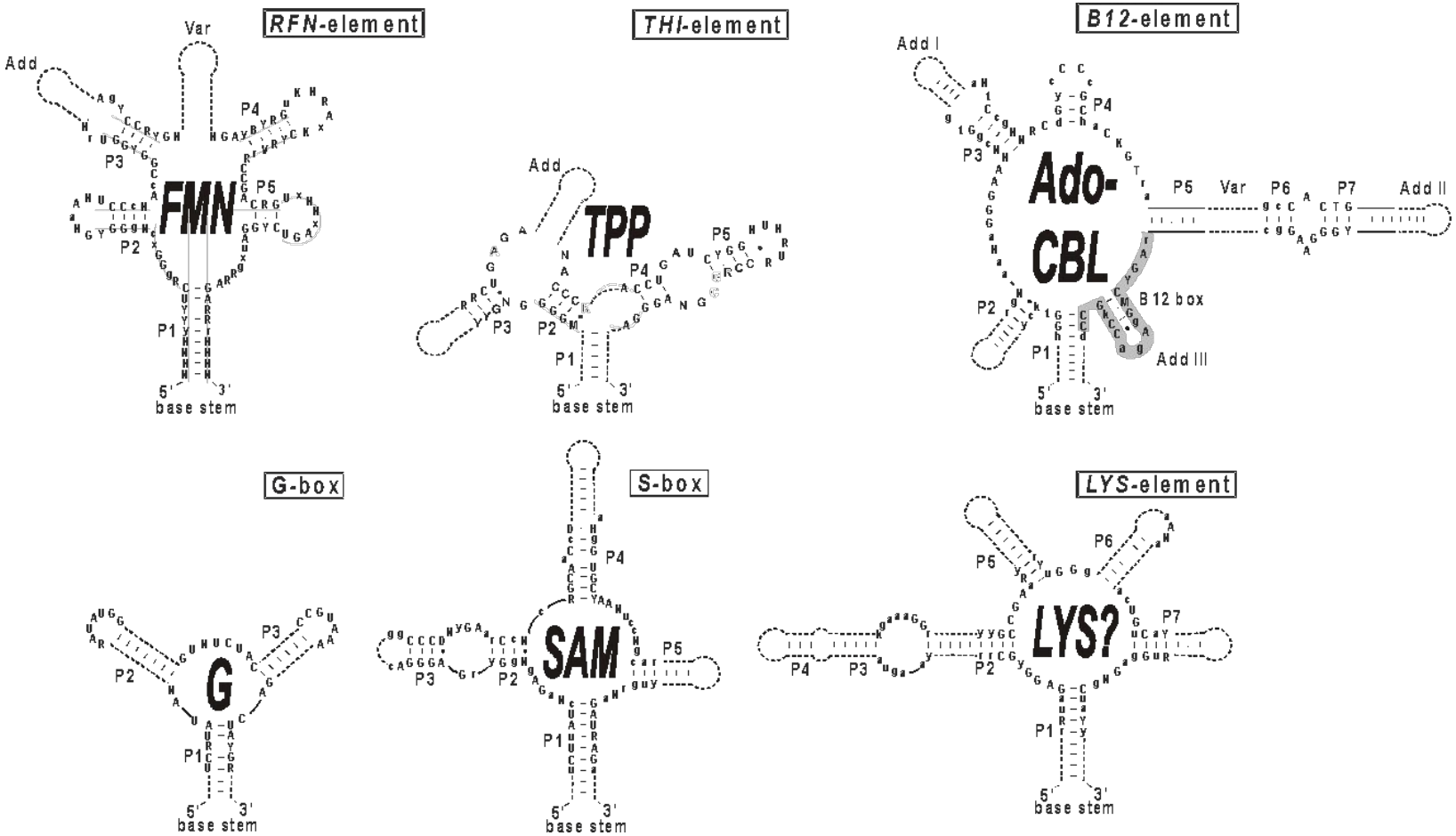
5'-нетранслируемые области бактериальных генов биосинтеза рибофлавина

```
      1                2          2'     3     Add.      3'      Variable        4              4'     5          5'              1'
      ==========>      ==>      <==     ==>   -><-    <===      ->     <-      ====>         <====    ==>      <==      <=========
BS    TTGTATCTTCGGGG-CAGGGTGGAAATCCCGACCGGCGGT  21 AGCCCGTGAC--  8  4  8 -----TGGATTCAGTTTAA-GCTGAAGCCGACAGTGAA-AGTCTGGAT-GGGAGAAGGATGAT
BQ    AGCATCCTTCGGGG-TCGGGTGAAATTCCCAACCGGCGGT  19 AGTCCGTGAC--  8  5  8 -----TGGATCTAGTGAAACTCTAGGGGCCGACAGTTAA-AGTCTGGAT-GGGAGAAGGATATG
BE    TGCATCCTTCGGGG-CAGGGTGAAATTCCCGACCGGCGGT  20 AGCCCGCGA---  3  4  3 -----AGGATCCGGTGCGATTCCGGAGCCGACAGT-AT-AGTCTGGAT-GGGAGAAGGATGCC
HD    TTTATCCTTCGGGG-CTGGGTGGAAATCCCGACCGGCGGT  19 AGTCCGTGAC-- 10  4 10 -----TGGACCTGGTGAAAATCCGGGGACCGACAGTGAA-AGTCTGGAT-GGGAGAAGGAAACG
Bam   TGTATCCTTCGGGG-CTGGGTGAAAATCCCGACCGGCGGT  23 AGCCCGTGAC--  8  4  8 -----TGGATTCAGTGAAAAGCTGAAGCCGACAGTGAA-AGTCTGGAT-GGGAGAAGGATGAG
CA    GATGTTCTTCAGGG-ATGGGTGAAATTCCCAATCGGCGGT   2 AGCCCGCAA---  3  4  3 ------AGATCCGGTTAAACTCCGGGGCCGACAGTTAA-AGTCTGGAT-GAAAGAAGAAATAG
DF    CTTAATCTTCGGGG-TAGGGTGAAATTCCCAATCGGCGGT   2 AGCCCGCG----  7  6  7 --------ATTTGGTTAAATTCCAAAGCCGACAGT-AA-AGTCTGGAT-GGAAGAAGATATTT
SA    TAATTCTTTCGGGG-CAGGGTGAAATTCCCAACCGGCAGT   6 AGCCTGCGAC-- 11  3 11 -----CTGATCTAGTGAGATTCTAGAGCCGACAGTTAA-AGTCTGGAT-GGGAGAAAGAATGT
LLX   ATAAATCTTCAGGG-CAGGGTGTAATTCCCTACCGGCGGT   2 AGCCCGCGA---  4  4  4 -----ATGATTCGGTGAAACTCCGAGGCCGACAGT-AT-AGTCTGGAT-GAAAGAAGATAATA
