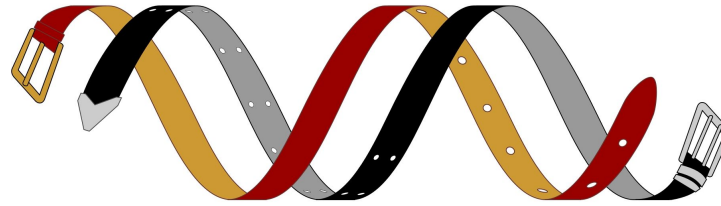


*Учебно-Научный Центр*



*Биоинформатика*

# **Биоинформатика, или молекулярная биология *in silico***

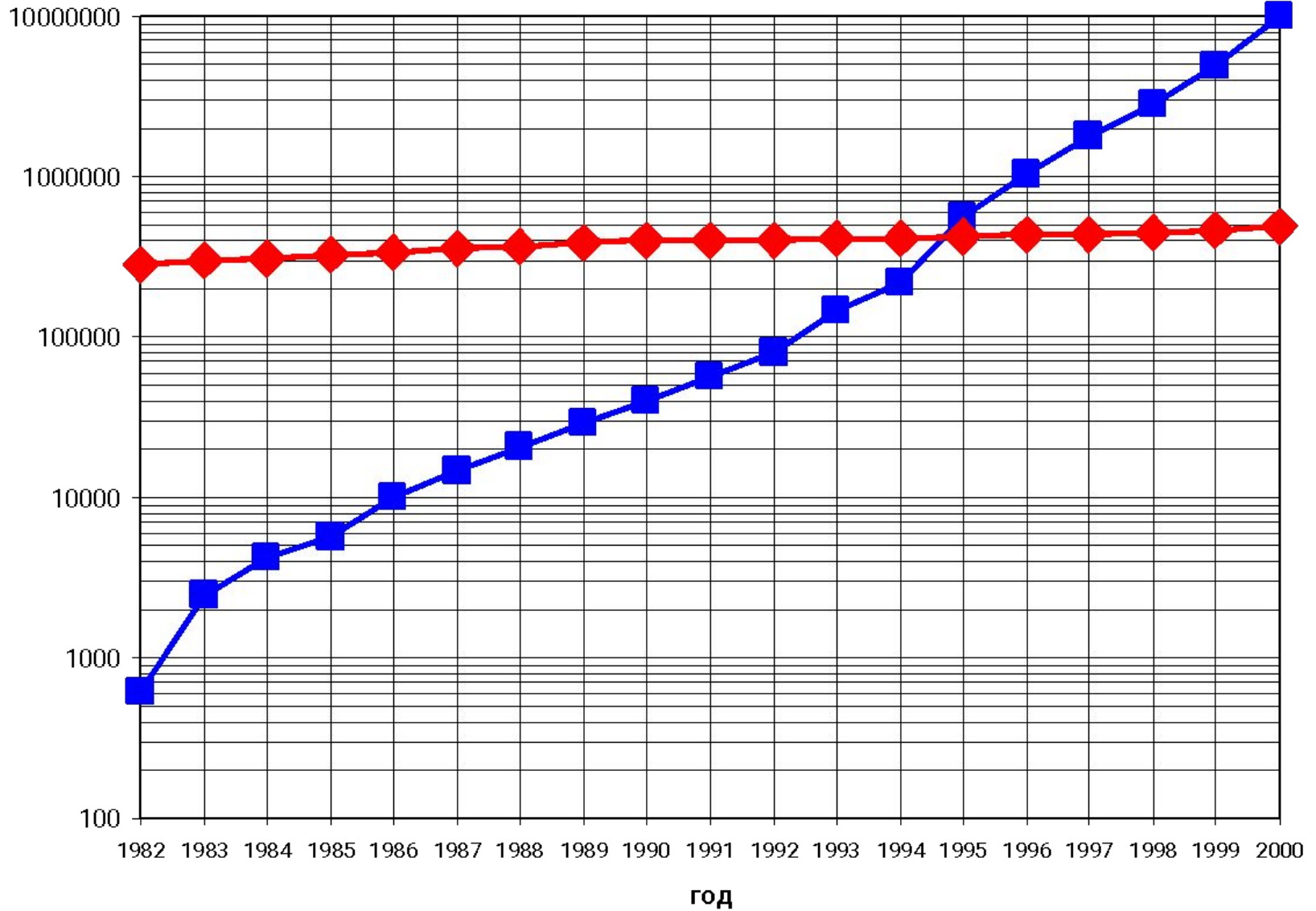
М.Гельфанд

Семинар в ИППИ 7 апреля 2006

# Пропаганда 1

красный: статьи

синий: последовательности



# Анализ индивидуальных генов

- Поиск родственных белков в банках последовательностей – перенос функции от гомологов
- Функциональные сайты (каталитические центры)
- Функциональные участки (трансмембранные сегменты, сигнальные пептиды и т.п.)

- Анализ на уровне индивидуальных генов даёт возможность охарактеризовать 50-75% генов в новом геноме

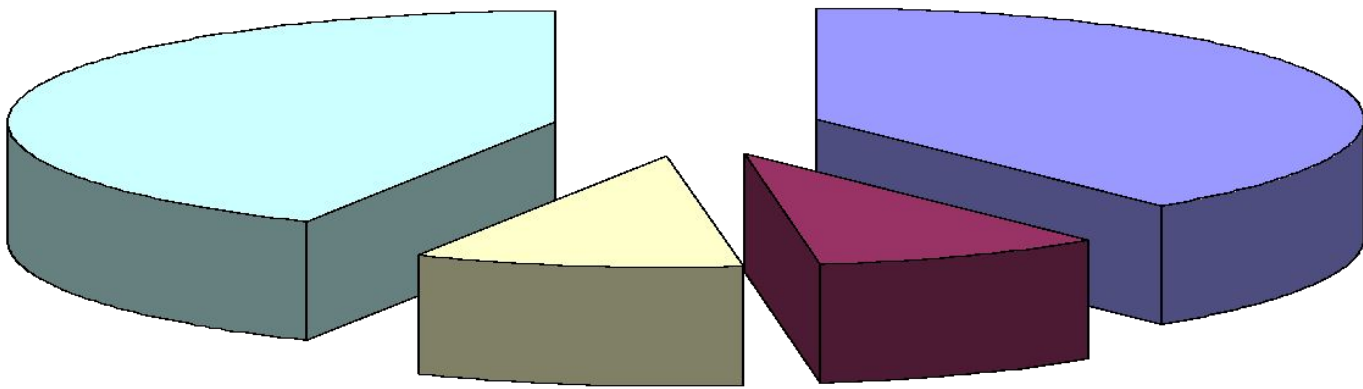
## Но:

- ~100 универсально отсутствующих генов (нет ни одного известного гена для известной функции)
- множество функций, для которых неизвестны представители в больших таксонах
- в каждом геноме ~5-10% консервативных генов с неизвестной функцией
- трудно предсказывать специфичность в мультигенных семействах (транспортёры, факторы транскрипции)
- нельзя найти что-то принципиально новое

# How much do we know about the *Escherichia coli* proteome?

**Characterized experimentally**

**Function inferred by similarity only**

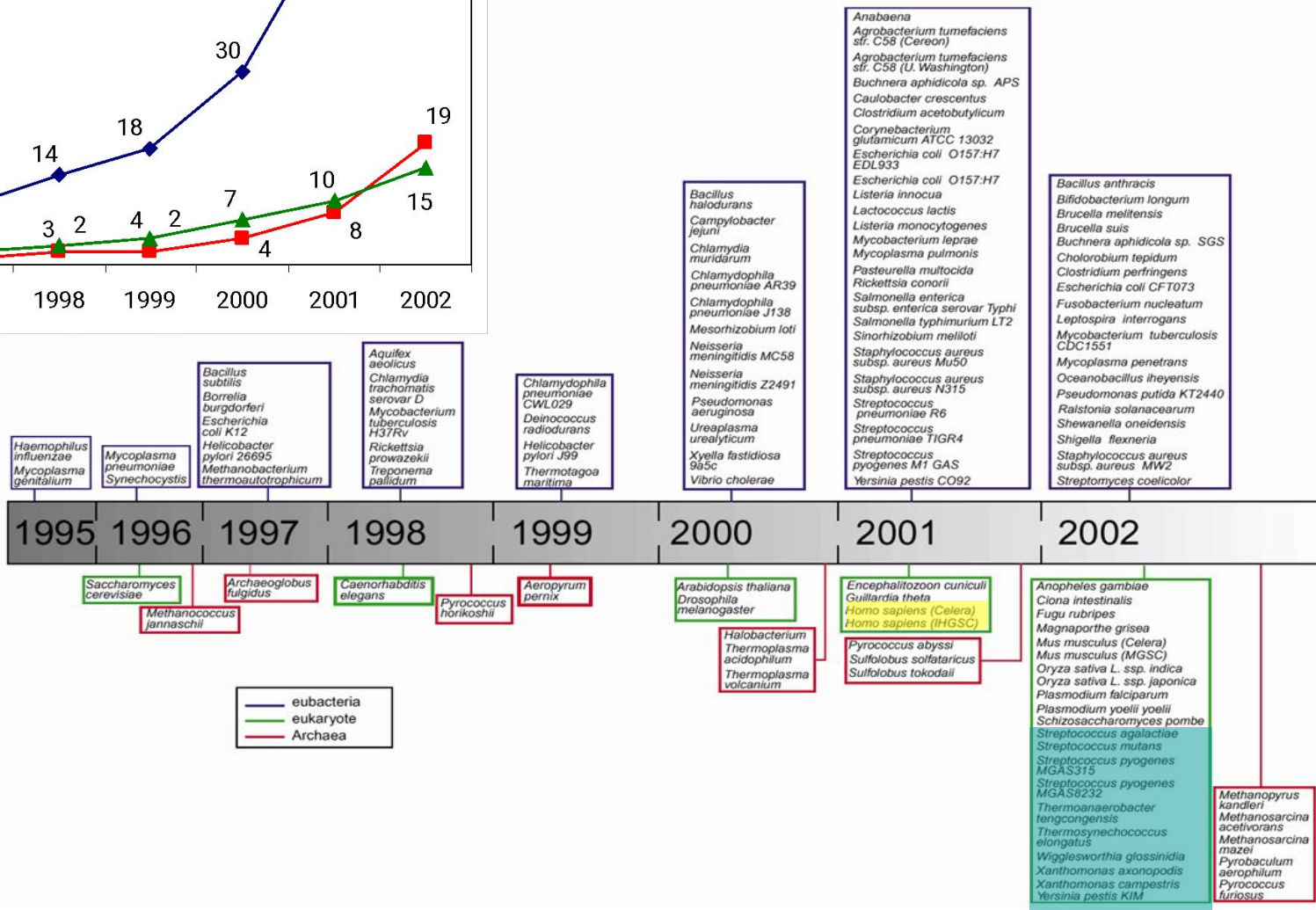
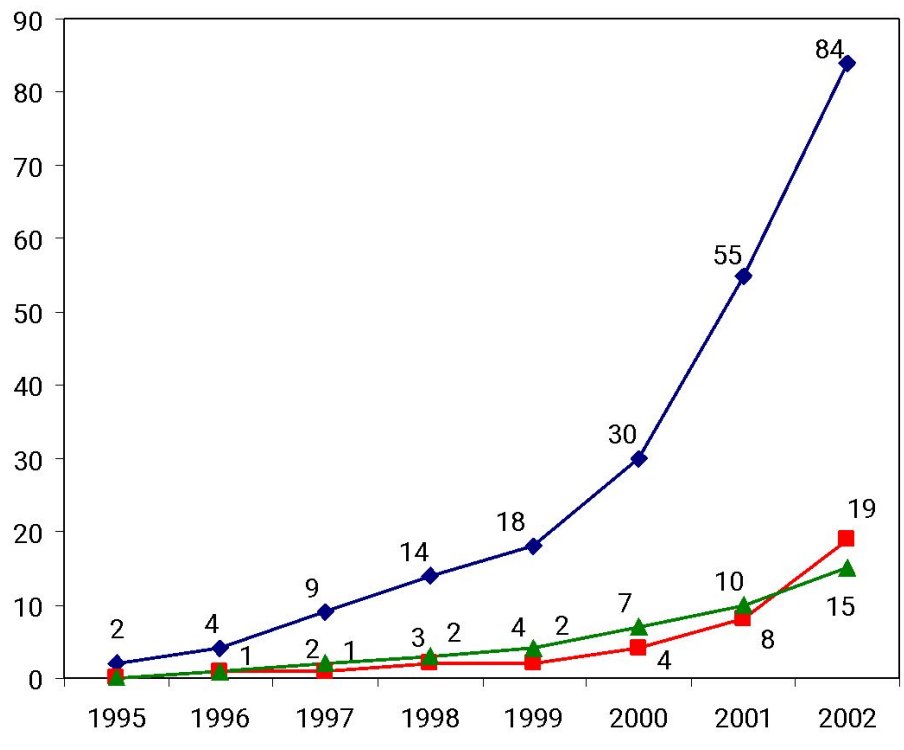


