

# **Биоинформатика: биологические тексты**

**М.С.Гельфанд**

29 октября 2006

Первый фестиваль науки МГУ  
Факультет биоинженерии и биоинформатики

# Расшифрован геном!

расшифрован геном человека - Поиск в Google - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Refresh Print

Address http://www.google.ru/search?hl=ru&q=%D1%80%D0%B0%D1%81%D1%88%D0%... Go Links

Войти

Google Веб Картинки Группы Каталог Дополнительно >

расшифрован геном человека Поиск Расширенный поиск Настройки

Поиск в Интернете  Поиск страниц на русском

Веб Результаты 1 - 10 из примерно 93 000 для **расшифрован геном человека**. (0,22 секунд)

**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, ...  
[www.n-t.ru/tp/in/rs.htm](#) - 12k - Сохранено в кэше - Похожие страницы

14.04.2003 **Геном человека расшифровали до конца / Новости ...**  
Программа «**Геном человека**» завершена. Сейчас исследователям известен практически весь генетический код **человека**, а дальнейшая его расшифровка, по их словам, ...  
[83.222.3.160:8101/mednovosti/news/2003/04/14/genome/](#) - 21k - Сохранено в кэше - Похожие страницы

**Расшифрован геном мыши: отличие от человека** - не более процента / ROL  
**Расшифрован геном** мыши: отличие от **человека** - не более процента. 5 декабря 2002 г. С точки зрения генетиков, **человек** мало чем отличается от мыши. ...  
[www.rol.ru/news/med/news/02/12/05\\_001.htm](#) - 20k - Сохранено в кэше - Похожие страницы

Журнальный зал | Знамя, 2000 N10 | Лев Киселев / Валерий Сойфер ...  
Закончилось начало — **геном человека расшифрован**. Дальше — самое главное, самое интересное, самое трудное — как же все-таки наши **гены** через свои продукты ...  
[magazines.russ.ru/znamia/2000/10/konfer.html](#) - 97k - Сохранено в кэше - Похожие страницы

**Расшифрован геном человека**  
**Расшифрован геном человека** ... После этого возможна публикация окончательной версии **генома человека**. Из заявления Вентера следует, что полные данные ...  
[www.ma-ma.ru/r/library/news/28823.php](#) - 26k - Сохранено в кэше - Похожие страницы

Элементы - новости науки: Полностью **расшифрован геном собаки**  
Полностью **расшифрован геном** собаки ... Генетическое сходство с **человеком** (а **геном** собаки ближе к человеческому, чем, например, **геном** мыши) позволит ...  
[elementy.ru/news/165035](#) - 29k - Сохранено в кэше - Похожие страницы

Mass Media : **Расшифрован геном** черного тополя  
**Расшифрован геном** черного тополя. 20.09.2006 14:33 | Радио Свобода ... В США объявлен конкурс на доступный способ расшифровки **генома человека** (NEWSru.com) ...  
[www.rambler.ru/news/science/dolly/8731606.html?i\\_detail=1](#) - 41k - Сохранено в кэше - Похожие страницы

ALTEY Laboratory. Система автоматизации лабораторных исследований

https://www.google.com/accounts/Login?continue=http://www.g Internet

расшифрован геном - Поиск в Google - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Refresh Print

Address http://www.google.ru/search?hl=ru&q=%D1%80%D0%B0%D1%81%D1%88%D0%B8%D1... Go Links

Войти

Google Веб Картинки Группы Каталог Дополнительно >

расшифрован геном Поиск Расширенный поиск Настройки

Поиск в Интернете  Поиск страниц на русском

Веб Результаты 1 - 10 из примерно 228 000 для **расшифрован геном**. (0,25 секунд)

Элементы - новости науки: Полностью **расшифрован геном** собаки  
Консорциум генетиков завершил расшифровку генетического кода собаки — породистого боксера. В процессе работы была также составлена карта однонуклеотидных ...  
[elementy.ru/news/165035](#) - 29k - Сохранено в кэше - Похожие страницы

MEMBRANA | Полужизнь: **расшифрован геном** самого загадочного ...  
Mimi, или Mimivirus, номинально, самый большой вирус в мире. Но фактически учёные не могут подобрать для него самую подходящую ветку на древе жизни.  
[www.membrana.ru/articles/global/2004/10/15/214500.html](#) - 54k - Сохранено в кэше - Похожие страницы

Mass Media : **Расшифрован геном** черного тополя  
**Расшифрован геном** черного тополя. 20.09.2006 14:33 | Радио Свобода.  
Международный коллектив генетиков, под руководством профессора Джеральда Таскэна (Gerald ...  
[www.rambler.ru/news/science/dolly/8731606.html](#) - 39k - Сохранено в кэше - Похожие страницы

"Аккумулятор новостей": самые свежие события и факты — интересно и ...  
США: ученые впервые **расшифровали геном** курицы. Ученые впервые опубликовали детальный отчет о расшифровке **генома** курицы. Об этом пишет журнал Nature. ...  
[news.battery.ru/dossier/?dsrid=346&from\\_m=dossier&from\\_n=0&newsId=23532482](#) - 43k - Сохранено в кэше - Похожие страницы

"Аккумулятор новостей": самые свежие события и факты — интересно и ...  
Франция: **расшифрован геном** самого загадочного биологического объекта. Mimi, или Mimivirus, номинально, самый большой вирус в мире. ...  
[news.battery.ru/dossier/?dsrid=346&from\\_m=dossier&from\\_n=0&newsId=20429953](#) - 47k - Сохранено в кэше - Похожие страницы  
[ [Дополнительные результаты с news.battery.ru](#) ]

**Расшифрован геном** шимпанзе - Компьюлента  
**Геном** наших ближайших родственников состоит из 2,8 миллиардов пар оснований ДНК. Теперь генетики будут пытаться отыскать различия между геномом шимпанзе и ...  
[science.compulenta.ru/225318/](#) - 35k - Сохранено в кэше - Похожие страницы

Lenta.ru: Прогресс: **Расшифрован геном** собаки  
В американском институте Брода **расшифровали геном** собаки, сообщает Reuters. Генетики считают, что их открытие может пригодиться при поиске лекарств от ...  
[lenta.ru/news/2005/12/08/dog/](#) - 18k - Сохранено в кэше - Похожие страницы

Russian America Online - dating. iournal. news. discussions ...

http://64.233.183.104/search?q=cache:YcRTcfU9MsJ:elementy.ru/news, Internet

# Расшифрован ли геном?

- Перехватить зашифрованное сообщение – еще не значит его понять





# Фрагмент генома (0.0001% генома человека)

cgtgcac ttctgaaggacttcaggtac cggcgtgccc cggctcctac tgtcc gceetgctcgc gtcctgggtgccc ccttgagtagggcgggagagg  
CAGCCAAAGGCGGAGCTGATGGCTGCGC CGAGGGCGGGGCGGGGTGCAAGGCTGGAGCC TTCGGGCATGGCGGGCTTTGGGGGGCATT CGCTG GGGGAGGAG  
ACCCCGTTTGAC CCCTGACCTC CGGGC CCTGC TGACGTCAGGA AACTT CTGAC CCCC GGGCCC GAGTGACTTA TGGGACCCCC AGTC TCTGGGCCCGGTTG  
TCTGTTGGGTC ACTGAACCCC GAGCATGCCTGACGT CTGGGACCCC GGGTC CCCGGGCACA ACTGACTGCGGTGAC CCCAGATAC CAGGA CCCGGGAGG  
CCTCAGAGA AACTCTGGAACCCGTTCCG CCGCGTGGCTGGCGGTGGCGCTGGGCGCTGGGGGGCAGT GCTGTTGTTGTTGTGGGC GGGGGTCCGGGTCC  
TCCGGCC GTCCTCGCCG CCGTC CCTAG CCCC GC CCCC GCTTC TCCCC GGAGT CAGTA CAACTTCATC GCAGA TGTGGTGGAGAAGA CAGCA CCTGC CGTG  
GTCTATA TCGAGATCCTGGACC Ggtaa tgggtg ggggt agacc gggag gcaact gaagc cacag gctgg agggc gggcgggtag gagggggtcagagcc tcct  
cttatctgtgctttccc tccatttcag GCACC CTTTC TTGGG CCGCGAGGTC CCTAT CTCGAACGGC TCAGGATTCGTGGTG GCTGCCGATGGGCTCATT  
GTCACCAACGCC CATGTGGTGGCTGATCGGCG CAGAGTCCGTGTGAGACTGCTAAGC GGCAG CACGTATGAGGCCGTGGTCA CAGCTGTGGATCCC GTGG  
CAGACATCGCAACGCTGAGGATTCAGACTAAG gtggggggctg ggggta gggca ggtct ggttg gagct gctta tttgc tcgca tcttcagatgacaggtct  
cttttac ccattctccc ttagg agcctctccc cacgc tgcctctgggacgctcagctgatgtccggc aaggggagtt tgttg ttgc catgggaagtccct  
ttgcaactgcagaacacgatcac atccggcattgttagctctgtctcagcgtcc agccagagac ctgggactcc cccaa accaa tgtggaata cattc aaac  
tgatgcagctat tgatgtgcgtcctga tagga gagaaatgac aatgatgggggaggggggagggc tgtgt ggtac aagca ccaactgat atatggtgg  
atgagcc tatatagagc ttaggctgcaaaaatgtggc cacttattca tgggc tgaga aagaagagaa tttgg aaaa gtaacc taca tccgtggtatgcccc  
cagacttagaatccccagatctcttcc atgtttctct cttgtcctac agTTTGGAAA CTCTGAGGT CCCCTGGTTAACCTG gtgagtgagacatc ctcc  
cttccaagaatc cctgc cccaggtcag tgtgg gaagggtaggtttcc cctaa ttcaaggatgtttgg tcaagttctgagcagttc tttgt tggctatct  
ctcaata tccaaccaga tctcc ccaac acttgctggtacttt tgttc gggtgcccc atccc ctactat ttgtttaggttagggaaactgggggctgtatc  
cctgcag GATGGGGAGGTGATTGGAGTGAACA CCATGAAGGT CACAGCTGGAATCTC CTTTGCCATC CCTTC TGATC GTCTTCGAGAGTTTCTGCATCGT  
GGGGAAAAGAAG Agtga gctgcctta tgggg aaacgggttc ctta atgtggtgga aataggggaagggca ttcag tgggacttc ctgga ggggtggtct  
actgggagaaga gggcagggaa ggaag gatgt agctgggtgg ggtc atttgtccctctgtc acagATTCCCTCCTCC GGAATCAGTGGGTC CCAGC GGC  
CTACATTGGGGT GATGATGCTGACCCTGAGTC CCAGgtatgagcttt agggacagtgacatg taatgtgacc agtgtaatcagaggggggc acctc tatt  
gagctttgttctcatttctgtc tttatctaag atgaa ctgtg tcacacttga aataa tcaca agagc tgtctccctt catca tcttgacttctta tccc  
actccac tttgtacacc tgtcaccaga ttgat ttcactctgt tactgctttgatttc aagcc ttcaatccat taact tggca tttta agggc cattt tcca  
tctgtctgtaaatcaac tttctagacc tggctgtaat accttccat atgaa tactc agcca actgatttc ctactccc atgttt tttat ttatgctgt  
tcttgttttata tgggttaaatat gct  
attgatttaacactgtttgtca cat  
tgtat tttgttagtacctagccc ttt  
ggcta atagggtgatctgtgta aag  
tcactctgggctccccctgcaca cct  
tccttctctctgtccat ttttctctatag GGTGGTCTGCGGCCTGGTGATGTGATTTTGGCCATTGGGGAGCAGATGGTACAAAATGCTGAAGATGTT  
TATGAAGCTGTT CGAAC CCAATCCCAGTTGGC AGTGCAGATC CCGCGGGGAC GAGAAACACTGACCTTATATGTGAC CCCTGAGGT CACAGAATGAATAG  
ATCACCAAGAGTATGAGGCTCC TGCTC TGATTTCCTC CTTGC CTTTC TGGCTGAGGT TCTGAGGGCA CCGAGACAGAGGGTTAAATGAACCAGTGGGGGC  
AGGTCCC TCCAA CCACC AGCAC TGACT CCTGGGCTCTGAAgAATCAC AGAAA CACTTTTTATATAAAAATAAAAATTATACCTAGCaacatatatagtaaa

