

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 (0.1% генома *E. coli*)

aacgggcaaatatgtctctgtgtggattaaaaaagagtgctctgatagcagctcttgaactggttacc tgcctgtagtaaataaaaattta ttgac ttaggtcactaaatactttaaccaatataagga
tagcgcaagacagataaaaat tacagagtacacaac atcca tgaac cgc at tagca cccacc attac caccaccacc accat tacc acaggtaacggtgctggggctgacgggtacaggaaac acagaaaa
aagcccgcacctgacagtgccgggtttttttcgacc aaaggtaacgaggtaacaac catgc gagggttgaa gttcggcgggtacatcagtgggcaaaa tgcagAACgtttctgccgggtgccc gatattct
ggaagcaaatgyc caggc agggg cagggtggcgaaccgtctctct tgcctccggcc aaaaat taccaccatctggttagcga tgaattgaaaaaaccc attagcggcc aagatgctttacccaaatatc agcga tgc
cgaacgtatttt tgcgcaactctgac gggac tgcctccggcc ccagc cgggaattcc gctgg cacaattgaaaaccttctgtc gacc aggaatttgc ccaaaa taaaa catgtcctgc atggc attagttt
gttggggcagtgcccggatagc atcaacgctgctgctgatttggcctggtggagaaaaatgtcga tgcct attatggcggcgtgtttagaacgctggtggtcacaacggtaccgttatcgatccgggtcgaaaa
actgcttgacgtggggc attac ctgcaatctaccgttgatattgtctgagtc acccggcgtattggggcaacggcattccggctgaccac atgggtctga tggctggtttactgcccggtaaatgaaaa
aggcgagctggtggttc tgggacgcaacggttccgac tactcctgctgcccgtgctggcggcctgtttacgccc cgaattgttgc gaga tctggacggga tgttgacgggtgttta tactgcatccggctca
ggtgccc gatgc gagggtgttgaagtc gatgtctatcagggaagoga tggagctttc ttaacttoggc gctaaagtcttccac cccc gcacc atcac cccca tgcctcagtttcagatacccttgcctgat
taaaaaatccggaaatcctcaagctcc aggtacgctc attgggtgccaagccgtgatgaagcaaatcaccggtcaagggcatttccaatctgaaataa catggcaatgttccagcgttccggcc cgggggat
gaaaaggatggttggca tggcggcggcgtctcttggcaagcagatgtcac gcccgtat tccgtgtgtgctgat tacgc aatcacttccgaa tacagatcagtttccgctccggc aaagc gactggt
gcgagctgaacggcaatgcaggaagatcttactggaactgaaagaaggc ttactggagc cgttggcgggtgacggaacggctggccattatctc ggtgg taggtgatggtatgc gcacc ttactgg
gatctcggcgaatcttggcc gggctggccc gggcc aatatcaaca ttgtc gccat tgcctc aggga tcttc tgaac gctcaatctctgtc tgggtcaataacgatgatgc gaccactggc gtgcgct
tactcatcagatgctgt tcaat accga tcaagg ttatc gaagtgtttg tgaattggcgtcgggtggcgttggcgggtggc tgcctgagc aactgaagcgtcagc aaagc tgggtgaagaataaacatatacga
cttactgctcgggtgttggtaactc gaaggcactgctcac caatgtacatggccttaatc tggaa aactggcaggaagaa ctggcgcgaagccaa agagc cgtttaatctcgggc gcttaattcgcct
cgtgaaa gaata tcatc tgcgtgaaccc ggtca ttgttgactgtacttccagc caggc agtggcggatcaata tgcggacttctcgc gcgaa ggtttccacgttggtaacggc gaaca aaaaggccaacac
ctcgtcagatggaattactaccatcagttgcgttatgcygggaaaaatcgcggcgtaaaattcc tctatgacac caacgttggggcgtgattaccgggtatcggagaac ctgcaaaatc tgcctc aatgc tgg
tgatgaa ttgat gaagtctcc ggcattctttcaggtctgcctttcttataatc ttcggcaagt tagac gaaggcatgagttctccgaggccagccac actggcgggg gaaatgggttatacc gaacc gga
cccgcgagatgactcttctggatgga tgtggcggcgaagctattgattctc gctcgtgaaa cgggacgtgaactggagctgcccggatattgaaat tgaac ctgtgctgccc cgcagagtttaacgc cga
gggtgatgtcgc cgttttatggcgaa tctgtcacagctcgaagatc tctttggccc gctgtg tggcgaaggc cctgtgatgaa gaaaagttttggcctatgttggc aatat tgatgaaagatggcgtctg
cccgctgaagat tgcgaagtgatgg taatgatccgctgttcaag tgaaaaatggcgaaa acgcc ctggccttctatagc cactattatcagcc gctgc cgttgg tactgcccggatattgggtc ggg
caatgac gttac agctgcgggtgctctt tgcgtatctgtctacgtaccctctca tggaa gtttagagtc tgaca tgggttaagt ttatgccc ggcttccagtgccaa tatgagcgtc ggggttgatgtgc
tcggggc ggcgg tgaca cctgt tgatgtgca ttgctcggagatgtagtcac ggttgaggcg ccagagacat cagtcctcaacaac ctggacgcttggcc gataa gctgc cgtcagagcc acgggaaa
atatacgtttatc agtgc tgggagcgtttttgc cagga gcttggcaagcaaatccag tggcg atgac tctggaaaagaatata gccgatcgg ttcgggcttaaggctc cagcgcctgt tcaagtggctcgg
cgtgatggcgaatgaaca ctggcgaagcgttaatgacactcgtttgctggcctttg atgggcaggt tggaa gggcgtatc tccggcagcattcat tacga caacgtggcaccgtgttttct tgg
gtggtatgacgttgatgattgaagaaa acgac atcatcagtc agcaagtgcc aggggtttgat gaggtygctgt ggggtgctggc gtatccggggatta aagtc tgcac ggcagaagcc agggc tattt tac
cggcgcagatc gccgc cagga ttgca ttggccacggcggac atctggcaggcttca ttcac gctcgtattcccgctcagcc tgagcttgc cgcga agctgatgaa agatgttctc gctga accctacc
gtgaacgggttac tggca ggttccggc agggcggcga gggcgttggcgaatcggcgggta gcgagcggta tctcc ggctc cggc ccgac ttgt tgcctctgtg tgaca agccggatc cggcc agc
gcgttgc cgaactggttgggtbaaa aactacctgcaaaa tcaaggaa ggtttgt tcaata tttgc cggctggatacggcggggcgc acgagtactggaaa actaa atgaa actctacaatctgaa agatcaca
atgagcaggtcagctttgycga agccg taacc cagggttgggcaaaa aatca ggggc tgttt tccc gcacgacctgcccga atcc agcctgactgaaattgatgagatgc tgaagctgga tttgtca
cccgcag tgcgaagatcctctc ggcgtttattggtgatgaaa tcccgcagga aatcc tggaa gagcgcgtac gcccggcgtttgccc tccc ggctc cggtc gccaa tgttgaaagc gatgtcggttgct
tgyaattgttcc acgggccaac gctggcatttaaaaat ttcggcgggtcgttttatggcacaataactgaccatattggggcgttaagcc agtgaacattctgaccccagacatcc ggtgactcggag
cggcagtggtc atgcttt
tcgacggcgatttcgatgc
actacttgaagctggttcg
gttttatgtctcgaccaa
gtgtggaagagtgttccg
ctggcgtagcttatcgtg
ctgctgaccaggttgaatcc
caggcgaataggtctgttc
ctcggcaccgcgc atccggc
gaaat taaaga gagcgtg
gaa gcgat tctcggtgaa
acgttggatc tgc
caaaagagctggcagaac
gtgc tgaattaccc ttgct
ttgc ataac ctgcccggc
gattttgctgc gttgc
gtaaa ttgat gatgaa
tca tcaatcattcatta
tctcaatcaggccgggt
tgcttttatgca gccc
gctttttatgaagaaata
tggagaaaaa cga
caggaa aaaggagaaa
ttctc aataa atg
cggtaa attagagatt
aggat tgcggaga
aatacaactgcc gttct
cat

