

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 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search

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

Google.ru/расшифрован геном человека

Веб Картинки Карты Новости Группы Почта Gmail

Войти

Google.ru/расшифрован геном человека

Поиск в Интернете Поиск страниц

Веб Результаты 1 - 10 из примерно 73 200 для **расшифрован геном человека**

[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> <div class="textLooks" >  
<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 -

genome deciphered - Поиск в Google - Microsoft Internet Explorer

View Favorites Tools Help

Back Forward Stop Home Search Favorites

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

Google.ru/human genome deciphered

Картинки Карты Новости Группы Почта Gmail

Войти

Google.ru/human genome deciphered

Поиск в Интернете Поиск страниц

результаты 1 - 10 из примерно 373 000 для **human genome deciphered**

[e.gov | 2000 Release: Two-Thirds Human Genome Sequenced](#) - [ Перевести эту страницу ]  
Is of **Human DNA Script Deciphered** by Human Genome Project. Public Consortium To Complete "Working Draft" of Human Genome. March 2000 ...  
genome.gov/10002080 - 15k -  
Сохранено в кэше - Похожие страницы

[Ma Mania — Deciphering the human genome: what does it mean](#)  
[ Перевести эту страницу ]  
... gave extraordinary attention to the recent announcement of the draft sequence of the **human genome** by the Human Genome Project. ...  
www.wersingenesi.org/docs2001/0309\_genome.asp - 31k -  
Сохранено в кэше - Похожие страницы

[NewsHour: Breaking the Code -- December 2000](#)  
[ Перевести эту страницу ]  
... genetic sequences would be comparatively easy to decipher with the aid of computers. If you printed out the whole **human genome** on pages-- and I don't know if ...  
www.newshour.com/bb/health/july-dec99/dna\_12-2.html - 27k -  
Сохранено в кэше - Похожие страницы

[The Genome Deciphered, May Herald Medical News](#)  
[ Перевести эту страницу ]  
... **Genome Deciphered**; May Herald Medical News ... Compared to the **human genome**, the **genome** is only about 1.5 percent different, ...  
nationalgeographic.com/news/2007/04/070412-macaque-genome.html - 27k -  
Сохранено в кэше - Похожие страницы

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

расшифрован геном - Поиск в Google - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites

Address http://www.google.ru/search?complete=1&hl=ru&rlz=1T4GFRC\_ruRU2... Go Links

Google.ru/расшифрован геном

Веб Картинки Карты Новости Группы Почта Gmail

еще

Войти

Google.ru/расшифрован геном

Поиск в Интернете Поиск страниц на русском

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

[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.shtm - 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. ...  
md.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 Guerin) ...

# Перехватить зашифрованное сообщение -

## еще не значит его понять

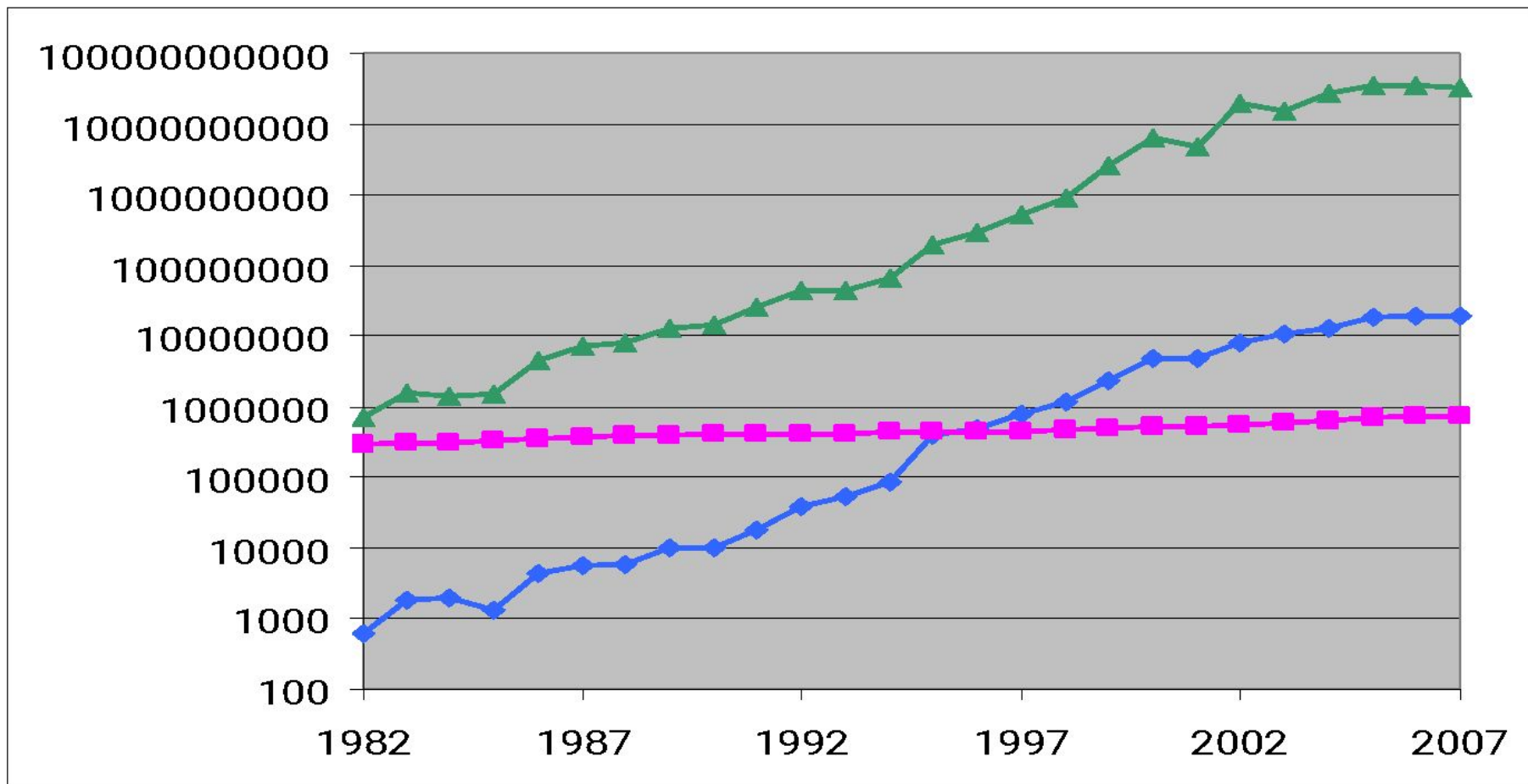
aacgggcaatatgtctctgtgtggattaaaaaagagtgctgatagcagctcttgaactggttacc tggcgtgagtaaataaaa ttttatggac ttaggtcactaaatactttaaccaat atagga  
tagcgcaagacagataaaaat tacagagtac acaac atcca tgaac cgc at tagca cccacc attac caccaccatc accat tacc acaggtaacggtgctggggctgacgggtacaggaaac acagaaa  
aagcccgccacctgacagtgccgggctttttttcgacc aaaggtaacagaggta acaac catgc gagggttgaa gttcggcgggtacatcagtgggcaaaa tgcagaacgt tttctgccgggtgccc gatatct  
ggaaagc aatgc caggc agggg caggt ggcgaccgtc ctctc tgc cc ccgcc aaaa tacca accatctggt agcga tgatt gaaaaaaccc attagcggcc aagatgcttt accca atatc agcga tgc  
cgaaacgtat ttttggccgaactctgac gggac tgc cc gccgc ccagc cggga tttcc gctgg cacaatgaaaaccttctgtc gacc aggaattgccc aaaa taaaa catgtcctgc atggc attag ttt  
gttggggcagtgcccggatagc atcaacgctgccgtgatttggcctggtggagaaaaatgtcga tgc cc attatggcggcgtgttgaagcgcgtggtcacaacgtt accgt tatcgatccgggtcga aaa  
actgcttgcagtggggc attac ctogaatctaccgttgatattgctgagtc acccgccgtat tggcggcaagccgcattccggctgaccac atgggtctga tggctggttt cactgcccgtaatgaaa  
aggcgagctggtggttc tgggacgcaacggttccgac tactc cgtcgcgggtgctggc ggccctgttaccgcgc cgtattgtgc gaga tctggacggga tgttgacgggtgttta tactgcgatccggctca  
ggtgccc gatgc gaggt tgttgaagtc gatgtcctatcagga agoga tggagctttc tta ct tggc gctaaagttc ttcac cccc gcacc atcac cccca tgc cc cagttcaga tccc tgcctgat  
taaaaaat accggaatcctcaagctcc aggtacgctc attgggtgccagccgtgatga agacgaatta ccgggtcaagggcatt tccaatctgaataa catggcaatg ttcagcgttccggcc cggggat  
gaaaaggatggt tggca tggcggcgcgctcttggcagc gatgcaac gcgcc gbtat tccg tgggtgctgat tacgc aatcacttccgaa tacag tatcagtttc tgcgt tccgc aaagc gactggt  
gcgagctgaacggcaatgcaggaaga gttctacctggaactgaaagaaggc ttactggagc cgttggcgggtgacggaacggctggccattatctc ggtgg taggt gatggtatgc gcacc ttactgg  
gatctcggcga aattcttggcc ggcgtggccc ggc cc aatatcaaca ttgtc gccat tgcctc aggga tcttc tgaac gctcaatctctgtc tgggtcaataacgat gatgc gaccactggc gtcgcgt  
tactcatcagatgctgt tcaat accga tcaagg ttatc gaagtgtttg tgattggcgt cgggtg cgttggcgg tgcgc tgc tggagc aactgaagcgtcagc aaagc tgggtgaagaataaacat atcga  
cttacgtgctcgggtgttgctaac tgaaggcactgctcac caatgtacat ggcc taatc tggaa aactggcaggaagaa ctggcgcgaagccaa agagc cgtttaatc tgggc gcttaattgcct  
cgtgaaa gaata tcatc tgc tgaaccc ggtca ttgttgactgtacttccagc caggc agtggcggatcaata tggcgcacttctgc gcgaa ggtttccacgtt gttacgcc gaaca aaaaggccaacac  
ctcgtc gatgga tta ctaccatcagttgcgttatgcggcggaaaaatcgcggcgtaaattcc tctatgacac caacgttggggcgtgattaccgggtatcggagaac ctgca aaatc tgcctc aatgc tgg  
tgatgaa ttgat gaagtctcc ggcatctcttcaggtcgcctttctt atatc ttcggcaaggtagac gaaggcatgagtttc tccgagggcaccac actggcggg gaaatgggttatacc gaacc gga  
cccgcgagatgactcttctggatgga tgtggcgcgt aagctattgattctc gctcgtgaaa cgggacgtgaaactggagctgcccggatattgaaat tgaac ctgtgctgccc cgcagagtttaacgc cga  
gggtgatgctgc cgttttatggcgaa tctgtcacagctcgcagatc tctttggccgc cgtgtggcgaaaggc ccgtgatgaa ggaaaagttt tggcgtatg ttggc aatat tgatgaaagatggcgtctg  
cccgctgaagat tggcgaagtgatgg taatgatccgctgttcaaag tgaaaaatggcgaaa acgcc ctggccttctatagc cactattatcagcc gctgc cgttggtactgcccggatattggtgc ggg  
caatgac gttac agctgcgggtgctt tgc t gatc tgc t acgtacc tctca tggaa gttag gagtc tgaca tgggttaagttatgccc ggcttccagtgccaa tatgagcgtc ggggttgatgtgc  
tcggggc ggcgg tgaca cctgt t gatg tgc a ttgctcggagatgta gtcac ggttgaggcg gcagagacat tca gtc tcaaacac ctggacgct ttgcc gataa gctgc cgtcagagcc acgggaaa  
atctcgtttatc agtgc tgggagcgtttttgc cagga cttggcaagcaaat tccag tggcg atgac tctggaaaagaat at gccgatcgg ttcgggcttaaggctc cagcgcctgt tca gttggtcgg  
cgtgatggcgatgaaatgaacactgcccgaagccgcttaatgacactcgtttgctggcctttg atgggcgagttggaaaggcgtatc tccggcagcattcat tacga caacgtggcaccgtgttttc tgg  
gtggtatgcagttgatgattgaagaaaacgac atcatcagtc agcaagtgcc aggggtttgat gagggtgctgtgggtgctggc gtatccggggattaaagtc tgcac ggcagaagcc agggc tattttac  
cggcgcagtatc gccgc cagga ttgca ttggc cagggcgcac atctggcaggcttca ttcac gccctgctattcccgctcagcc tgagcttgc cgcga agctgatgaa agatgttatc gctga accctacc  
gtgaacggttac tggca ggttccggc agggcggcga gggcgttggc gaaatcggcgggtga ggcagcggta tctcc ggctc cggc ccgac tttgt tgcctctgtg tgaca agccggatc cccc agc  
gcgttgc cga ctggttgggtbaa aaactacctgcaaaa tcaaggaaggtttgt tcatatttgc cggctggatacggcggggcgc acgagta ctggaaa actaa atgaa actctacaatctgaa agatcaca  
atgagcaggtcagctttgycga agccgt aacc caggggttgggcaaaa atca ggggc tgttt tccc gcacgacctgcccga attc agcctgactgaaattgatgagatgc tgaagctggat tttgtca  
cccgcag tgcga agatcctctc ggcgtttattgggtgatgaaa tcccgcagga aatcc tggaa gagcgcgtac ggcggcgtttggc tttccc ggctc cggtc gccaa tgttgaaagc gatgtcgggtgct  
tgyaattgttcc acgggccaac gctggcat ttaaaaat ttcggcgggtcgttttatggcaca aatactgaccatattggggcgttaagcc agtgaacattctgacccggacatcc ggttatctggag  
cggcagtggtc atgcttt  
tcgacggcgatttcgatgc  
actacttgaagctggttcg  
gttttatgctcgcaccaaa  
gtgtggaagagtgttccg  
ctgccgtagcttatcgtgc  
caaaagagctggcagaacg  
tgcttttatgcagccggc