**Геном человека: 3 000 000 000 нуклеотидов**  
**Примерно 25 тысяч генов, < 5% генома кодирует белки**

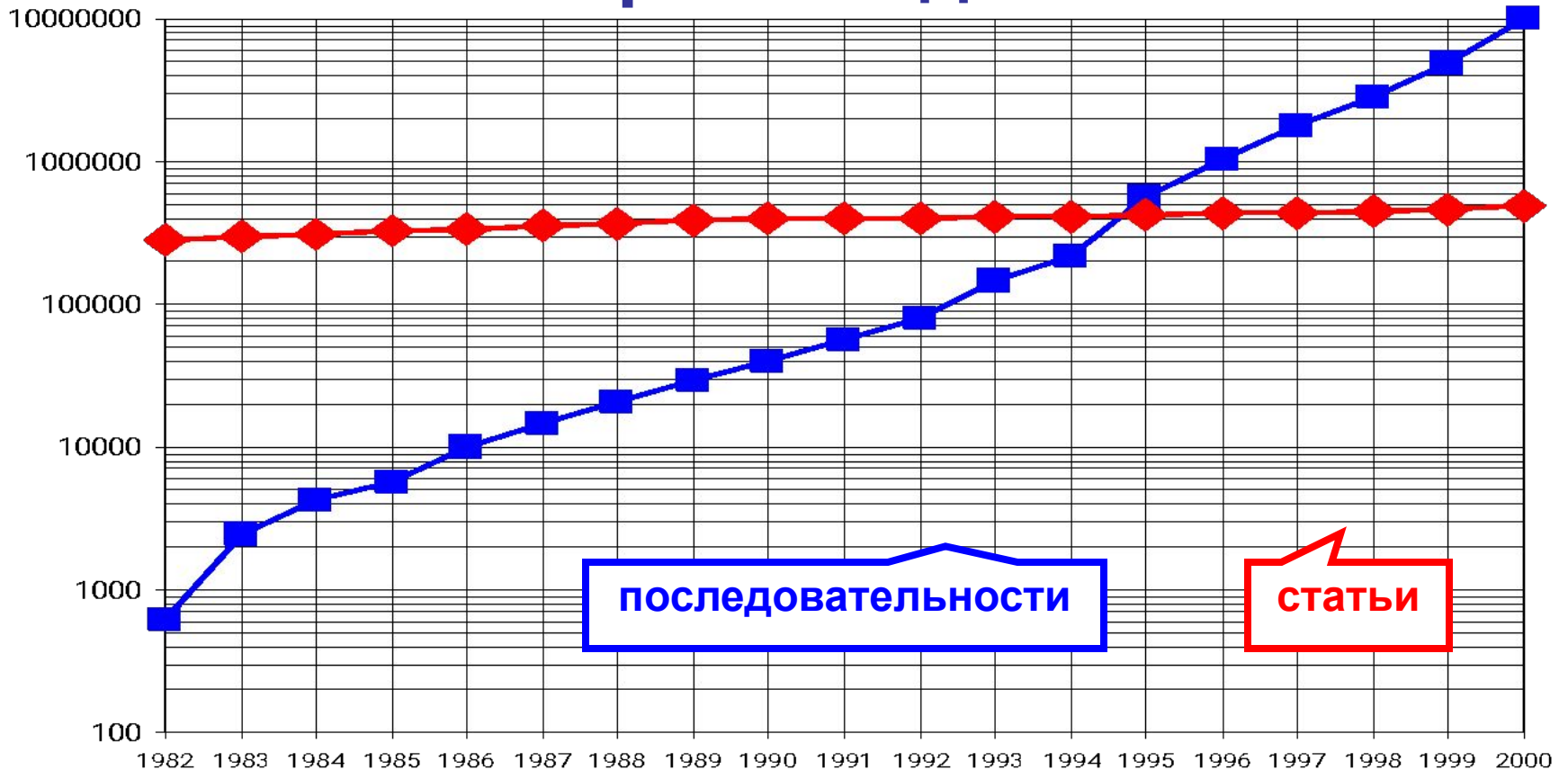
# Что же мы хотим понять?

- Где
  - картировать гены в геноме и определить аминокислотные последовательности кодируемых белков
- Что
  - предсказать функции генов (кодируемых белков)
- Когда
  - описать регуляцию генов, зависимость экспрессии от внешних условий и внутреннего состояния клетки
- Где – 2
  - определить локализацию белка в клетке (или вне её)

TTT	F	TCT	S	TAT	Y	TGT	C
TTC	F	TCC	S	TAC	Y	TGC	C
TTA	L	TCA	S	TAA	stop	TGA	stop
TTG	L	TCG	S	TAG	stop	TGG	W
CTT	L	CCT	P	CAT	H	CGT	R
CTC	L	CCC	P	CAC	H	CGC	R
CTA	L	CCA	P	CAA	Q	CGA	R
CTG	L	CCG	P	CAG	Q	CGG	R
ATT	I	ACT	T	AAT	N	AGT	S
ATC	I	ACC	T	AAC	N	AGC	S
ATA	I	ACA	T	AAA	K	AGA	R
ATG	M start	ACG	T	AAG	K	AGG	R
GTT	V	GCT	A	GAT	D	GGT	G
GTC	V	GCC	A	GAC	D	GGC	G
GTA	V	GCA	A	GAA	E	GGA	G
GTG	V	GCG	A	GAG	E	GGG	G

Таблица генетического кода

# Пропаганда



- Проблемы: нет возможности исследовать <sup>год</sup> все экспериментально
- Возможности: можно использовать методы сравнительной геномики
  - > 1000 геномов бактерий (~400 полных)
  - простейшие: малярийный плазмодий, инфузория, лейшмания, ...
  - растения: арабидопсис, тополь, рис
  - пара дюжин дрожжей и другие грибы
  - насекомые: дюжина дрозофил, комар, пчела, ...
  - позвоночные: человек, шимпанзе, мышь, крыса, собака, курица, 2 рыбы, ...

