

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

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

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.wersingogenesis.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 Human Genome Deciphered, May Herald Medical News](#)  
[ Перевести эту страницу ]  
... Genome Deciphered; May Herald Medical News ... Compared to the human genome, the dog 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

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 Gedin) ...

# Перехватить зашифрованное сообщение -

## еще не значит его понять

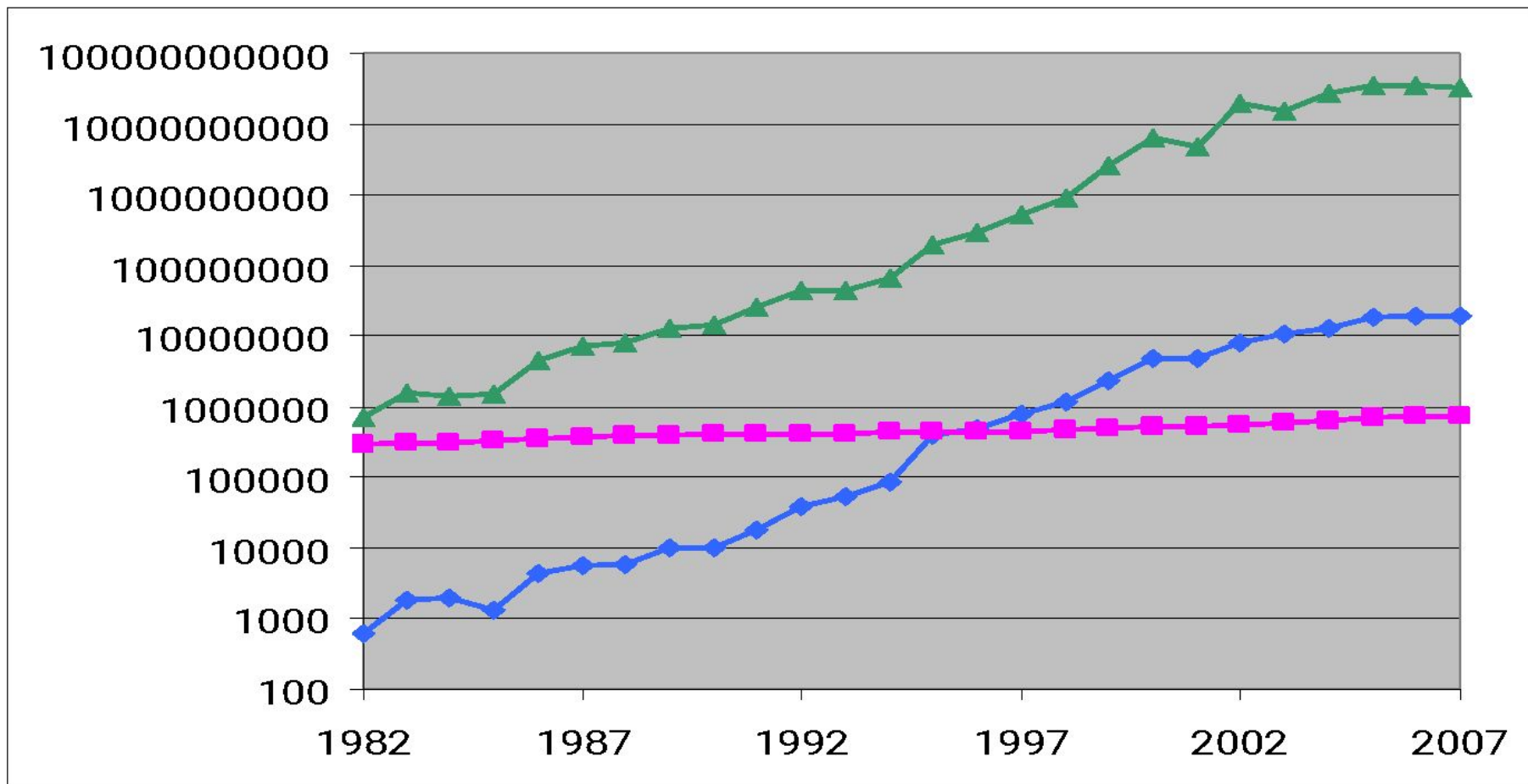
aacgggcaatatgtctctgtgtggattaaaaaagagtgctgatagcagctcttgaactggttacc tggcgtgagtaaataaaa ttttat ttagac ttaggtcactaaatactttaaccaat atagga tagcgcaagac agataaaaat tacagagtac acaac atcca tgaac cgc at tagca cccacc attac caccaccacc accat tacc acaggtaacgtgtcggggctgacgggtacaggaaac acagaaaa aagcccgccacctgacagtgccgggcttt tttttcgacc aaaggtaacgaggta acaac catgc gagggttgaa gttcggcgggtacatcagtgggcaaaa tgcagaacgt tttctgccgggtgccc gatattct ggaagc aatgc caggc agggg caggt ggcgaccgtc ctctc tgc cc ccgcc aaaaat taccaccatctggt agcga t gatt gaaaaaaccc attagcggcc aagat gcttt accca atatc agcga tgc cgaacgtat ttt tggcgaactctgac gggac tgc cc gccgc ccagc cgggatttcc gctgg cacaat tga aactt tgc tgc gacc aggaattgccc aaaa taaaa catgt cctgc atggc attag ttt gttggggcagtgcccggatagc atcaacgctgccgtgatttggcgtgcccagaaaaatgtcga tgc cc attatggccggcgtgttagaagcgcgtggtcacaacgtt accgt tatc gatccgggtcga aaaa actgcttgcagtggggc attac ctogaatctaccgttgatattgctgagtc acccggcgtat tggcggcaagccgcattccggctgaccac atgggtctga tggctggttt cactgcccgtaatgaaaa aggcgagctggtggttc tgggacgcaacggttccgac tactcgcctgcccgtgctggcggcctgtttacgcgc c gattgttgc gaga tctggacggga tgttgacgggtgttta tactgcgatccggctca ggtgccc gatgc gaggt tgttgaagtc gatgtcctatcaggaagoga tggagctttc t tact tggc gctaaagttc ttcac cccc gcacc atcac cccca tgc cc cagtt tca ga tccc t gccc gat ta aaaaat accggaatc ctcaagctcc aggtacgctc attgggtgccagccgt gatga agacga aatta cgggtcaagggcatt tccaatctga aataca tggcaatg ttcagcgttccggc cgggg gat gaaaagggatggt tggca tggcggcgcgctct tggcagc gatgcaac ggc cc gbtat tccg tgggtgctgat tacgc aatca tctccgaa tacag taca gtttc tgcgt tccgc aaagc gactggt gcgagctgaacgggcaatgcaggaaga gttctacctggaactgaaagaaggc ttactggagc cgttggcgggtgacggaacggctggccattatctc ggtgg taggt gatggtatgc gcacc ttactgg gatctcggcga aattct tggc ggcgtggccc ggc cc aatatcaaca ttgct gccat tgc tcc aggga tcttc tgaac gctcaatctctgtc tgggtcaataacgat gatgc gaccactggc gtcgcgt tactcatcagatgctgt tcaat accga tcaagg ttatc gaagtgtttg t gattggcgt cgggtg cgttggcgg tgcgc tgc tggagc aactgaagcgtcagc aaagc tgggtgaagaataaacat atcga ctta cgtgtctcgggtttgctaac tcc gaaggcactgctcac caatgtacatggcct taatc tggaa aactggcaggaagaa ctggcgc aagccaa agagc cgtttaatctcgggc gcttaattcgcct cgtgaaagaatactatc tgc tgaaccc ggtca ttgttgactgtacttccagc caggc agtggcggatcaata tggcgcacttctgc gcgaa ggtttccacgtt gttacgcc gaaca aaaaaggccaacac ctgc tgc gatgga t taccatcagttgcgttatgcggcga aaaaatcgcggcgtaaattcc tctatgacac caacgttggggcgtgattaccgggtatcgcgagaac ctgca aaatc tgc tcc aatgc tgg t gatgaa ttgat gaagt tctcc ggcat tctttcaggt tgc tttctt atatc ttcggcaaggtagac gaaggcatgagtttc tccgagggcaccac actggcggg gaaatgggttatacc gaacc gga cccgcgagatgactcttctggatgga tgtggcgcgt aagctattgattctc gctcgtgaaa cgggacgtgaaactggagctgcccggatattgaaat tgaac ctgtgctgccc gcagagtttaacgc cga ggggtgatgctgc cgttttatggcgaatctgtcacagctcgaagatc tctttggcgc gctgtggcgaaggc cgtgatgaa gaaaagtt tggcgtatg ttggc aatat t gatgaa gatggcgtctg cccgctgaagat tggcgaagtgatgg taatgatccgctgttcaaag tga aaaaatggcga aacgcc ctggccttctatagc cactattatcagcc gctgc cgttggtactgcccggatattggtgc ggg caatgac gttac agctgcgggtgtctt tgc t gatctgctacgtacc tctca tggaa gttag gagtc tgaca tgggttaagtttatgccc ggcttccagtgccaa tatgagcgtc ggggttgatgtgc tgggggc ggcgg tgaca cctgt t gatg tgc a ttgctcggagatgta gtcac ggttgaggcgcagagacat tca gtc tcaaacac ctggcagcct tggc gataa gctgc cgtcagagcc accgggaaa atatcgtttatc agtgc tgggagcgtttttgc cagga cttggcaagcaaat tccag tggcg atgac tctgg aaaaagaaat gccgatcgg ttcgggcttaaggctc agcgcctgt tca gttggtcggc cgtgatggc gatgaatgaaca ctgcccgaagccgct taatgacactcgtttgctggcctttg atgggcgagttggaa gggcgtatc tccggcagcattcat tacga caacgtggcaccgtgttttc ttg gttggtatgcagttgatgattgaagaaaacgac atcatcagtc agcaagtgcc aggggtttgat gagggtgctgtgggtgctggc gtatccggggattaaagtc tgcac ggcagaagcc agggc tttttac cggcgcagtatc gccgc cagga ttgca ttggcgcagggcgcac atctggcaggcttca ttcac gccctgctattcccgctcagcc tgagcttgc cgcga agctgatgaa agatgttatc gctga accctacc gtagaacggttac tggca ggttccggc aggcgcggca ggcgg ttggcga aatcggcgcggta ggcagcggta tctcc ggctc cggc ccgac tttgt tgcctctgtg tgaca agccggatc cccc agc gcgttgc c gactggttgggtbaa aaaaactacctgcaaaa tcaaggaaggtttgt tcaata tttgc cggctggatacggcggggcgc acgag tactggaaa actaa atgaa actctacaatctgaa agatcaca atgagcaggtcagctttgcgca agccgt aacc caggggttgggcaaaa aatca ggggc tgttt tccc gcacg accctggcggaa atc agcctgactgaaattgatgagatgc tgaagctgga tttgtca cccgcag tgcga agatcctctc ggcgt ttattggtgatgaaa tcccgcagga aatcc tggaa gagcgcgtac ggcgcgctttgccc tccc ggctc cggtc gccaa tgttgaaagc gatgtcggttgct tggaaattgtcc accggccaac gctggcat ttaaaaat ttcggcgggtcgtttatggcaca aatactgaccatattggggcgttaagcc agtgaacattctgaccccagacatcc ggttatctggc cggcagtggtc atgcttt tgcacggcgatttcgatgc actacttgaagctggttgc gttttatgctgcgacc aa gttgtggaagagtgttccg ctgcccgtagcttatcgtgc caaaagagctggcagaacg tgc ttttatgca gcccggc