PN    AACTATCTTCAGGG-CAGGGTGAAATTCCCTACCGGTGGT   2 AGCCCACGA---  3  4  3 -----ATGATTTGGTGAAATTCCAAAGCCGACAGT-AT-AGTCTGGAT-GAAAGAAGATAAAA
TM    AAACGCTCTCGGGG-CAGGGTGGAATTCCCGACCGGCGGT   3 AGCCCGCGAG--  5  4  5 -----TTGACCCGGTGGAATTCCGGGGCCGACGGTGAA-AGTCCGGAT-GGGAGAGAGCGTGA
DR    GACCTCTTTCGGGG-CGGGGCGAAATTCCCCACCGGCGGT  15 AGCCCGCGAA--  8 12  9 -----CCGATGCCGCGCAACTCGGCAGCCGACGGTCAC-AGTCCGGAC-GAAAGAAGGAGGAG
TQ    CACCTCCTTCGGGG-CGGGGTGGAAGTCCCCACCGGCGGT   3 AGCCCGCGAA--  5  4  5 -----CCGACCCGGTGGAATTCCGGGGCCGACGGTGAA-AGTCCGGAT-GGGAGAAGGAGGGC
AO    AATAATCTTCAGGG-CAGGGTGAAATTCCCGATCGGCGGT   2 AGTCCGCGA---  7  7  7 -----AGGAACCGGTGAGATTCCGGTACCGACAGT-AT-AGTCTGGAT-GGGAGAAGATGAAA
DU    TTTAATCTTCAGGG-CAGGGTGAAATTCCCGATCGGTGGT   2 AGTCCGCGA--- 13  4 12 -----AGGAACTAGTGAAATTCTAGTACCGACAGT-AT-AGTCTGGAT-GGAAGAAGAGCAGA
CAU   GAAGACCTTCGGGG-CAAGGTGAAATTCCTGATCGGCGGT  20 AGCCCGCGA---  3  4  3 -----AGGACCCGGTGTGATTCCGGGGCCGACGGT-AT-AGTCCGGAT-GGGAGAAGGTCGGC
EN    TAAAGTCTTCAGGG-CAGGGTGAAATTCCCGACCGGTGGT   2 AGTCCACG----  5  4  5 -------GATTTGGTGAAATTCCAAAACCGACAGT-AG-AGTCTGGAT-GGGAGAAGAATTAG
TFU   ACGCGTGCTCCGGG-GTCGGTGAAAGTCCGAACCGGCGGT   3 AGTCCGCGAC-- 8  5  8 -----TGGAACCGGTGAAACTCCGGTAACCGACGGTGAA-AGTCCGGAT-GGGAGGTAGTACGTG
SX    -AGCGCACTCCGGG-GTCGGTGAAAGTCCGAACCGGCGGT   3 AGTCCGCGAC-- 8  5  8 -----TTGACCAGGTGAAATTCCTGGACCGACGGTTAA-AGTCCGGAT-GGGAGGCAGTGCGCG
BU    GTGCGTCTTCAGGG-CGGGGTGAAATTCCCCACCGGCGGT  30 AGCCCGCGAGCG    137   GTCAGCAGATCTGGTGAGAAGCCAGAGCCGACGGTTAG-AGTCCGGAT-GGAAGAAGATGTGC
BPS   GTGCGTCTTCAGGG-CGGGGCGAAATTCCCCACCGGCGGT  21 AGCCCGCGAGCG  8  4  8 GTCAGCAGATCTGGTCCGATGCCAGAGCCGACGGTCAT-AGTCCGGAT-GAAAGAAGATGTGC
REU   TTACGTCTTCAGGG-CGGGGTGCAATTCCCCACCGGCGGT  31 AGCCCGCGAGCG  7  5  7 GTCAGCAGATCTGGTGAGAGGCCAGGGCCGACGGTTAA-AGTCCGGAT-GAAAGAAGATGGGC