# Поиск генов если известен белок: просто

aaacgccttggccttctatagccactattatcagccgctgcccgttggtactgcgcggata

tggtgcgggcaatgacggttacagctgccgggtgtctttgctgatctgctacgtaccctctc

M V K V Y A P A S S A N M

atggaagtaggagctctgacatgggttaaagttatgccccggcttccagtgccaatatga

S V G F D V L G A A V T P V D G A L L G

gcgteggtttgatgtgctcgggggcggcggtgacacctgttgatgggtgcattgctcggag

D V V T V E A A E T F S L N N L G R F G

atgtagtcacggttgaggcggcagagacattcagctcaacaacctcggacgctttgggt

A D K L P S E P R E N V Y Q C W E R F C

ccgataagctgccgctcagagccacgggaaaatgtttatcagtgctgggagcgtttttgcc

Q E L G K Q I P V A M T L E K N M P I G

aggagcttggcaagcaaattccagtggcgatgactctggaaaagaatatgccgatcgggt

F V H I C R L D T A G A R V L E N

ttgttcatatttgccggctggatagggcgggcgacgagtactggaaaactaaatgaaac

tctacaatctgaaagatcacaatgagcaggtcagctttgcgcaagccgtaaccagggt

tgggcaaaaatcaggggctgtttttcccgcacgacctgccggaattcagcctgactgaaa



# ... или родственный белок: тоже просто

aaacgccctggccttctatagccactattatcagccgctgccggttggtactgcgcgata

tggtgcgggcaatgacggttacagctgccggtgtctttgctgatctgctacgtaccctctc

M V V V Y A P A S I G N V

atggaagtaggagctctgacatgggttaaagtttatgccccggcttccagtgccaatatga

S V G F D V L G A A V S P I D G S L L G

gcgtcgggtttgatgtgctcggggcgggcggtgacacctggtgatggatgcattgctcggag

D R V E V A A G E Q P F T L K C V G D F

atgtagtcacgggttgaggcggcagagaca---ttcagttctcaacaacctcggacgctttg

V A K L P V E Q E E N V Y H C W L V F A

ccgataagctgccgtcagagccacgggaaaatgtttatcagtgctgggagcgtttttgcc

R E L D K K V P V S M T L E K N M P I G

aggagcttggcaagcaaattccagtggcgatgactctggaaaagaatatgccgatcgggt

F V H V C R L D S T G S K V L E N

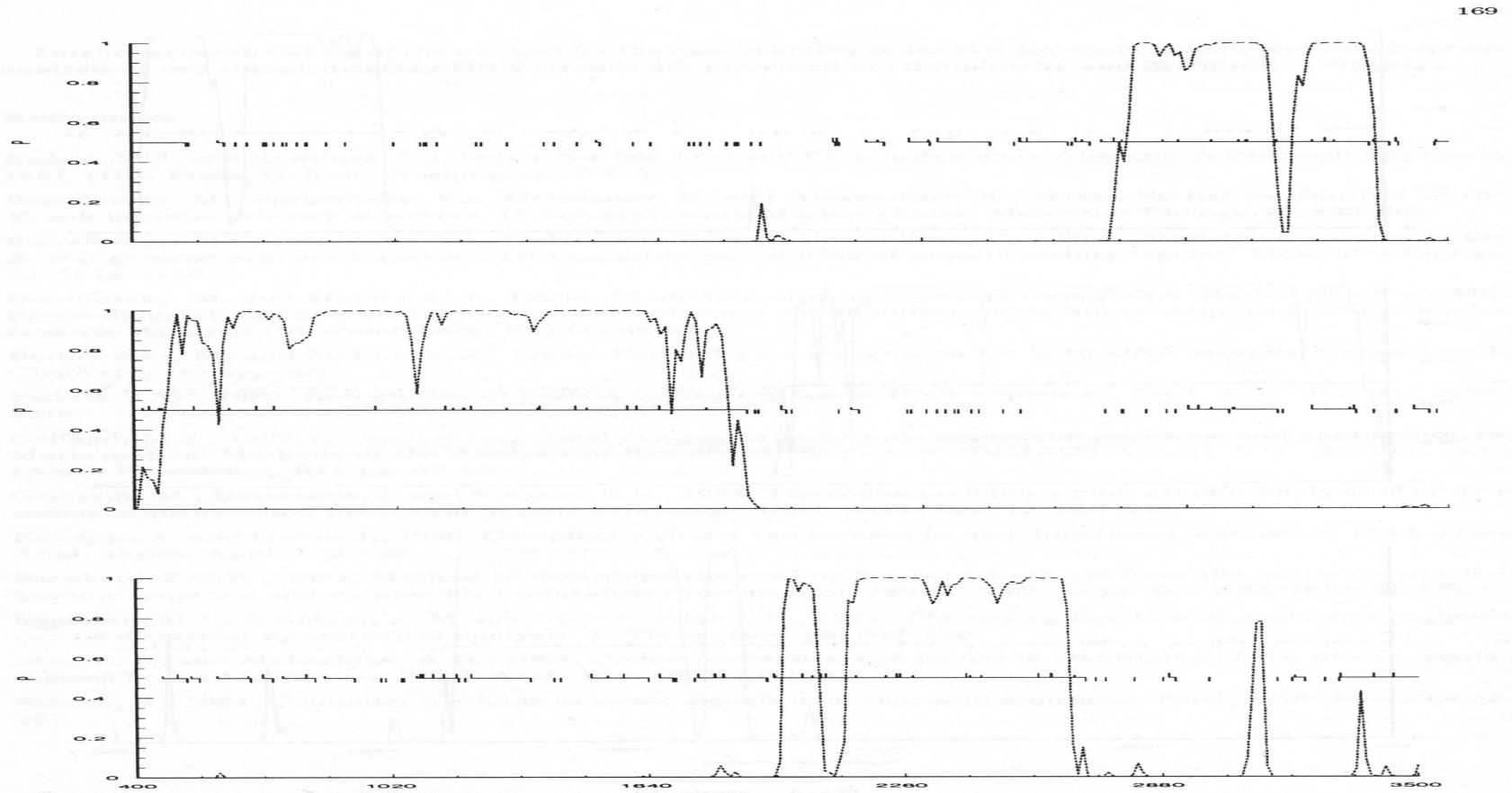
ttgttcatatttgccggctggatcggcgggcgcacgagtagtggaaaactaaatgaaac

tctacaatctgaaagatcacaatgagcaggtcagctttgcgcaagccgtaacccaggggt

tgggcaaaaatcaggggctgtttttcccgcacgacctgccggaattcagcctgactgaaa

# Статистические особенности

- Отличия в частотах олигонуклеотидов в кодирующих и не кодирующих областях



- Стартовые кодоны

# Начала генов *Bacillus subtilis*

<i>dnaN</i>	ACATTATCCGTTAGGAGGATAAAAA <b>ATG</b>
<i>gyrA</i>	GTGATACTTCAGGGAGGTTTTTTA <b>ATG</b>
<i>serS</i>	TCAATAAAAAAAGGAGTGTTTCGC <b>ATG</b>
<i>bofA</i>	CAAGCGAAGGAGATGAGAAGATTC <b>ATG</b>
<i>csfB</i>	GCTAACTGTACGGAGGTGGAGAAG <b>ATG</b>
<i>xpaC</i>	ATAGACACAGGAGTCGATTATCTC <b>ATG</b>
<i>metS</i>	ACATTCTGATTAGGAGGTTTCAAG <b>ATG</b>
<i>gcaD</i>	AAAAGGGATATTGGAGGCCAATAA <b>ATG</b>
<i>spoVC</i>	TATGTGACTAAGGGAGGATTCGCC <b>ATG</b>
<i>ftsH</i>	GCTTACTGTGGGAGGAGGTAAGGA <b>ATG</b>
<i>pabB</i>	AAAGAAAATAGAGGAATGATACAA <b>ATG</b>
<i>rplJ</i>	CAAGAATCTACAGGAGGTGTAACC <b>ATG</b>
<i>tufA</i>	AAAGCTCTTAAGGAGGATTTTAGA <b>ATG</b>
<i>rpsJ</i>	TGTAGGCGAAAAGGAGGGAAAATA <b>ATG</b>
<i>rpoA</i>	CGTTTTGAAGGAGGGTTTTAAGTA <b>ATG</b>
<i>rplM</i>	AGATCATTTAGGAGGGGAAATTCA <b>ATG</b>

# Участок связывания рибосом

<i>dnaN</i>	ACATTATCCGTTAGGAGGATAAAAATG
<i>gyrA</i>	GTGATACTTCA <del>G</del> GGAGGTTTTTTAATG
<i>serS</i>	TCAATAAAAAAAGGAGT <del>T</del> GTTTCGCATG
<i>bofA</i>	CAAGCGAAGGAG <del>A</del> TGAGAAGATTCATG
<i>csfB</i>	GCTAACTGTAC <del>C</del> GGAGGTGGAGAAGATG
<i>xpaC</i>	ATAGACACAGGAG <del>T</del> CGATTATCTCATG
<i>metS</i>	ACATTCTGATTAGGAGGTTTCAAGATG
<i>gcaD</i>	AAAAGGGATAT <del>T</del> GGAGGCCAATAAATG
<i>spoVC</i>	TATGTGACTAA <del>G</del> GGAGGATTCGCCATG
<i>ftsH</i>	GCTTACTGTGGGAGGAGGTAAGGAATG
<i>pabB</i>	AAAGAAA <del>T</del> AGAGGAATGATACAAATG
<i>rplJ</i>	CAAGAATCTACAGGAGGTGTAACCATG
<i>tufA</i>	AAAGCTCTTAAGGAGGATTTTAGAATG
<i>rpsJ</i>	TGTAGGCGAAAAGGAGGGAAAATAATG
<i>rpoA</i>	CGTTTTGAAGGAGGGTTTTAAGTAATG
<i>rplM</i>	AGATCATTTAGGAGGGGAAATTCAATG

# Сравнение генов в родственных геномах

```

sty TCGCTCG--CAGCGGAAAGAGGATTACGCCCTTCGCCTGGAGGCTGTGCAGGGGC---GCCGGAGATGGGATGCATAATT
stm TCGCTCG--CAGCGGAAAGAGGATTACGCCCTTCGCCTGGAGGCTGTGCAGGGGC---GCCGGAGATGGGATGCATAATT
sen TCGCTCG--CAGCGGAAAGAGGATTACGCCCTTCGCCTGGAGGCTGTGCAGGGGC---GCCGGAGATGGGATGCATAATT
eco TTGCCCG--TGCCAGACGGCAGATTATCTCCCTGACCTGGTGGTTGCCAGGAGGAGGGCCGGAAATAGGTTGTATCATT
kpn ----CGG--TGGCGCAGTGCCTGATGGG-CCTCGCCCTGGAGGACGGTCTGGCAT--ATCAGCAAGGGGGTGCCTCATG
ype TTGTTAGAACAGGGGAAAACGGTAAACAGTGTGGCATTAGATGTCGGTTATAGCT----CCGCCTCTGCTTTTATCGCC
      *           *                   * * * * *                   * *           * * *
  