**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  
"bioinforma\* OR computa\*"

16 тыс. "bioinforma\*"

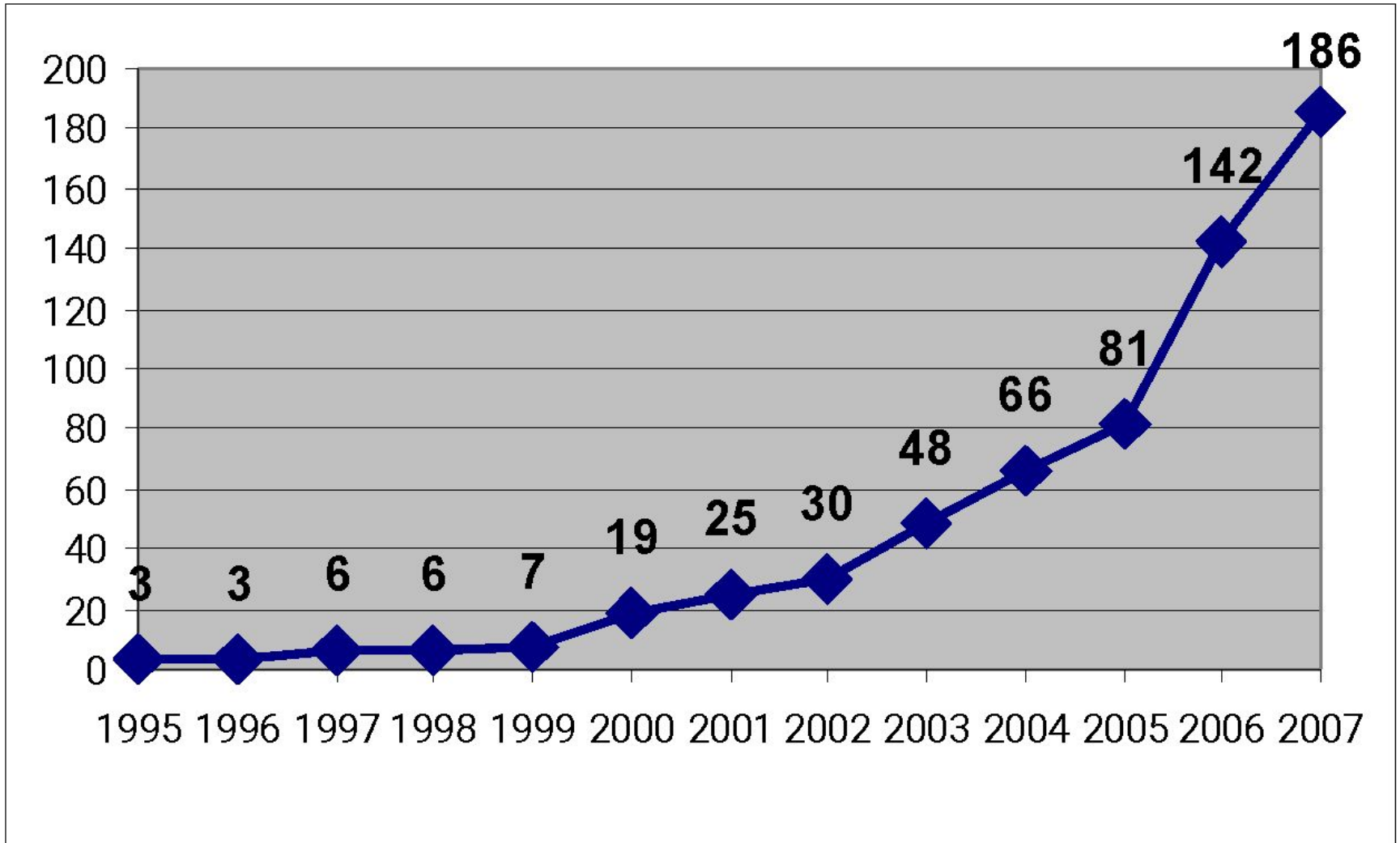
65 тыс. "bioinforma\* OR computa\*"