**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\*"

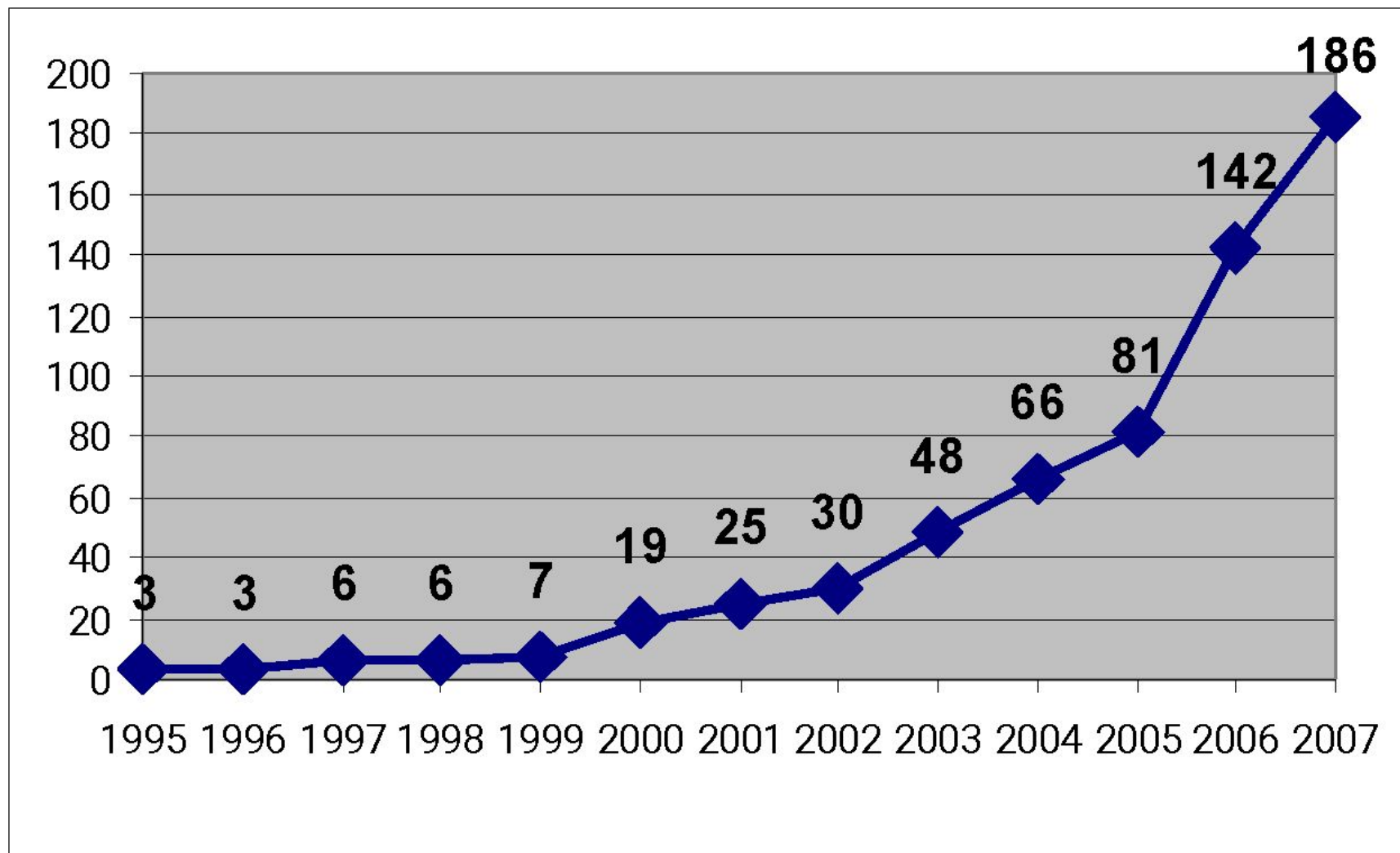
65 тыс. "bioinforma\* OR computa\*"