RSO   GTACGTCTTCAGGG-CGGGGTGGAATTCCCCACCGGCGGT  21 AGCCCGCGAGCG 11  3 11 GTCAGCAGATCCGGTGAGATGCCGGGGCCGACGGTCAG-AGTCCGGAT-GGAAGAAGATGTGC
EC    GCTTATTCTCAGGG-CGGGGCGAAATTCCCCACCGGCGGT  17 AGCCCGCGAGCG  8  4  8 GACAGCAGATCCGGTGTAATTCCGGGGCCGACGGTTAG-AGTCCGGAT-GGGAGAGAGTAACG
TY    GCTTATTCTCAGGG-CGGGGCGAAATTCCCCACCGGCGGT  67 AGCCCGCGAGCG  8  3  8 GTCAGCAGATCCGGTGTAATTCCGGGGCCGACGGTTAA-AGTCCGGAT-GGGAGAGGGTAACG
KP    GCTTATTCTCAGGG-CGGGGCGAAATTCCCCACCGGCGGT  20 AGCCCGCGAGCG  8  4  8 GTCAGCAGATCCGGTGTAATTCCGGGGCCGACGGTTAA-AGTCCGGAT-GGGAGAGAGTAACG
HI    TCGCATTCTCAGGG-CAGGGTGAAATTCCCTACCGGTGGT   2 AGCCCACGAGCG 26  9 30 GTCAGCAGATTTGGTGAAATCCAAAGCCGACAGT-AA-AGTCTGGAT-GAAAGAGAATAAAA
VK    GCGCATTCTCAGGG-CAGGGTGAAATTCCCTACCGGTGGT  14 AGCCCACGAGCG 11  9 11 GTCAGCAGATTTGGTGAGAATCCAAAGCCGACAGT-AT-AGTCTGGAT-GAAAGAGAATAAGC
VC    CAATATTCTCAGGG-CGGGGCGAAATTCCCCACCGGTGGT  13 AGCCCACGAGCG  5  4  5 GTCAGCAGATCTGGTGAGAAGCCAGGGCCGACGGTTAC-AGTCCGGAT-GAGAGAGAATGACA
YP    GCTTATTCTCAGGG-CGGGGTGAAAGTCCCCACCGGCGGT  40 AGCCCGCGAGCG 16  6 16 GTCAGCAGACCCGGTGTAATTCCGGGGCCGACGGTTAT-AGTCCGGAT-GGGAGAGAGTAACG
AB    GCGCATTCTCAGGG-CAGGGTGAAAGTCCCTACCGGTGGT  25 AGCCCACGAGCG 16  4 27 GTCAGCAGATTTGGTGCGAATCCAAAGCCGACAGTGAC-AGTCTGGAT-GAAAGAGAATAAAA
BP    GTACGTCTTCAGGG-CGGGGTGCAATTCCCCACCGGCGGT  18 AGCCCGCGAGCG 10  4 10 GTCAGCAGACCTGGTGAGATGCCAGGGCCGACGGTCAT-AGTCCGGAT-GAGAGAAGATGTGC
AC    ACATCGCTTCAGGG-CGGGGCGTAATTCCCCACCGGCGGT  16 AGCCCGCGAGCA 10  3 11 ---CGCAGATCTGGTGTAAATCCAGAGCCGACGGT-AT-AGTCCGGAT-GAAAGAAGACGACG
Spu   AACAATTCTCAGGG-CGGGGTGAAACTCCCCACCGGCGGT  34 AGCCCGCGAGCG  6  6  6 GTCAGCAGATCTGGTG 52 TCCAGAGCCGACGGT 31 AGTCCGGAT-GGAAGAGAATGTAA