# 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

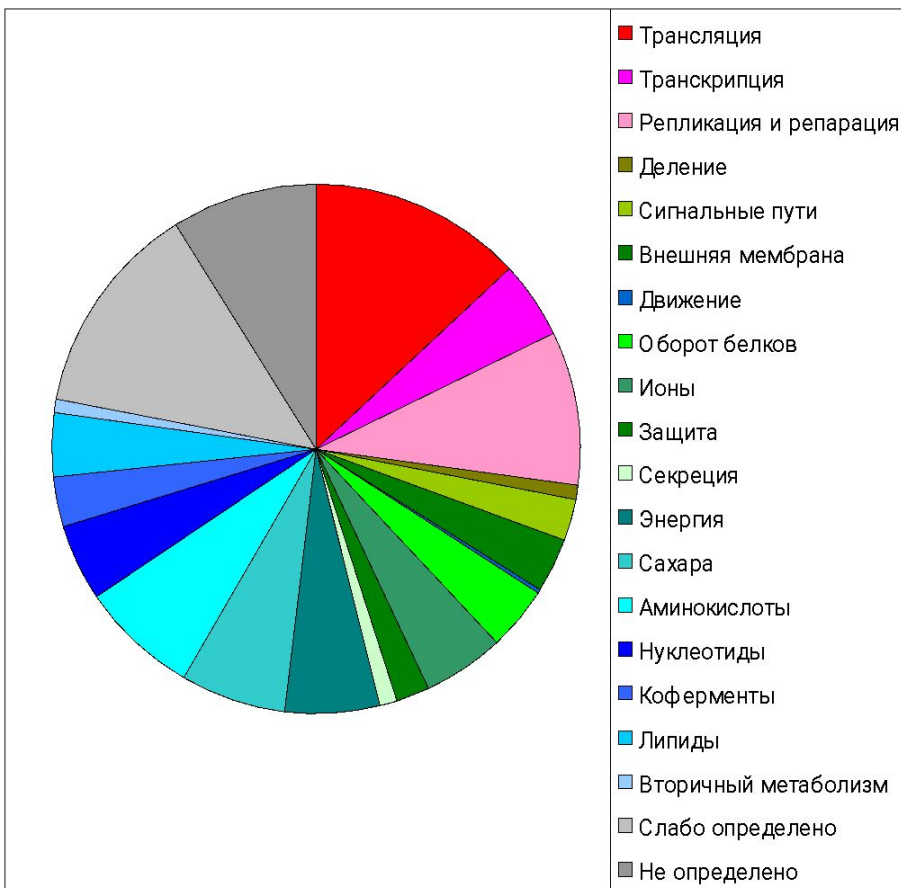
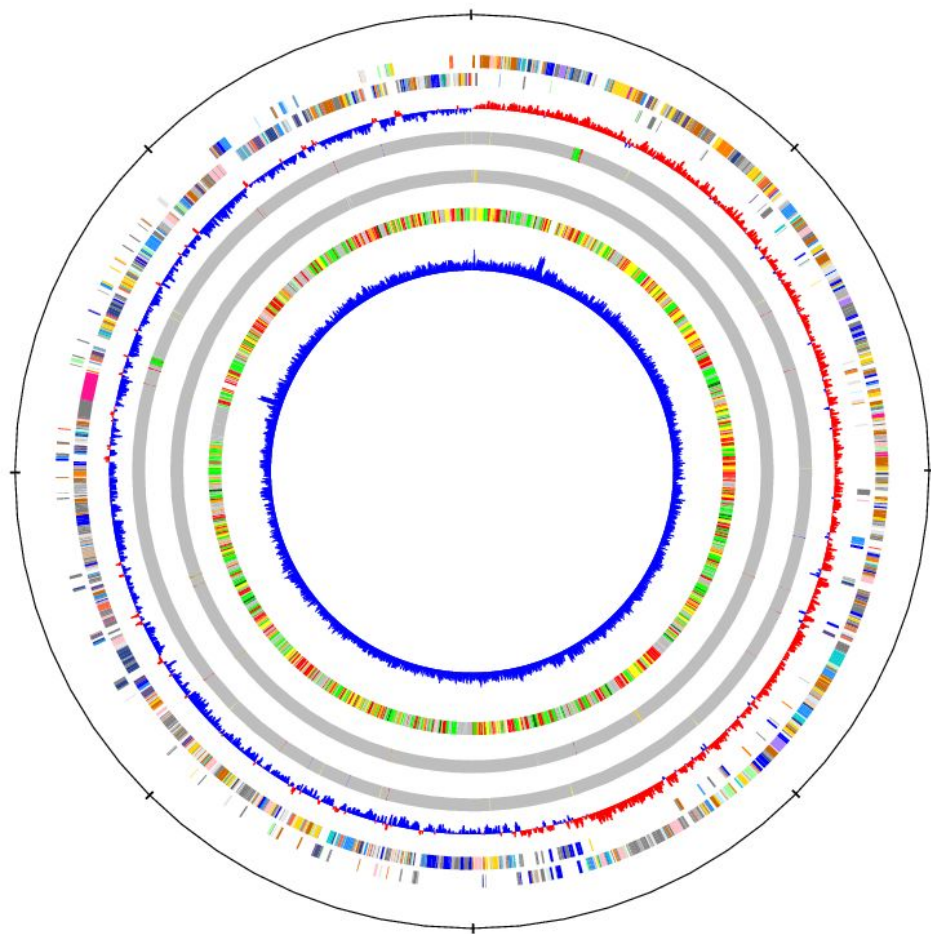


# Математические и алгоритмические проблемы

- **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 быстроедействие 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: ИППТИ РАН

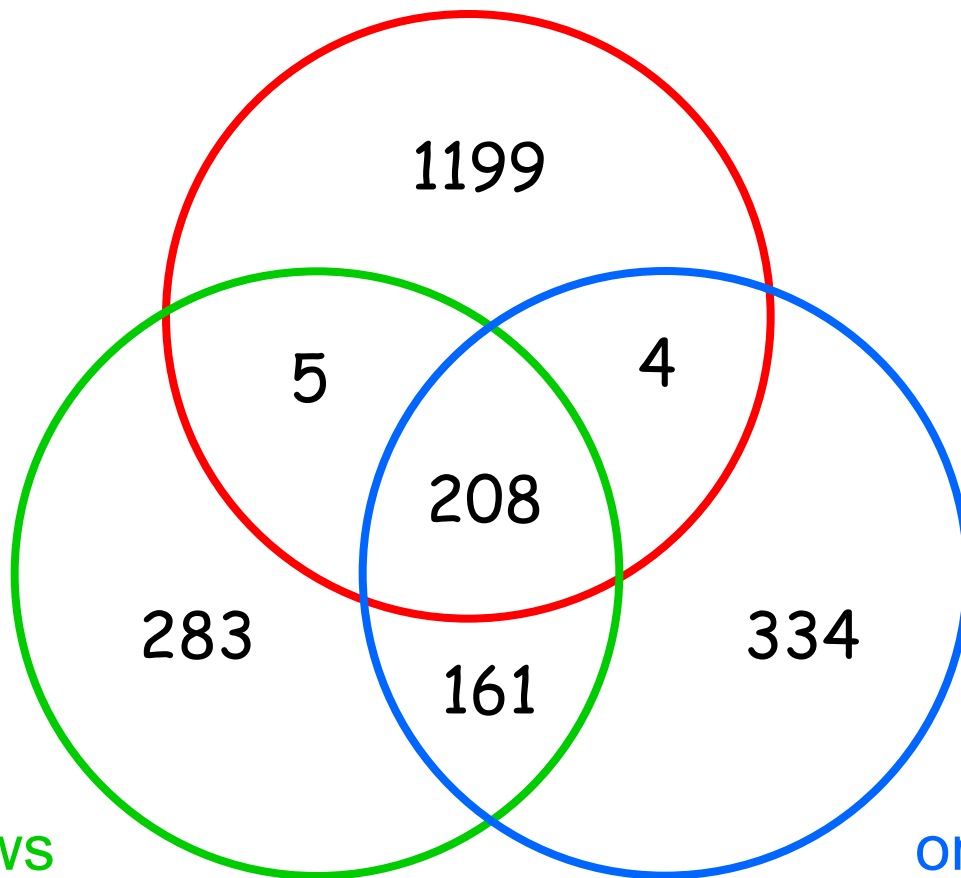


~1,5 Мб; ~1400 генов.