Геном бактерии: несколько миллионов нуклеотидов

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

Фрагмент генома (0.0001% генома человека)

cgtgcac ttctgaaggacttcaggtac cggcgtgccc cggctcctac tgtcc gcectgctcgc gtcctgggtgccc ccttgagtagggcgggagagg
CAGCCAAAGGCGGAGCTGATGGCTGCGC CGAGGGCGGGGCGGGGTGCAAGGCTGGAGCC TTCGGGCATGGCGGGCTTTGGGGGGCATT CGCTGGGGGAGGAG
ACCCCGTTTGAC CCCTGACCTC CGGGC CCTGC TGACGTCAGGAACCTT CTGAC CCCC GGCCCG GAGTGACTTATGGGACCCCCAGTC TCTGGGCCCGGTTG
TCTGTTGGGTC ACTGAACCCC GAGCATGCCTGACGT CTGGGACCCC GGGTC CCCGGGCACA ACTGACTGCGGTGAC CCCAGATAC CAGGACCCGGGAGG
CCTCAGAGA AACTCTGGAACCCGTTCCG CCGCGTGGCTGGCGGTGGCGCTGGGCGCTGGGGGGCAGT GCTGTTGTTGTTGTGGGC GGGGGTCCGGGTCC
TCCGGCC GTCCTCGCCG CCGTC CCTAG CCCC GC CCCC GCTTC TCCCC GGAGT CAGTA CAACTTCATC GCAGA TGTGGTGGAGAAGACAGCA CCTGC CGTG
GTCTATATCGAGATCCTGGACC 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 GGCGA CACGTATGAGGCCGTGGTCA CAGCTGTGGATCCC GTGG
CAGACATCGCAACGCTGAGGATTCAGACTAAG gtggggggctg ggggta gggca ggtct ggttg gagct gctta tttgc tcgca tcttcagatgacaggtct
cttttac ccatttccc ttagg agcct ccccc cacgc tgcctctgggacgctcagctgatgtccggc aaggggagtt tgttgtgcatgggaagtccct
ttgcaactgcagaacacgatcac atccggcattgttagctctgtcagcgtcc agccagagac ctgggactcc cccaa accaa tgtggaatacattc aaac
tgatgcagctat tgatgtgctcctga 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 CCTTCGATC GTCTTCGAGAGTTTCTGCATCGT
GGGGAAAAGAAG Agtga gctgcctta tgggg aaacgggttc ctta atgtggtgga aataggggaagggca ttcag tgggacttc ctgga ggggtggtct
actgggagaaga gggcagggaa ggaag gatgt agctgggtgg ggtc atttgtccctctgtc acagATTCCTCCTCC GGAATCAGTGGGTC CCAGC GGCG
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 agccaactgatttc ctactccc atgttt tttat ttatgctgt
tcttgttttata tgggttaaatat gct
attgatttaacactgtttgtca cat
tgtat tttgttagtacctagccc ttt
ggcta atagggtgatctgtgta aag
tcactctgggctccccctgcaca cccgctcagggg agagggc cggag cggcagcctcaggggcaagggtg cggcagcctcagggc cccgcaac cct
tccttctcctctgtccat tttttctctatag GGCTGGTCTGCGGCCTGGTGATGTGATTTTGGCCATTGGGGAGCAGATGGTACAAAATGCTGAAGATGTT
TATGAAGCTGTT CGAAC CCAATCCCAGTTGGC AGTGCAGATC CCGCGGGGAC GAGAAACACTGACCTTATATGTGAC CCCTGAGGT CACAGAATGAATAG
ATCACCAAGAGTATGAGGCTCC TGCTC TGATTTCCTC CTTGC CTTTC TGGCTGAGGTTCTGAGGGCA CCGAGACAGAGGGTTAAATGAACCAGTGGGGGC
AGGTCCC TCCAAC CACCAGCAC TGA CTCTGGGCTCTGAAgAATCACAGAAA CACTTTTTATATAAAAATAAAAATTATACCTAGCaacatatatagtaaa