# 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

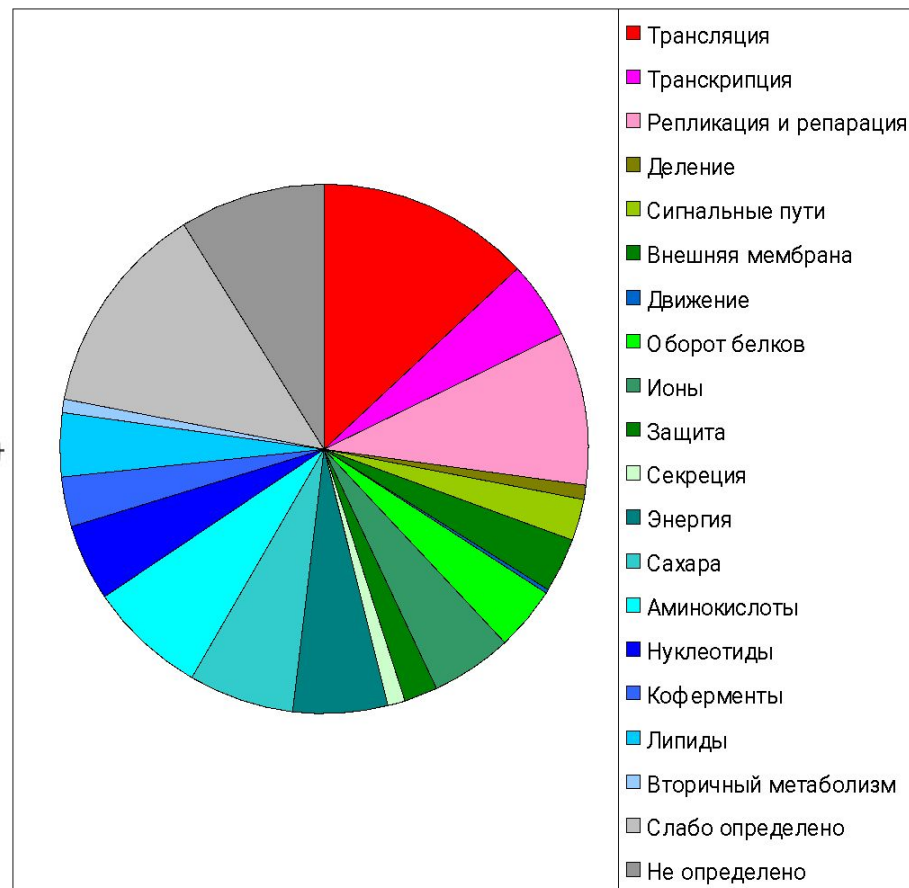
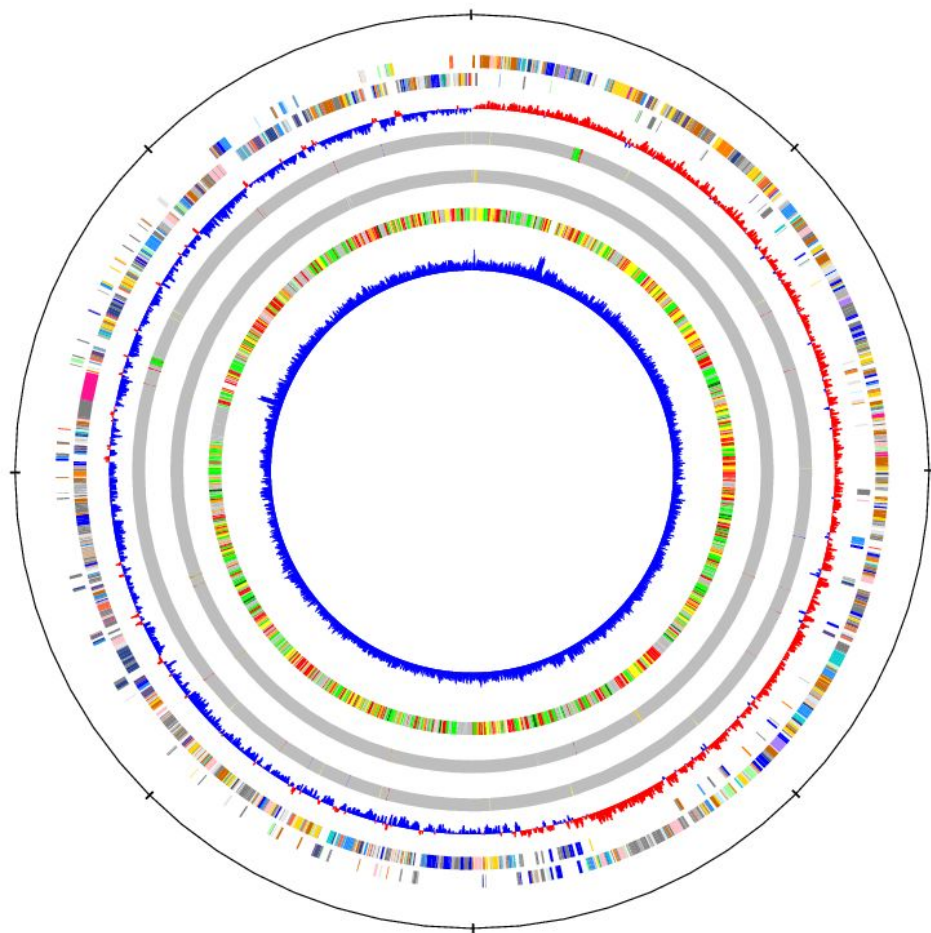


# Математические и алгоритмические проблемы

- **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 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: ИППТИ РАН

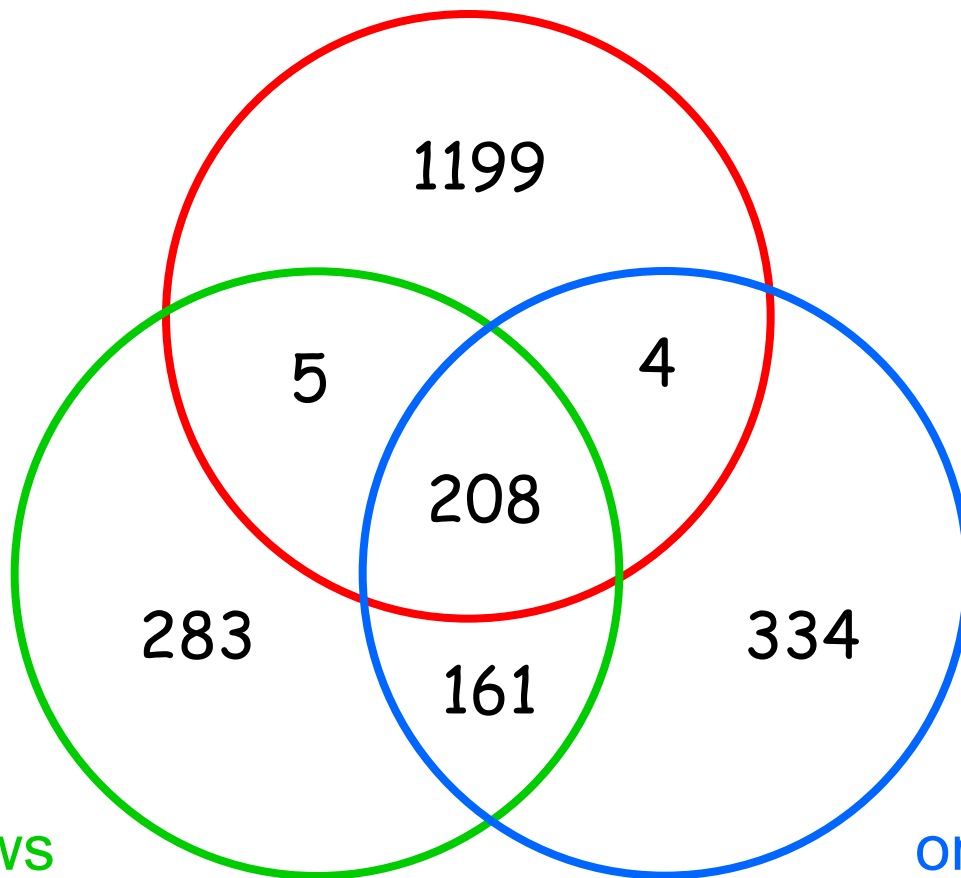


~1,5 Мб; ~1400 генов.