```

```

sty AATTATCCTTTAAC-----CATAAATCTGAGCAATA-TATGCTTGGCGGCCAGATTATGGC--ACACTTGTCCGG
stm AATTATCCTTTAAC-----CATAAATCTGAGCAATA-TATGCCTGGCGGCCAGATTATGGC--ACACTTGTCCGG
sen AATTATCCTTTAAC-----CATAAATCTGAGCAATA-TATGCCTGGCGGCCAGATTATGGC--ACACTTGTCCGG
eco ACGTATCCTTATAC-----CTGAAATCTTCGCAAG--TATGCCTGGCCGCGAGATTATGGC--ACACTTGTCCGG
kpn ATTCATCCTTTCGATATCGCGGTGCTGGAACCAGGTGATGAGTATGCCTGGCGGCCAGATTATGGC--ACACTTCCCAG
ype ATGTTTCAGCAAATAT-----CGGGTACCA-CGCCTGAGCGTTTCCGGCGGGGCAATAGTGGCTTATACTAAGCCCC
      *   **           *   * * *                   *   *** *   **   *****   *   ***   **
  
```

```

sty TTAACTCTCGTT-CTCAACAG-----GTACGACAGTC--GTGAAAATTCTCGTTGATGAAAATATGCCTTACGCCCGC
stm TTAACTCTCGTT-CTCAACAG-----GTACGACAGTC--GTGAAAATTCTCGTTGATGAAAATATGCCTTACGCCCGC
sen TTAACTCTCGTT-CTCAACAG-----GTACGACAGTC--GTGAAAATTCTCGTTGATGAAAATATGCCTTACGCCCGC
eco TTAACTCTCGT--CTCATAACAG-----GTAACACAAAC--GTGAAAATCCTTGTGTTGATGAAAATATGCCTTATGCCCGC
kpn TTAACTCTCGTT-CTCAGACAG-----GTAAGTAACT--GTGAAAATCCTCGTTGATGAAAATATGCCCTATGCCCGT
ype CTGTTTTTCATCTGTATGGCAGTTCGCTGTGGAGAGTAAAGTAAAATTTCTGGTTGATGAAAATATGCCGTACGCTGAG
      *   * * * *   *   ***           **           *   ***** * * ***** * * * *
  
```

- Гены консервативнее, чем межгенные области (точнее, особенности эволюции другие)

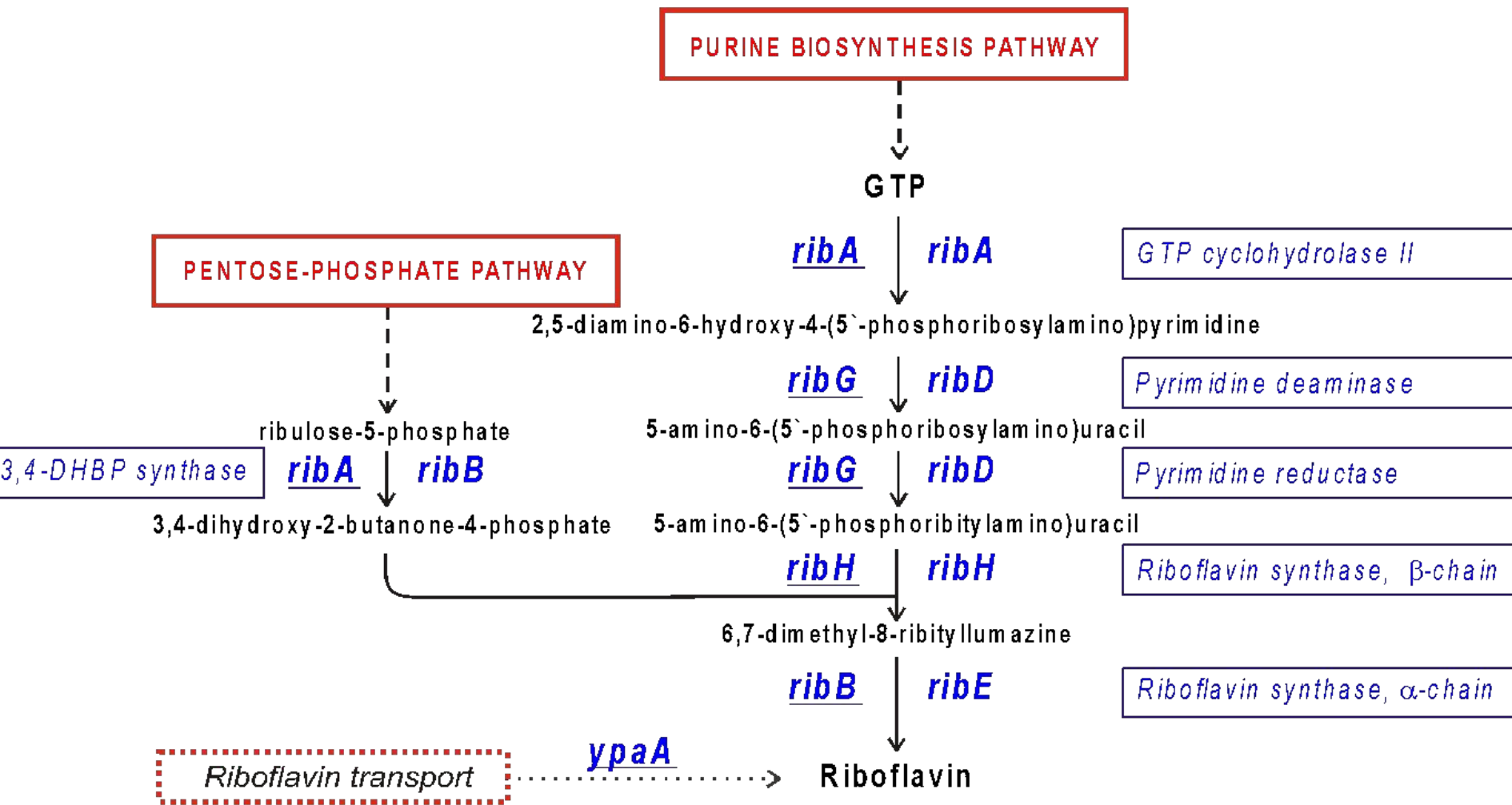
# Мораль

- ***Комплексный подход:***  
использование многих разнородных соображений, каждое из которых по отдельности – слабое
- ***Сравнительный подход:***  
одновременный анализ множества геномов (находящихся на различных эволюционных расстояниях друг от друга)

# Как предсказывать функции

- Белки, похожие по последовательности, имеют сходные функции
- Если нет родственных белков с известной функцией, то:
  - предсказав структурные особенности, можно определить функциональный класс
  - изучение геномного контекста позволяет отнести белок к функциональной подсистеме

# Метаболический путь биосинтеза рибофлавина (витамина В<sub>2</sub>)

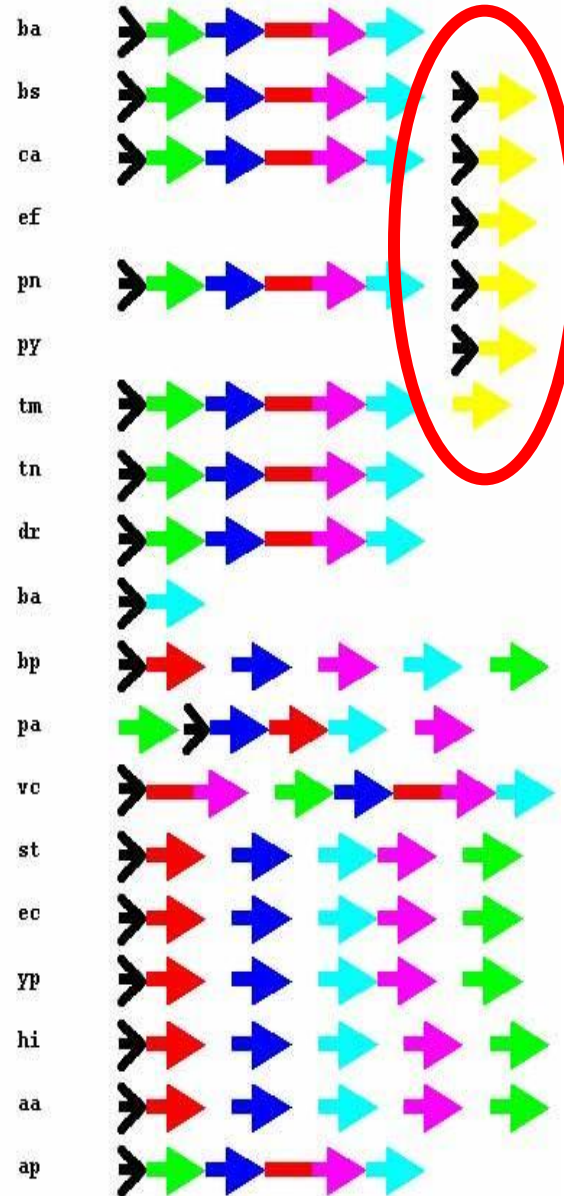
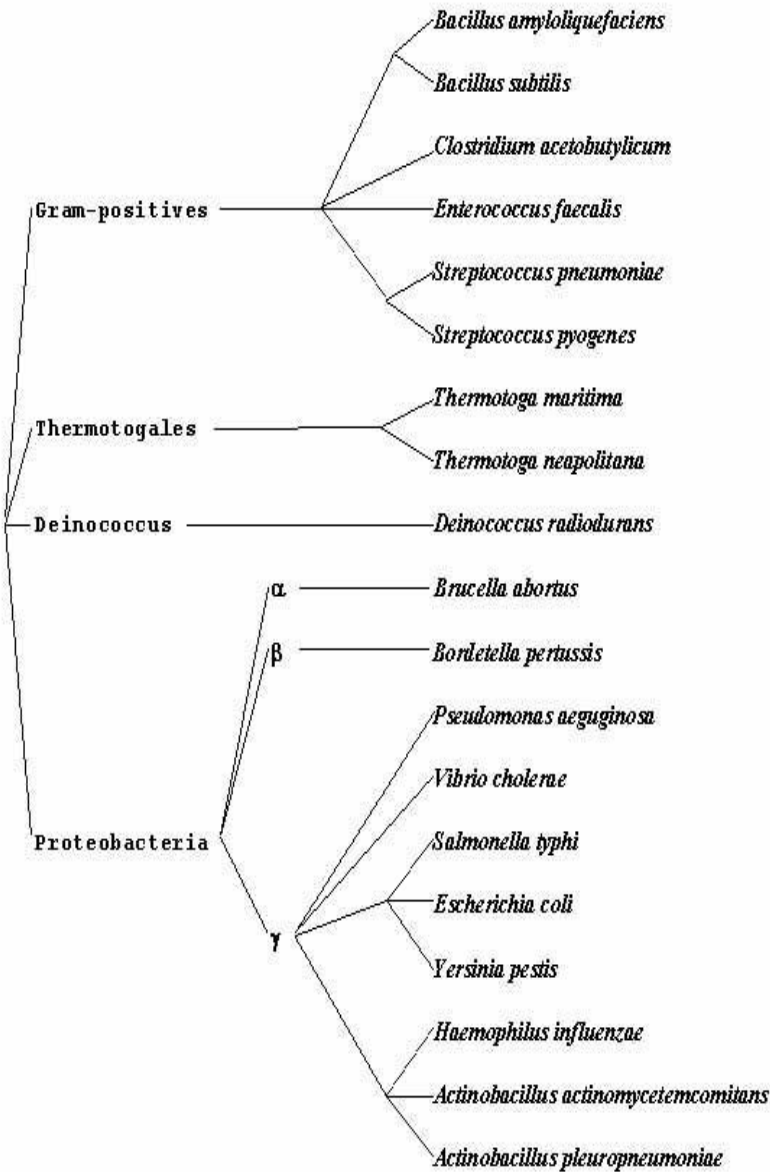