Установлены функции ~80% генов; проведена метаболическая реконструкция

# Сравнение с родственными геномами

*Acholeplasma laidlawii*



aster yellows  
*Phytoplasma*

onion yellows  
*Phytoplasma*

## Сравнительная геномика - 2

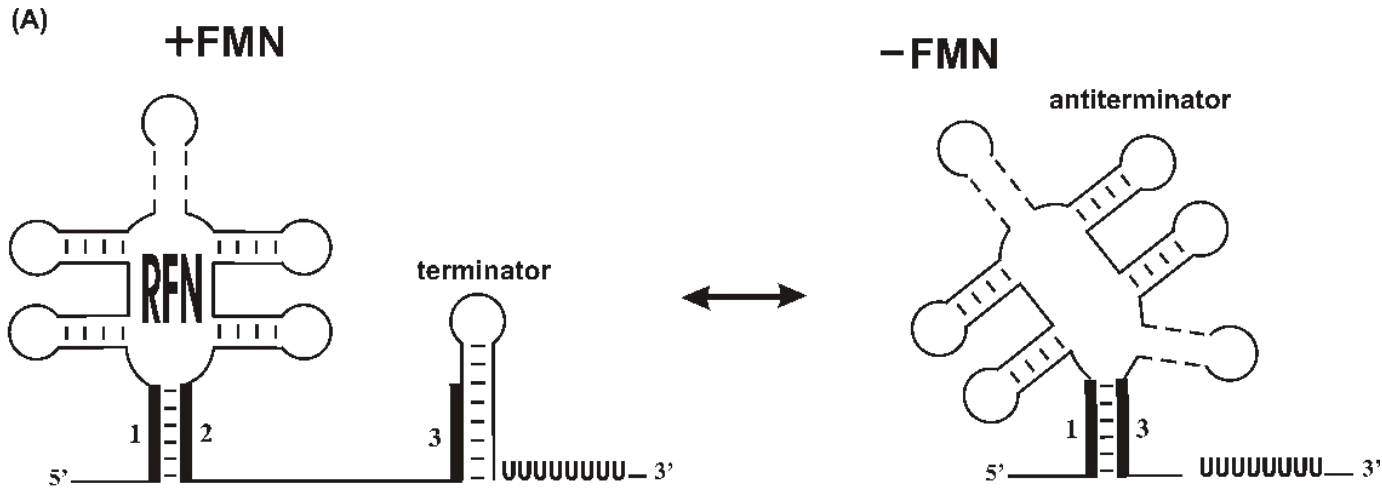
- Не обязательно последовательности:
  - структура белка и РНК
  - расположение генов на хромосоме (ко-локализация)
  - ко-регуляция и ко-экспрессия генов
  - филогенетические образцы (совместное появление в геномах)
- Предсказав структурные особенности белка, можно определить его функциональный класс
- 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

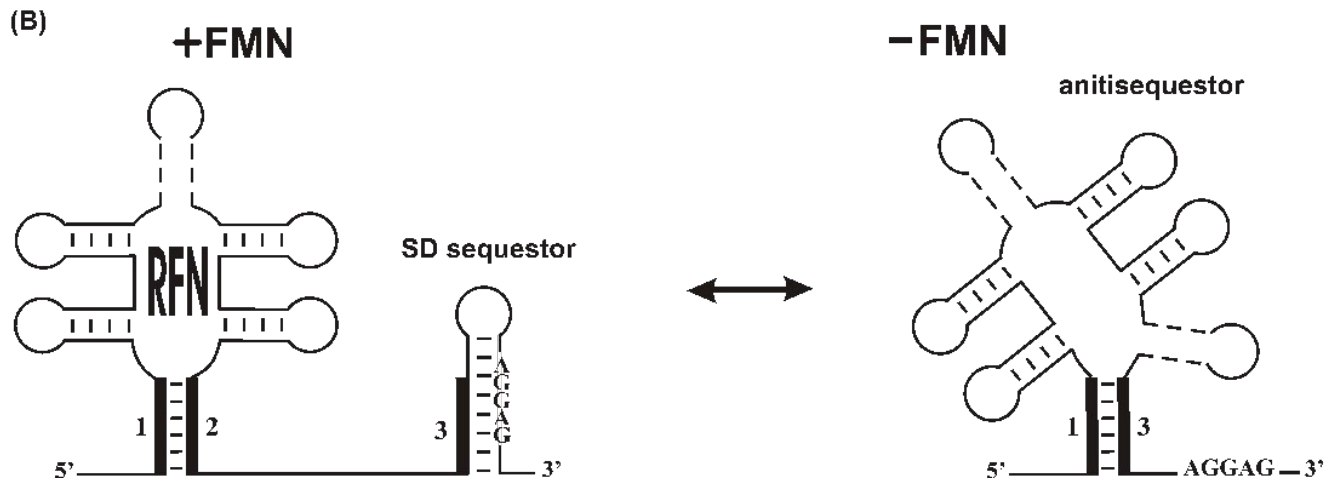
# Регуляция экспрессии генов за счет формирования альтернативных структур РНК