**“Hypothetical”**

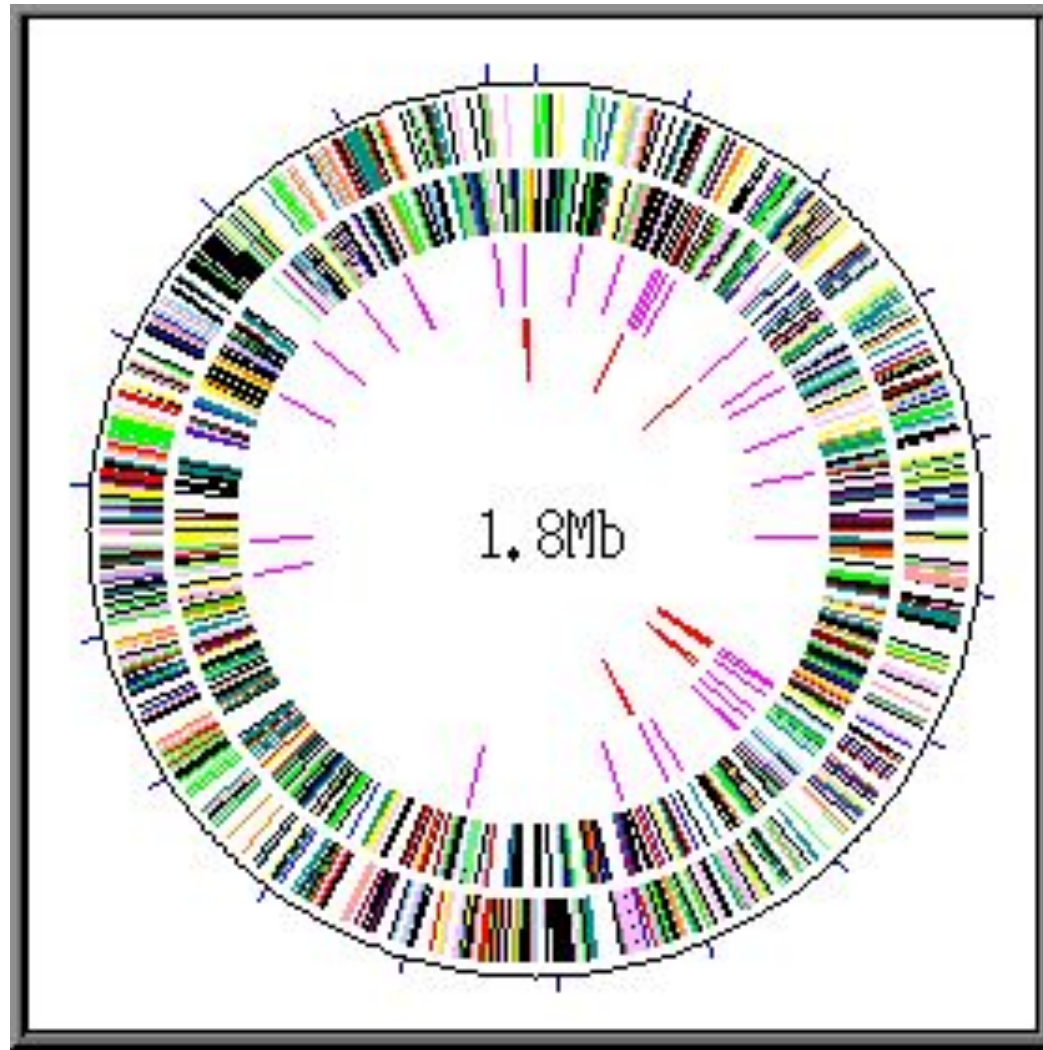
**“Conserved hypothetical”**

# Пропаганда – 2

## Полные геномы



# *Haemophilus influenzae*, 1995



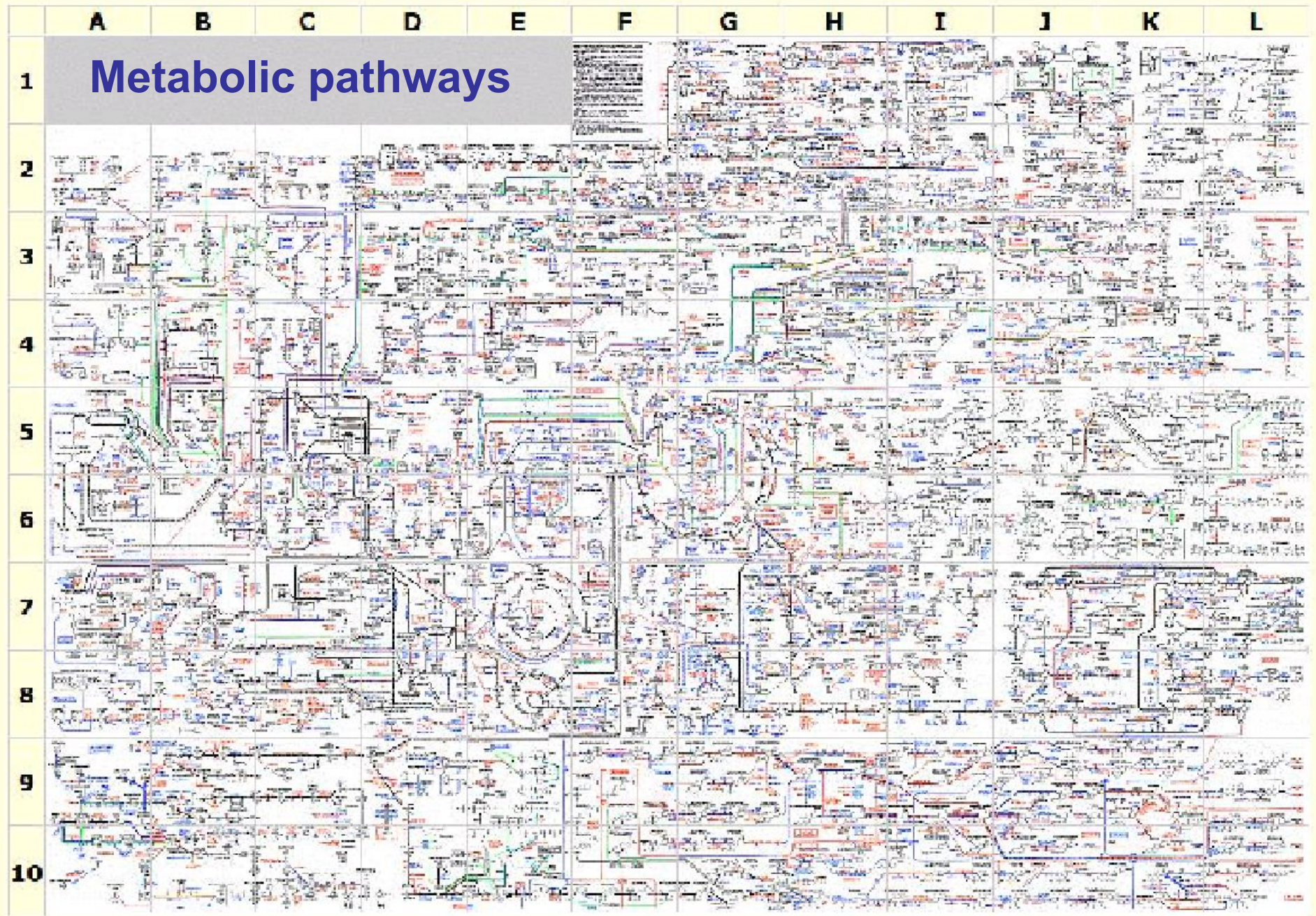






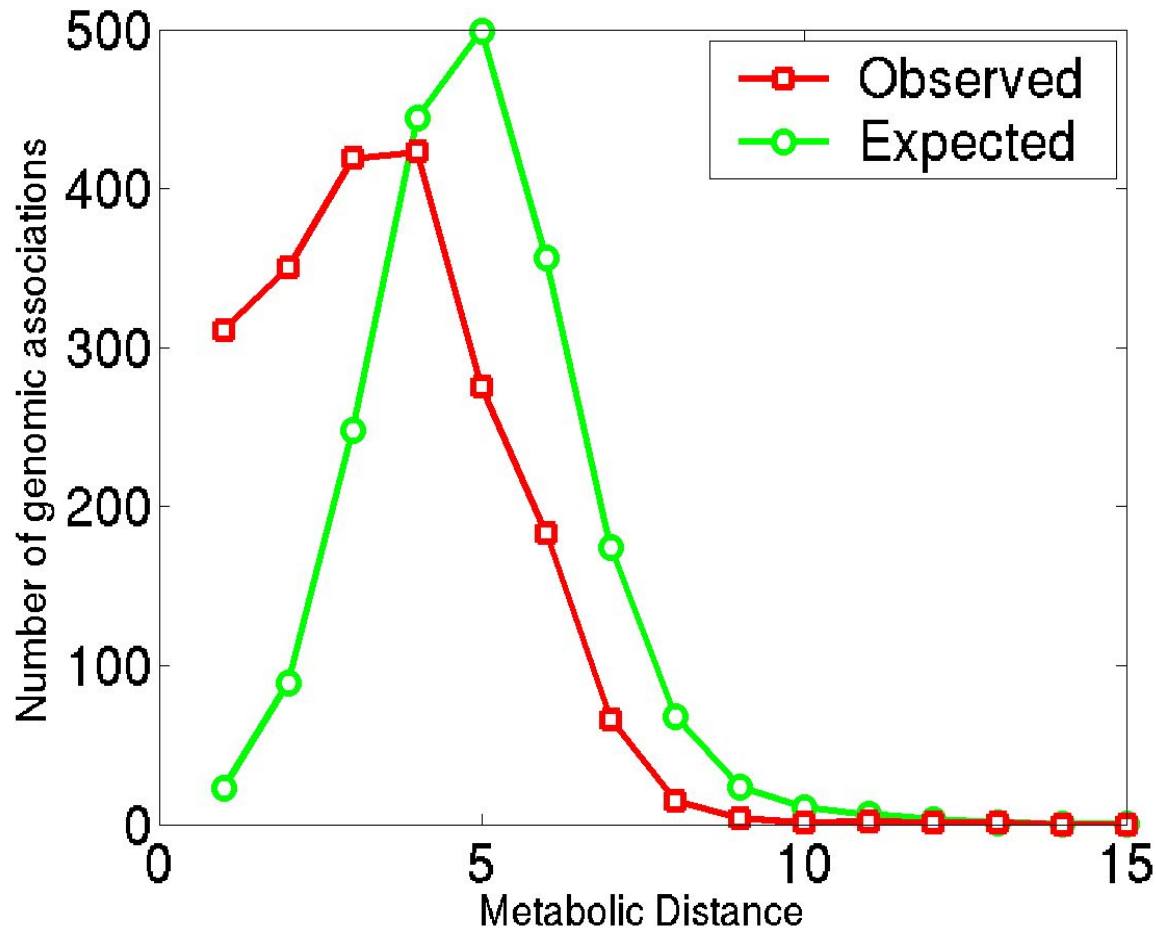
# Сравнительно-геномные подходы

- **Positional clustering**
- **Phylogenetic profiling**
- **Gene fusions**

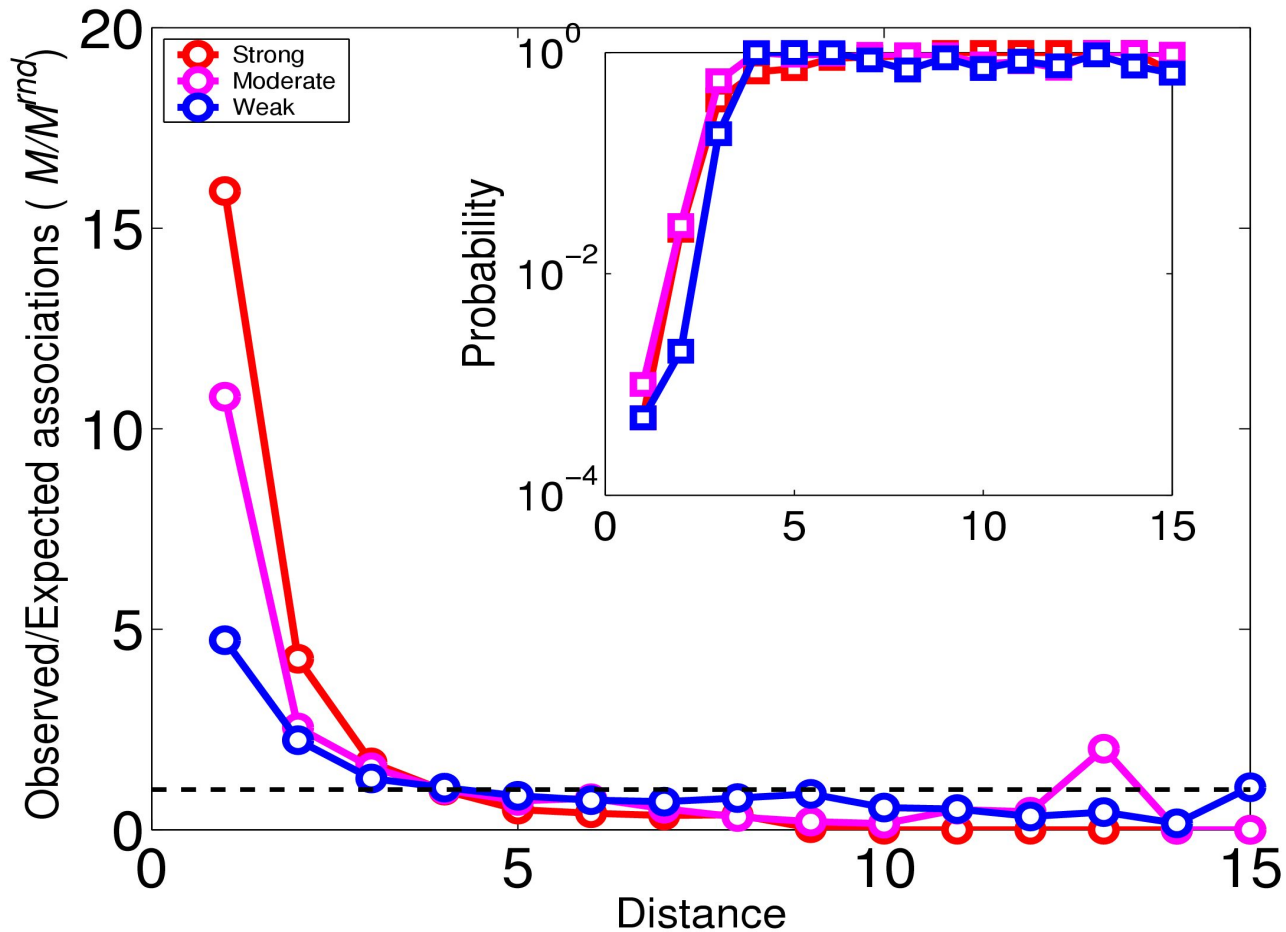