PP    GTCGGTCTTCAGGG-CGGGGTGTAAGTCCCCACCGGCGGT  13 AGCCCGCGAGCG  7  3  7 GTCAGCAGATCTGGTGCAACTCCGAGAGCCGACGGTCAT-AGTCCGGAT-GAAAGAAGGCGTCA
AU    GGTTGTTCTCAGGG-CGGGGTGCAATTCCCCACCGGCGGT  17 AGCCCGCGAGCG  7  9  7 GTCAGCAGATCCGGTGAGAGGCCGGAGCCGACGGT-AT-AGTCCGGAT-GGAAGAGGACAAGG
PU    AAACGTTCTCAGGG-CGGGGTGCAATTCCCCACCGGCGGT  19 AGCCCGCGAGCG 19  4 18 GTCAGCAGACCCGGTGTGATTCCGGGGCCGACGGTCAC-AGTCCGGATGAAGAGAGAACGGGA
PY    TAACGTTCTCAGGG-CGGGGTGCAACTCCCCACCGGCGGT  19 AGCCCGCGAGCG 15  4 16 GTCAGCAGACCCGGTGTGATTCCGGGGCCGACGGTCAT-AGTCCGGATGAAGAGAGAGCGGGA
PA    TAACGTTCTCAGGG-CGGGGTGAAAGTCCCCACCGGCGGT  19 AGCCCGCGAGCG 14  4 13 GTCAGCAGACCCGGTGCGATTCCGGGGCCGACGGTCAT-AGTCCGGATAAAGAGAGAACGGGA
MLO   TAAAGTTCTCAGGG-CGGGGTGAAAGTCCCCACCGGCGGT  16 AGCCCGCGAGCG  8  5  8 GTCAGCAGATCCGGTGTGATTCCGGAGCCGACGGTTAG-AGTCCGGAT-GAAAGAGGACGAAA
SM    AAGCGTTCTCAGGG-CGGGGTGAAATTCCCCACCGGCGGT  34 AGCCCGCGAGCG  8  3  8 GTCAGCAGATCCGGTCGAATTCCGGAGCCGACGGTTAT-AGTCCGGAT-GGAAGAGAGCAAGC
BME   GCTTGTTCTCGGGG-CGGGGTGAAACTCCCCACCGGCGGT  17 AGCCCGCGAGCG 10 15 10 GTCAGCAGATCCGGTGAGATGCCGGAGCCGACGGTTAA-AGTCCGGAT-GGAAGAGAGCGAAT
BS    ATCAATCTTCGGGG-CAGGGTGAAATTCCCTACCGGCGGT  18 AGCCCGCGA---  5  4  5 -----AGGATTCGGTGAGATTCCGGAGCCGACAGT-AC-AGTCTGGAT-GGGAGAAGATGGAG
BQ    GTCTATCTTCGGGG-CAGGGTGAAAATCCCGACCGGCGGT  27 AGCCCGCGA---  3  5  3 -----AGGATTTGGTGTGATTCCAAAGCCGACAGT-AT-AGTCTGGAT-GGGAGAAGATGGAG
BE    ATTCATCTTCGGGG-CAGGGTGAAATTCCCGACCGGCGGT  20 AGCCCGCGA---  3  4  3 -----AGGATCCGGTGCGAGTCCGGAGCCGACAGT-AT-AGTCTGGAT-GGGAGAAGATGAAG
CA    AATGATCTTCAGGG-CAGGGTGAAATTCCCTACCGGCGGT   2 AGCCCGCGAG--  3  4  3 ----TATGATCCGGTTTGATTCCGGAGCCGACAGT-AA-AGTCTGGAT-GAAAGAAGATATAT