Установлены функции ~80% генов; проведена метаболическая реконструкция

# Сравнение с родственными геномами

*Acholeplasma laidlawii*



aster yellows  
*Phytoplasma*

onion yellows  
*Phytoplasma*

## Сравнительная геномика - 2

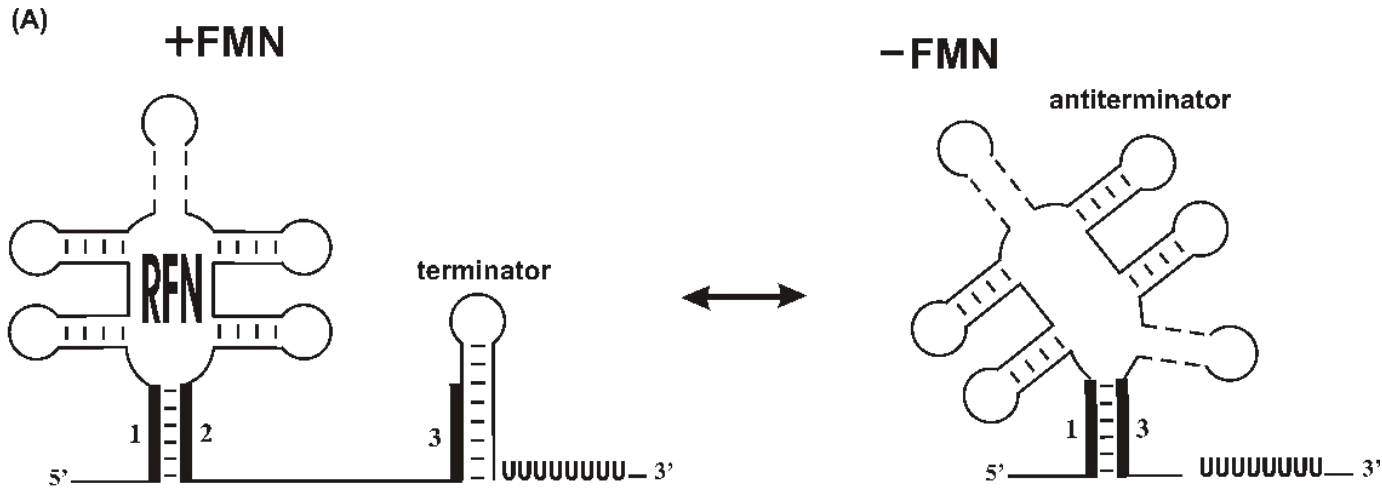
- Не обязательно последовательности:
  - структура белка и РНК
  - расположение генов на хромосоме (ко-локализация)
  - ко-регуляция и ко-экспрессия генов
  - филогенетические образцы (совместное появление в геномах)
- Предсказав структурные особенности белка, можно определить его функциональный класс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

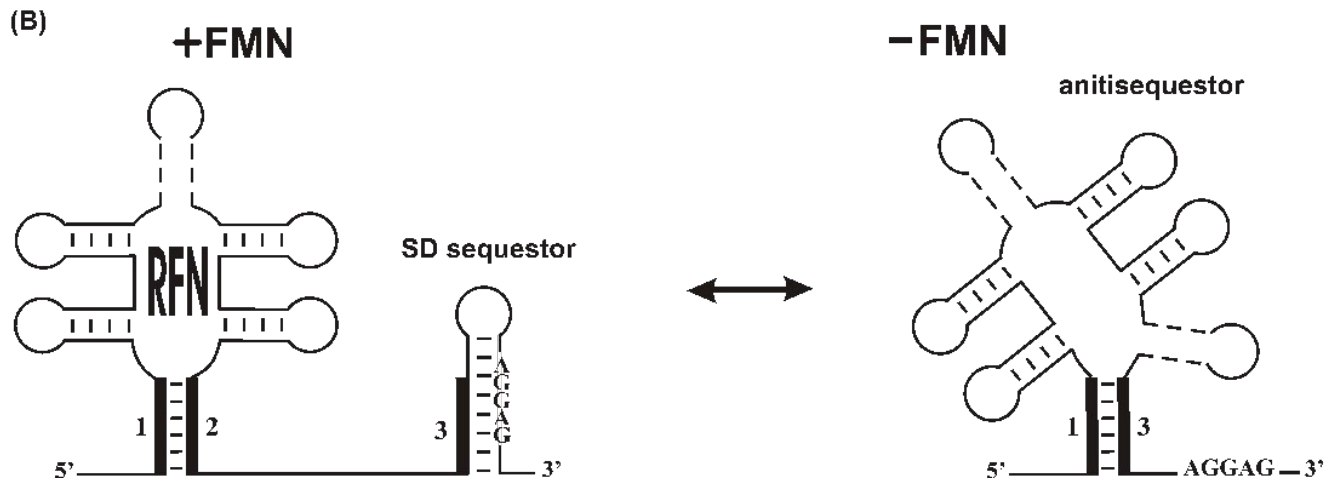
# Регуляция экспрессии генов за счет формирования альтернативных структур РНК