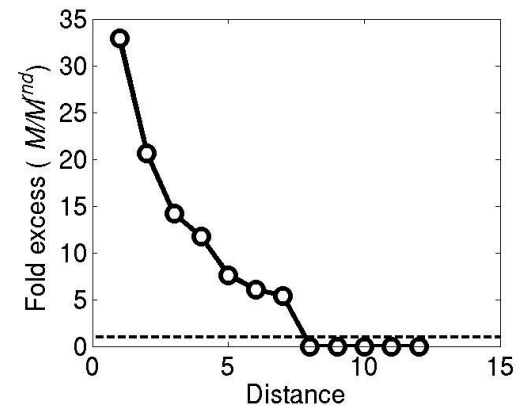
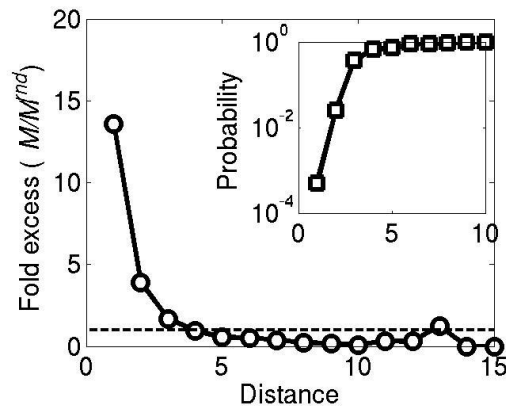
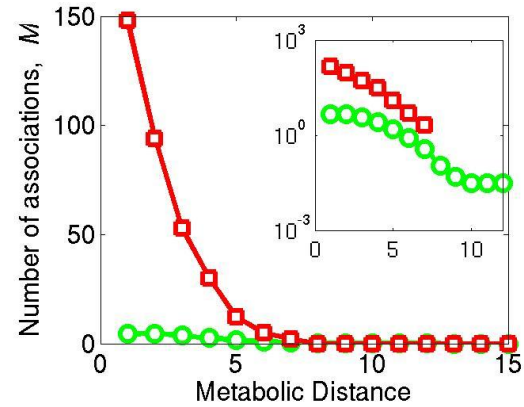
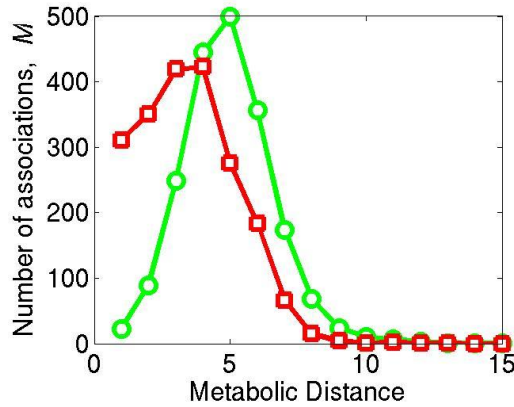
# Functionally dependent genes tend to cluster on chromosomes in many different organisms



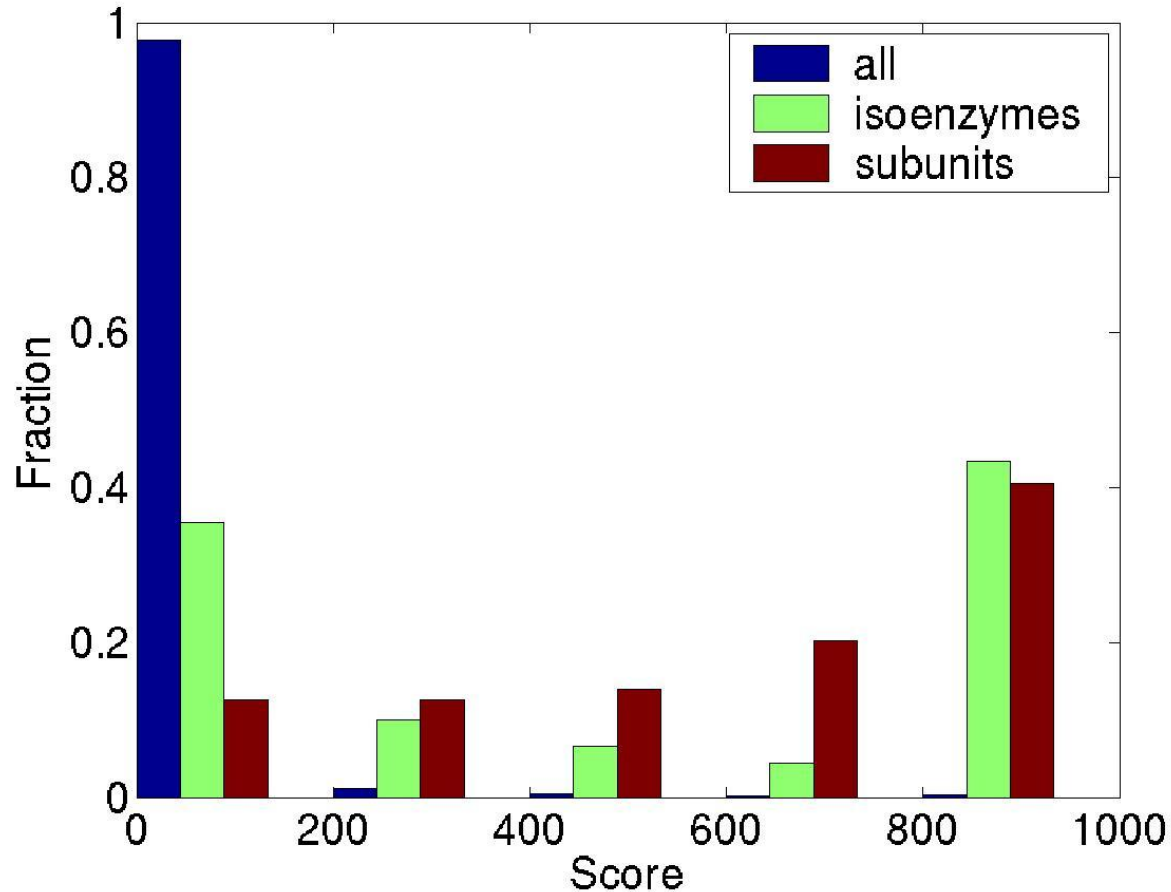
**More genomes (stronger links)  
=> highly significant clustering**