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

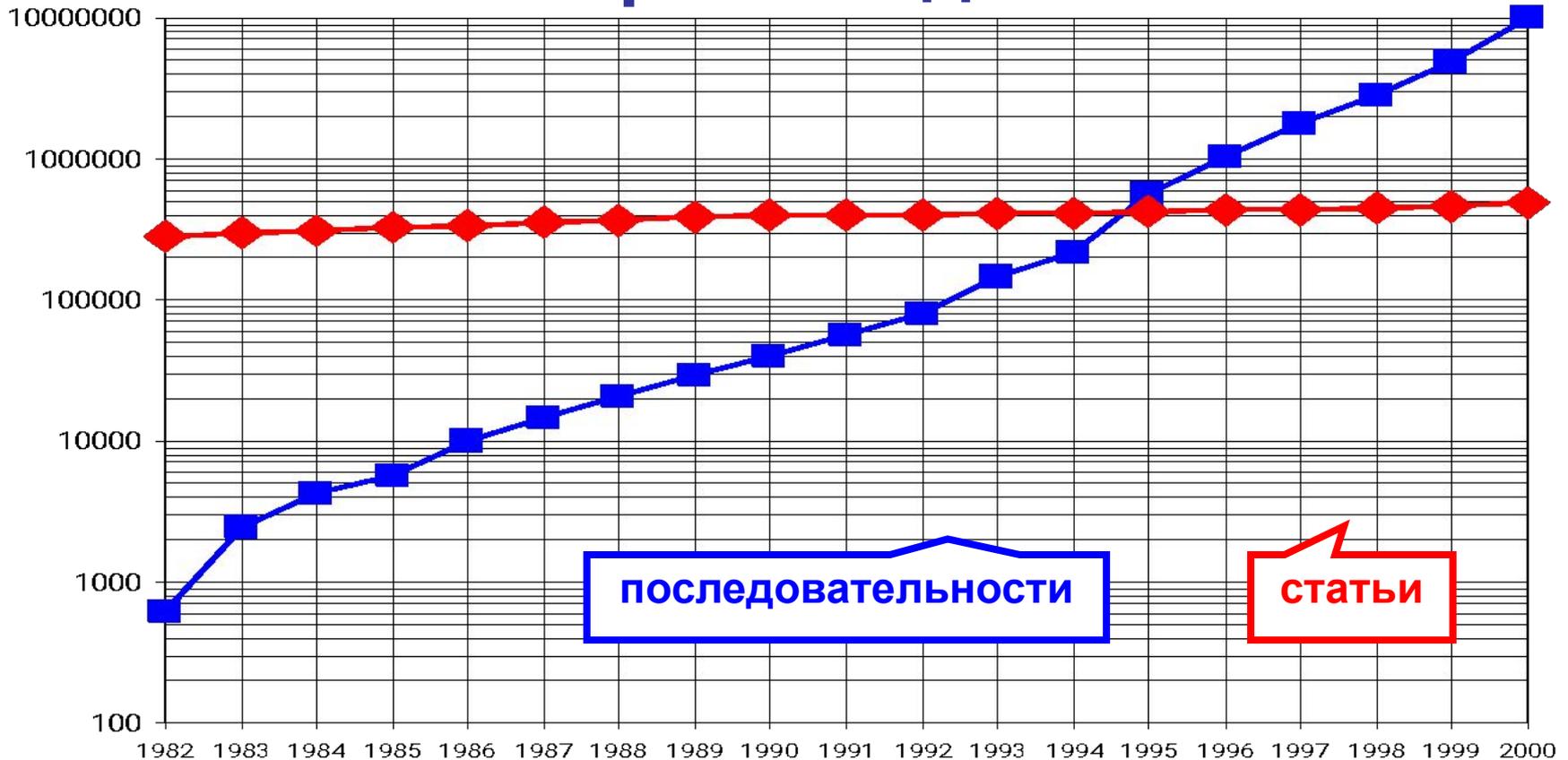
Что же мы хотим понять?

- 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 ^{год} 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
gcgteggtttgatgtgctcggggcggcggtgacacctgttgatgggtgcattgctcggag
D V V T V E A A E T F S L N N L G R F G
atgtagtcacggttgaggcggcagagacattcagttcaacaacctcggacgctttgggt
A D K L P S E P R E N V Y Q C W E R F C
ccgataagctgccgtcagagccacgggaaaatgtttatcagtgctgggagcgtttttgcc
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
ttgttcatatttgccggctggatagggcgggcgcacgagtactggaaaactaaatgaaac
tctacaatctgaaagatcacaatgagcaggtcagctttgcgcaagccgtaaccagggtt
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