## • 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-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	21	AGC	CCCTGAC	--	8	4	8	----TGGATTTCAGTTTAA-G	CTGAAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATGAT
BQ	AGCATCTTTC	GGGG-TC	GGGTGAAAAT	CCC	ACC	GGCCGGT	19	AGT	CCCTGAC	--	8	5	8	----TGGATCTAGTGAACT	CTAGGGCCGACAGT-AT-AGT	CTGGAT-GGGAGAGGATATG
BE	TGCATCTTTC	GGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	20	AGC	CCCTGAC	--	3	4	3	----AGGATCCGTGCAACT	CTGGAGCCGACAGT-AT-AGT	CTGGAT-GGGAGAGGATGGC
HD	TTTATCTTTC	GGGG-CT	GGGTGAAAAT	CCC	ACC	GGCCGGT	19	AGT	CCCTGAC	--	10	4	10	----TGGACTCGTGAAAAT	CTGGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGAAACG
Bam	TGTATCTTTC	GGGG-CT	GGGTGAAAAT	CCC	ACC	GGCCGGT	23	AGC	CCCTGAC	--	8	4	8	----TGGATTCTAGTAAAAG	CTGAAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATGAG
CA	GATGCTTTC	AGGG-AT	GGGTGAAAAT	CCC	ACC	GGCCGGT	2	AGC	CCCTGAC	--	3	4	3	----AGATCCGTGTAAT	CTGGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGAAATAG
DF	CTTAACTTTC	GGGG-TA	GGGTGAAAAT	CCC	ACC	GGCCGGT	2	AGC	CCCTGAC	--	7	6	7	----ATTGCGTTAAAAT	CTGAAGCCGCACAGT-AA-AGT	CTGGAT-GGGAGAGGATATTT
SA	TAAATCTTTC	GGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCAGT	6	AGC	CTCCGAC	--	11	3	11	----CTGACTAGTGGAAAT	CTAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGAAATGT
LLX	ATAAATCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	2	AGC	CCCTGAC	--	4	4	4	----ATGATTCGTGAAAAT	CTGAAGCCGCACAGT-AT-AGT	CTGGAT-GGGAGAGGAAATAA
PN	AACTATCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	2	AGC	CCCTGAC	--	3	4	3	----ATGATTTGCTGAAAAT	CTGAAGCCGCACAGT-AT-AGT	CTGGAT-GGGAGAGGATAAAA
TM	AAACCTCTTC	GGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	3	AGC	CCCTGAC	--	5	4	5	----TTGACTCGTGAAAAT	CTGGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGAGCTGA
DR	GACCTCTTTC	GGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	15	AGC	CCCTGAC	--	8	12	9	----CCGATCCGTCGCAACT	CTGGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGAGGAG
TQ	CACCTCTTTC	GGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	3	AGC	CCCTGAC	--	5	4	5	----CCGACTCGTGAAAAT	CTGGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGAGGAG
AO	AATAATCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	2	AGT	CCCTGAC	--	7	7	7	----AGGAACCGTGAAAAT	CTGGTACCAGCAGT-AT-AGT	CTGGAT-GGGAGAGGATGAAA
DU	TTTAACTTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	2	AGT	CCCTGAC	--	13	4	12	----AGGAACCTAGTAAAAT	CTAGTACCAGCAGT-AT-AGT	CTGGAT-GGGAGAGGAGCAGA
CAU	GAAACCTTTC	GGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	20	AGC	CCCTGAC	--	3	4	3	----AGGACCCGCTGTGAT	CTGGAGCCGCACAGT-AT-AGT	CTGGAT-GGGAGAGGATGGC
EN	TAAAGCTTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	2	AGT	CCCTGAC	--	5	4	5	----GATTTGCTGAAAAT	CTGAAGCCGCACAGT-AG-AGT	CTGGAT-GGGAGAGGATTAAG
TFU	ACGGCTCTTC	CGGG-CT	GGGTGAAAAT	CCC	ACC	GGCCGGT	3	AGT	CCCTGAC	--	8	5	8	----TGGAACTCGTGAAAAT	CTGGTACCAGCAGT-AT-AGT	CTGGAT-GGGAGAGGATAGCTGC
SX	-AGCCACTTC	CGGG-CT	GGGTGAAAAT	CCC	ACC	GGCCGGT	3	AGT	CCCTGAC	--	8	5	8	----TTGACCAGCTGAAAAT	CTGGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGAGTCCGC
BU	GTCCCTCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	30	AGC	CCCTGAC	--	137	8	8	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
BPS	GTCCCTCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	21	AGC	CCCTGAC	--	8	4	8	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
REU	TTACCTCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	31	AGC	CCCTGAC	--	7	5	7	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
RSO	GTACCTCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	21	AGC	CCCTGAC	--	11	3	11	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
EC	GCTTATCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	17	AGC	CCCTGAC	--	8	4	8	GACACAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
TY	GCTTATCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	67	AGC	CCCTGAC	--	8	3	8	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
KP	GCTTATCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	20	AGC	CCCTGAC	--	8	4	8	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
HI	TCGCTCTTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	2	AGC	CCCTGAC	--	26	9	30	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
VK	GCCTATCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	14	AGC	CCCTGAC	--	11	9	11	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
VC	CAATATCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	13	AGC	CCCTGAC	--	5	4	5	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
YP	GCCTATCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	40	AGC	CCCTGAC	--	16	6	16	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
AB	GCCTATCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	25	AGC	CCCTGAC	--	16	4	27	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
BP	GTACCTCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	18	AGC	CCCTGAC	--	10	4	10	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
AC	ACATCTCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	16	AGC	CCCTGAC	--	10	3	11	---CGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
Spu	AACTATCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	34	AGC	CCCTGAC	--	6	6	6	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
PP	GTCCCTCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	13	AGC	CCCTGAC	--	7	3	7	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
AU	GGTTGCTTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	17	AGC	CCCTGAC	--	7	9	7	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
PU	AAACCTCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	19	AGC	CCCTGAC	--	19	4	18	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
PY	TAACTCTTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	19	AGC	CCCTGAC	--	15	4	16	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
PA	TAACTCTTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	19	AGC	CCCTGAC	--	14	4	13	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
ML0	TAAAGTCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	16	AGC	CCCTGAC	--	8	5	8	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
SM	AACTCTCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	34	AGC	CCCTGAC	--	8	3	8	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
BME	GCTTCTCTTC	GGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	17	AGC	CCCTGAC	--	10	15	10	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC
BS	ATCAATCTTC	GGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	18	AGC	CCCTGAC	--	5	4	5	----AGGATTCGCTGAGAT	CTGGAGCCGCACAGT-AC-AGT	CTGGAT-GGGAGAGGATGGAG
BQ	GTCTATCTTC	GGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	27	AGC	CCCTGAC	--	3	5	3	----AGGATTCGCTGAGAT	CTGAAGCCGCACAGT-AT-AGT	CTGGAT-GGGAGAGGATGGAG
BE	ATTCATCTTC	GGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	20	AGC	CCCTGAC	--	3	4	3	----AGGATTCGCTGAGAT	CTGGAGCCGCACAGT-AT-AGT	CTGGAT-GGGAGAGGATGGAG
CA	AATGATCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	2	AGC	CCCTGAC	--	3	4	3	----TATGATCCGCTTTGAT	CTGGAGCCGCACAGT-AT-AGT	CTGGAT-GGGAGAGGATATAT
DF	GAAATCTTTC	GGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	2	AGC	CCCTGAC	--	6	4	6	----GATTTGCTGAGAT	CTGAAGCCGCACAGT-AT-AGT	CTGGAT-GGGAGAGGATATTT
EF	GCTTCTCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	3	AGT	CCCTGAC	--	5	3	5	----ATTGAAATTCGCTGAAAT	CTGAAGCCGCACAGT-AT-AGT	CTGGAT-GGGAGAGGATGGAG
LLX	AAATATCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	21	ACT	CCCTGAC	--	4	4	4	----TTGAAAGCAGTGGAAAT	CTGCTAGCCGACAGT-AT-AGT	CTGGAT-GGGAGAGGATGAA
LO	GTCTATCTTC	GGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	3	AGT	CCCTGAC	--	3	10	3	----TTGACTCTGCTGAAAT	CTGGAGCCGCACAGT-AT-AGT	CTGGAT-GGGAGAGGATGTTG
PN	AAGAATCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	125	AGT	CCCTGAC	--	3	4	3	----GATGCTGCTGAGAT	CTCAAGCCGCACAGT-AT-AGT	CTGGAT-GGGAGAGGATGAAA
ST	AAGTCTCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	14	AGT	CCCTGAC	--	3	4	3	----GATGCTGCTGAAAT	CTCAAGCCGCACAGT-AT-AGT	CTGGAT-GGGAGAGGATGAAA
MN	AAGTCTCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	104	AGT	CCCTGAC	--	3	4	3	----GATGCTGCTGAAAT	CTCAAGCCGCACAGT-AT-AGT	CTGGAT-GGGAGAGGATGAAA
SA	ATTCATCTTC	GGGG-TC	GGGTGAAAAT	CCC	ACC	GGCCAGT	6	AGC	CTCCGAC	--	11	3	11	----CTGACTAGTGGAAAT	CTAGAGCCGCACAGT-AT-AGT	CTGGAT-GGGAGAGGATGGAG
AMI	TCACAGTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	14	AGC	CCCTGAC	--	5	5	5	----TGAATCTGCTGCAAT	CTAGAGCCGCACAGT-AT-AGT	CTGGAT-GGGAGAGGATGAAA
DFA	ACGAACCTTC	GGGG-TC	GGGTGAAAAT	CCC	ACC	GGCCGGT	20	AGC	CCCTGAC	--	11	4	11	----CGACTGCTGCTGCAAT	CTGAAGCCGCACAGT-AT-AGT	CTGGAT-GGGAGAGGATGAAA
EN	AATAATCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	2	AGT	CCCTGAC	--	4	6	4	----GATTTGCTGAAAAT	CTGAAGCCGCACAGT-AG-AGT	CTGGAT-GGGAGAGGATGAAA
GLU	---TGTCTTC	AGGG-CA	GGGTGAAAAT	CCC	ACC	GGCCGGT	28	AGC	CCCTGAC	--	10	4	10	GTCAAGCAGATCTGCTGAGAG	CCAGAGCCGCACAGTCAA-AGT	CTGGAT-GGGAGAGGATAGTGC