DF    GAAGATCTTCGGGG-CAGGGTGAAATTCCCTACCGGCGGT   2 AGCCCGCG----  6  4  6 -------GATTTGGTGAGATTCCAAAGCCGACAGT-AA-AGTCTGGAT-GAGAGAAGATATTT
EF    GTTCGTCTTCAGGGGCAGGGTGTAATTCCCGACCGTGGT   3 AGTCCACGAC-- 5  3  5 ----ATTGAATTGGTGTAATTCCAATACCGACAGT-AT-AGTCTGGAT--AAAGAAGATAGGG
LLX   AAATATCTTCAGGG-CACCGTGTAATTCGGGACCGGCGGT  21 ACTCCGCGAT-- 4  4  4 -----TTGAAGCAGTGAGAATCTGCTAGCGACAGT-AA-AGTCTGGAT-GGAAGAAGATGAAC
LO    GTTCATCTTCGGGG-CAGGGTGCAATTCCCGACCGGTGGT   3 AGTCCACGAT-- 3 10  3 ----TTGACTCTGGTGTAATTCCAGGACCGACAGT-AT-AGTCTGGAT-GGGAGAAGATGTTG
PN    AAGAGTCTTCAGGG-CAGGGTGAAATTCCCGACCGGCGGT 125 AGTCCGTG---- 3  4  3 -------GATGTGGTGAGATTCCACAACCGACAGT-AT-AGTCTGGAT-GGGAGAAGACGAAA
ST    AAGTGTCTTCAGGG-CAGGGTGTGATTCCCGACCGGCGGT  14 AGTCCGCG---- 3  4  3 -------GATGTGGTGTAACTCCACAACCGACAGT-AT-AGTCTGGAT-GAGAGAAGACCGGG
MN    AAGTGTCTTCAGGG-CAGGGTGAGATTCCCGACCGGCGGT 104 AGTCCGCG---- 3  4  3 -------GATGTGGTGAAATTCCACAACCGACAGT-AA-AGTCTGGAT-GGGAGAAGACTGAG
SA    ATTCATCTTCGGGG-TCGGGTGTAATTCCCAACCGGCAGT   6 AGCCTGCGAC-- 11  3 11 -----CTGATCTAGTGAGATTCTAGAGCCGACAGT-AT-AGTCTGGAT-GGGAGAAGATGGAG
AMI   TCACAGTTTCAGGG-CGGGGTGCAATTCCCCACTGGCGGT  14 AGCCCGCGC--- 5  5  5 ------TGATCTGGTGCAAATCCAGAGCCAACGGT-AT-AGTCCGGAT-GGAAGAAACGGAGC
DHA   ACGAACCTTCGAGG-TAGGGTGAAATTCCCGACCGGCGGT  20 AGCCCGCAAC-- 11  4 11 --CGACTGACTTGGTGAGACTCCAAGGCCGACGGT-AT-AGTCCGGAT-GGGAGAAGGTACAA
EN    AATAATCTTCGGGG-CAGGGTGAAATTCCCGACCGGTGGT   2 AGTCCACG---- 4  6  4 -------GATTTGGTGAAATTCCAAAACCGACAGT-AG-AGTCTGGAT-GAGAGAAGAAAAGA
GLU   ---TGTTCTCAGGG-CGGGGCGAAATTCCCCACCGGCGGT  28 AGCCCGCGAGCG 10  4 10 GTCAGCAGATCCGGTTAAATTCCGGAGCCGACGGTCAT-AGTCCGGAT-GCAAGAGAACC---
```