gcgtcggggttgatgtgctcggggcgggcggtgacacctggtgatggatgcattgctcggag

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

ccgataaagctgccgtcagagccacgggaaaatgtttatcagtgctgggagcgtttttgcc

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

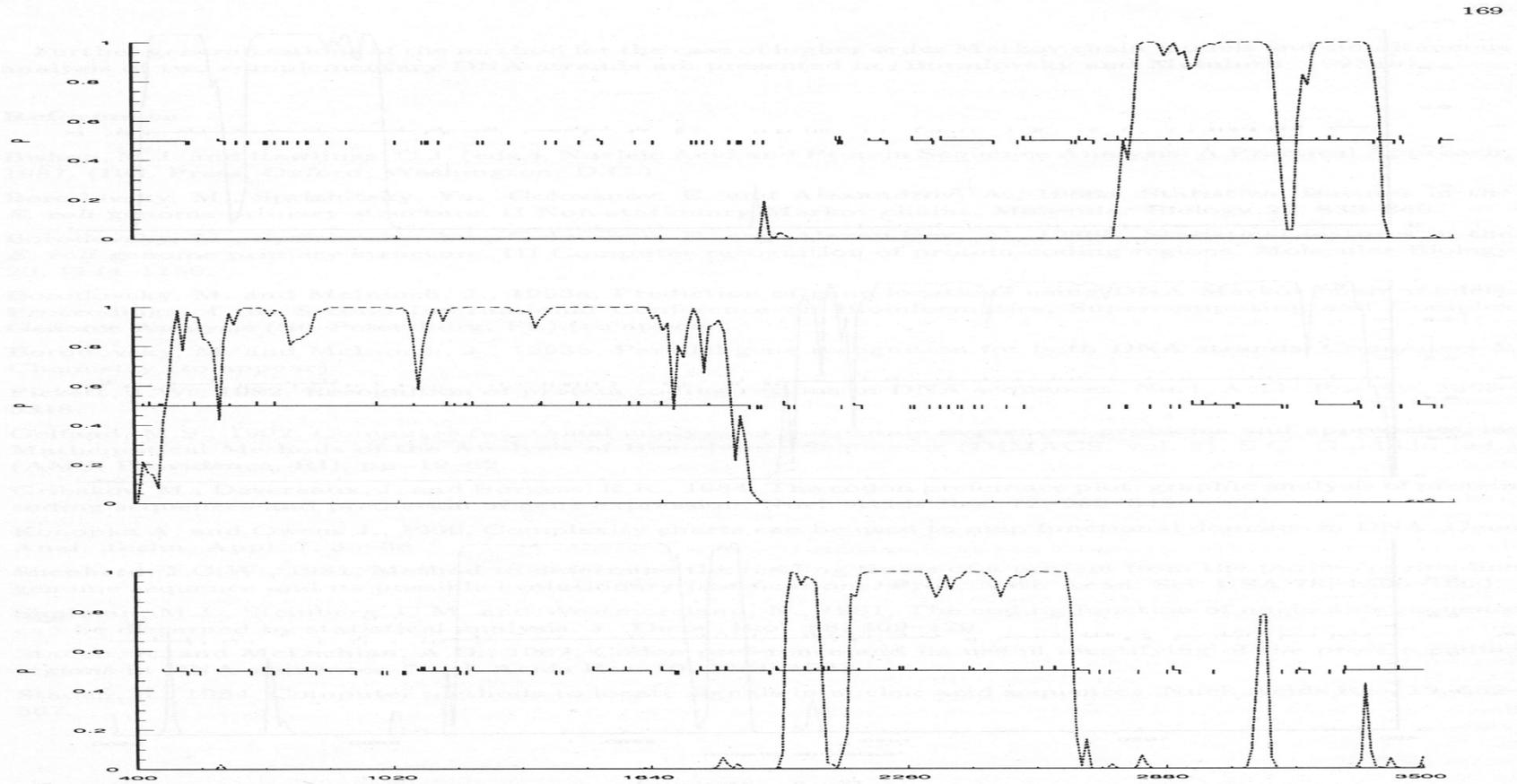
ttgttcatatttgccggctggatcggcgggcgcacgagtagtggaaaactaaatgaaac