# Аттенюация транскрипции

Antiterminator

The *RFN* element

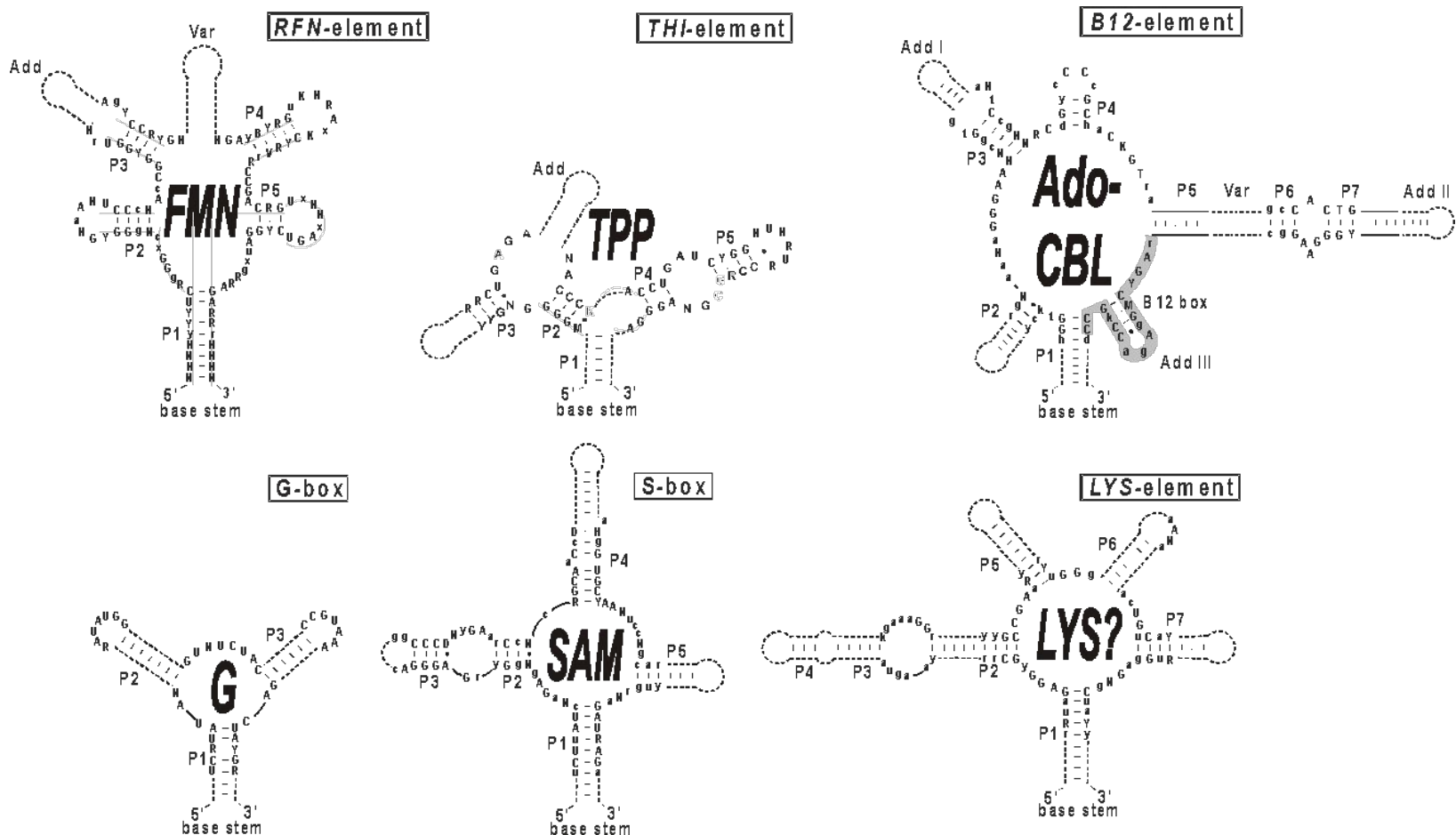
Terminator

<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--	ATCCTTGGGGTTTTTGTTTTTTTT
<b>BE</b>	ACATAACGATATAGTGAT <b>GCACTCCTTCGGGGC</b> AGGGTG	---	TCTGGATGGGGAAGGATGC	155	-----TG <b>AGCCCGGGG</b> CAT-----	CCCGGGGTTTCATTTTTATTG
<b>HD</b>	AAATTGAATAATTAATTTTTC <b>TCCCTTCGGGGC</b> TGGGTG	---	TCTGGATGGGAGAAGGAAAAC	148	-----AT <b>CCCCCGT</b> GAGAACAAAA--	TCTCTGGGGCTTTTTTGCGCGC
<b>CA</b>	TAATGGTAATTTAATAGGAT <b>TCTTCAGGGAT</b> TGGGTG	---	TCTGGATGAAAGAAGAAATA	34	----- <b>AATCTCGAAGGA</b> TTACC----	TTTCTTTGAGATTTTTTTATTG
<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>	ATCATCTGTAAATGAATAACTAT <b>TTCAGGGC</b> AGGGTG	---	TCTGGATGAAAGAAGATAAA	81	-----TGTAT <b>GCCTTGAG</b> TAGTCCCC--	TATTCAGGTATATTTTTTTGGAGG
<b>PN*</b>	ATCATCTGTAAATGAATAACTAT <b>TTCAGGGC</b> AGGGTG	---	TCTGGATGAAAGAAGATAAA	19	-----CG <b>TGCTGAAATGA</b> TTACTTG	TCATTTTCAGAGCATTTTTGTTAATC
<b>TM</b>	AAAAC <b>TGAATACAAAGAAACGCTCCTTCGGGGC</b> AGGGTG	---	TCCGGATGGGAGAGAGCGTG	13	-----ATGGG <b>ACCCGAGA</b> -----	<b>GGTCCCTTTTC</b> TTTACA
<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>TGCTCAGAC</b> TTT-----	GTTTGAGCATTTTTTTATTAA
<b>SA</b>	TAATAACAATTTCAATATAAT <b>TCTTCGGGGC</b> AGGGTG	---	TCTGGATGGGAGAAAGAAATG	74	----- <b>TTTTCTCC</b> TTGCACTTAATT--	GATGTGAGGATTTTTGTTTATA
<b>DHA</b>	ACTCTTTTTAGATGAATACGA <b>ACTTCGAGGTA</b> AGGGTG	---	TCCGGATGGGAGAAGGTACA	43	-----GTTT <b>ATGCTCGAGGA</b> CACCAT <b>TTCTC</b> CGAGGCATTTTTGTTCTTTC	
<b>FN</b>	GAAAAATAAATATTAATAAATA <b>TCTTCGGGGC</b> AGGGTG	---	TCTGGATGAGAGAAGAAAG	40	-----CTT <b>ACCCGAAT</b> TTCTAT-----	AATTCGGTTTTTTTTATTTT
<b>CA</b>	AAATAAAAAAATAAAGAATGAT <b>TTCAGGGC</b> AGGGTG	---	TCTGGATGAAAGAAGATAATA	19	-----TAT <b>GCCTTGA</b> CGTTTTT-----	CGTTGGGGCTTTTTTAATGCT
<b>DF</b>	AAAATTAAAAAATCAAGAAG <b>ATCTTCGGGGC</b> AGGGTG	---	TCTGGATGAGAGAAGATAAT	45	-----ATAAAAA <b>CTCGAAGAT</b> AGGG--	TCTTCGAGTTTTTTGTTTTTCCTAA
<b>BS</b>	TAATTAATTTCAATATGATCAAT <b>TCTTCGGGGC</b> AGGGTG	---	TCTGGATGGGGAAGATGGA	103	--AAAGAAC <b>CTTTCGGTTTTTC</b> GAATAAGATGTGA <b>TCGAAAAGGAG</b> AGAATGA <b>AGTG</b> AAA	
<b>BQ</b>	GGGAAAATAGAAATATCGG <b>TCTATCTTCGGGGC</b> AGGGTG	---	TCTGGATGGGAGAGATGGA	54	-----AT <b>TTCTCC</b> TTTGTGTAAA--	ACACA <b>AGGG</b> TTTTTTTCGTTCT <b>ATG</b>
<b>BE</b>	ATAAAAAATGATAAGCGA <b>TTTCACTCTCGGGC</b> AGGGTG	---	TCTGGATGGGAGAGATGAA	114	-----GGCAGC <b>CTTCTCTT</b> GTGAGGATGAAT <b>TCACGAGAAGGGAGGAG</b> AACAAGC <b>ATG</b>	
<b>PN</b>	GTTTTTTGTTATGATAAAAGAG <b>TCTTCAGGGC</b> AGGGTG	---	TCTGGATGGGAGAGACGAA	137	--AA <b>CTTCTCT</b> GTATTTTAG-----	AAAAT <b>TGGAG</b> GAACCTGTT <b>ATG</b> ACA
<b>ST</b>	TAAATCTGCTATGCTAGAAG <b>TCTTCAGGGC</b> AGGGTG	---	TCTGGATGGGAGAGACCGG	130	--GG <b>ACTTCTTTT</b> CAATTTGAAA--	AAAT <b>TGGAG</b> GAATTTTTTA <b>ATG</b> TC
<b>MN</b>	ATTTTTTGATAATGCTATAAG <b>TCTTCAGGGC</b> AGGGTG	---	TCTGGATGGGAGAGACTGA	138	--GG <b>CTTCTTTT</b> CGATTTGTAA--	AAAT <b>TGGAG</b> GAATTTTTTT <b>ATG</b> AA
<b>SA</b>	AAATTTAATAATGATAAA <b>TTTCACTCTTCGGGGC</b> AGGGTG	---	TCTGGATGGGAGAGATGGA	17	----- <b>TTCTCTATTC</b> TTACG-----	AGATGAATGGA <b>AGGAG</b> AAAAATGAAT <b>ATG</b>
<b>EF</b>	AAAAAATAAATAAAGGTT <b>TCTTCAGGGC</b> AGGGTG	---	GTCTGGATAAAGAAGATAGG	33	-----CTACTCTATTTTT <b>CCCTGCAG</b> A-----	AAAA <b>TGGGG</b> TTTTTTTTGT <b>ATG</b> A
<b>LLX</b>	TTTTTGTGCTATAATAAAAA <b>TCTTCAGGGC</b> ACCGTG	---	TCTGGATGGAAGAAGATGAA	66	--TCAACT <b>CTCTCGAA</b> TT <b>TGAAGA</b> AT-TATT <b>TTCTCATA</b> TT <b>TGGAG</b> GTTTTTTTT <b>ATG</b> T	
<b>LO</b>	ATTGTAAGAAAATATTCGTT <b>CATCTTCGGGGC</b> AGGGTG	---	TCTGGATGGGAGAGATGTTG	79	---ATGCACAAAC <b>TCTCC</b> TCAACTTTTTTTA-----	<b>GTGAGG</b> TTTTTTATTTCG