Аттенюация транскрипции



Другие РНК-переключатели, найденные методами сравнительной геномики



- Есть во всех трех основных царствах (бактерии, археи, эукариоты)
- Древнейшие регуляторные элементы: реликт «РНКового мира»?

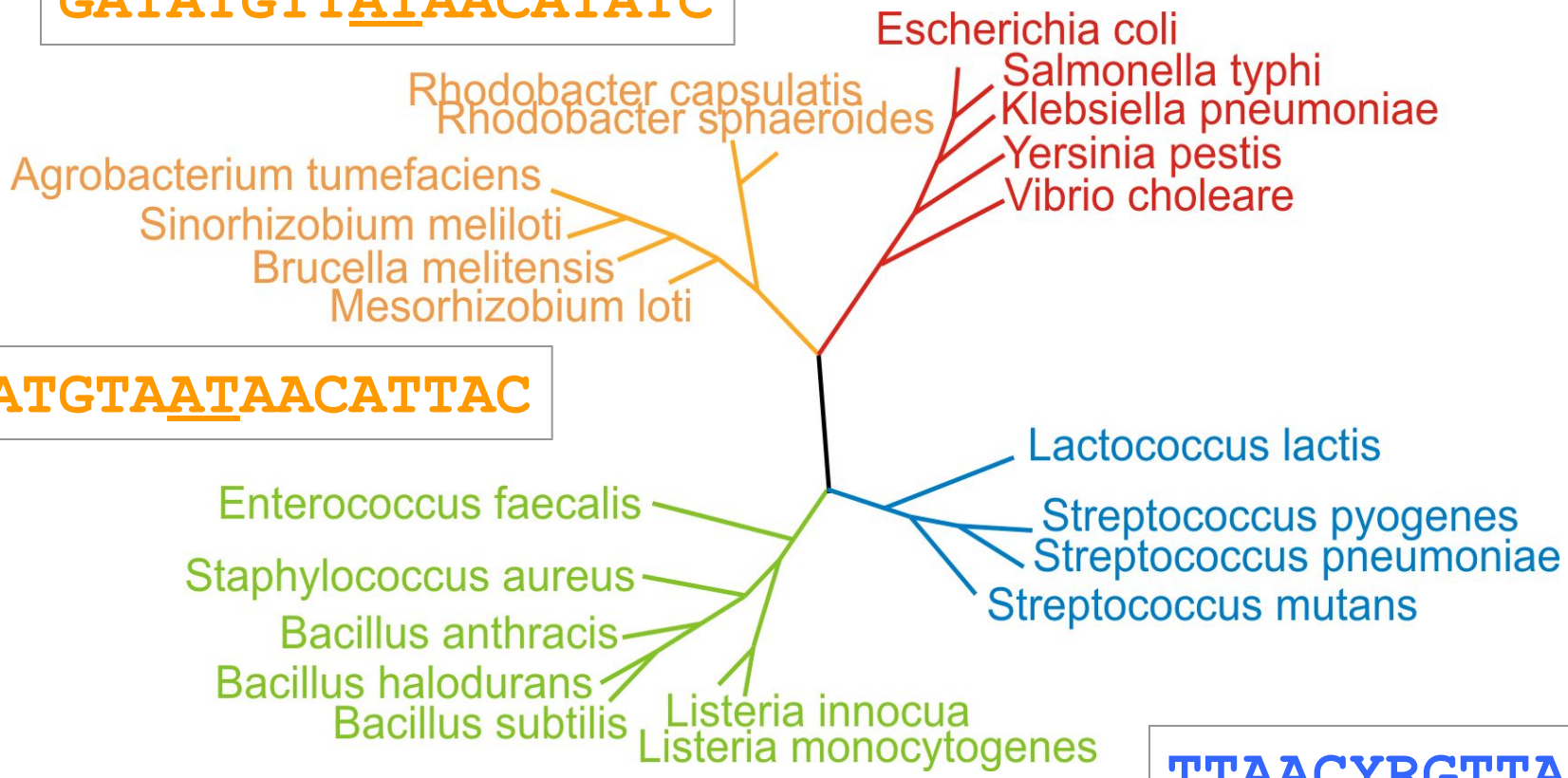
Регуляторы гомеостаза цинка

nZUR-α

nZUR-γ

GATATGTTATAACATATC

GAAATGTTATANTATAACATTTC



GTAATGTAATAACATTAC

TTAACYRGTTAA

pZUR

TAAATCGTAATNATTACGATTTA

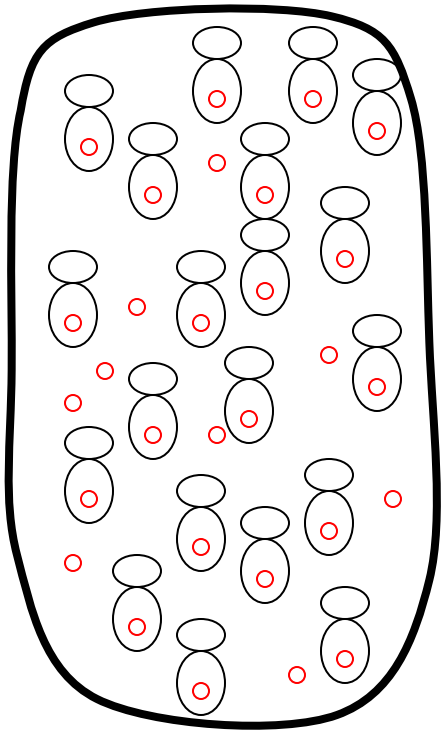
AdcR

Регуляция гомологов рибосомальных белков

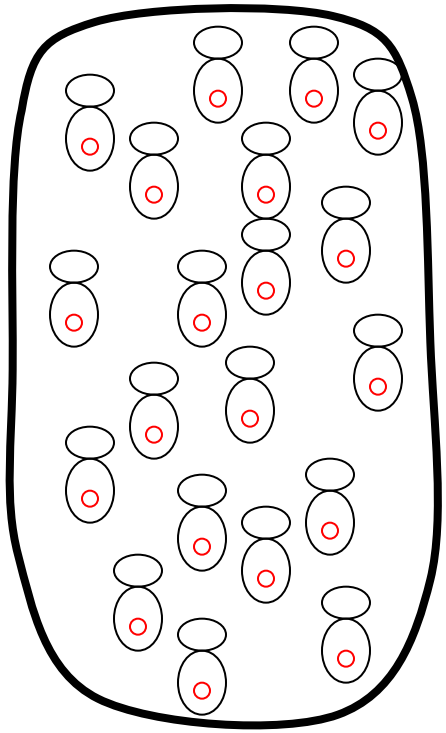
		L36	L33	L31	S14
nZUR	*E. coli, S.typhi*	(–)	–	(–) +	–
	K. pneumoniae	(–)	–	(–) –	–
	Y. pestis,V. cholerae	(–) ×	–	(–) +	–
pZUR	*B subtilis*	(–)	(–) + –	(–) +	(–) +
	S. aureus	(–)	(–) – –	–	(–) +
	Listeria spp.	(–)	(–) –	–	(–) +
	E. faecalis	(–)	(–) × – –	–	(–) + –
AdcR	*S. pne., S. mutans*	(–)	(–) – –	–	(–)
	S. pyo., L. lactis	(–)	(–) – –	–	(–) +