## • 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 рибосафина

	1	2	2'	3	Add.	3'	Variable	4	4'	5	5'	1'
BS	TTGTATCTTC	GGGG-CAGGGT	GGAATCCC	GACCGCCGT	21	AGCCCGTGCAC--	8 4 8	----TGGATTCA	TTTAA-GCTGAAGCCGC	ACGTGAA-AGT	CTGGAT-GGGAAG	GGATGAT
BQ	AGCATCTTTC	GGGG-TCGGGT	GAAATCCC	AACCGCCGT	19	AGTCCGTGCAC--	8 5 8	----TGGATCTAGT	GAAACTCTAGGGCCGC	ACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAG
BE	TGCATGCTTC	GGGG-CAGGGT	GAAATCCC	GACCGCCGT	20	AGCCCGTGCAC--	3 4 3	----AGGATCCGT	GAAACTCTAGGGCCGC	ACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAG
HD	TTTATGCTTC	GGGG-CTGGGT	GGAATCCC	GACCGCCGT	19	AGTCCGTGCAC--	10 4 10	----TGGACTCGT	GAAAACTGGGACC	GCACAGT-AGT	CTGGAT-GGGAAG	GAAGATAG
Bam	TGTATCTTTC	GGGG-CTGGGT	GAAATCCC	GACCGCCGT	23	AGCCCGTGCAC--	8 4 8	----TGGATTCTAGT	GAAAAGCTGAAGCCGC	ACAGT-AGT	CTGGAT-GGGAAG	GAAGATAG
CA	GATGCTTTC	AGGG-ATGGGT	GAAATCCC	AATCGCCGT	2	AGCCCGCAA---	3 4 3	----AGATCCGT	TAAACTGGGACC	GCACAGT-AGT	CTGGAT-GGGAAG	GAAGATAG
DF	CTTAACTTTC	GGGG-TAGGGT	GGAATCCC	AATCGCCGT	2	AGCCCGCAA---	7 6 7	----ATTTGCTT	TAAACTCAAAGCCGC	ACAGT-AA-AGT	CTGGAT-GGGAAG	GAAGATAGT
SA	TAAATCTTTC	GGGG-CAGGGT	GAAATCCC	AACCGCCAGT	6	AGCTCCGCAC--	11 3 11	----CTGACTAGT	GAGACTCTAGAGCCGC	ACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
LLX	ATAAATCTTC	AGGG-CAGGGT	GTAATCCC	ACCAGCCGGT	2	AGCCCGCAA---	4 4 4	----ATGATCCGT	GAAACTCAAAGCCGC	ACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
PN	AACTATCTTC	AGGG-CAGGGT	GAAATCCC	ACCAGCCGGT	2	AGCCCGCAA---	3 4 3	----ATGATTTG	TGAAACTCAAAGCCGC	ACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
TM	AAACCTCTTC	GGGG-CAGGGT	GGAATCCC	GACCGCCGGT	3	AGCCCGCAA---	5 4 5	----TTGACTCCGT	GGAATCAAAGCCGC	ACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
DR	GACCTCTTTC	GGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	15	AGCCCGCAA---	8 12 9	----CCGATCCG	CTGAAACTGGGACC	GCACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
TQ	CACCTCTTTC	GGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	3	AGCCCGCAA---	5 4 5	----CCGACCCG	CTGGAATCAAAGCCGC	ACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
AO	AATAATCTTC	AGGG-CAGGGT	GGAATCCC	GATCCGGT	2	AGCCCGCAA---	7 7 7	----AGGAACCCGT	GGAATCAAAGCCGC	ACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
DU	TTTAACTTTC	AGGG-CAGGGT	GGAATCCC	GATCCGGT	2	AGCCCGCAA---	13 4 12	----AGGAACCT	AGTGAATCTAGT	ACCAGCAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
CAU	GAAACCTTTC	GGGG-CAGGGT	GGAATCCC	GATCCGGT	20	AGCCCGCAA---	3 4 3	----AGGAACCCG	CTGGAATCAAAGCCGC	ACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
EN	TAAAGCTTTC	AGGG-CAGGGT	GGAATCCC	GACCGT	2	AGCCCGCAA---	5 4 5	----GATTTG	CTGGAATCAAAGCCGC	ACAGT-AG-AGT	CTGGAT-GGGAAG	GAAGATAGT
TFU	ACGGTCTTC	CGGG-GTCGGT	GTAAGTCCC	GAAGCCGGT	3	AGTCCCGTGCAC--	8 5 8	----TGGAACTCG	TGAAACTCCGCT	ACCAGCAGT-AG-AGT	CTGGAT-GGGAAG	GAAGATAGT
SX	-AGCCACTTC	CGGG-GTCGGT	GTAAGTCCC	GAAGCCGGT	3	AGTCCCGTGCAC--	8 5 8	----TTGACCA	CTGGAATCAAAGCCGC	ACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
BU	GTGCTCTTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	30	AGCCCGCAAAGCC	137	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AG-AGT	CTGGAT-GGGAAG	GAAGATAGT
BPS	GTGCTCTTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	21	AGCCCGCAAAGCC	8 4 8	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
REU	TTACGTCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	31	AGCCCGCAAAGCC	7 5 7	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
RSO	GTACGTCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	21	AGCCCGCAAAGCC	11 3 11	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
EC	GCTTATCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	17	AGCCCGCAAAGCC	8 4 8	GACACAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
TY	GCTTATCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	67	AGCCCGCAAAGCC	8 3 8	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
KP	GCTTATCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	20	AGCCCGCAAAGCC	8 4 8	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
HI	TCGCTATCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	2	AGCCCGCAAAGCC	26 9 30	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AA-AGT	CTGGAT-GGGAAG	GAAGATAGT
VK	GCCTATTCTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	14	AGCCCGCAAAGCC	11 9 11	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
VC	CAATATCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	13	AGCCCGCAAAGCC	5 4 5	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
YP	GCCTTATCTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	40	AGCCCGCAAAGCC	16 6 16	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
AB	GCCTATTCTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	25	AGCCCGCAAAGCC	16 4 27	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
BP	GTACGTCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	18	AGCCCGCAAAGCC	10 4 10	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
AC	ACATGCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	16	AGCCCGCAAAGCA	10 3 11	---CGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
Spu	AACTATCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	34	AGCCCGCAAAGCC	6 6 6	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
PP	GTGCTCTTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	13	AGCCCGCAAAGCC	7 3 7	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
AU	GGTTGCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	17	AGCCCGCAAAGCC	7 9 7	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
PU	AAACGTCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	19	AGCCCGCAAAGCC	19 4 18	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
PY	TAACTCTTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	19	AGCCCGCAAAGCC	15 4 16	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
PA	TAACTCTTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	19	AGCCCGCAAAGCC	14 4 13	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
ML0	TAAAGTCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	16	AGCCCGCAAAGCC	8 5 8	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
SM	AACTGCTTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	34	AGCCCGCAAAGCC	8 3 8	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
BME	GCTTGTCTTC	GGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	17	AGCCCGCAAAGCC	10 15 10	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT
BS	ATCAATCTTC	GGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	18	AGCCCGCAA---	5 4 5	----AGGATTC	CTGCTGAGAAAGCCAG	AGCCGACAGT-AC-AGT	CTGGAT-GGGAAG	GAAGATAGT
BQ	GTCTATCTTC	GGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	27	AGCCCGCAA---	3 5 3	----AGGATTC	CTGCTGAGAAAGCCAG	AGCCGACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
BE	ATTCATCTTC	GGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	20	AGCCCGCAA---	3 4 3	----AGGATTC	CTGCTGAGAAAGCCAG	AGCCGACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
CA	AATGATCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	2	AGCCCGCAA---	3 4 3	----TATGATC	CTGCTGAGAAAGCCAG	AGCCGACAGT-AA-AGT	CTGGAT-GGGAAG	GAAGATAGT
DF	GAAATCTTTC	GGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	2	AGCCCGCAA---	6 4 6	----GATTTG	CTGCTGAGAAAGCCAG	AGCCGACAGT-AA-AGT	CTGGAT-GGGAAG	GAAGATAGT
EF	GTCTCTTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	3	AGTCCAGCAC--	5 3 5	----ATTTGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
LLX	AAATATCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	21	ACTCCCGTGCAC--	4 4 4	----TTGAAAG	CTGCTGAGAAAGCCAG	AGCCGACAGT-AA-AGT	CTGGAT-GGGAAG	GAAGATAGT
LO	GTCTATCTTC	GGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	3	AGTCCAGCAC--	3 10 3	----TTGACT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
PN	AAAGATCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	12	AGTCCCGTGCAC--	3 4 3	----GATGCT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
ST	AAATGCTTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	14	AGTCCCGTGCAC--	3 4 3	----GATGCT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
MN	AAATGCTTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	10	AGTCCCGTGCAC--	3 4 3	----GATGCT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AA-AGT	CTGGAT-GGGAAG	GAAGATAGT
SA	ATTCATCTTC	GGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	6	AGCTCCGCAC--	11 3 11	----CTGACT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
AMI	TCACAGTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	14	AGCCCGCAA---	5 5 5	----TGACTC	CTGCTGAGAAAGCCAG	AGCCGACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
DFA	ACGAACCTTC	GGGG-TAGGGT	GGAATCCC	ACCAGCCGGT	20	AGCCCGCAAAC--	11 4 11	---GACTC	CTGCTGAGAAAGCCAG	AGCCGACAGT-AT-AGT	CTGGAT-GGGAAG	GAAGATAGT
EN	AATAATCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	2	AGTCCAGCAC--	4 6 4	----GATTTG	CTGCTGAGAAAGCCAG	AGCCGACAGT-AG-AGT	CTGGAT-GGGAAG	GAAGATAGT
GLU	---TGTCTTC	AGGG-CAGGGT	GGAATCCC	ACCAGCCGGT	28	AGCCCGCAAAGCC	10 4 10	GTCAAGCAGAT	CTGCTGAGAAAGCCAG	AGCCGACAGT-AGT	CTGGAT-GGGAAG	GAAGATAGT