tctacaatctgaaagatcacaatgagcaggtcagctttgcgcaagccgtaaccaggggt

tgggcaaaaatcaggggctgtttttcccgcacgacctgccggaattcagcctgactgaaa

Статистические особенности

- Отличия в частотах олигонуклеотидов в кодирующих и не кодирующих областях



- Стартовые кодоны

Начала генов *Bacillus subtilis*

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

Участок связывания рибосом

<i>dnaN</i>	ACATTATCCGTTAGGAGGATAAAAATG
<i>gyrA</i>	GTGATACTTCA G GGAGGTTTTTTAATG
<i>serS</i>	TCAATAAAAAAAGGAGT T GTTTCGCATG
<i>bofA</i>	CAAGCGAAGGAG A TGAGAAGATTCATG
<i>csfB</i>	GCTAACTGTAC C GGAGGTGGAGAAGATG
<i>xpaC</i>	ATAGACACAGGAG T CGATTATCTCATG
<i>metS</i>	ACATTCTGATTAGGAGGTTTCAAGATG
<i>gcaD</i>	AAAAGGGATAT T GGAGGCCAATAAATG
<i>spoVC</i>	TATGTGACTAA G GGAGGATTCGCCATG
<i>ftsH</i>	GCTTACTGTGGGAGGAGGTAAGGAATG
<i>pabB</i>	AAAGAAA T 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--TGGCGCAGTGCTGATGGG-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 ATTCATCCTTTTCGATATCGCGGTGCTGGAACCAGGTGATGAGTATGCCTGGCGGCCAGATTATGGC--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 CTGTTTTTCATCTGTATGGCAGTTCGCTGTGGAGAGTAAAGTGAATAATTCTGGTTGATGAAAATATGCCGTACGCTGAG
      *   * * * *   *   ***           **           *   ***** * * ***** * * * * *
  
```

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

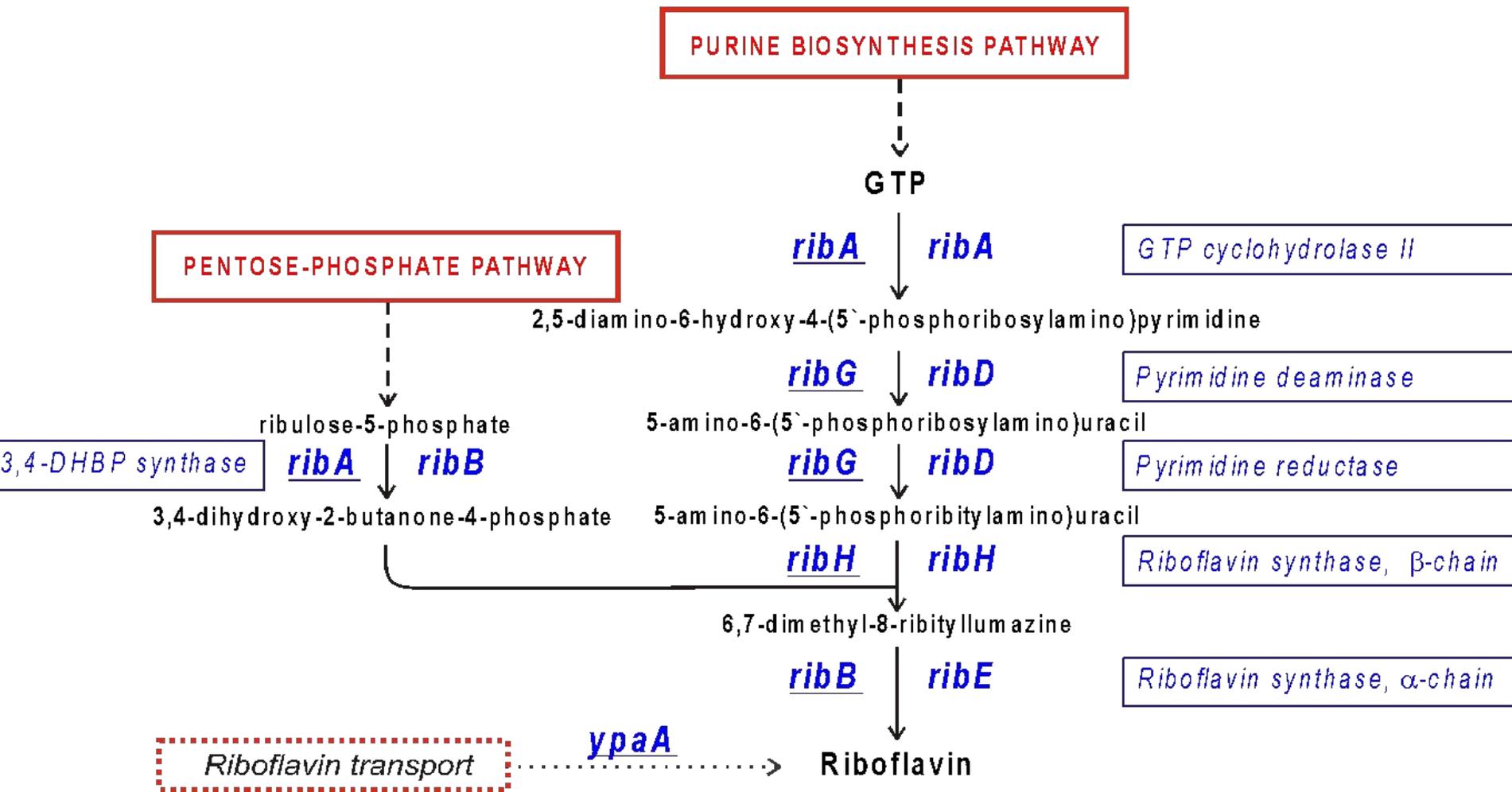
Мораль

- ***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

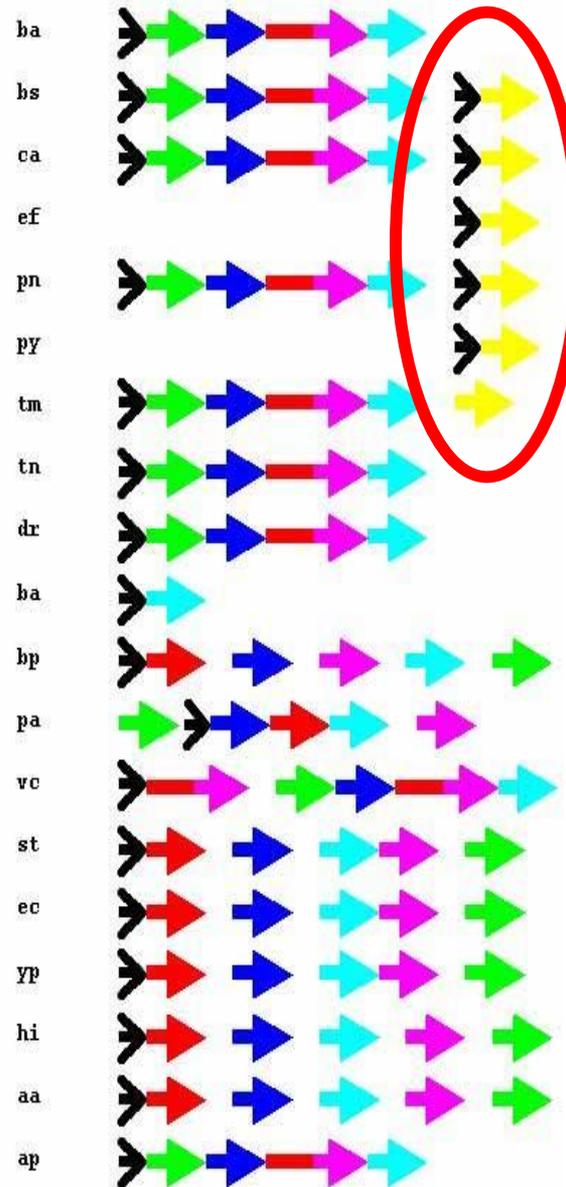
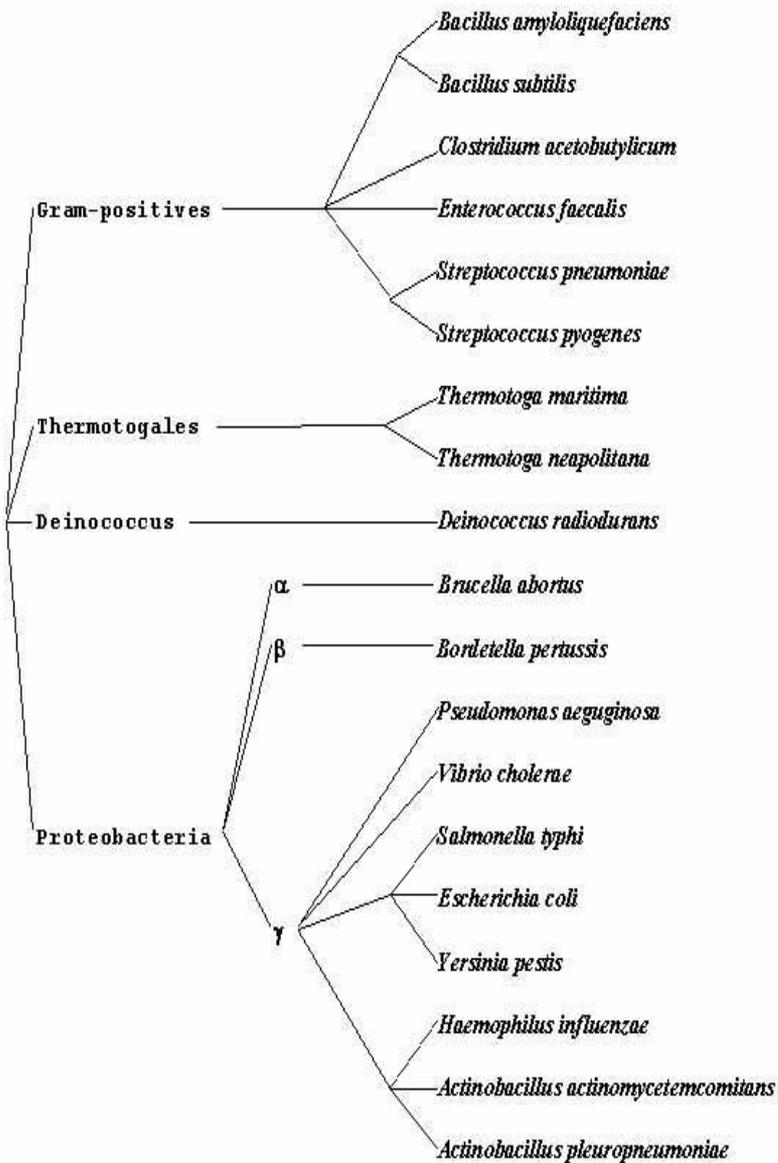
Метаболический путь биосинтеза рибофлавина (витамина В₂)



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	AGCATCCTTCGGGG-TCGGGTGAAATCCCACCCGGCGGT	19	AGTCCGTGAC--	8	5	8	-----TGGATCTAGT GAAACTCTAGGCCCGACAGT -AT-AGTCTGGAT-GGGAGAAGGATATG					
BE	TGGATCCTTCGGGG-CAGGGTGGAAATCCCACCCGGCGGT	20	AGCCCGGGA---	3	4	3	-----AGGATCCGGT GCGAATCCGGAGCCGACAGT -AT-AGTCTGGAT-GGGAGAAGGATGCC					
BD	TTTTATCCTTCGGGG-CTGGGTGGAAATCCCACCCGGCGGT	19	AGTCCGTGAC--	10	4	10	-----TGGACTCTGT GAAACTCCGGGACCGACAGT GAA-AGTCTGGAT-GGGAGAAGGAAACC					
Bam	TGTATCCTTCGGGG-CTGGGTGGAAATCCCACCCGGCGGT	23	AGCCCGTGAC--	8	4	8	-----TGGATTTCAGT GAAAGCTGAAGCCGACAGT GAA-AGTCTGGAT-GGGAGAAGGATGAG					
CA	GATGTTCTTCAGGG-ATGGGTGAAATCCCACCCGGCGGT	2	AGCCCGCAA---	3	4	3	-----AGATCCGGT TAAACTCCGGGGCCGACAGT AA-AGTCTGGAT-GAAAGAAGAAATAG					
DE	CTTAATCTTCGGGG-TAGGGTGGAAATCCCACCCGGCGGT	2	AGCCCGGCG---	7	6	7	-----ATTGTGTTAAACTCCAAAGCCGACAGT AA-AGTCTGGAT-GAAAGAAGATATTT					