Плохой сценарий

достаточно цинка

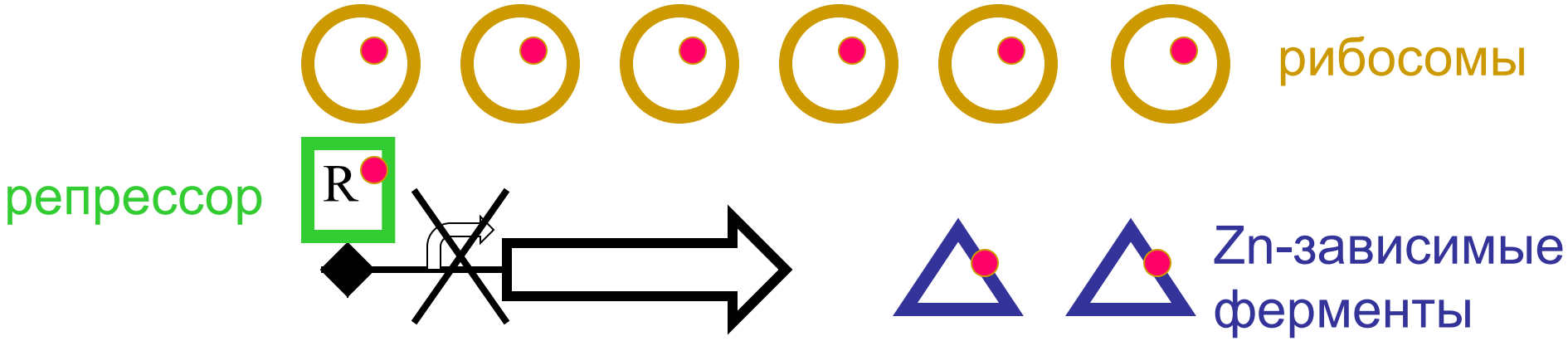


недостаточно цинка: весь цинк использован рибосомами, не хватает цинка для ферментов

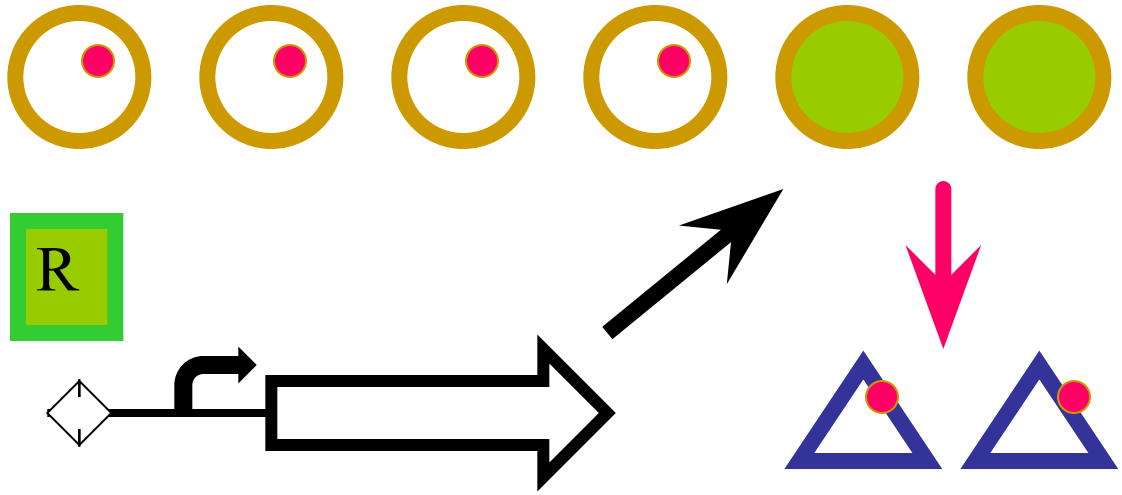


Регуляторный механизм

Достаточно Zn



Голодание по Zn



Предсказание …

(Proc Natl Acad Sci U S A. 2003 Aug 19;100(17):9912-7.)