# Аттенюация транскрипции

Antiterminator

The *RFN* element

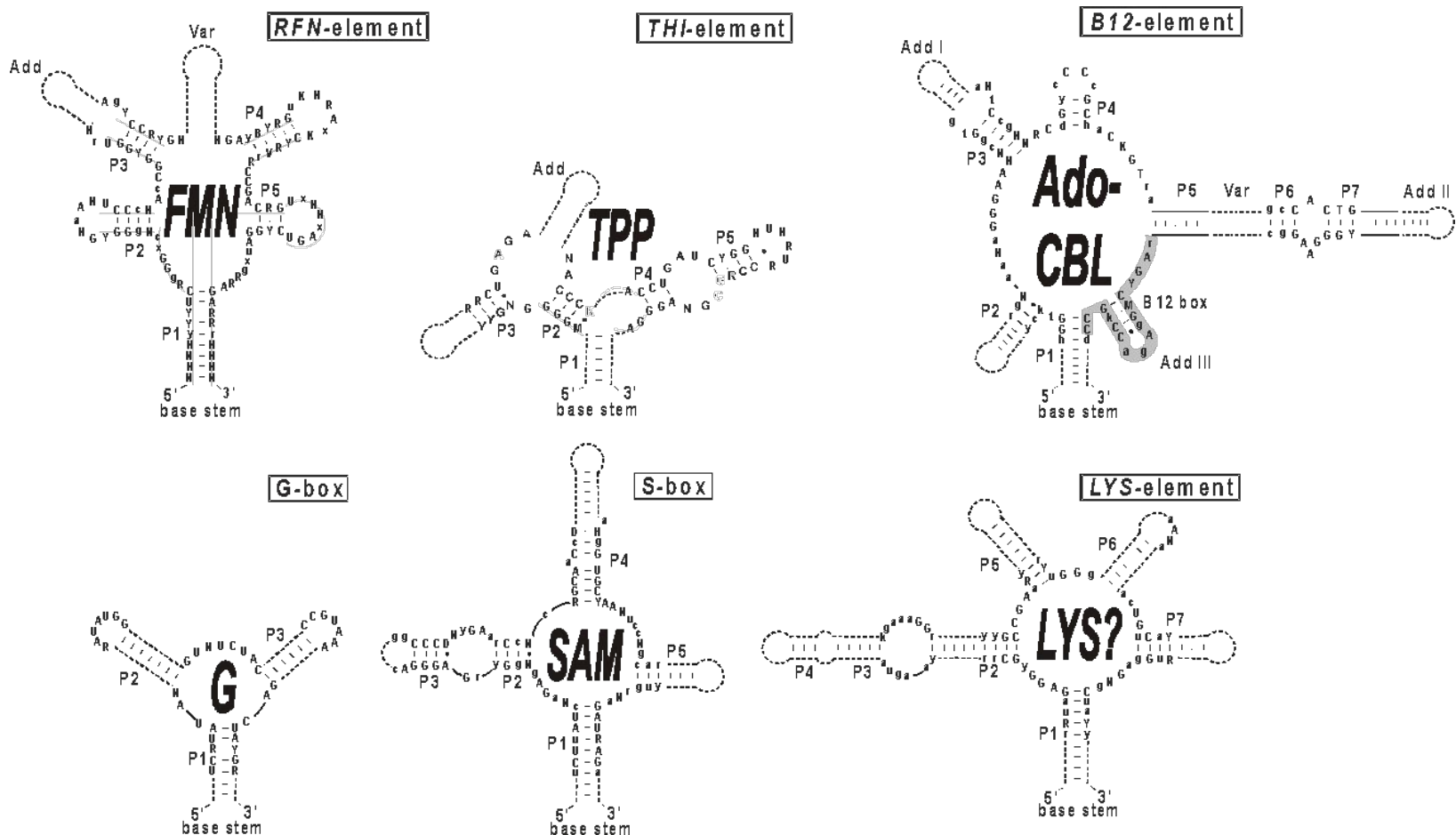
Terminator

<b>Bam</b>	GACAAAAAATATTTGATTGTA <b>TCTTCGGGGCT</b> TGGGTG	---	TCTGGATGGGGAAGGATGA	59	-----GTAA <b>AGCCCGAA</b> TGTGTAA--	ACATTCGGGGCTTTTTGACGCCAAAT
<b>BS</b>	GGACAAATGAATAAAGATTGTAT <b>TTCGGGGC</b> AGGGTG	---	TCTGGATGGGAGAAGGATGA	59	-----CTAA <b>AGCCCGAA</b> TTTTTTA--	TAAATTCGGGGCTTTTTGACGGTAAA
<b>BQ</b>	CTATAAATTTGAGCAAACAGC <b>ATCCTTCGGGGT</b> CGGGTG	---	TCTGGATGGGAGAAGGATAT	250	-----CCAA <b>AGCCCAAGGAT</b> ATTAAA--	ATCCTTGGGGTTTTTTGTTTTTTTT
<b>BE</b>	ACATAACGATATAGTGAT <b>GCACTCCTTCGGGGC</b> AGGGTG	---	TCTGGATGGGGAAGGATGC	155	-----TG <b>AGCCCGGGG</b> CAT-----	CCCGGGGTTTCATTTTTATTG
<b>HD</b>	AAATTTGAATAATTAATTTTTC <b>TCCCTTCGGGGC</b> TGGGTG	---	TCTGGATGGGAGAAGGAAAAC	148	-----AT <b>CCCCCGT</b> GAGAACAAAA--	TCTCTGGGGCTTTTTTGCGCGC
<b>CA</b>	TAATGGTAATTTAATAGGAT <b>TCTTCAGGGAT</b> TGGGTG	---	TCTGGATGAAAGAAGAAATA	34	----- <b>AATCTC</b> GAAGGATTACC---	TTTCTTTGGAGATTTTTTTATTG
<b>DF</b>	TAAATATAAATTTAACTTTAACT <b>TTCGGGGT</b> AGGGTG	---	TCTGGATGGAAGAAGATAAT	63	-----TAA <b>ACCTGAG</b> TTAATT-----	CTCAGGGTTTTTTGTTTTAAAAA
<b>LLX</b>	ACTTTAGCTACAAATGAATAAAT <b>TTCAGGGC</b> CAGGGTG	---	TCTGGATGAAAGAAGATAAT	127	-----AAAAG <b>ACCTGAA</b> ATTTT-----	ATTTTAGGGTCTTATTTTTTATTAG
<b>PN*</b>	ATCATCTGTAAATTTGAATAACTAT <b>TTCAGGGC</b> AGGGTG	---	TCTGGATGAAAGAAGATAAA	81	-----TGTAT <b>GCCTTGAG</b> TAGTCCCC--	TATTTCAAGGTATTTTTTTGGAGG
<b>PN*</b>	ATCATCTGTAAATTTGAATAACTAT <b>TTCAGGGC</b> AGGGTG	---	TCTGGATGAAAGAAGATAAA	19	-----CG <b>TGCTTGAATGA</b> TTACTTG	TCATTTTCAGAGCATTTTTGTTAATC
<b>TM</b>	AAAAC <b>TGAATACAAAGAAACGCTCCTTCGGGGC</b> CAGGGTG	---	TCCGGATGGGAGAGAGCGTG	13	-----ATGGG <b>ACCCGAGA</b> -----	<b>CGG</b> TCCCTTTTCTTTTACA
<b>AO</b>	ATTTGCAACAATTTTTTAATA <b>ATCTTCAGGGC</b> AGGGTG	---	TCTGGATGGAAGAAGATGAA	33	-----TTTACA <b>AGCCTTGAGAT</b> CGAAAG--	ATTTCAAGGCTTTTTTCATCATT
<b>DU</b>	AATTTTTTTAACTACTATTTTA <b>ATCTTCAGGGC</b> AGGGTG	---	TCTGGATGGAAGAAGAAGAG	47	-----TGCATA <b>AGCCTTGAGAT</b> CTTAG--	GATTTCAAGGCTTTTTCATAGTTA
<b>FN</b>	TAATCGAATAATGATAAATAAG <b>TCTTCAGGGC</b> AGGGTG	---	TCTGGATGGGAGAAGAATTA	18	-----ATAT <b>TGCTC</b> AGACTTT-----	GTTTGAGCATTTTTTTATTAA
<b>SA</b>	TATAACAATTTCACTATAAAT <b>TCTTTCGGGGC</b> AGGGTG	---	TCTGGATGGGAGAAAGAAATG	74	----- <b>TTTTCTCC</b> TTGCACTTAATT--	GATGTGAGGATTTTTGTTTATA
<b>DHA</b>	ACTCTTTTTAGATGAATACGA <b>ACTTCGAGGTA</b> AGGGTG	---	TCCGGATGGGAGAAGGTACA	43	-----GTTTT <b>ATGCCCTCAGGAA</b> CACCAT	TTCTCTGAGGCATTTTTGTTCTTTC
<b>FN</b>	GAAAAATAAATATTAATAAATA <b>TCTTCGGGGC</b> CAGGGTG	---	TCTGGATGGAGAGAAGAAAG	40	-----CTT <b>ACCCGAAT</b> TTCTAT-----	AATTCGGTTTTTTTATTTTT
<b>CA</b>	AAATATAAAAAATAAAGAATGAT <b>TTCAGGGC</b> AGGGTG	---	TCTGGATGAAAGAAGATAATA	19	-----TAT <b>GCCTTGA</b> CGTTTTT-----	CGTTGGGGCTTTTTTAATGCT
<b>DF</b>	AAAATTAAAAAATCAAGAAG <b>ATCTTCGGGGC</b> CAGGGTG	---	TCTGGATGGAGAGAAGATAAT	45	-----ATAAAAA <b>CTCGAAGAT</b> AGGG--	TCTTCGAGTTTTTTGTTTTTCTTAA
<b>BS</b>	TAATTTAAATTTCACTATGATCAAT <b>TCTTCGGGGC</b> AGGGTG	---	TCTGGATGGGGAAGGATGGA	103	--AAAGAAC <b>CTTTCGGTTTTTC</b> GA	GTAAGATGTGATCGGAAAAGGAGAGAATGA <b>AGTG</b> AAA
<b>BQ</b>	GGGAAAATAGAAATATCGG <b>TCTATCTTCGGGGC</b> AGGGTG	---	TCTGGATGGGAGAGATGGA	54	-----AT <b>TTCTCC</b> TTTTGTGTAAA--	ACACAAGGGTTTTTTTCGTTCT <b>ATG</b>
<b>BE</b>	ATAAAAAATGATAAGCGA <b>TTTCACTCTTCGGGGC</b> AGGGTG	---	TCTGGATGGGAGAGATGAA	114	-----GGCAGC <b>CTTCTCTT</b> TTGTA	GGATGAATCAGGAGAAAGGGAGGAGAAACAAGC <b>ATG</b>
<b>PN</b>	GTTTTTTGTTATGATAAAAGAG <b>TCTTTCAGGGC</b> CAGGGTG	---	TCTGGATGGAGAGAACGAA	137	--AA <b>CTTCTCTC</b> TGATTTTAG-----	AAAAT <b>TGGAG</b> GAACCTGTT <b>ATG</b> ACA
<b>ST</b>	TAAATCTGCTATGCTAGAAG <b>TGCTTTCAGGGC</b> CAGGGTG	---	TCTGGATGGAGAGAACCGG	130	--GGAA <b>CTTCTTTT</b> CAATTTGAAA--	AAAT <b>TGGAG</b> GAATTTTTT <b>ATG</b> TC
<b>MN</b>	ATTTTTTGATATGCTATAAG <b>TGCTTTCAGGGC</b> CAGGGTG	---	TCTGGATGGGAGAGACTGA	138	--GG <b>CTTCTTTT</b> CGATTTGTAA--	AAAT <b>TGGAG</b> GAATTTTTT <b>ATG</b> AA
<b>SA</b>	AAATTTAATAATGATAAA <b>TTTCACTTTCGGGGC</b> CAGGGTG	---	TCTGGATGGGAGAGATGGA	17	----- <b>TTCTCTTATTC</b> TTACG-----	AGATGAATGGA <b>AGGAG</b> AAAAATGAAT <b>ATG</b> A
<b>EF</b>	AAAAAATAAATAAAGGTT <b>TCTTTCAGGGC</b> CAGGGTG	---	GTCTGGATAAAGAAGATAGG	33	-----CTACTCTATTTTT <b>CCCTG</b> CAGA--	AAAA <b>TGGGG</b> TTTTTTTTT <b>ATG</b> A
<b>LLX</b>	TTTTTTGTGCTATAATAAAAA <b>TCTTTCAGGGC</b> ACCGTG	---	TCTGGATGGAAGAAGATGAA	66	--TCAACT <b>CTCTC</b> GAAAT <b>TGAAGA</b> A	T-TATT <b>TTCTC</b> ATAT <b>TTGGAG</b> GTTTTTTTT <b>ATG</b> T
<b>LO</b>	ATTGTAAGAAAATATTCGTT <b>CATCTTTCGGGGC</b> AGGGTG	---	TCTGGATGGGAGAGATGTTG	79	---ATGCACAAAC <b>TCTCC</b> TCAACTTTTTTTA-----	<b>GTTGAGG</b> TTTTTTTATTTCG