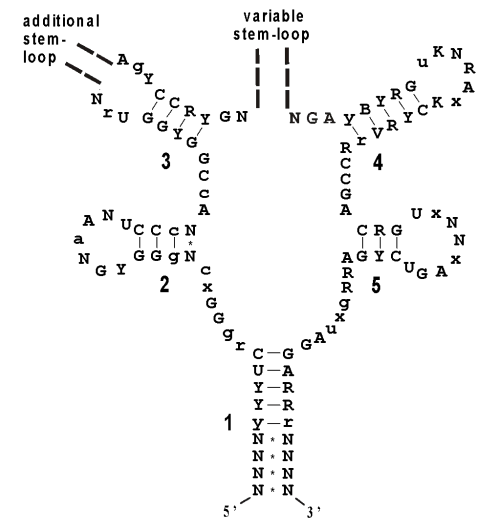
# Консервативная последовательность перед генами рибофлавинового пути

	1	2	2'	3	Add.	3'	Variable	4	4'	5	5'	1'
BS	TTGTATCTTGGGG-CAGGGTGGAAATCCCACCCGGCGGT	21	AGCCCGTGAC--	8	4	8	-----TGGATTTCAGTTTAA-GCTGAAGCCGACAGT GAA-AGTCTGGAT-GGGAGAAGGATGAT					
BQ	AGCATCCTTGGGG-TCGGGTGAAATCCCACCCGGCGGT	19	AGTCCGTGAC--	8	5	8	-----TGGATCTAGT GAAACTCTAGGCCCGACAGT -AT-AGTCTGGAT-GGGAGAAGGATATG					
BE	TGGATCCTTGGGG-CAGGGTGGAAATCCCACCCGGCGGT	20	AGCCCGGGA--	3	4	3	-----AGGATCCGGT GCGAATCCGGAGCCGACAGT -AT-AGTCTGGAT-GGGAGAAGGATGCG					
BD	TTTTATCCTTGGGG-CTGGGTGGAAATCCCACCCGGCGGT	19	AGTCCGTGAC--	10	4	10	-----TGGACTCTGT GAAACTCCGGGACCGACAGT GAA-AGTCTGGAT-GGGAGAAGGAAACC					
Bam	TGTATCCTTGGGG-CTGGGTGAAATCCCACCCGGCGGT	23	AGCCCGTGAC--	8	4	8	-----TGGATTTCAGT GAAAGCTGAAGCCGACAGT GAA-AGTCTGGAT-GGGAGAAGGATGAG					
CA	GATGTTCTTCAAGG-ATGGGTGAAATCCCACCCGGCGGT	2	AGCCCGCAA--	3	4	3	-----AGATCCGGT TAAACTCCGGGGCCGACAGT AA-AGTCTGGAT-GAAAGAAGAAATAG					
DE	CTTAATCTTGGGG-TAGGGTGGAAATCCCACCCGGCGGT	2	AGCCCGGG--	7	6	7	-----ATTGTGTTAACTCCAAAGCCGACAGT AA-AGTCTGGAT-GGAAGAAGATATTT					
SA	TAAATCCTTGGGG-CAGGGTGGAAATCCCACCCGGCGGT	6	AGCCCTGGAC--	11	3	11	-----CTGATCTAGT GAGATTCTAGAGCCGACAGT TAA-AGTCTGGAT-GGGAGAAGCAATGT					
LLX	ATAAATCTTCAAGG-CAGGGTGTAAATCCCACCCGGCGGT	2	AGCCCGGGA--	4	4	4	-----ATGATTTGGT GAAACTCCAGGCCCGACAGT -AT-AGTCTGGAT-GAAAGAAGATTAATA					
PN	AACATCTTCAAGG-CAGGGTGGAAATCCCACCCGGCGGT	2	AGCCCGACGA--	3	4	3	-----ATGATTTGGT GAAACTCCAAAGCCGACAGT -AT-AGTCTGGAT-GAAAGAAGATAAAA					
TM	AAACGCTCTTGGGG-CAGGGTGGAAATCCCACCCGGCGGT	3	AGCCCGGAG--	5	4	5	-----TTGACCCGGT GGAATCCGGGGCCGACGGT GAA-AGTCTGGAT-GGGAGAGAGCGTGA					
DR	GACCTCCTTGGGG-CGGGGCGAAATCCCACCCGGCGGT	15	AGCCCGGAA--	8	12	9	-----CCGATGCCCGGCAACTCCGGAGCCGACGGT CAC-AGTCTGGAT-GAAAGAAGGAGGAG					
TQ	CACCTCCTTGGGG-CGGGGTGGAAATCCCACCCGGCGGT	3	AGCCCGGAA--	5	4	5	-----CCGACCCGGT GGAATCCGGGGCCGACGGT GAA-AGTCTGGAT-GGGAGAAGGAGGGC					
AO	AATAATCTTCAAGG-CAGGGTGGAAATCCCACCCGGCGGT	2	AGTCCGGGA--	7	7	7	-----AGGAACCCGT GAGATTCCGGTACCGACAGT -AT-AGTCTGGAT-GGAAGAAGATGAAA					
DU	TTTAATCTTCAAGG-CAGGGTGGAAATCCCACCCGGCGGT	2	AGTCCGGGA--	13	4	12	-----AGGAACCTAGT GAAATCTAGTACCGACAGT -AT-AGTCTGGAT-GGAAGAAGAGCAGA					
CAU	GAAAGCCTTGGGG-CAGGGTGGAAATCCCACCCGGCGGT	20	AGCCCGGGA--	3	4	3	-----AGGACCCGGT GTGATTCGGGGCCGACGGT -AT-AGTCTGGAT-GGGAGAAGGTGGCC					
FN	TAAAGTCTTCAAGG-CAGGGTGGAAATCCCACCCGGCGGT	2	AGTCCAG--	5	4	5	-----GATTTGGT GAAACTCCAAAGCCGACAGT -AG-AGTCTGGAT-GGGAGAAGAAATAG					
TEU	ACCGTCTTGGGG-GTGGGTGAAAGTCCGACCCGGCGGT	3	AGTCCGGAC--	8	5	8	-----TGGAACCCGT GAAACTCCGGTACCGACGGT GAA-AGTCTGGAT-GGGAGGTAGTACGG					
SX	-AGCCCACTTGGGG-GTGGGTGAAAGTCCGACCCGGCGGT	3	AGTCCGGAC--	8	5	8	-----TTGACCCAGT GAAATCTGGACCGACGGT TAA-AGTCTGGAT-GGGAGGCACTGGCCG					
BU	GTCCGCTTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	30	AGCCCGGGAGCG	13	7		GT CAGCAGATCTGGT GAGAAGCAGAGCCGACGGT TAG-AGTCTGGAT-GGAAGAAGATGTGC					
BPS	GTCCGCTTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	21	AGCCCGGGAGCG	8	4	8	GT CAGCAGATCTGGT CCGATTCCAGAGCCGACGGT CAT-AGTCTGGAT-GAAAGAAGATGTGC					
RBV	TTACGCTTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	31	AGCCCGGGAGCG	7	5	7	GT CAGCAGATCTGGT GAGATCCAGGGCCGACGGT TAA-AGTCTGGAT-GAAAGAAGATGGCC					
RSO	GTACGCTTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	21	AGCCCGGGAGCG	11	3	11	GT CAGCAGATCCGGT GAGATCCGGGGCCGACGGT CAG-AGTCTGGAT-GGAAGAAGATGTGC					
EC	GCTTATCTTCAAGG-CGGGGTGGAAATCCCACCCGGCGGT	17	AGCCCGGAGCG	8	4	8	GACAGCAGATCCGGT GTAATTCGGGGCCGACGGT TAG-AGTCTGGAT-GGGAGAGAGTAAAG					
TY	GCTTATCTTCAAGG-CGGGGTGGAAATCCCACCCGGCGGT	67	AGCCCGGAGCG	8	3	8	GT CAGCAGATCCGGT GTAATTCGGGGCCGACGGT TAA-AGTCTGGAT-GGGAGAGAGTAAAG					
KP	GCTTATCTTCAAGG-CGGGGTGGAAATCCCACCCGGCGGT	20	AGCCCGGAGCG	8	4	8	GT CAGCAGATCCGGT GTAATTCGGGGCCGACGGT TAA-AGTCTGGAT-GGGAGAGAGTAAAG					
HI	TCGCACTTTCAGGG-CAGGGTGGAAATCCCACCCGGCGGT	2	AGCCCGGAGCG	26	9	30	GT CAGCAGATTTGGT GAAATCCAAAGCCGACAGT -AA-AGTCTGGAT-GAAAGAGAAATAAA					
VK	GCGCATCTTCAAGG-CAGGGTGGAAATCCCACCCGGCGGT	14	AGCCCGGAGCG	11	9	11	GT CAGCAGATTTGGT GAGAATCCAAAGCCGACAGT -AT-AGTCTGGAT-GAAAGAGAAATAAAG					
VC	CAATATCTTCAAGG-CGGGGTGGAAATCCCACCCGGCGGT	13	AGCCCGGAGCG	5	4	5	GT CAGCAGATCTGGT GAGAAGCAGGGCCGACGGT TAC-AGTCTGGAT-GAGAGAGAAATGACA					
YP	GCTTATCTTCAAGG-CGGGGTGGAAAGTCCCACCCGGCGGT	40	AGCCCGGAGCG	16	6	16	GT CAGCAGATCCGGT GTAATTCGGGGCCGACGGT TAT-AGTCTGGAT-GGGAGAGAGTAAAG					
AB	GCGCATCTTCAAGG-CAGGGTGGAAATCCCACCCGGCGGT	25	AGCCCGGAGCG	16	4	27	GT CAGCAGATTTGGT GCGAATCCAAAGCCGACAGT GAC-AGTCTGGAT-GAAAGAGAAATAAA					