Comparative genomics of bacterial zinc regulons: Enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins

Ekaterina M. Panina*†, Andrey A. Mironov*‡, and Mikhail S. Gelfand*‡§

… и подтверждение

(Mol Microbiol. 2004 Apr;52(1):273-83.)

Zinc is a key factor in controlling alternation of two types of L31 protein in the *Bacillus subtilis* ribosome

Hideaki Nanamiya,[1†] Genki Akanuma,[1†]
Yousuke Natori,[1] Rikinori Murayama,[1] Saori Kosono,[2]
Toshiaki Kudo,[2] Kazuo Kobayashi,[3]
Naotake Ogasawara,[3] Seung-Moon Park,[4] Kozo Ochi[4]
and Fujio Kawamura[1*]

(+ еще пять статей в последующие годы):
другие гены, другие бактерии

Сводка подтвердившихся предсказаний

- Регуляторы
  - РНК-переключатели
    - витамины: рибофлавин, тиамин, кобаламин
    - аминокислоты: лизин, метионин
  - Факторы транскрипции
    - NrdR: рибонуклеотид-редуктазы
    - MtaR, CmbR: метионин и цистеин
    - NiaR, NrtR: метаболизм NAD
    - NsrR, NnrA: нитрозативный стресс
- Регуляторные взаимодействия
  - регуляторные мотивы в ДНК (>10)
  - отдельные сайты связывания (>20)
- Ферменты
  - ThiN и TenA (биосинтез тиамина)
  - CobX, CobZ (биосинтез кобаламина)
  - FadE (синтез жирных кислот)
  - AbnA, Xca (катаболизм арабинозы)
  - NagK, NagBII (катаболизм N-ацетилглюкозамина)
- Транспортеры
  - витамины и кофакторы
    - YpaA и RibM: рибофлавин
    - BioMNY: биотин
    - ThiXYZ: тиамин
    - NiaP: ниацин
    - Vng1369-71: корриноиды
  - сахара и полисахариды
    - OgtABCD: продукты деградации пектина
    - NagP: N-ацетилглюкозамин
  - аминокислоты
    - MetD: метионин
    - SteT: треонин
  - ионы металлов
    - CbiMNQO, HoxN: кобальт
    - NikMNQO: никель
  - нуклеотиды:
    - YicE: ксантин

Не только тексты

Другие типы массовых экспериментов:

- Транскриптомика
  - «выстилающие массивы»: полная карта транскриптов
  - уровень экспрессии и время жизни мРНК
  - ДНК-белковые взаимодействия
- Протеомика
  - концентрации белков
  - белок-белковые взаимодействия, белковые комплексы
  - структуры белков
- Эпигенетика
  - метилирование ДНК
  - положение и модификации нуклеосом
- Генетика
  - летальность мутаций
  - фенотипы
  - синтетические летали

«Неприкладная» биоинформатика

- Молекулярная эволюция
  - филогения генов
  - таксономия организмов
  - горизонтальные переносы и т.п.
  - положительный и отрицательный отбор
    - что сделало нас людьми?
    - лекарственная устойчивость
  - эволюция геномов
- Системная биология
  - строение геномов
  - сети взаимодействий
    - белок-белковые
    - регуляция транскрипции
    - сигнальные пути

Перспективы

- Индивидуальные геномы
  - персональные человеческие геномы
  - геномы штаммов
- Метагеномы
  - некультивируемые бактерии
  - метагеномы экологических ниш
    - бактериальный метагеном человека
- Другие виды данных
  - интеграция
- Моделирование (пока рано)
  - построение полных карт

Биоинформатика в России

- Сильные устойчивые школы
  - Москва+Пущино: содружество лабораторий в разных учреждениях
  - Новосибирск: вертикальная структура
- Высокий уровень
  - конкурентоспособный в мире
- Интеграция с образованием
  - Факультет биоинженерии и биоинформатики МГУ
  - Кафедра информационной биологии ФЕН НГУ
- Перспективная область:
  - относительно дешево
  - общедоступная информация – можно использовать чужие данные
  - связь с экспериментальными исследованиями, ведущимися на современном уровне
- Как развивать:
  - технические разработки (базы данных, пакеты программ») – контракты, наличие потребителя
  - интегрированные компоненты в медицинских и молекулярно-биологических проектах (крупные проекты *должны* иметь биоинформатическую поддержку)
  - самостоятельные исследовательские работы – гранты

Учебно-Научный Центр

Биоинформатика

- Российский фонд фундаментальных исследований
- **РАН, программа «Молекулярная и клеточная биология»**
- INTAS
- Howard Hughes Medical Institute