# ... особенно в линейных путях (справа)

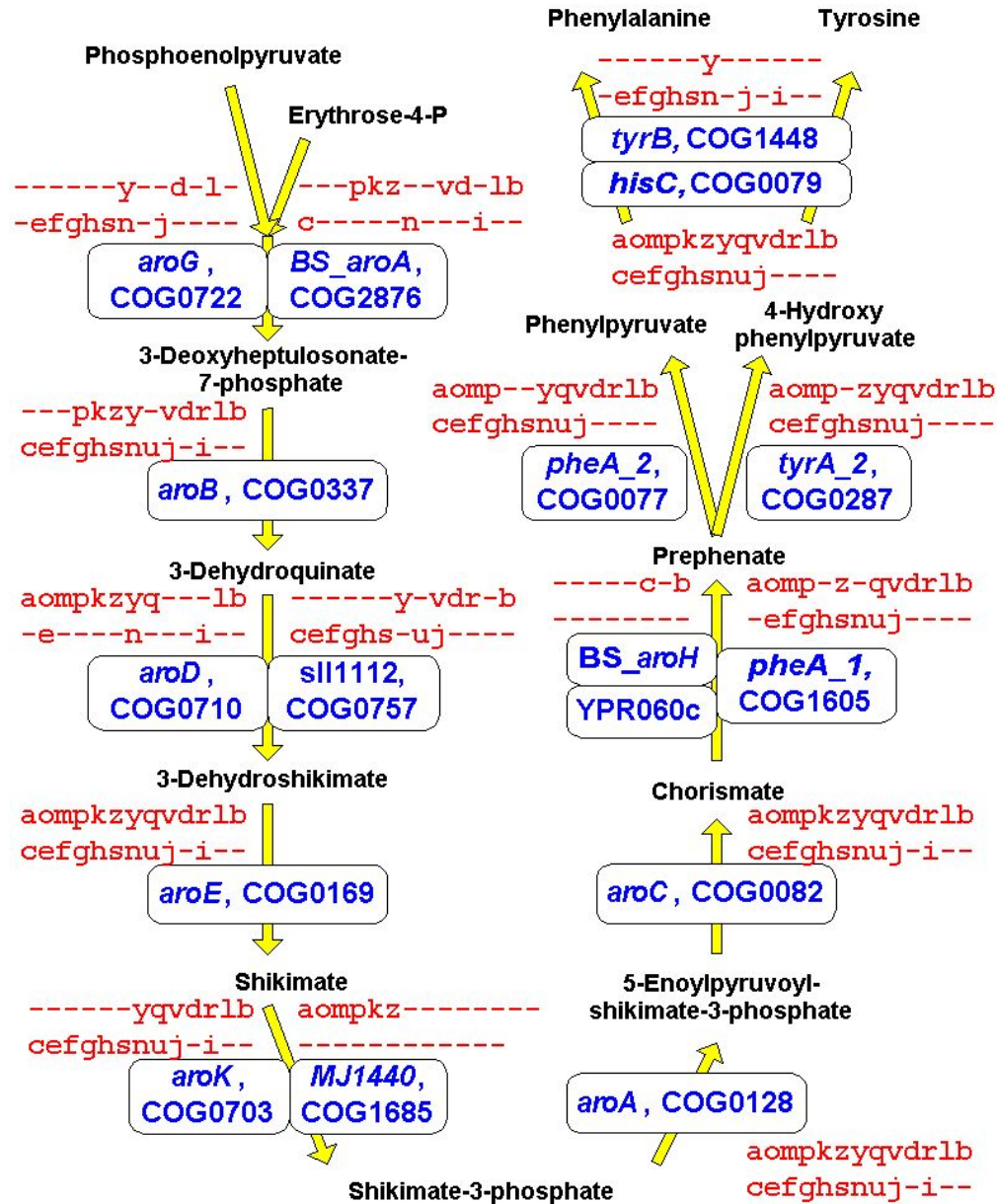


# Распределение уровней связи (бимодальное для изоферментов, монотонное для субъединиц)





# Phyletic profiles in the Phe/Tyr pathway



# Arithmetics of phyletic patterns

## 3-dehydroquinate dehydratase (EC 4.2.1.10):

Class I (AroD) COG0710 aompkzyq---lb-e-----n---i--

Class II (AroQ) COG0757 ------y-vdr-bcefghs-uj-----

Two forms combined

aompkzyqvdr**l**bcefghsnuj-i--

## Shikimate dehydrogenase (EC 1.1.1.25):

AroE COG0169 aompkzyqvdr**l**bcefghsnuj-i--

## Shikimate kinase (EC 2.7.1.71):

Typical (AroK) COG0703 -----yqvdr**l**bcefghsnuj-i--

Archaeal-type COG1685 aompkz-----

Two forms combined

aompkzyqvdr**l**bcefghsnuj-i--

## 5-enolpyruvylshikimate 3-phosphate synthase (EC 2.5.1.19)

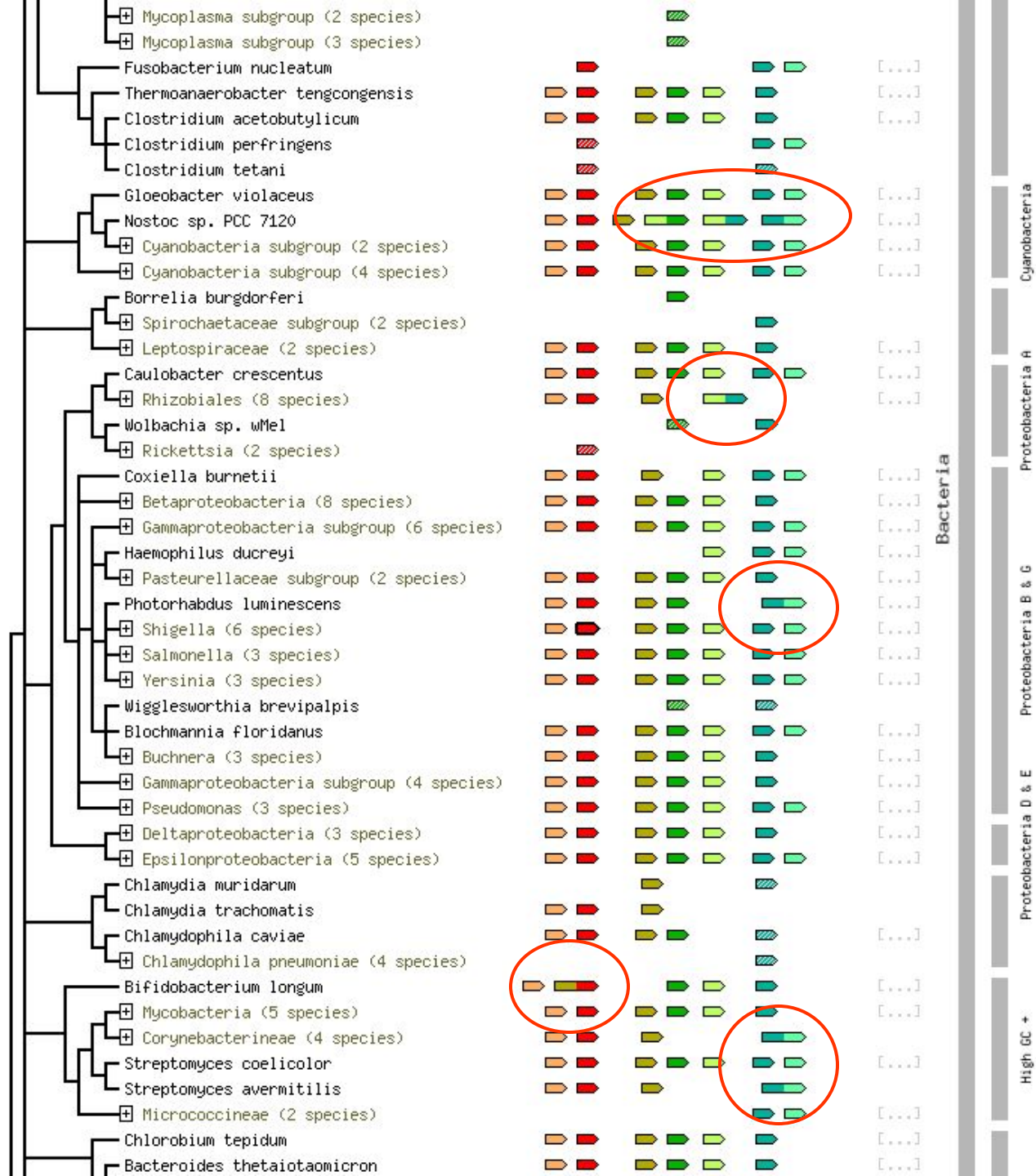
AroA COG0128 aompkzyqvdr**l**bcefghsnuj-i--