Antiterminator



# Другие РНК-переключатели, найденные методами сравнительной геномики



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

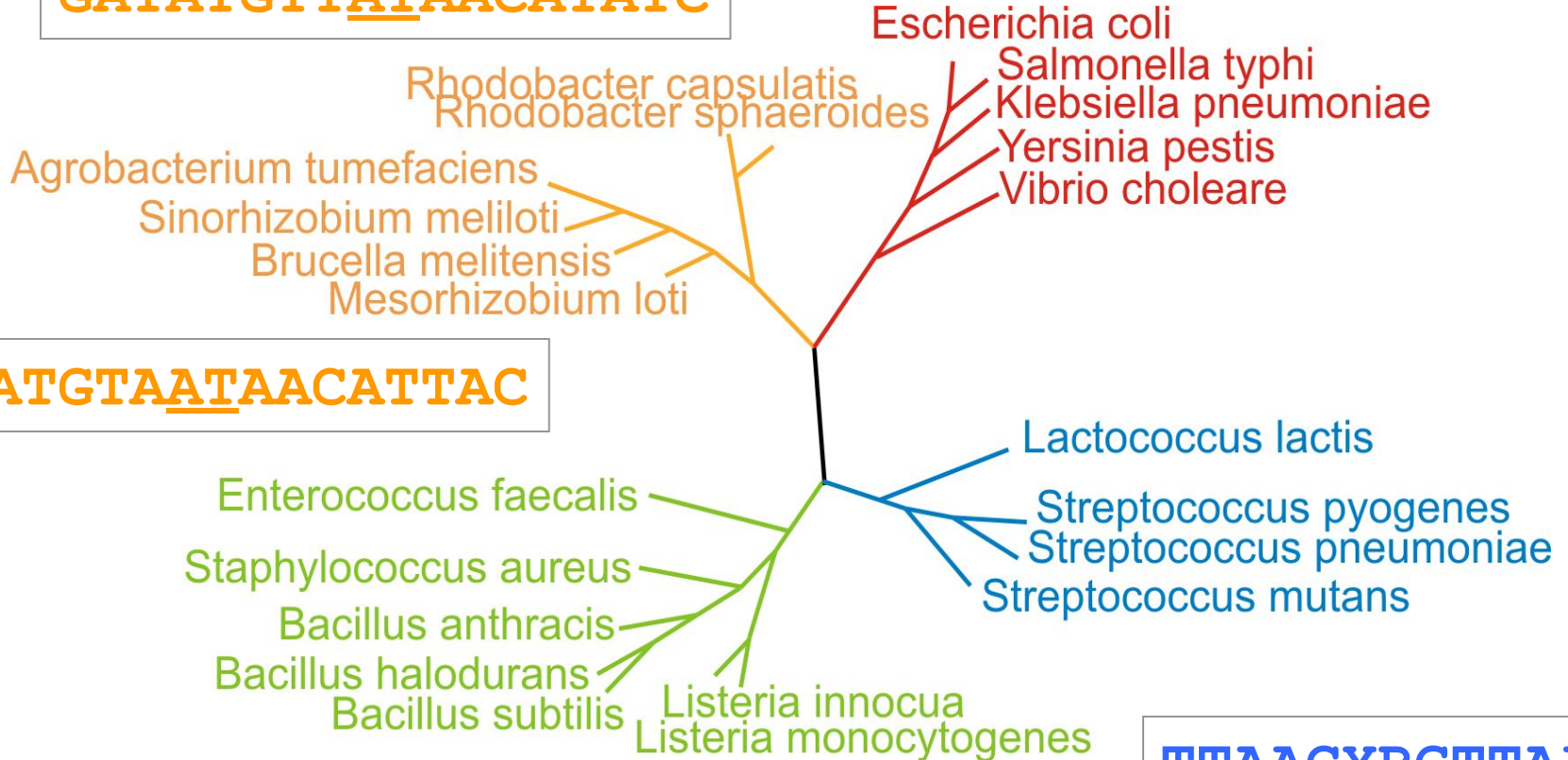
# Регуляторы гомеостаза цинка

nZUR- $\alpha$

nZUR- $\gamma$

GATATGTTATAACATATC

GAAATGTTATATAACATTC



GTAATGTAATAACATTAC

TTAACYRGTTAA

pZUR

TAAATCGTAATNATTACGATTTA

AdcR

# Регуляция гомологов рибосомальных белков

	L36	L33	L31	S14
<i>E. coli, S.typhi</i>	(-)	-	(-) +	-
<i>K. pneumoniae</i>	(-)	-	(-) -	-
<i>Y. pestis, V. cholerae</i>	(-) x	-	(-) +	-
<i>B subtilis</i>	(-)	(-) + -	(-) +	(-) +
<i>S. aureus</i>	(-)	(-) - -	-	(-) +
<i>Listeria spp.</i>	(-)	(-) -	-	(-) +
<i>E. faecalis</i>	(-)	(-) x - -	-	(-) + -
<i>S. pne., S. mutans</i>	(-)	(-) - -	-	(-)
<i>S. pyo., L. lactis</i>	(-)	(-) - -	-	(-) +