Antiterminator



# Другие РНК-переключатели, найденные методами сравнительной геномики



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

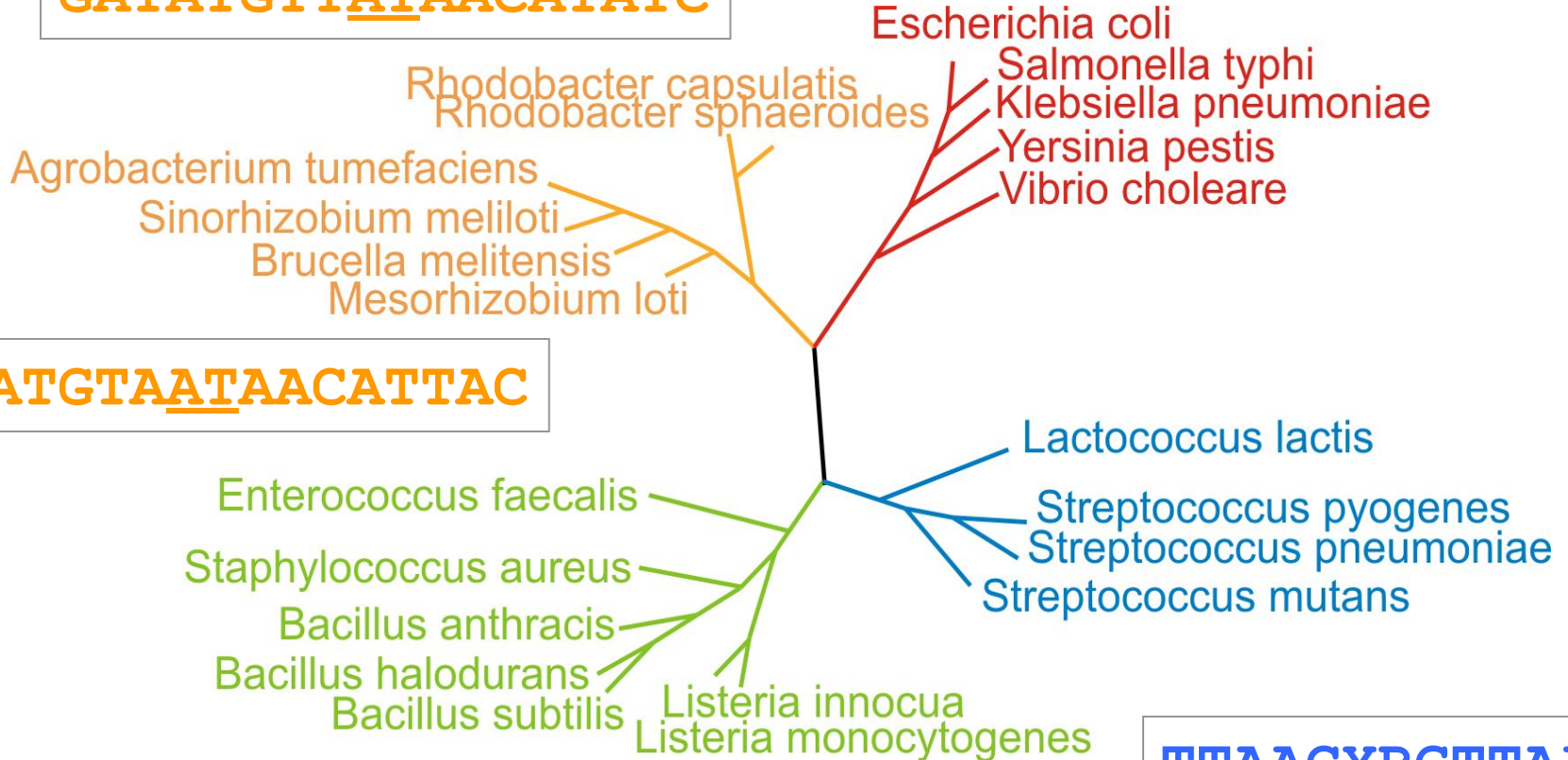
# Регуляторы гомеостаза цинка

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>	(-)	(-) - -	-	(-) +

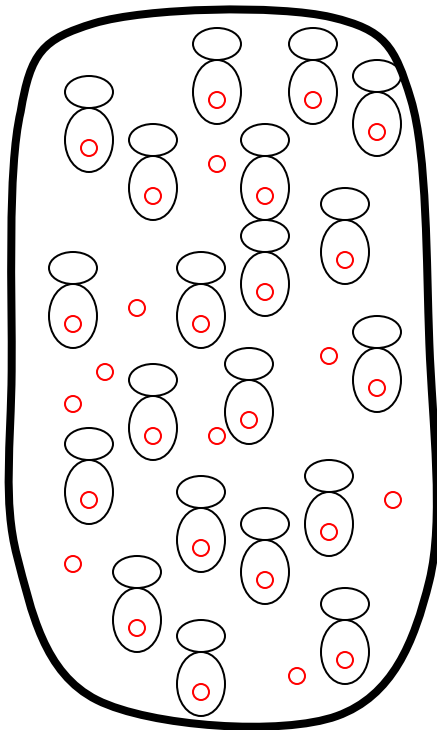
nZUR

pZUR

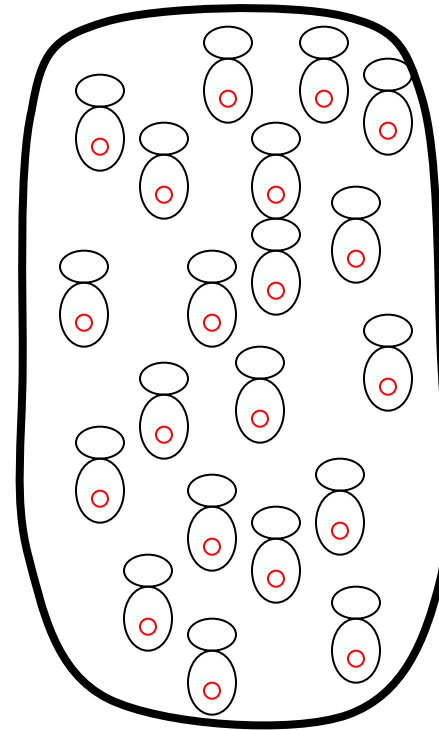
AdcR

# Плохой сценарий

достаточно  
цинка

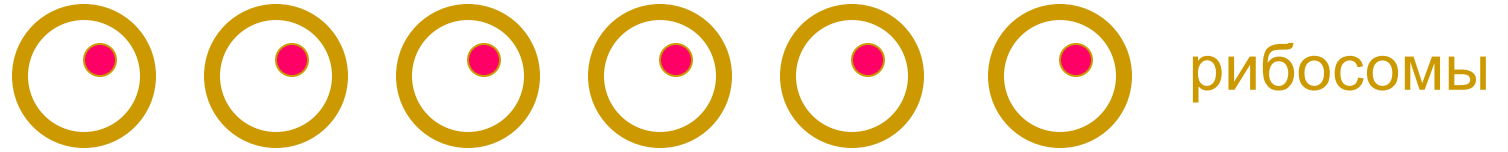


недостаточно цинка: весь цинк  
использован рибосомами, не  
хватает цинка для ферментов

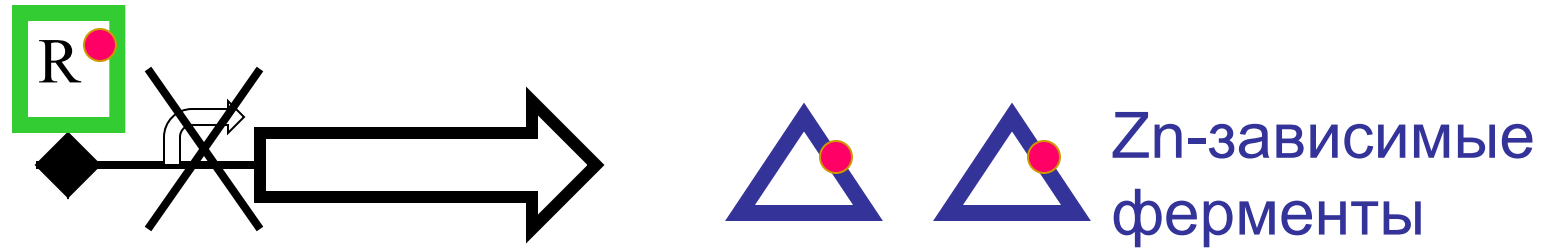


# Регуляторный механизм

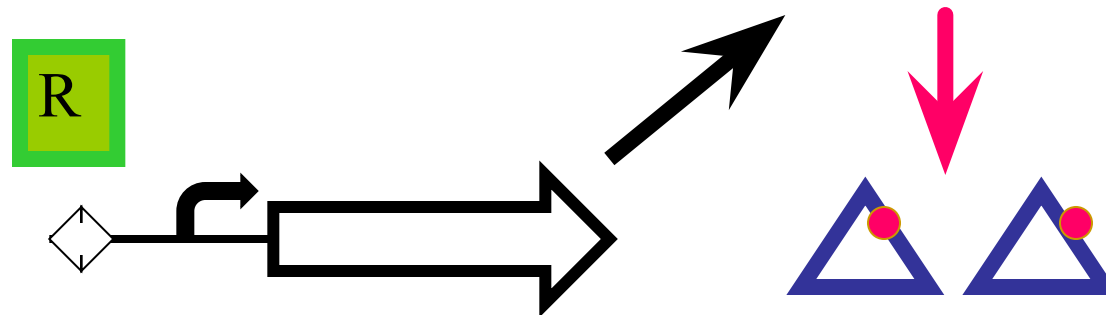
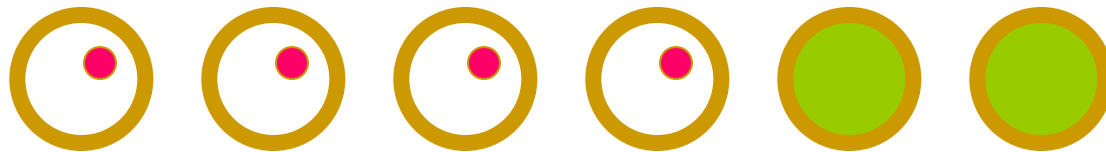
Достаточно **Zn**



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\*†, 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,<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