## Chorismate synthase (EC 2.5.1.19)

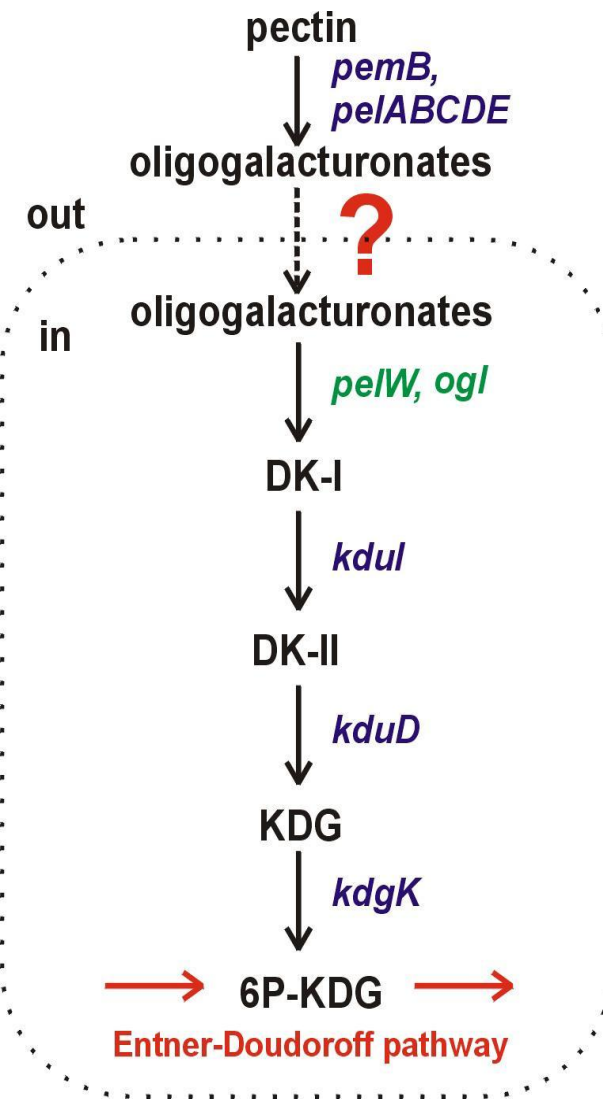
AroC

COG0082 aompkzyqvdr**l**bcefghsnuj-i--

# STRING: *trpB* – fusions



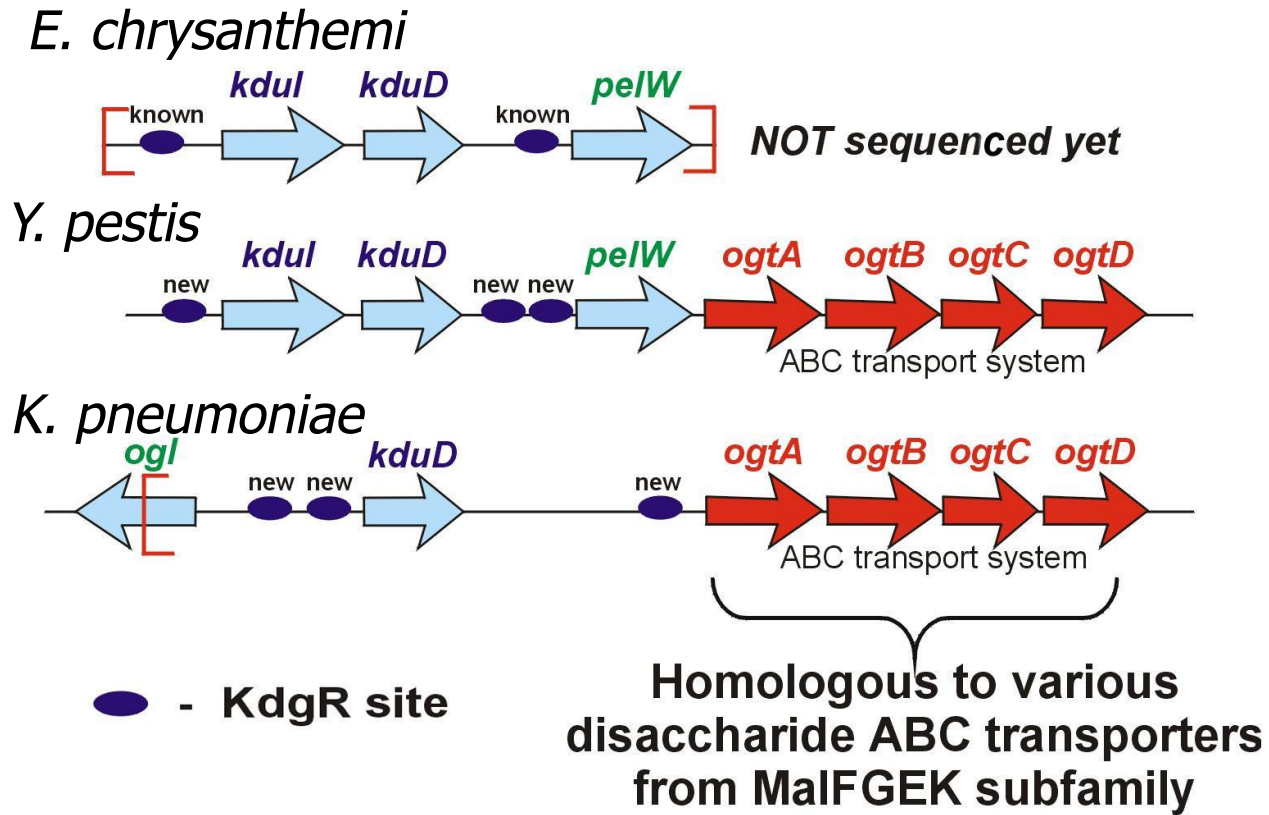
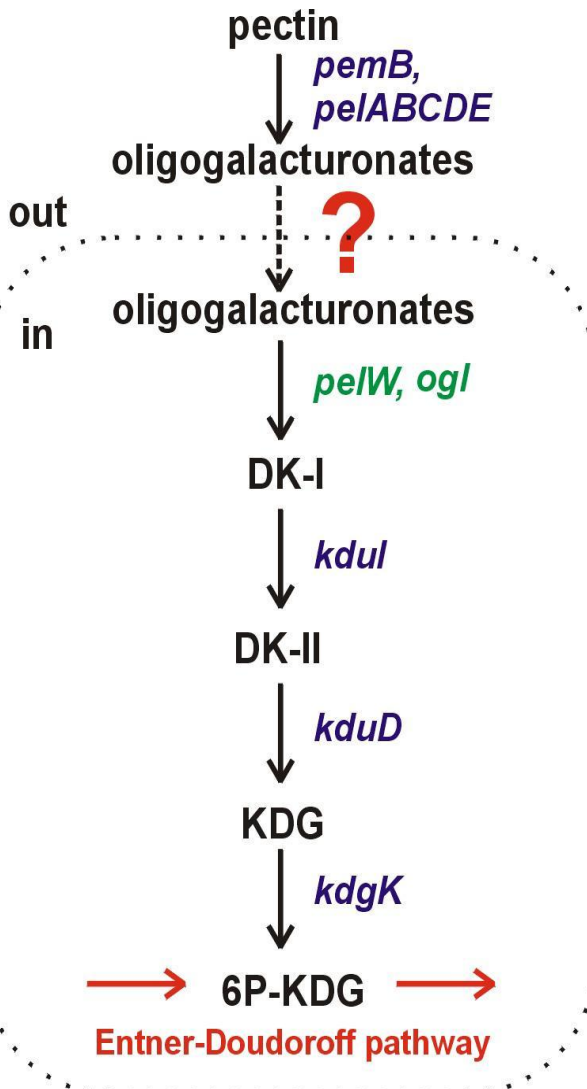
# Утилизация пектина



*E. chrysanthemi*

Entner-Doudoroff pathway

# ... и транспорт олигогалактуронатов





# УраА: транспортёр рибофлавина

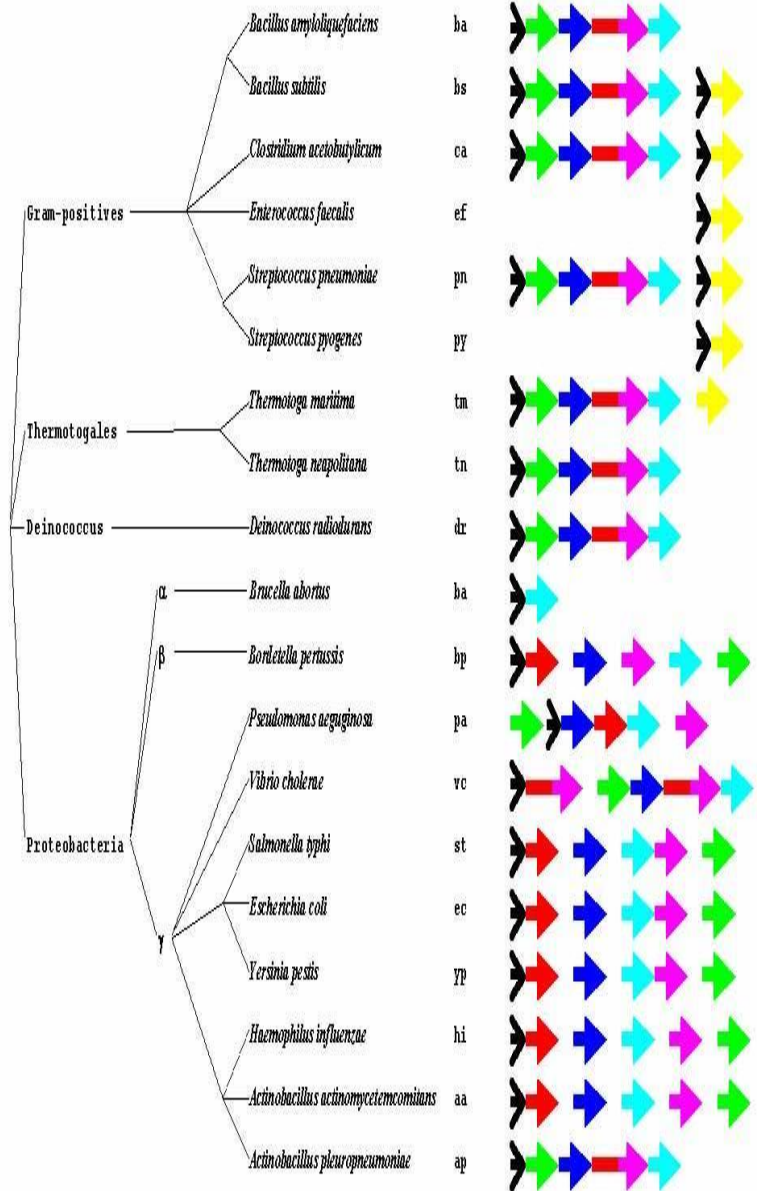
- 5 предсказанных ТМ-сегментов => потенциальный транспортёр
- регуляторный RFN-элемент => ко-регуляция с генами метаболизма рибофлавина => транспорт рибофлавина или предшественника
- *S. pyogenes*, *E. faecalis*, *Listeria*: есть ураА, нет генов биосинтеза рибофлавина => транспорт рибофлавина

## Предсказание:

УраА – рибофлавиновый транспортёр (Gelfand et al., 1999)

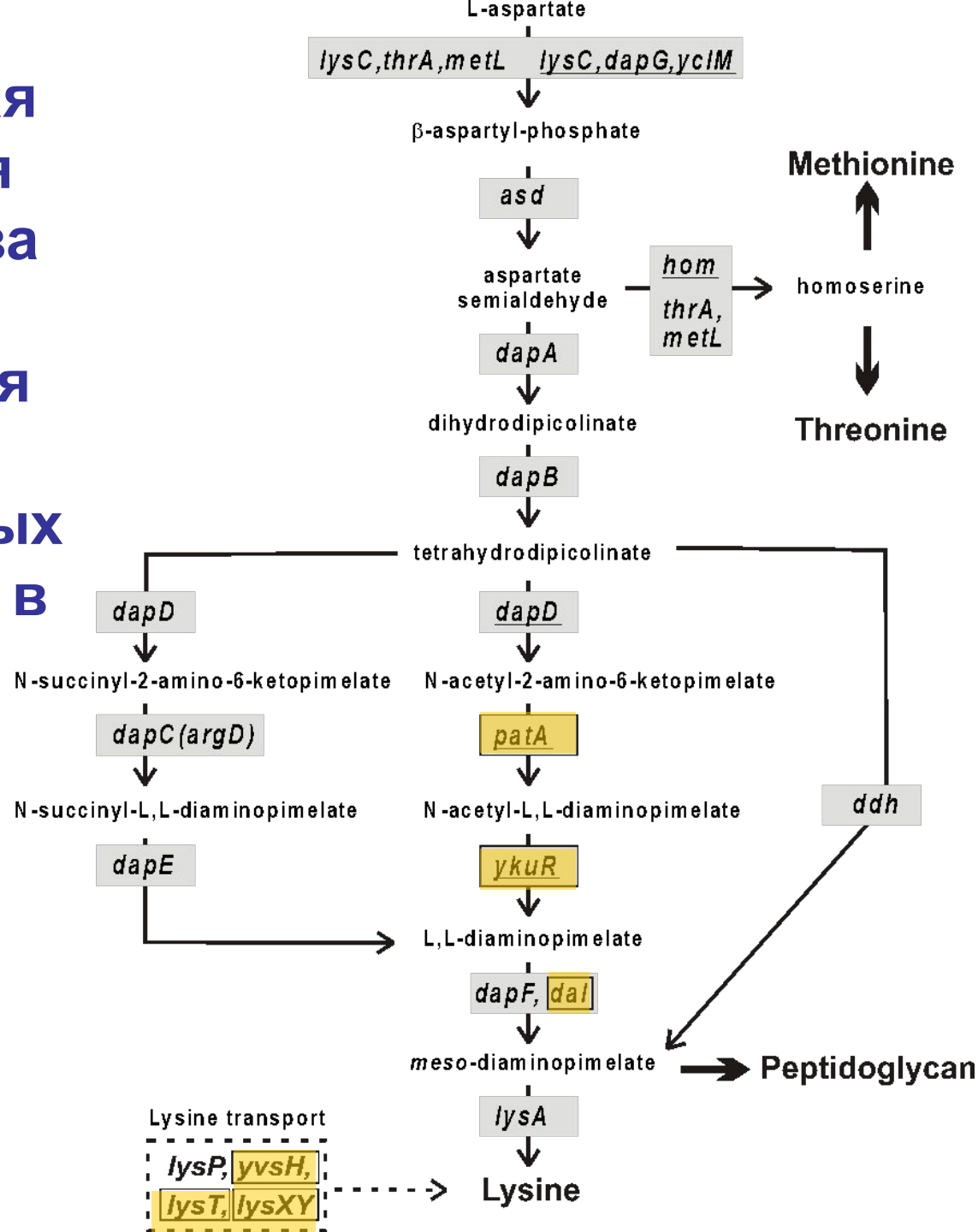
## Проверка:

- УраА переносит рибофлавин (генетический анализ, Кренева и др., 2000)
- ураА регулируется рибофлавином (анализ экспрессии на микрошпотах





**Метаболическая  
 реконструкция  
 пути биосинтеза  
 лизина:  
 Идентификация  
 пути  
 ацетилированных  
 интермедиатов в  
*B. subtilis* и  
 родственных  
 бактериях**



# Идентификация пути ацетилированных интермедиатов - 0

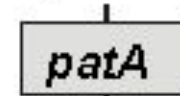
*dapD* (*yquQ*):