SA	TAAATCCTTCGGGG-CAGGGTGGAAATCCCACCCGGCGGT	6	AGCCCTGGAC---	11	3	11	-----CTGATCTAGT GAGATTCTAGAGCCGACAGT TAA-AGTCTGGAT-GGGAGAAGCAATGT					
LLX	ATAAATCTTCAGGG-CAGGGTGTAAATCCCACCCGGCGGT	2	AGCCCGGGA---	4	4	4	-----ATGATTTGGT GAAACTCCAGGCCCGACAGT -AT-AGTCTGGAT-GAAAGAAGATTAATA					
PN	AACATCTTCAGGG-CAGGGTGGAAATCCCACCCGGCGGT	2	AGCCCGACGA---	3	4	3	-----ATGATTTGGT GAAACTCCAAAGCCGACAGT -AT-AGTCTGGAT-GAAAGAAGATAAAA					
TM	AAACGCTCTTCGGGG-CAGGGTGGAAATCCCACCCGGCGGT	3	AGCCCGGCGAG---	5	4	5	-----TTGACCCGGT GGAATCCGGGGCCGACGGT GAA-AGTCTGGAT-GGGAGAGAGCGTGA					
DR	GACCTCCTTCGGGG-CGGGGCGAAATCCCACCCGGCGGT	15	AGCCCGGGA---	8	12	9	-----CCGATGCCCGGCAACTCCGGAGCCGACGGT CAC-AGTCTGGAT-GAAAGAAGGAGGAG					
TQ	CACCTCCTTCGGGG-CGGGGTGGAAATCCCACCCGGCGGT	3	AGCCCGGGA---	5	4	5	-----CCGACCCGGT GGAATCCGGGGCCGACGGT GAA-AGTCTGGAT-GGGAGAAGGAGGGC					
AO	AATAATCTTCAGGG-CAGGGTGGAAATCCCACCCGGCGGT	2	AGTCCGGGA---	7	7	7	-----AGGAACCCGT GAGATTCCGGTACCGACAGT -AT-AGTCTGGAT-GGAAGAAGATGAAA					
DU	TTTAATCTTCAGGG-CAGGGTGGAAATCCCACCCGGCGGT	2	AGTCCGGGA---	13	4	12	-----AGGAACCTAGT GAAATCTAGTACCGACAGT -AT-AGTCTGGAT-GGAAGAAGAGCAGA					
CAU	GAAAGCCTTCGGGG-CAGGGTGGAAATCCCACCCGGCGGT	20	AGCCCGGGA---	3	4	3	-----AGGACCCGGT GTGATTCGGGGCCGACGGT -AT-AGTCTGGAT-GGGAGAAGGTGGCC					
FN	TAAAGTCTTCAGGG-CAGGGTGGAAATCCCACCCGGCGGT	2	AGTCCAGC---	5	4	5	-----GATTTGGT GAAACTCCAAAGCCGACAGT -AG-AGTCTGGAT-GGGAGAAGAAATAG					
TEU	ACCGTCTTCGGGG-GTGGGTGAAAGTCCGAACCCGGCGGT	3	AGTCCGGAC---	8	5	8	-----TGGAACCCGT GAAACTCCGGTACCGACGGT GAA-AGTCTGGAT-GGGAGGTAGTACGG					
SX	-AGCCCACTTCGGGG-GTGGGTGAAAGTCCGAACCCGGCGGT	3	AGTCCGGAC---	8	5	8	-----TTGACCCAGT GAAATCTGGACCGACGGT TAA-AGTCTGGAT-GGGAGGCACTGCCGG					
BU	GTCCGCTCTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	30	AGCCCGGGAGCC	13	7		GT CAGCAGATCTGGT GAGAAGCAGAGCCGACGGT TAG-AGTCTGGAT-GGAAGAAGATGTGC					
BPS	GTCCGCTCTTCAGGG-CGGGGCGAAATCCCACCCGGCGGT	21	AGCCCGGAGCC	8	4	8	GT CAGCAGATCTGGT CCGATTCCAGAGCCGACGGT CAT-AGTCTGGAT-GAAAGAAGATGTGC					
RBV	TTACGCTCTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	31	AGCCCGGGAGCC	7	5	7	GT CAGCAGATCTGGT GAGATCCAGGGCCGACGGT TAA-AGTCTGGAT-GAAAGAAGATGGCC					
RSO	GTACGCTCTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	21	AGCCCGGGAGCC	11	3	11	GT CAGCAGATCCGGT GAGATCCGGGGCCGACGGT CAG-AGTCTGGAT-GGAAGAAGATGTGC					
EC	GCTTATTCTCAGGG-CGGGGCGAAATCCCACCCGGCGGT	17	AGCCCGGAGCC	8	4	8	GACAGCAGATCCGGT GTAATCCGGGGCCGACGGT TAG-AGTCTGGAT-GGGAGAGATTAACC					