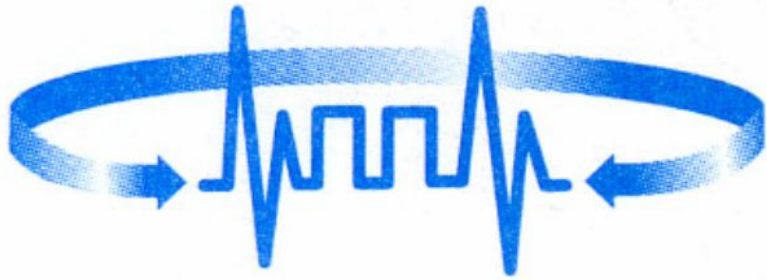
- Молекулярная эволюция
  - филогения генов
  - таксономия организмов
  - горизонтальные переносы и т.п.
  - положительный и отрицательный отбор
    - что сделало нас людьми?
    - лекарственная устойчивость
  - эволюция геномов
- Системная биология
  - строение геномов
  - сети взаимодействий
    -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

# Перспективы

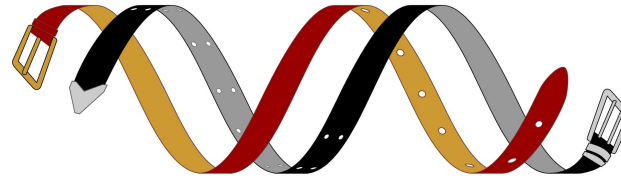
- **Индивидуальные геномы**
  - персональные человеческие геномы
  - геномы штаммов
- **Метагеномы**
  - некультивируемые бактерии
  - метагеномы экологических ниш
    - бактериальный метагеном человека
- **Другие виды данных**
  - интеграция
- **Моделирование (пока рано)**
  - построение полных карт

# 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