- ортолог известного гена *E. coli*

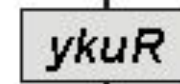
tetrahydrodipicolinate



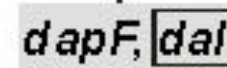
N-acetyl-2-amino-6-ketopimelate



N-acetyl-L,L-diaminopimelate

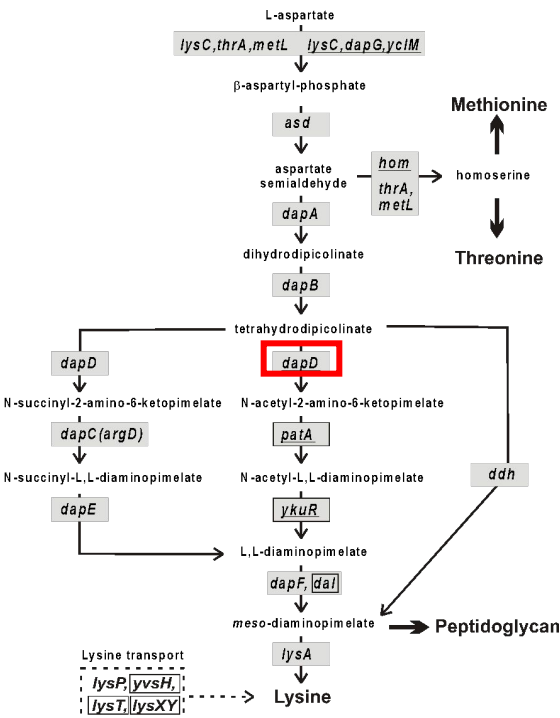


L,L-diaminopimelate



meso-diaminopimelate

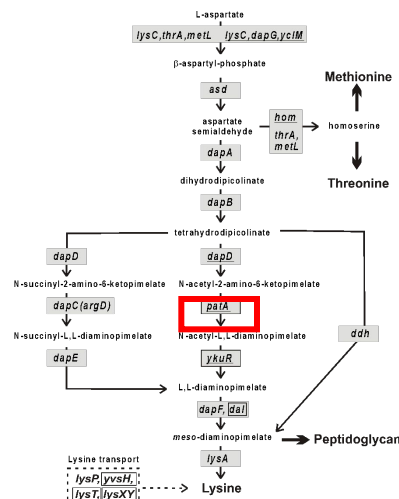
↓



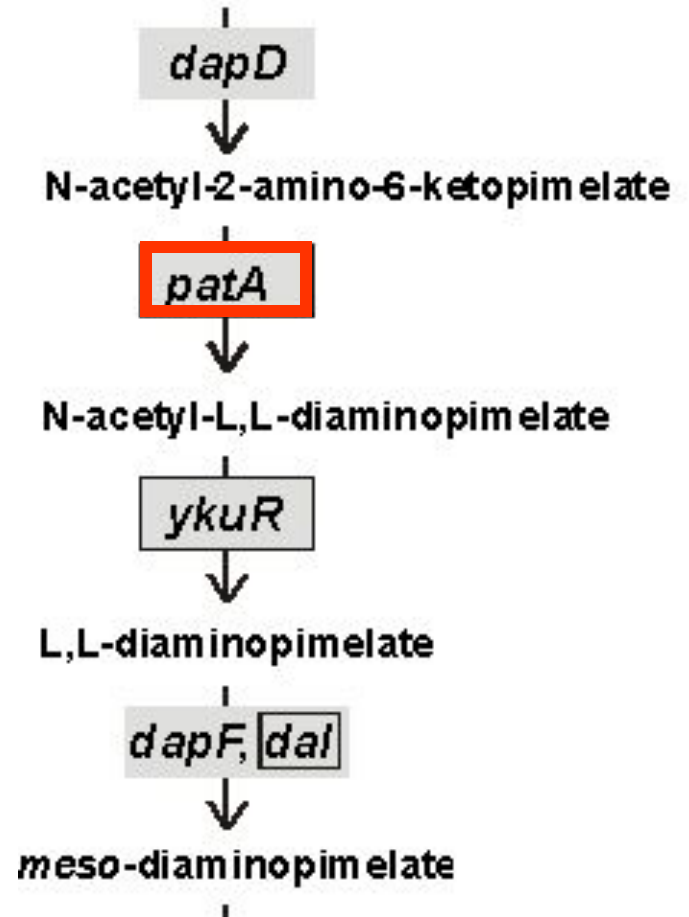
# Идентификация пути ацетилированных интермедиатов - 1

*patA*:

- пиридоксаль-фосфат-зависимая аминотрансфераза (по гомологии)
- ко-локализуется и ко-регулируется с генами биосинтеза лизина во многих грам-положительных бактериях



tetrahydrodipicolinate

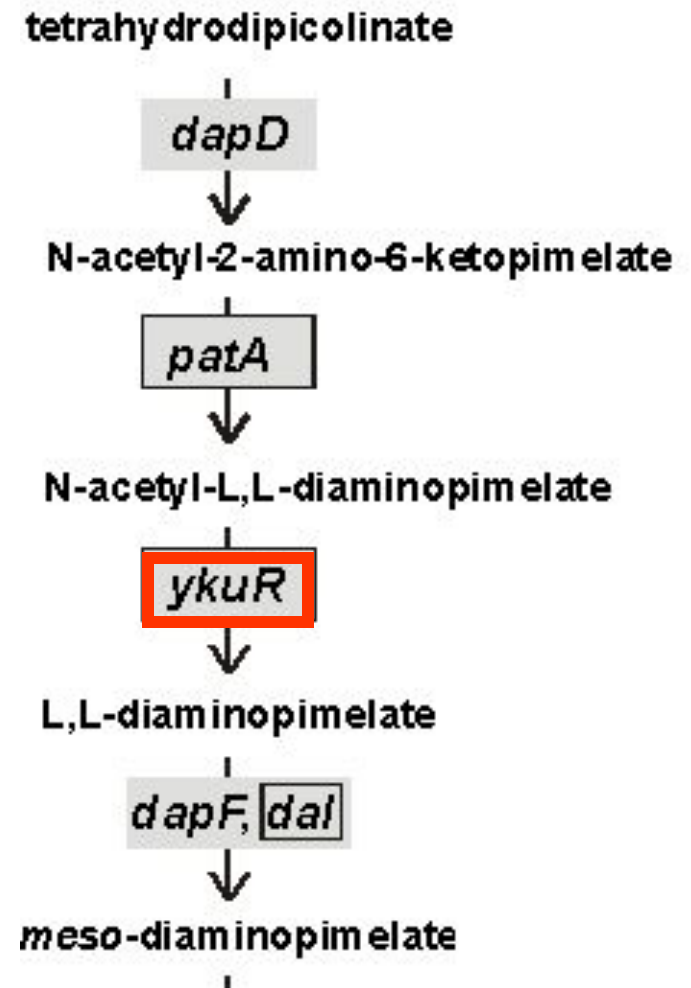


meso-diaminopimelate

# Идентификация пути ацетилированных интермедиатов - 2

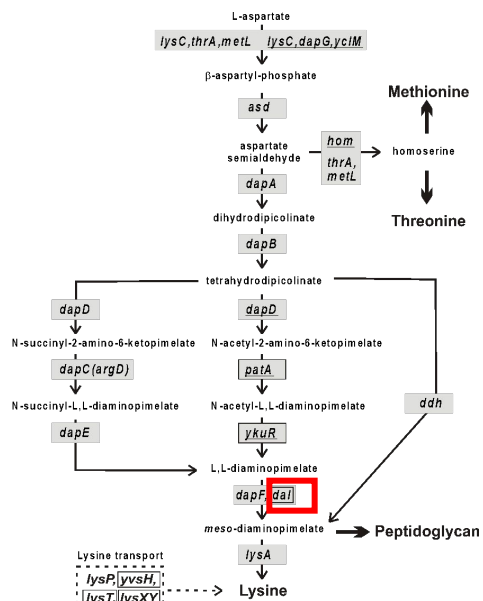
*ykuR*:

- N-ацил-L-аминокислота амидогидролаза (по гомологии)
- ко-локализуется и ко-регулируется с геном биосинтеза лизина *dapD* во многих грам-положительных бактериях
- в некоторых случаях принадлежит к большому лизиновому оперону, регулируемому LYS-элементом

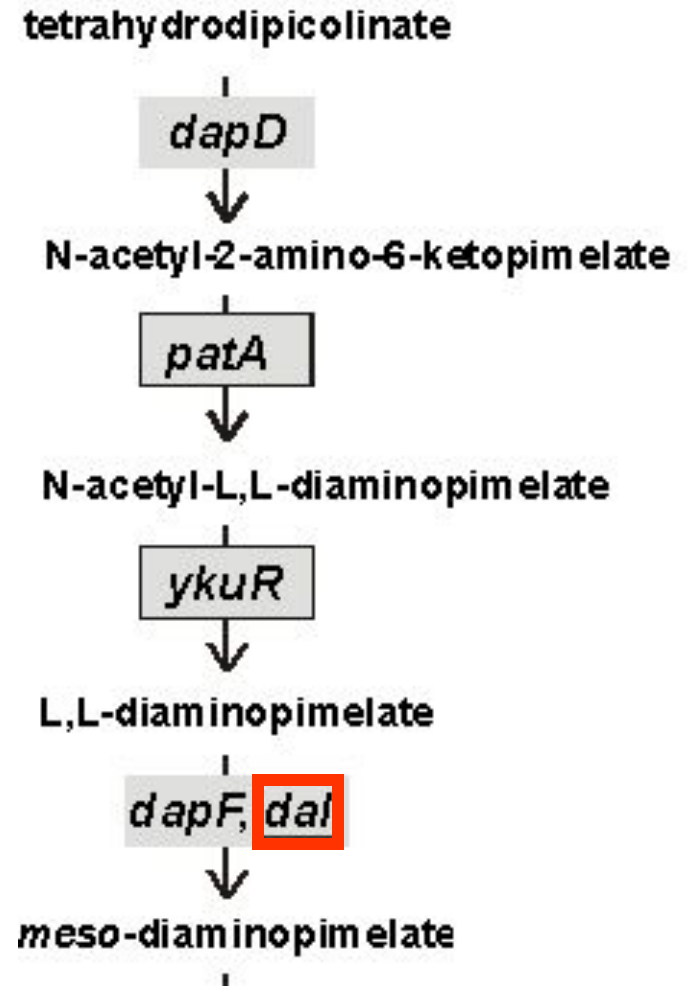


*dapX*:

- *dapF* отсутствует у некоторых бактерий (*Staphylococcus aureus*, *Opococcus oeni*, *Leuconostoc mesenteroides*)
- во **всех** этих геномах есть *dapX*, гомологичный аланиновой рацемазе и другим эпимеразам
- в *S. aureus* *dapX* принадлежит к большому лизиновому оперону
- в *O. oeni* оперон *dapX-asd* регулируется LYS-элементом



## Идентификация пути ацетилированных интермедиатов - 3



# Сравнительная геномика систем утилизации цинка

## Две роли цинка в бактериях:

- Структурная в ДНК-полимеразах, праймазах, рибосомных белках
- Каталитическая в протеазах и других белках