TY	GCTTATTCTCAGGG-CGGGGCGAAATCCCACCCGGCGGT	67	AGCCCGGGAGCC	8	3	8	GT CAGCAGATCCGGT GTAATCCGGGGCCGACGGT TAA-AGTCTGGAT-GGGAGAGGATTAACC					
KP	GCTTATTCTCAGGG-CGGGGCGAAATCCCACCCGGCGGT	20	AGCCCGGAGCC	8	4	8	GT CAGCAGATCCGGT GTAATCCGGGGCCGACGGT TAA-AGTCTGGAT-GGGAGAGATTAACC					
HI	TCGCACTCTCAGGG-CAGGGTGGAAATCCCACCCGGCGGT	2	AGCCCGGAGCC	26	9	30	GT CAGCAGATTTGGT GAAACTCCAAAGCCGACAGT -AA-AGTCTGGAT-GAAAGAGAAATAAA					
VK	GCGCACTCTCAGGG-CAGGGTGGAAATCCCACCCGGCGGT	14	AGCCCGGAGCC	11	9	11	GT CAGCAGATTTGGT GAGAATCCAAAGCCGACAGT -AT-AGTCTGGAT-GAAAGAGAAATTAACC					
VC	CAATATTCTCAGGG-CGGGGCGAAATCCCACCCGGCGGT	13	AGCCCGGAGCC	5	4	5	GT CAGCAGATCTGGT GAGAAGCAGGGCCGACGGT TAC-AGTCTGGAT-GAGAGAGAAATGACA					
YP	GCTTATTCTCAGGG-CGGGGTGGAAAGTCCCACCCGGCGGT	40	AGCCCGGGAGCC	16	6	16	GT CAGCAGATCCGGT GTAATCCGGGGCCGACGGT TAT-AGTCTGGAT-GGGAGAGATTAACC					
AB	GCGCACTCTCAGGG-CAGGGTGGAAATCCCACCCGGCGGT	25	AGCCCGGAGCC	16	4	27	GT CAGCAGATTTGGT GCGAATCCAAAGCCGACAGT GAC-AGTCTGGAT-GAAAGAGAAATAAA					
BP	GTACGCTCTTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	18	AGCCCGGAGCC	10	4	10	GT CAGCAGACTGGT GAGATCCAGGGCCGACGGT CAT-AGTCTGGAT-GAGAGAAGATGTGC					
AC	ACATCTCTCAGGG-CGGGGGTAAATCCCACCCGGCGGT	16	AGCCCGGAGCA	10	3	11	---CGCAGATCTGGT GTAATCCAGAGCCGACGGT -AT-AGTCTGGAT-GAAAGAAGACGACC					
Spu	AACAACTCTCAGGG-CGGGGTGGAAACTCCCACCCGGCGGT	34	AGCCCGGGAGCC	6	6	6	GT CAGCAGATCTGGT G 52 TCCAGAGCCGACGGT 31 AGTCTGGAT-GGAAGAGAAATGTA					
PP	GTCCGCTCTTCAGGG-CGGGGTGTAAATCCCACCCGGCGGT	13	AGCCCGGAGCC	7	3	7	GT CAGCAGATCTGGT GCAACTCCAGAGCCGACGGT CAT-AGTCTGGAT-GAAAGAAGGCGTCA					
AU	GGTGTCTCTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	17	AGCCCGGGAGCC	7	9	7	GT CAGCAGATCCGGT GAGAGCCCGGAGCCGACGGT -AT-AGTCTGGAT-GAAAGAGGCAAGG					
PU	AAACGCTCTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	19	AGCCCGGAGCC	19	4	18	GT CAGCAGACTCCGGT GTGATTCGGGGCCGACGGT CAC-AGTCTGGAT-GAAGAGAGAAAGGGA					
PY	TAAAGTCTCAGGG-CGGGGTGGAAACTCCCACCCGGCGGT	19	AGCCCGGGAGCC	15	4	16	GT CAGCAGATCCGGT GTGATTCGGGGCCGACGGT CAT-AGTCTGGAT-GAAGAGAGAGCGGGA					
PA	TAAAGTCTCAGGG-CGGGGTGGAAAGTCCCACCCGGCGGT	19	AGCCCGGGAGCC	14	4	13	GT CAGCAGACTCCGGT GCGAATCCGGGGCCGACGGT CAT-AGTCTGGAT-GAAGAGAGAAAGCGGA					
ML0	TAAAGTCTCAGGG-CGGGGTGGAAACTCCCACCCGGCGGT	16	AGCCCGGAGCC	8	5	8	GT CAGCAGATCCGGT GTGATTCGGAGCCGACGGT TAG-AGTCTGGAT-GAAAGAGGAGAAA					
SM	AAGCGTCTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	34	AGCCCGGGAGCC	8	3	8	GT CAGCAGATCCGGT GGAATCCGGAGCCGACGGT TAT-AGTCTGGAT-GGAAGAGAGCAAGC					