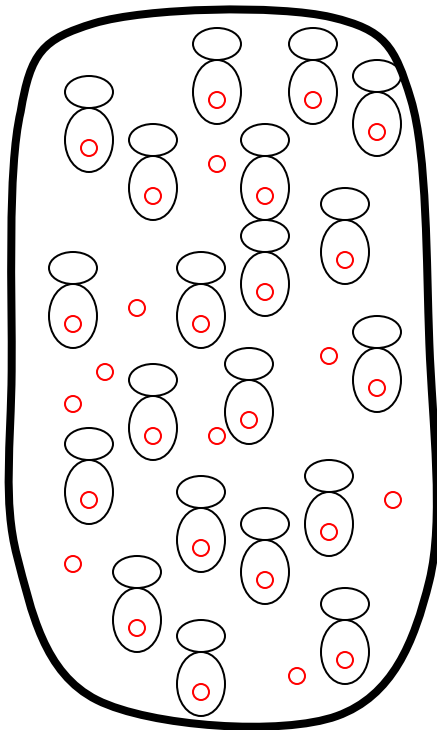
nZUR

pZUR

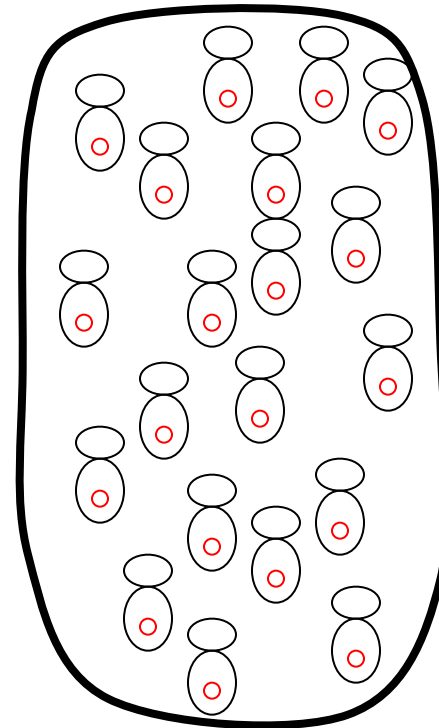
AdcR

# Плохой сценарий

достаточно  
цинка

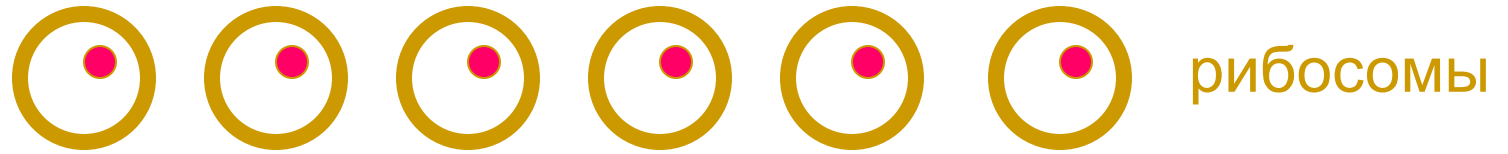


недостаточно цинка: весь цинк  
использован рибосомами, не  
хватает цинка для ферментов

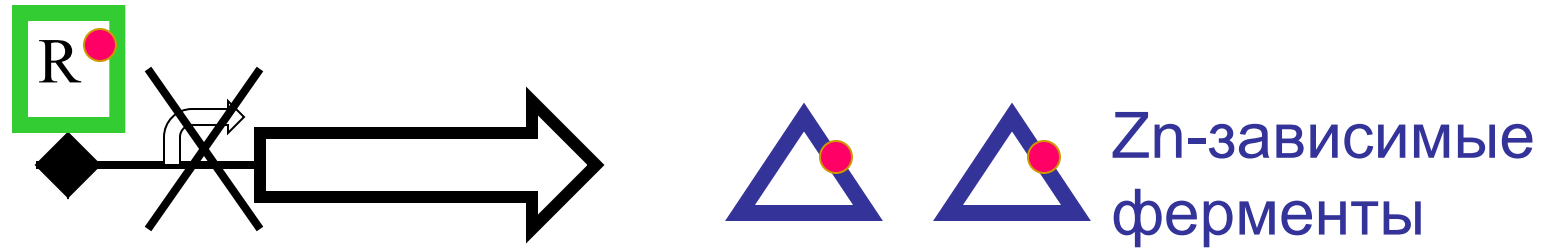


# Регуляторный механизм

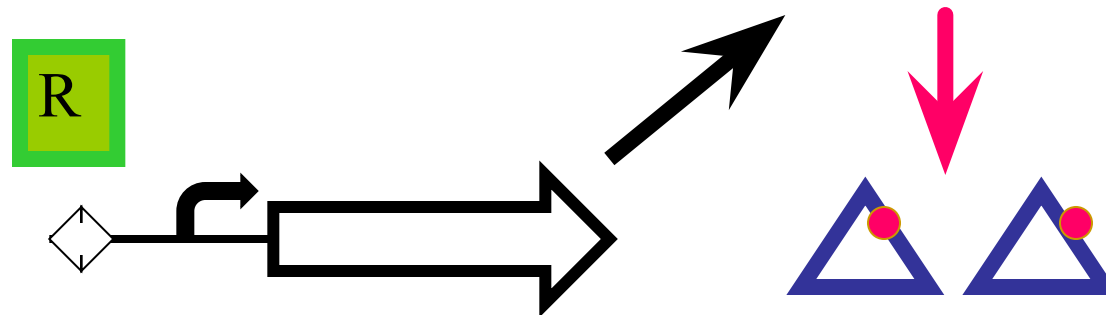
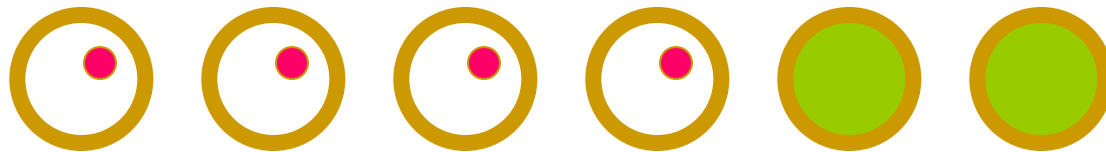
Достаточно **Zn**



репрессор



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\*<sup>†</sup>, Andrey A. Mironov\*<sup>‡</sup>, and Mikhail S. Gelfand\*<sup>‡§</sup>

## ... 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,<sup>1†</sup> Genki Akanuma,<sup>1†</sup>  
Yousuke Natori,<sup>1</sup> Rikinori Murayama,<sup>1</sup> Saori Kosono,<sup>2</sup>  
Toshiaki Kudo,<sup>2</sup> Kazuo Kobayashi,<sup>3</sup>  
Naotake Ogasawara,<sup>3</sup> Seung-Moon Park,<sup>4</sup> Kozo Ochi<sup>4</sup>  
and Fujio Kawamura<sup>1\*</sup>

(+ 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

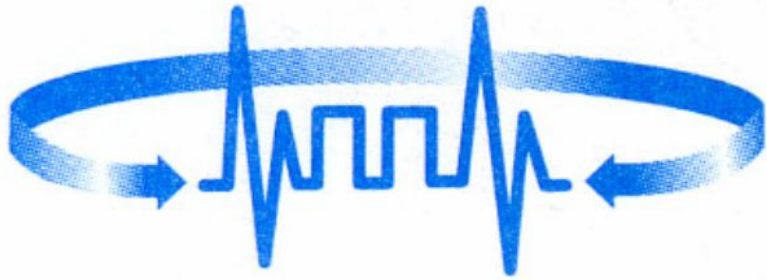
- Молекулярная эволюция
  - филогения генов
  - таксономия организмов
  - горизонтальные переносы и т.п.
  - положительный и отрицательный отбор
    - что сделало нас людьми?
    - лекарственная устойчивость
  - эволюция геномов
- Системная биология
  - строение геномов
  - сети взаимодействий
    -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

# Перспективы

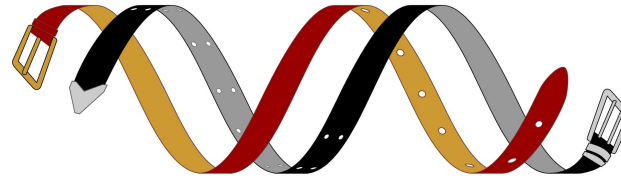
- **Индивидуальные геномы**
  - персональные человеческие геномы
  - геномы штаммов
- **Метагеномы**
  - некультивируемые бактерии
  - метагеномы экологических ниш
    - бактериальный метагеном человека
- **Другие виды данных**
  - интеграция
- **Моделирование (пока рано)**
  - построение полных карт

# 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