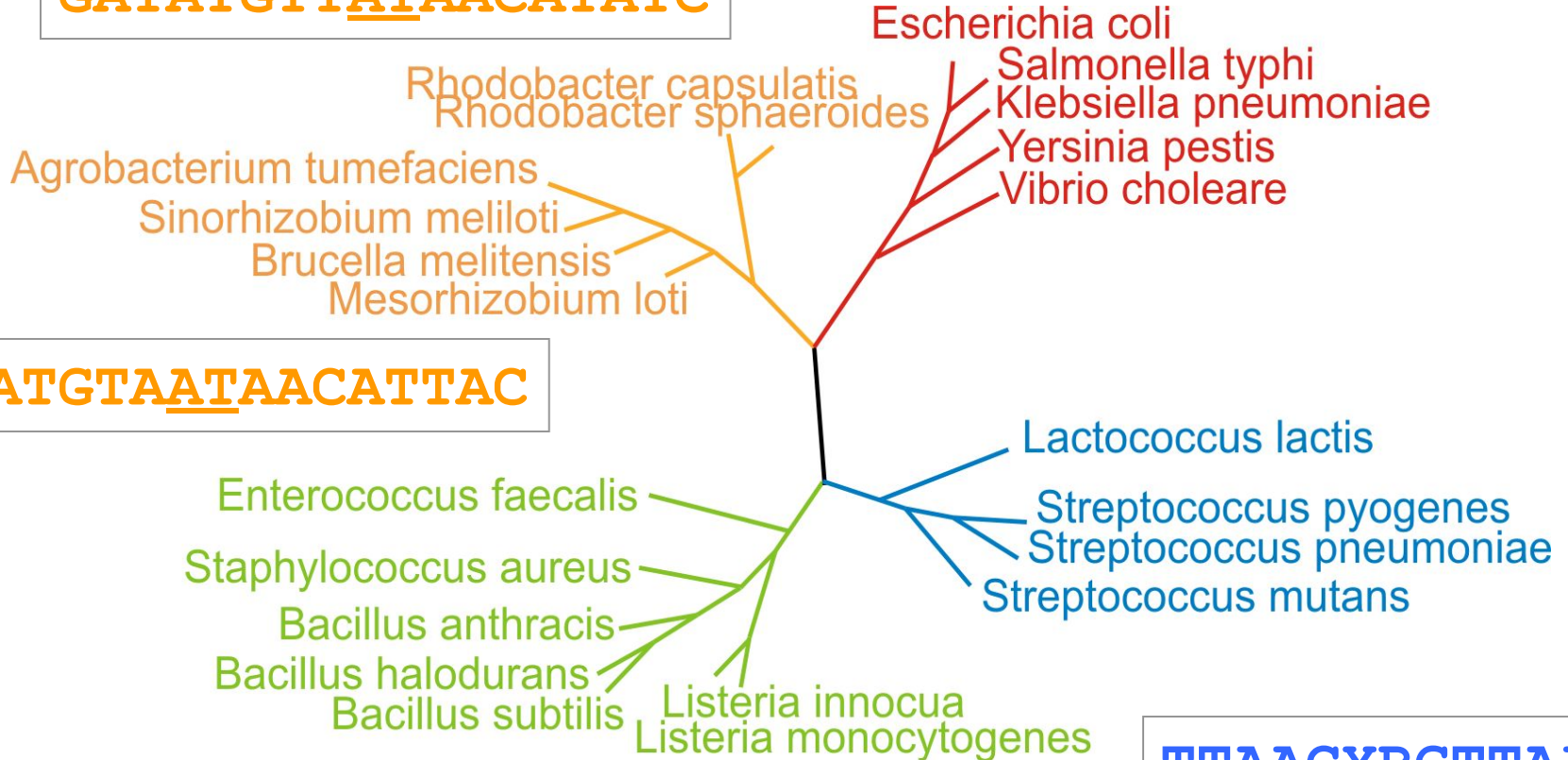
nZUR- $\alpha$

# Регуляторы и сигналы

nZUR- $\gamma$

GATATGTTATAACATATC

GAAATGTTATANTATAACATTTTC



GTAATGTAATAAACATTTAC

TTAACYRGTAA

pZUR

TAAATCGTAATNATTACGATTTA

AdcR

# Цинк и паралоги белков рибосом

	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. pneumoniae</i>	-	- - -	-	-

**nZUR**

**pZUR**

**AdCR**

(в скобках – мотив «цинковая лента»)

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

nZUR

pZUR

AdCR

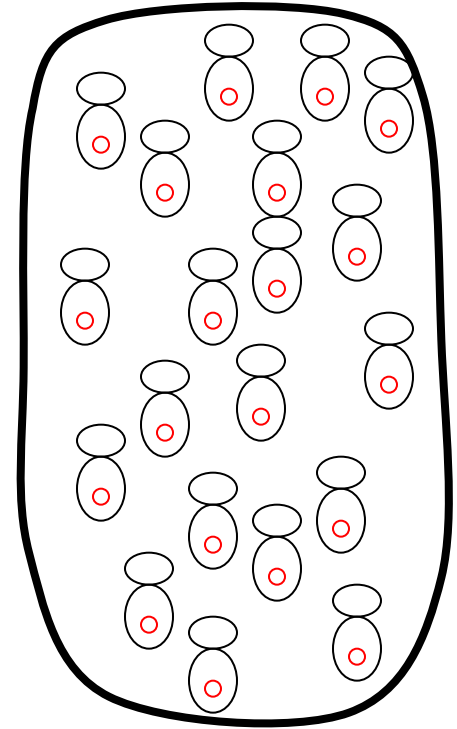
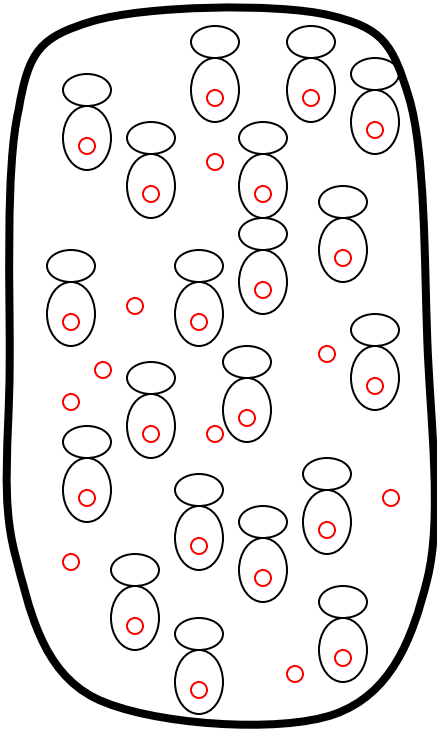
# Сводка наблюдений:

- Макарова-Романов-Купин, 2001:
  - L36, L33, L31, S14 – это единственные рибосомные белки, дублированные более, чем в одном геноме
  - L36, L33, L31, S14 – четыре из семи рибосомных белков, содержащих мотив цинковой ленты (четыре цистеина)
  - Из двух (или более) копий L36, L33, L31, S1, обычно одна содержит мотив цинковой ленты, а другая – нет
- Среди генов, кодирующих паралоги рибосомных белков, как правило один регулируется цинковым репрессором, а соответствующий белок никогда не имеет мотива цинковой ленты

# Плохой сценарий

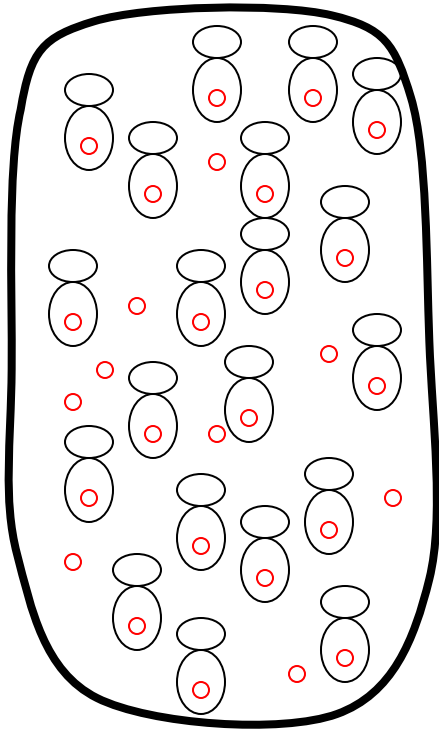
**недостаточно цинка:  
весь цинк потреблен  
рибосомами,  
ферменты голодают**

**достаточно цинка**

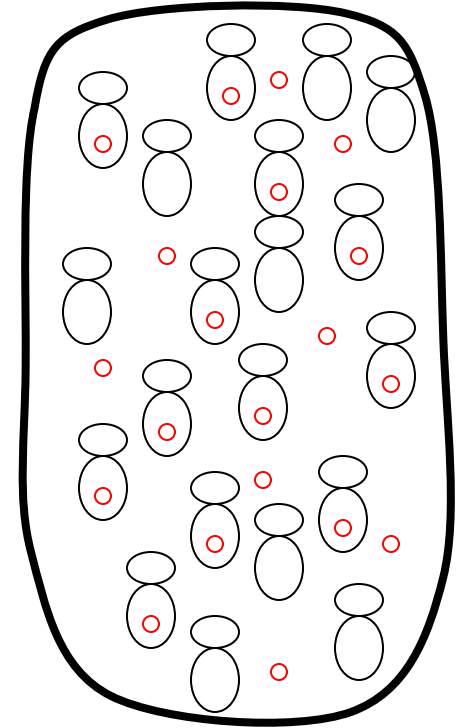


# Хороший сценарий

**достаточно  
цинка**



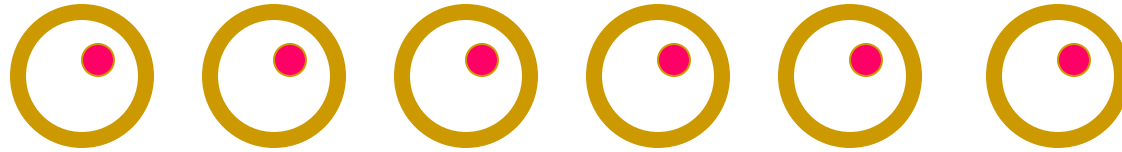
**недостаточно цинка:  
часть рибосом  
включает белки, не  
содержащие цинка –  
остается для  
ферментов**





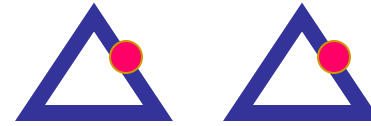
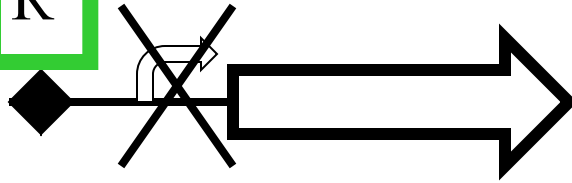
# Регуляторный механизм

Sufficient **Zn**



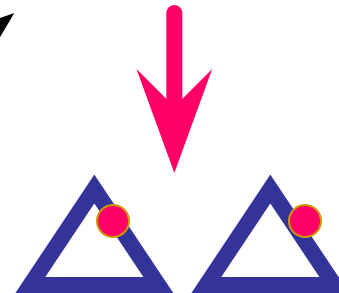
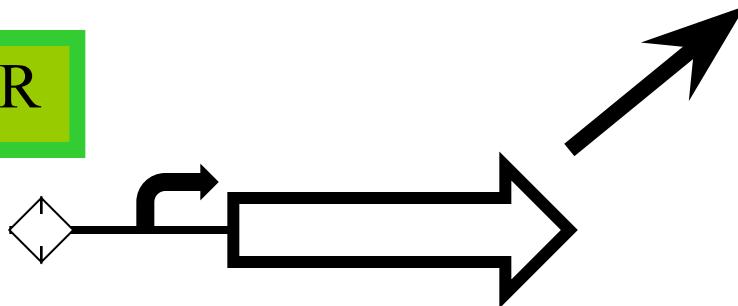
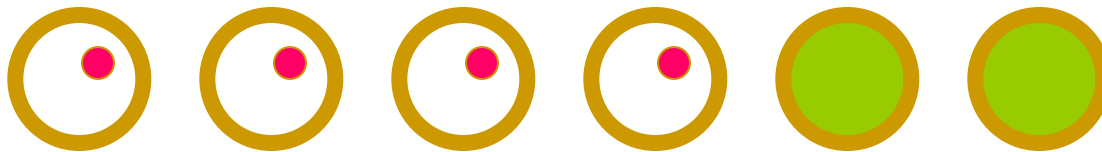
ribosomes

repressor



Zn-dependent enzymes

**Zn starvation**



# 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 ПОДТВЕРЖДЕНИЯ

### 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>

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

Journal of Bacteriology, April 2006, p. 2715-2720, Vol. 188, No. 7  
0021-9193/06/\$08.00+0 doi:10.1128/JB.188.7.2715-2720.2006

[Copyright © 2006, American Society for Microbiology](#). All Rights Reserved.