BP	GTACGCTTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	18	AGCCCGGAGCG	10	4	10	GT CAGCAGACTTGGT GAGATCCAGGGCCGACGGT CAT-AGTCTGGAT-GAGAGAAGATGTGC					
AC	ACATCTTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	16	AGCCCGGAGCG	10	3	11	---CGCAGATCTGGT GTAATTCAGAGCCGACGGT -AT-AGTCTGGAT-GAAAGAAGACGAGC					
Spu	AACATCTTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	34	AGCCCGGAGCG	6	6	6	GT CAGCAGATCTGGT G 52 TCCAGAGCCGACGGT 31 AGTCTGGAT-GGAAGAGAAATGTAA					
PP	GTCCGCTTTCAGGG-CGGGGTGTAAATCCCACCCGGCGGT	13	AGCCCGGAGCG	7	3	7	GT CAGCAGATCTGGT GCAACTCCAGAGCCGACGGT CAT-AGTCTGGAT-GAAAGAAGGCGTCA					
AU	GGTGTCTTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	17	AGCCCGGAGCG	7	9	7	GT CAGCAGATCCGGT GAGAGGCCAGAGCCGACGGT -AT-AGTCTGGAT-GAAAGAGGCAAGG					
PU	AAACGCTTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	19	AGCCCGGAGCG	19	4	18	GT CAGCAGACTCCGGT GTGATTCGGGGCCGACGGT CAC-AGTCTGGAT-GAAGAGAGAAAGGGA					
PY	TAAAGTCTTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	19	AGCCCGGAGCG	15	4	16	GT CAGCAGATCCGGT GTGATTCGGGGCCGACGGT CAT-AGTCTGGAT-GAAGAGAGAGCGGGA					
PA	TAAAGTCTTTCAGGG-CGGGGTGGAAAGTCCCACCCGGCGGT	19	AGCCCGGAGCG	14	4	13	GT CAGCAGACTCCGGT GCGAATTCGGGGCCGACGGT CAT-AGTCTGGAT-AAAGAGAGAAAGCGGA					
ML0	TAAAGTCTTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	16	AGCCCGGAGCG	8	5	8	GT CAGCAGATCCGGT GTGATTCGGAGCCGACGGT TAG-AGTCTGGAT-GAAAGAGGAGAAA					
SM	AAGCGTCTTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	34	AGCCCGGAGCG	8	3	8	GT CAGCAGATCCGGT GCAATTCGGAGCCGACGGT TAT-AGTCTGGAT-GGAAGAGAGCAAGC					
BMB	GCTTGTCTTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	17	AGCCCGGAGCG	10	15	10	GT CAGCAGATCCGGT GAGATGCCGGAGCCGACGGT TAA-AGTCTGGAT-GGAAGAGAGCGGATA					
BS	AICAATCTTGGGG-CAGGGTGGAAATCCCACCCGGCGGT	18	AGCCCGGGA--	5	4	5	-----AGGATTCGGT GAGATTCCGAGCCGACAGT -AA-AGTCTGGAT-GGGAGAAGATGGAG					
BQ	GTCTATCTTGGGG-CAGGGTGGAAATCCCACCCGGCGGT	27	AGCCCGGGA--	3	5	3	-----AGGATTTGGT GTGATTCAAAGCCGACAGT -AT-AGTCTGGAT-GGGAGAAGATGGAG					
BE	ATTCATCTTGGGG-CAGGGTGGAAATCCCACCCGGCGGT	20	AGCCCGGGA--	3	4	3	-----AGGATTCGGT GCGAGTCCGGAGCCGACAGT -AT-AGTCTGGAT-GGGAGAAGATGAAAG					
CA	AATGATCTTCAAGG-CAGGGTGGAAATCCCACCCGGCGGT	2	AGCCCGGAG--	3	4	3	-----TATGATCCGGT TTGATTCGGAGCCGACAGT -AA-AGTCTGGAT-GAAAGAAGATATAT					
DE	GAAATCTTTCAGGG-CAGGGTGGAAATCCCACCCGGCGGT	2	AGCCCGGG--	6	4	6	-----GATTTGGT GAGATTCCAAAGCCGACAGT -AA-AGTCTGGAT-GAGAGAAGATATTT					
EE	GTTCGCTTTCAGGGG-CAGGGTGGAAATCCCACCCGGCGGT	3	AGTCCAGAC--	5	3	5	-----ATTGATTTGGT GTAATTCATACCGACAGT -AT-AGTCTGGAT--AAAGAAGATAGGG					
LLX	AAATATCTTTCAGGG-CACCGTGTAAATCCCACCCGGCGGT	21	ACTCCGGAT--	4	4	4	-----TTGAAGCACT GAGAATCTGCTAGCCGACAGT -AA-AGTCTGGAT-GGAAGAAGATGAAC					
LO	GTTATCTTTCAGGG-CAGGGTGGAAATCCCACCCGGCGGT	3	AGTCCAGAT--	3	10	3	-----TTGACTCTGGT GTAATTCAGGAGCCGACAGT -AT-AGTCTGGAT-GGGAGAAGATGTTG					
PN	AAGAGCTTTCAGGG-CAGGGTGGAAATCCCACCCGGCGGT	12.5	AGTCCGTG--	3	4	3	-----GATGTGGT GAGATTCCACAGCCGACAGT -AT-AGTCTGGAT-GGGAGAAGAGGAAA					
ST	AAGTGTCTTTCAGGG-CAGGGTGTGATTCACCCGGCGGT	14	AGTCCGGG--	3	4	3	-----GATGTGGT GTAACTCCACAGCCGACAGT -AT-AGTCTGGAT-GAGAGAAGACCGGG					
MN	AAGTGTCTTTCAGGG-CAGGGTGTGATTCACCCGGCGGT	10.4	AGTCCGGG--	3	4	3	-----GATGTGGT GAAATTCACAGCCGACAGT -AA-AGTCTGGAT-GGGAGAAGACTGAG					
SA	ATTCATCTTTCAGGG-TCGGGTGTAAATCCCACCCGGCGGT	6	AGCTCTGGAC--	11	3	11	-----CTGATCTAGT GAGATTCTAGAGCCGACAGT -AT-AGTCTGGAT-GGGAGAAGATGGAG					
AMI	TCACAGTCTTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	14	AGCCCGGCG--	5	5	5	-----TGATCTGGT GCAACTCCAGAGCCAGCGGT -AT-AGTCTGGAT-GGAAGAAACGGAGC					
DHA	ACGAACTTTCAGGG-TAGGGTGGAAATCCCACCCGGCGGT	20	AGCCCGAAC--	11	4	11	---CGACTGACTTGGT GAGACTCCAAAGCCGACGGT -AT-AGTCTGGAT-GGGAGAAGTAAAGAA					
FN	AATAATCTTTCAGGG-CAGGGTGGAAATCCCACCCGGCGGT	2	AGTCCAG--	4	6	4	-----GATTTGGT GAAATTCAAAGCCGACAGT -AG-AGTCTGGAT-GAGAGAAGAAAGAA					
GLU	---TGTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	28	AGCCCGGAGCG	10	4	10	GT CAGCAGATCCGGT TAAATTCGGAGCCGACGGT CAT-AGTCTGGAT-GCAAGAGAAAC---					

# ... и еще перед одним геном (*uraA*)



цветные стрелки – гены пути  
 желтые стрелки – *uraA*, ген с неизвестной функцией  
 черные стрелки – регуляторный элемент



# 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* spp.:  
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 микрочипах, Lee et al., 2001; прямой эксперимент, Winkler et al., 2002).

# Молекулярная эволюция

Symp Soc Exp Biol. 1958;12:138-63.

**On protein synthesis.**

CRICK FH.

J Theor Biol. 1965 Mar;8(2):357-66.

**Molecules as documents of evolutionary history.**

Zuckerlandl E, Pauling L.