### Liberation of Zinc-Containing L31 (RpmE) from Ribosomes by Its Paralogous Gene Product, YtiA, in *Bacillus subtilis*

Genki Akanuma, Hideaki Nanamiya, Yousuke Natori, Naofumi Nomura, and Fujio Kawamura<sup>\*</sup>

# Регуляторная система «с нуля под ключ»

- Консервативный сигнал перед генами рибонуклеотид-редуктаз
- Потенциальный регулятор (через филогенетический паттерн + домены)

1	2	3	4	5
<u><i>α</i>-proteobacteria</u>		±	none in <i>Wolbachia</i> , <i>Rickettsia</i> spp.	
<u><i>β</i>-proteobacteria</u>		+	all	
<u><i>γ</i>-proteobacteria</u>		±	none in <i>Buchnera</i> , <i>Wigglesworthia</i>	<i>topA</i> in <i>Pseudomonas</i> spp.; <i>dnaA</i> in <i>Shewanella</i> spp.
<u><i>δ</i>-proteobacteria</u>		±	none in <i>Desulfovibrio</i> spp.	<i>dnaA</i> in <i>Mycrococcus xanthus</i> , <i>Desulfotalea psychrophila</i> ; <i>parA</i> in <i>Desulfuromonas</i> spp.
<u><i>ε</i>-proteobacteria</u>	n/a	-	none	
<u>Cyanobacteria</u>		±	none in <i>Nostoc</i> sp.	
<u>Bacteroidetes/Chlorobi</u>	n/a	-	none	
<u>Bacillus/Clostridium</u>		±	none in <i>Leuconostoc mesenteroides</i> , <i>Cenococcus oeri</i>	<i>dgt-pnuC</i> in lactobacilli; <i>nucA</i> in <i>Lactococcus lactis</i> ; <i>yrdD-yrdC</i> in <i>Bacillus</i> spp.; <i>ligA</i> in <i>Clostridium acetobutylicum</i>
<u>Mycoplasmatales</u>	n/a	-	none	
<u>Actinobacteria</u>		+	all	
<u>Spirochaetes</u>	n/a	±	only in <i>Treponema denticola</i>	
<u>Thermus/Deinococcus</u>		+	all	
<u>Thermotogales</u>		+	all	
<u>Chlamydiales</u>		+	all	
<u>Other</u>	n/a	±	in <i>Pirellula</i> sp., <i>Chloroflexus aurantiacus</i> , <i>Fusobacterium nucleatum</i> , none in <i>Aquifex aeolicus</i>	

- Реутилизация дезоксирибонуклеотидов

# Другие члены регулона

- Репликация (ДНК-лигазы, топоизомеразы, ДНК-полимеразы)

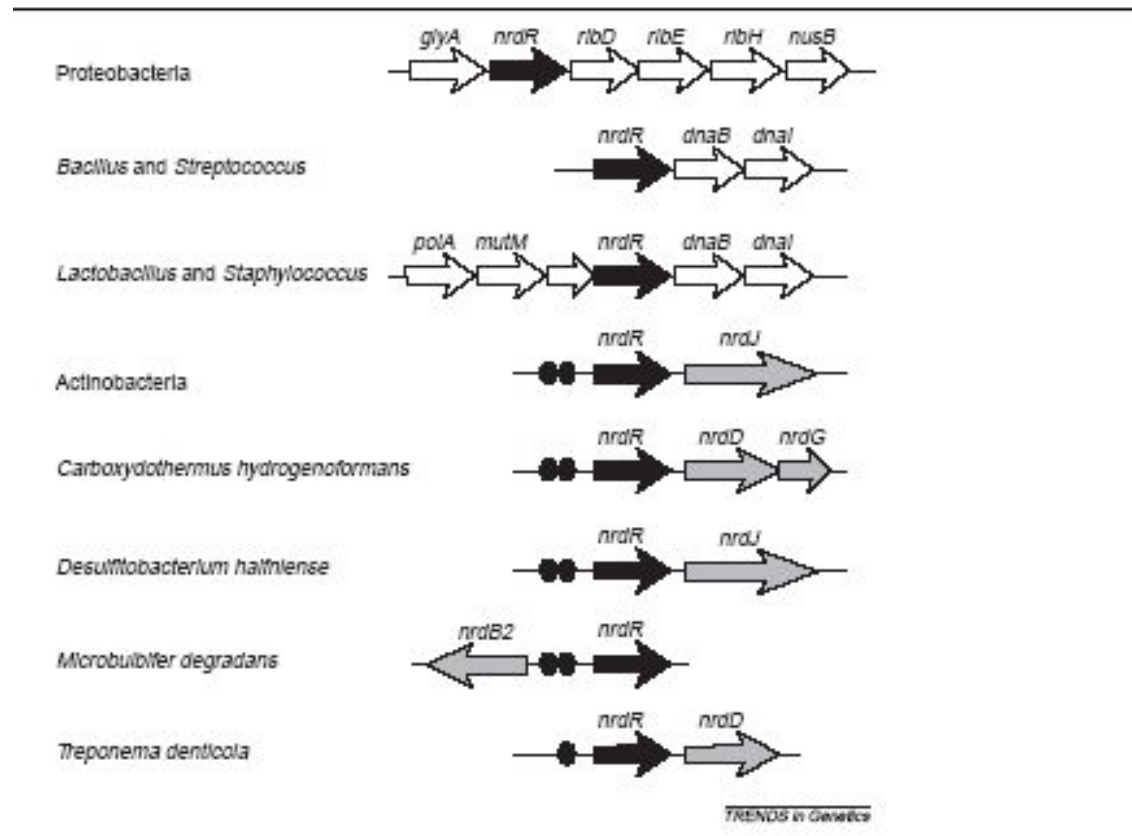
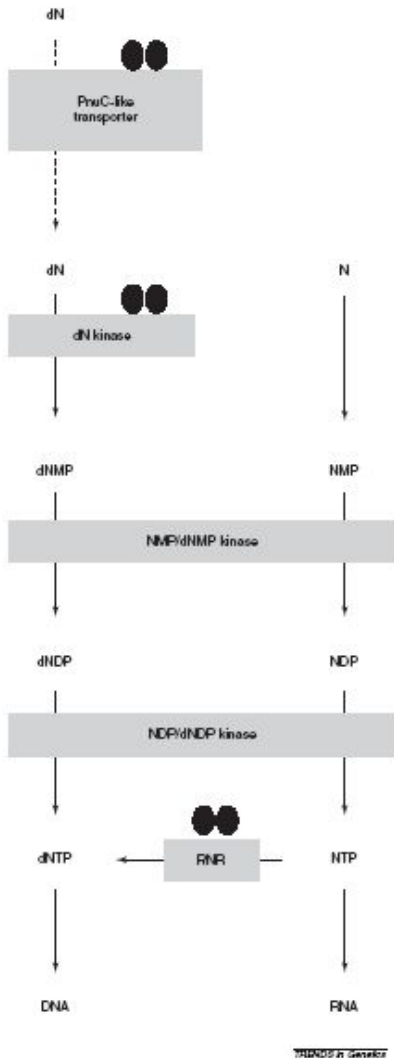


Figure 54. Two alternative pathways of the deoxyribonucleoside synthesis which are predicted to be regulated by NdrR in Gram-positive bacteria. dN (deoxynucleoside), dNMP/dNMP (deoxynucleoside monophosphate), dNDP/dNDP (deoxynucleoside diphosphate), dNTP/dNTP (deoxynucleoside triphosphate). Dots represent candidate NdrR-sites.

Figure 55. Genomic organization of the *nrdR*-containing loci in some bacterial genomes. Genes encoding the predicted ribonucleoside reductase regulator NdrR and the topoisomerase are shown in black and grey, respectively. The black circles indicate the predicted NdrR-sites. The direction of transcription is indicated.



# Как регулируется: репрессия в

## результате кооперативного связывания

A.

```

EC TGCTTTTTACTTTGAGCTACATCAAAAAAAGCTCAAACATCCTTGATGCAAGCACTTATATATAGACTTTAAAAATGCGTCCCAACCCAATATGTTGTATT
ST TGATTTTTACCTTGTTCTACATCAATAAATATGCAAACATCCTTGATGCAAATCACTACATATAGACTTTAAAAATGCACGCCGACCCAATATGTTGTATT
KP ACCTTTTTACCTTGTTCTGGGTCAATAAATATGCAAACATCCTTGATGCAAATCACTACATATAGAACTTAAAAATGCGCCTCGGCCCAACATATGTTGTATT
      *****  ***   **   *****  *****  *****  *****  *****  *****  *****  *   *****  **  *****
  
```

```

EC AATCGACTATAATTGCTACTACAGCTCCCCACG--AAAAAGGTGCGGCGTTGTGGATAAGC-GGATGGCGATTGCGGA-AAGCACCGGAAAACGAAACGA
ST AATTGACTACAATTGCTACAACACCTGTTCACT--CGACACAAGGTGAATTGTGGATAACCTGGGTCAGGATTGCGGG-AAGTCATTGGAAAAGAGATGA
KP AATCGTCTATTAT-GTCACCATATCTTGTTCGATGTCTGGCGGTGATGAGATGTGGATAAACGGGCCGGATCCGAAGGTAACAGCACGAGCCGTAGCGT
      *** * ***   ** *   ** *   ** *   *   *   *   *****  **   *   *   **   *   *   *
  
```

```

EC AAAAACCGGAAAACGCCTTTCCCAATTTTGTGGATAACCTGTTCTTAAAAATATGGAGCGATCATGACACCGCATGTGATGAAACGAGA
ST ATAAACCTGTTA-TGCTTTCCCGGCCTTGTGGATAACCTGTTCTTACAAATATGGAGTGATCATGACACCGCATGTGATGAAACGAGA
KP GCAGCGCCTTCG-GGATAACCTCCGCCTTGTGGATAACCTGTTCT---ATATATGGAGTGATCATGACACCGCATGTGATGAAACGAGA
      *   *   *   *   *   *   *****  *****  *   *****  *****  *****  *****  *****  **
  
```

B.

```

YP AACAGGGAATAACCC-TAACGCC--AATTTCCTTGTTCTAGGTCAACAATATTGGCTATCAGTTGACTGTCACTCATCCAGATACCCATATATAGTGTCT
YE AACAGGGAATAAACC-TAAAGCT--GATTTCCTTGTTCTAGGTCAATTAT-----GTTGACTGTCACTTCTGCCATTACCCATATATAGTGTCT
Eca AAGTTCGATTTATCTACTAGGGAGGAATTTCCTTGCTCTACATCAATTTTGCGAGCGATAAAAGTGCAAACACCCCTACGCAATTTCAATATATAGTGCCT
Ech AAGACTGATTTCTCTACGATGCCGAATTTCCTTGAGCCAGGTCAATTCTAACGCAATAAAACCGGGTCCCCCTCCAGGCGAATTTCAATATATAGTGTCT
      **   ** *   *   *   *   *****  ***   * *   *****  *   *   *   *   *   *   *   *****  **
  
```

```

YP ATAATATTTTAAGCATCTATATGTAGTAGTTATCCACAAAAGCAATCCACATCCCCCTCGCAGCCCTGATGTGCTGCGGGTTGC-CTTGTGGATAA-----
YE ATAGTAATTACGATACCTATATGTAGTAGTTATCCACAAAAGCAATCCACA-CCCCCTCGCAGCCCTGATGTGCTGCGGGTTGC-CTTGTGGATAAGATGG
Eca ATCCTGTAAACATTACCTACATATAGTGTTTATCCACAAAAGCAATCCACA-GCCCTCTGTAACCCTTGCCAGTTACGGTCTCGCCTGTGGATAAC----
Ech ATCCTTTTACAAGAACCTACATATGTTGTTTATCCACAAGAATCCACA-GCC-TCCGCACCCGTTGTCAGCCGCGGCTTCCGTCTGTGGATAAC----
      ** *   *   ***  ** *   **   *****  *****  **   *   *   ** *   *   *   *   *   *****  **
  
```

```

YP -----CCCTATGCGGCGGTATACAGGAGTGACATTGTGAAAACAGTAGTGATTAAACGGGACGGCTGCCAGGT
YE TTTTGGGGCTAATCCTACGCGGCAGGATACAGGAGCGACATTGTGAAAACAGTAGTGATTAAACGGGACGGTTGTCAGGT
Eca -----CTTTCCAG-----AGGAAGAA-AACGTGAAACCAGTAGTGATTAAACGGGACGGTTGCCAGGT
Ech -----ATCAACAAAGGAAGAACACCGAGGAACAAC---ATGAAACCAGTAGTGATTAAACGGGACGGATGTCAGGT
      *****  *   *****  *****  *****  *****  *****  *****  *****  *****
  
```

# Что осталось за кадром

- Эукариоты
- Структуры
- Молекулярная эволюция
  - Гены
  - Геномы
  - Метаболические и регуляторные системы
- Другие виды данных и что с ними делать
  - Экспрессия
  - Белок-ДНКовые взаимодействия
  - Белок-белковые взаимодействия
  - Структура хроматина (метилирование, гистоны и их модификации и т.д.)
- «Системная биология»