После расхождения видов изменения в генах накапливаются почти случайно  
=> гены в близких видах более похожи, чем в далеких  
=> степень различия генов можно использовать для датировки расхождений

# Молекулярная палеонтология: как видели динозавры

## Recreating a Functional Ancestral Archosaur Visual Pigment

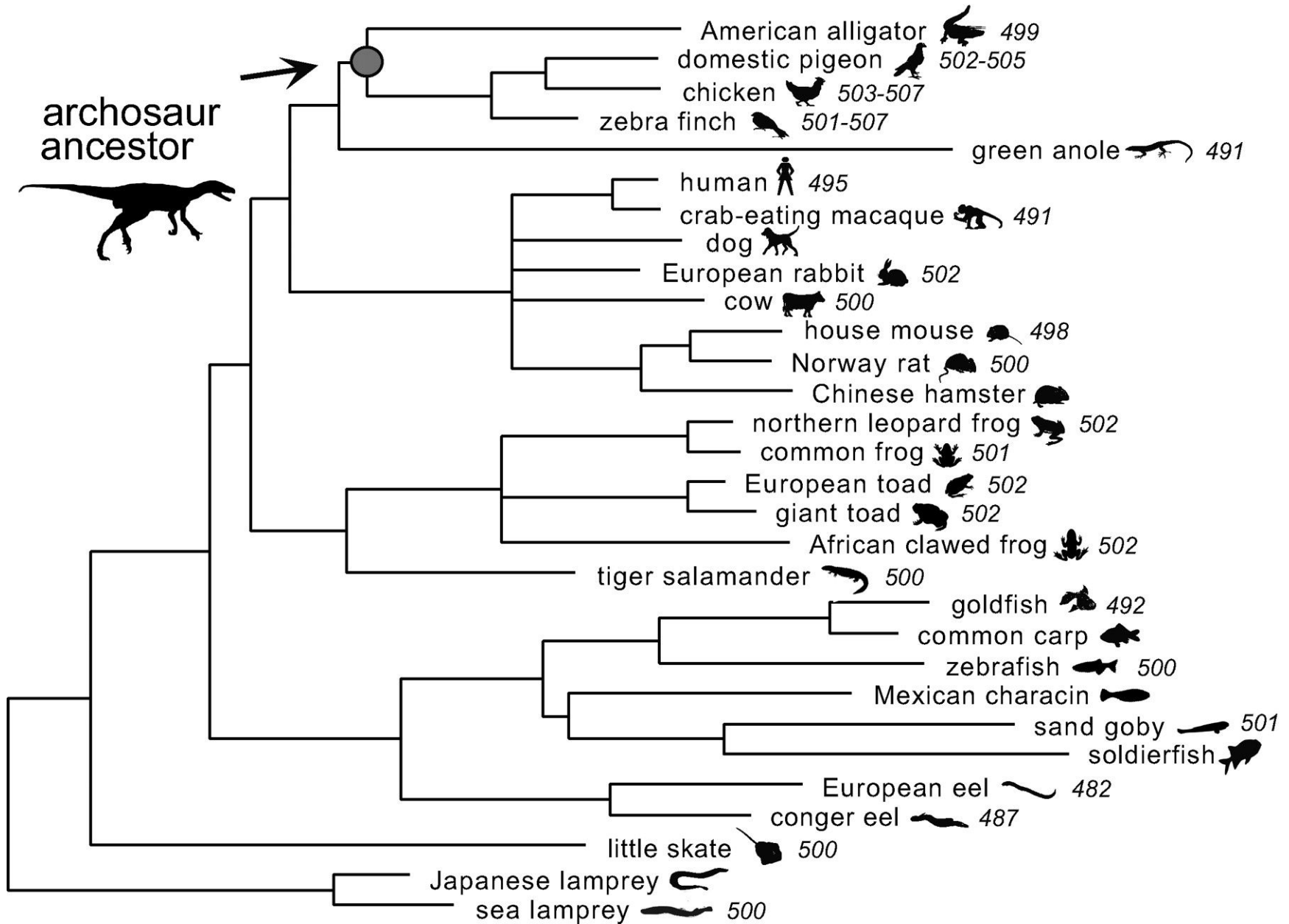
*Belinda S. W. Chang,\* Karolina Jönsson,\* Manija A. Kazmi,\* Michael J. Donoghue,† and Thomas P. Sakmar\**

\*Laboratory of Molecular Biology & Biochemistry, The Rockefeller University, New York; and †Department of Ecology & Evolutionary Biology, Yale University

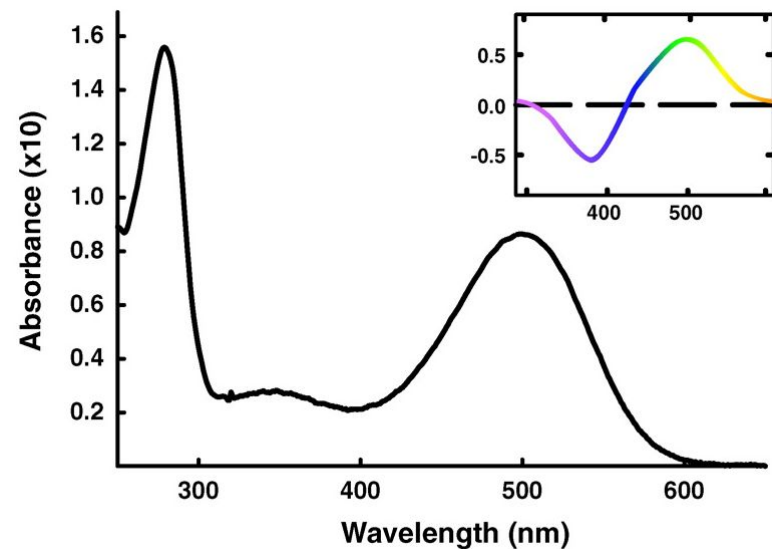
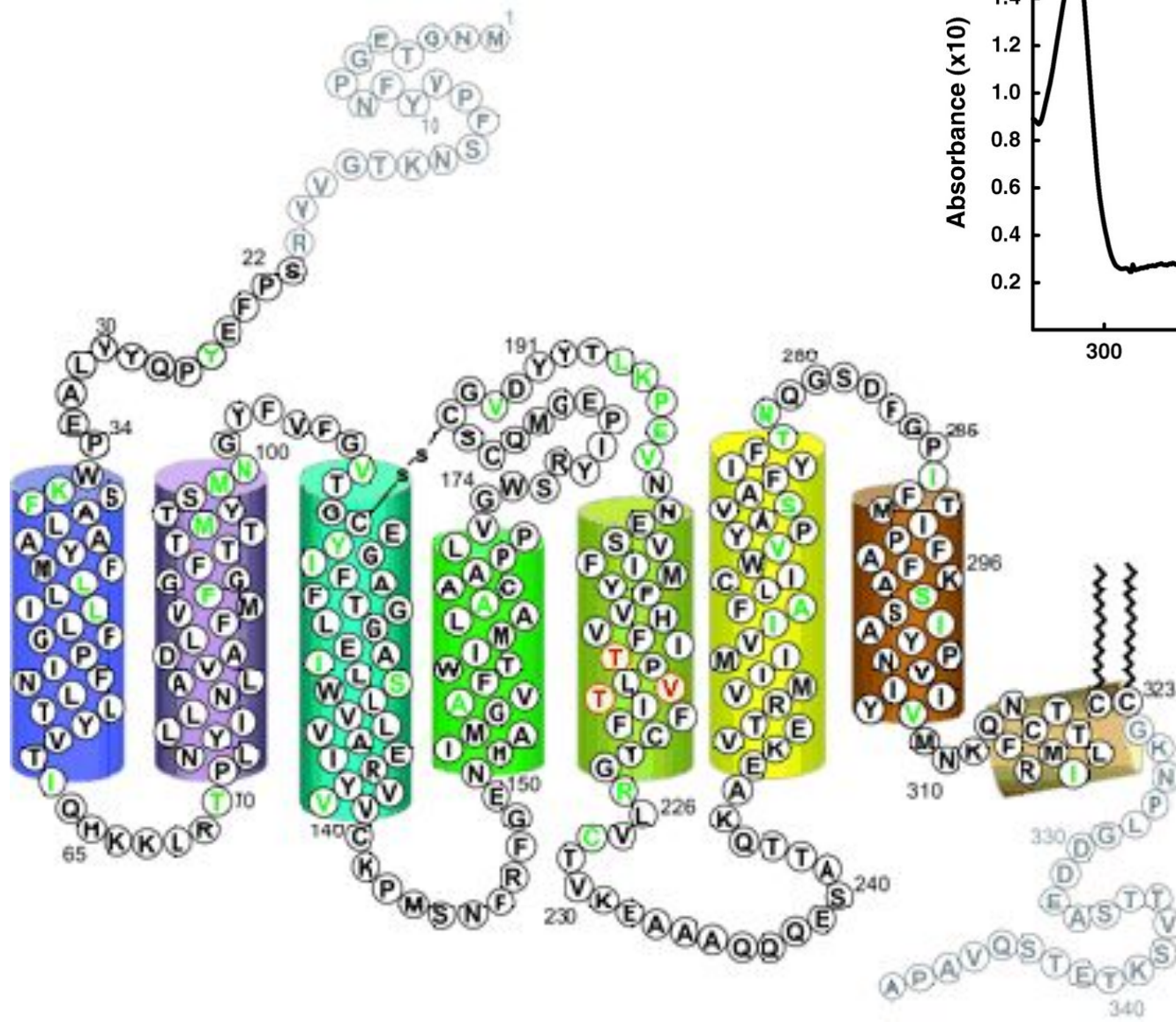
*Mol. Biol. Evol.* 19(9):1483–1489. 2002

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

# Таксономическое дерево



# Реконструированный белок и его свойства



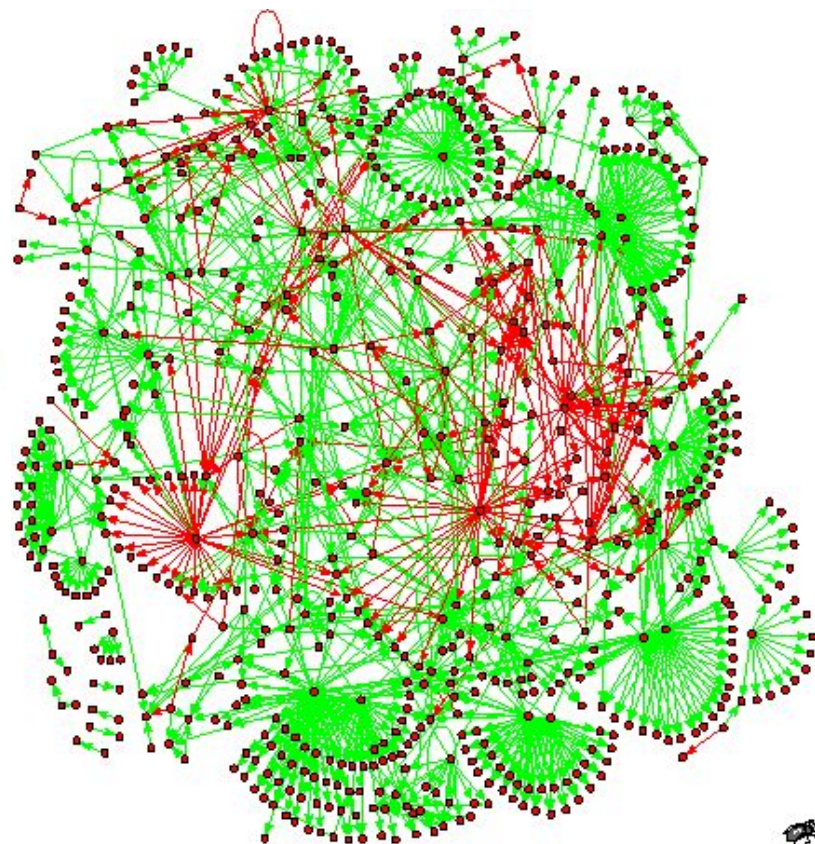
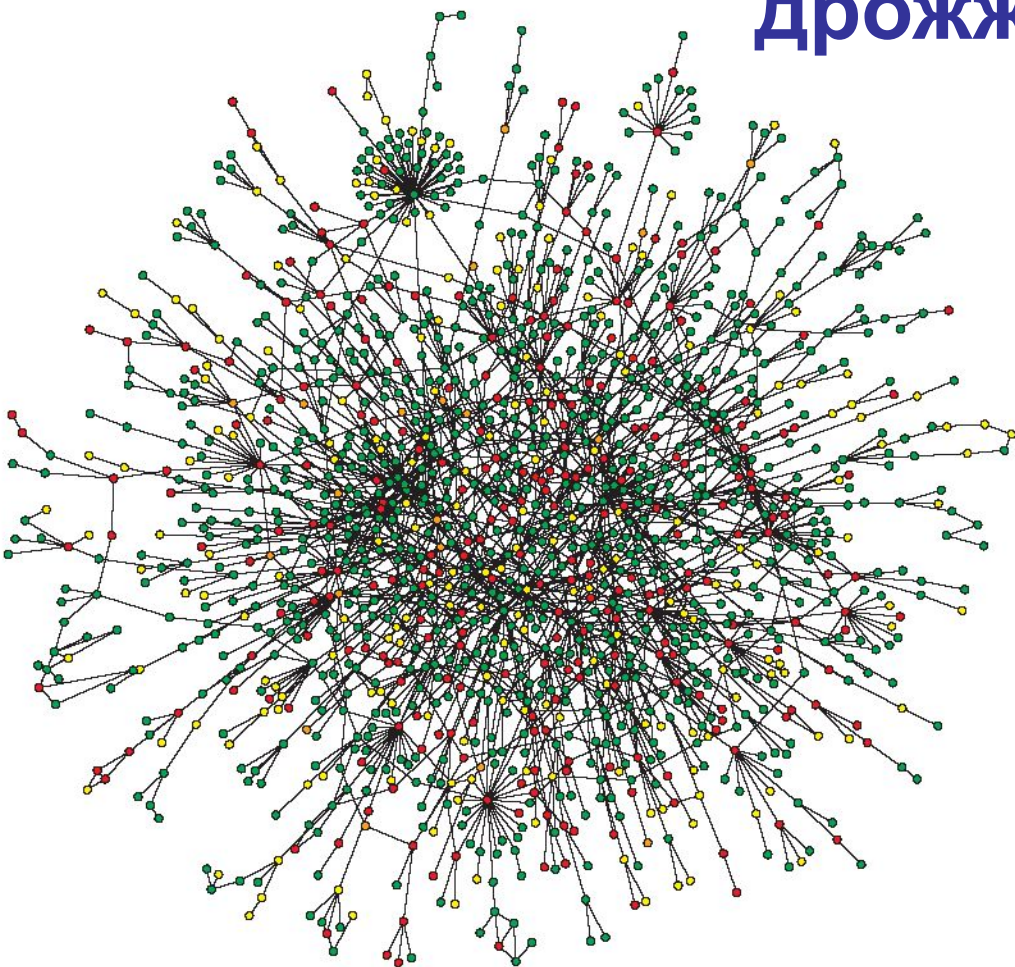
# Не только тексты

Можно использовать данные, которые порождаются другими типами массовых экспериментов

- Уровень экспрессии:
  - Концентрации мРНК
  - Концентрации белков
  - Время жизни мРНК и белков
- Взаимодействия:
  - Белок-ДНКовые
  - Белок-белковые
- Структура генома
  - Метилирование ДНК
  - Положение и модификация нуклеосом
- Функционально-генетические
  - Летальность мутаций
  - Фенотип
  - Синтетические летали



# Графы белок-белковых (структурных, сигнальных и др.) и белок-ДНКовых (регуляторных) взаимодействий в дрожжах



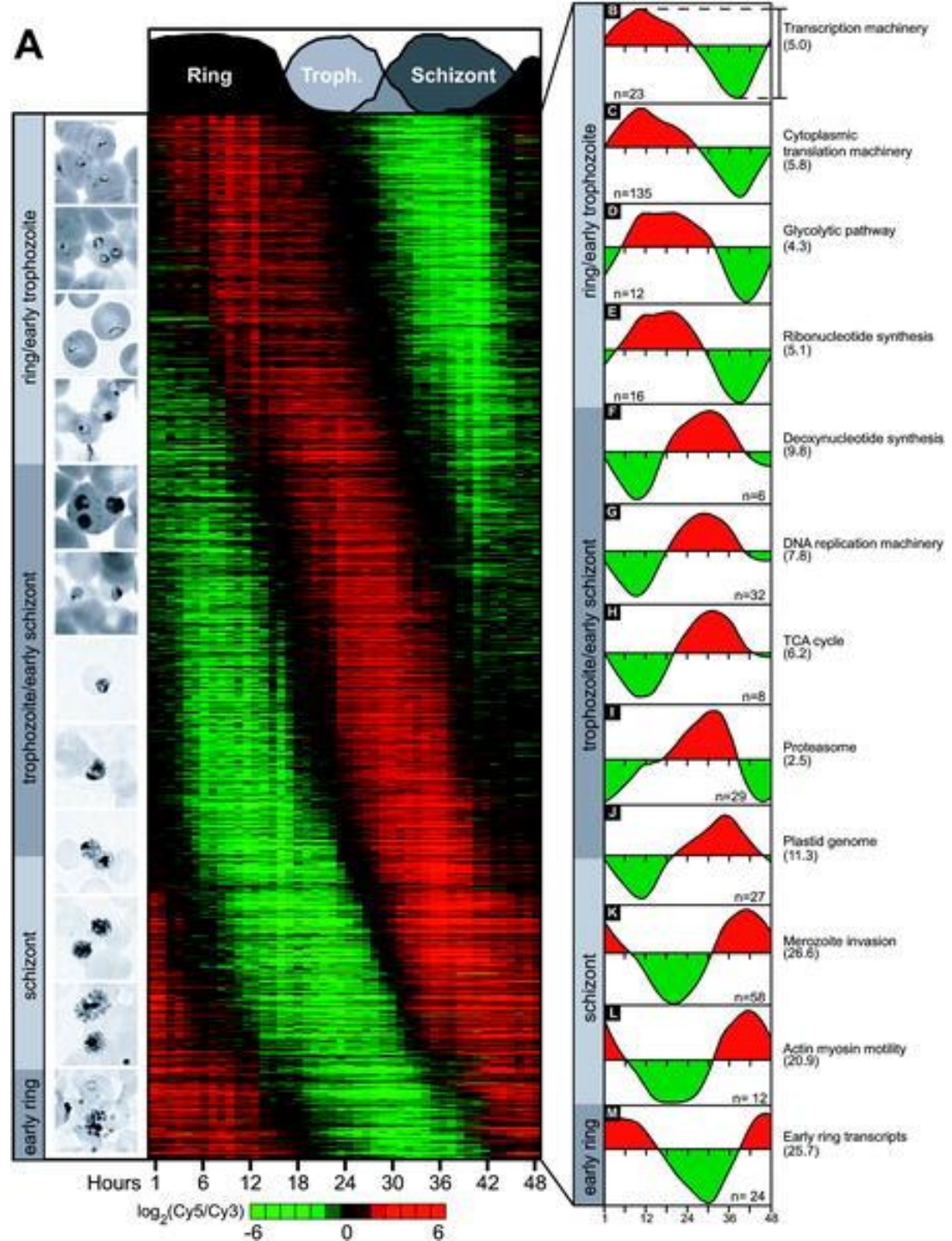
# Экспрессия (уровень работы) генов

Цикл развития  
малярийного плазмодия



The Transcriptome  
of the Intraerythrocytic Developmental Cycle  
of *Plasmodium falciparum*

Zbynek Bozdech<sup>1</sup>, Manuel Llinás<sup>1</sup>, Brian Lee Pulliam<sup>1</sup>, Edith D. Wong<sup>1</sup>, Jingchun Zhu<sup>2</sup>, Joseph L. DeRisi<sup>1\*</sup>



# Биоинформатика

- Биоинформатика – это биология *in silico*
- Кто этим занимается:
  - биологи, математики, физики, химики
  - ... биоинформатики
- Хорошая биоинформатика связана с биологией
  - обработка результатов
  - проверка предсказаний
  - в хороших современных проектах эти шаги повторяются несколько раз, начиная с этапа планирования

# Происхождение жизни

On the origin of genomes and cells within inorganic compartments

Eugene V. Koonin<sup>a</sup> and William Martin<sup>b</sup>

Trends in Genetics

Volume 21, Issue 12, December 2005, Pages 647-654