BMB	GCTTGTCTTCGGGG-CGGGGTGGAAACTCCCACCCGGCGGT	17	AGCCCGGGAGCC	10	15	10	GT CAGCAGATCCGGT GAGATGCCGGAGCCGACGGT TAA-AGTCTGGAT-GGAAGAGAGCGGTA					
BS	AICAATCTTCGGGG-CAGGGTGGAAATCCCACCCGGCGGT	18	AGCCCGGGA---	5	4	5	-----AGGATTCGGT GAGATTCCGAGCCGACAGT -AA-AGTCTGGAT-GGGAGAAGATGGAG					
BQ	GTCTATCTTCGGGG-CAGGGTGGAAATCCCACCCGGCGGT	27	AGCCCGGGA---	3	5	3	-----AGGATTTGGT GTGATTCAAAGCCGACAGT -AT-AGTCTGGAT-GGGAGAAGATGGAG					
BE	ATTCATCTTCGGGG-CAGGGTGGAAATCCCACCCGGCGGT	20	AGCCCGGGA---	3	4	3	-----AGGATTCGGT GCGAGTCCGGAGCCGACAGT -AT-AGTCTGGAT-GGGAGAAGATGAA					
CA	AATGATCTTCAGGG-CAGGGTGGAAATCCCACCCGGCGGT	2	AGCCCGGAG---	3	4	3	-----TATGATCCGGT TTGATTCGGAGCCGACAGT -AA-AGTCTGGAT-GAAAGAAGATATAT					
DE	GAAATCTTCGGGG-CAGGGTGGAAATCCCACCCGGCGGT	2	AGCCCGGCG---	6	4	6	-----GATTTGGT GAGATTCCAAAGCCGACAGT -AA-AGTCTGGAT-GAGAGAAGATATTT					
EF	GTTCGCTCTCAGGGGCAGGGTGTAAATCCCACCCGGCGGT	3	AGTCCAGAC---	5	3	5	-----ATTGATTTGGT GTAATCCAAATCCGACAGT -AT-AGTCTGGAT--AAAGAAGATAGGG					
LLX	AAATATCTTCAGGG-CACCGTGTAAATCCCACCCGGCGGT	21	ACTCCGGAT---	4	4	4	-----TTGAAGCACT GAGAATCTGCTAGCCGACAGT -AA-AGTCTGGAT-GGAAGAAGATGAAC					
LO	GTTATCTTCGGGG-CAGGGTGGAAATCCCACCCGGCGGT	3	AGTCCAGAT---	3	10	3	-----TTGACTCTGGT GTAATCCAGGACCCGACAGT -AT-AGTCTGGAT-GGGAGAAGATGTTG					
PN	AAGAGCTCTCAGGG-CAGGGTGGAAATCCCACCCGGCGGT	12.5	AGTCCGTC---	3	4	3	-----GATGTGGT GAGATTCCACAAACCGACAGT -AT-AGTCTGGAT-GGGAGAAGAGTAA					
ST	AAGTGTCTTCAGGG-CAGGGTGTGATTCACCCGGCGGT	14	AGTCCGGC---	3	4	3	-----GATGTGGT GTAACTCCACAAACCGACAGT -AT-AGTCTGGAT-GAGAGAAGACCGGG					
MN	AAGTGTCTTCAGGG-CAGGGTGTGATTCACCCGGCGGT	10.4	AGTCCGGC---	3	4	3	-----GATGTGGT GAAATCCACAAACCGACAGT -AA-AGTCTGGAT-GGGAGAAGACTGAG					
SA	ATTCATCTTCGGGG-TCGGGTGTAAATCCCACCCGGCGGT	6	AGCTCTGGAC---	11	3	11	-----CTGATCTAGT GAGATTCTAGAGCCGACAGT -AT-AGTCTGGAT-GGGAGAAGATGGAG					
AMI	TCACAGTCTCAGGG-CGGGGTGGAAATCCCACCCGGCGGT	14	AGCCCGGCG---	5	5	5	-----TGATCTGGT GCAACTCCAGAGCCAAAGCGGT -AT-AGTCTGGAT-GGAAGAAACGGAGC					
DHA	ACGAACTTCAGGG-TAGGGTGGAAATCCCACCCGGCGGT	20	AGCCCGCAA---	11	4	11	-----CGACTGCTGGT GAGACTCCAAAGCCGACGGT -AT-AGTCTGGAT-GGGAGAAGTAAACAA					
FN	AATAATCTTCGGGG-CAGGGTGGAAATCCCACCCGGCGGT	2	AGTCCAGC---	4	6	4	-----GATTTGGT GAAATCCAAAGCCGACAGT -AG-AGTCTGGAT-GAGAGAAGAAAGAA					
GLU	---TGTTCTCAGGG-CGGGGCGAAATCCCACCCGGCGGT	28	AGCCCGGGAGCC	10	4	10	GT CAGCAGATCCGGT TAAATCCGGAGCCGACGGT CAT-AGTCTGGAT-GCAAGAGAAACC---					

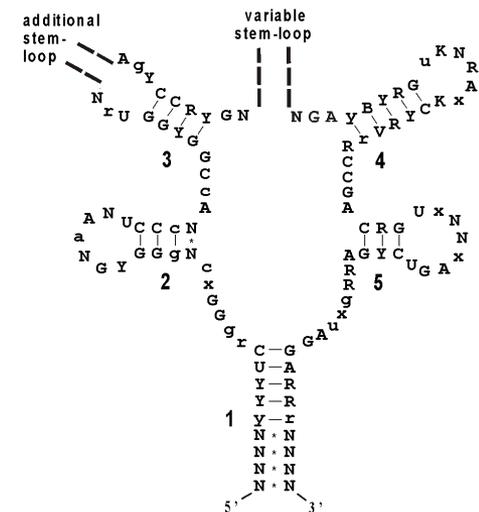
... и еще перед одним геном (*uraA*)



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

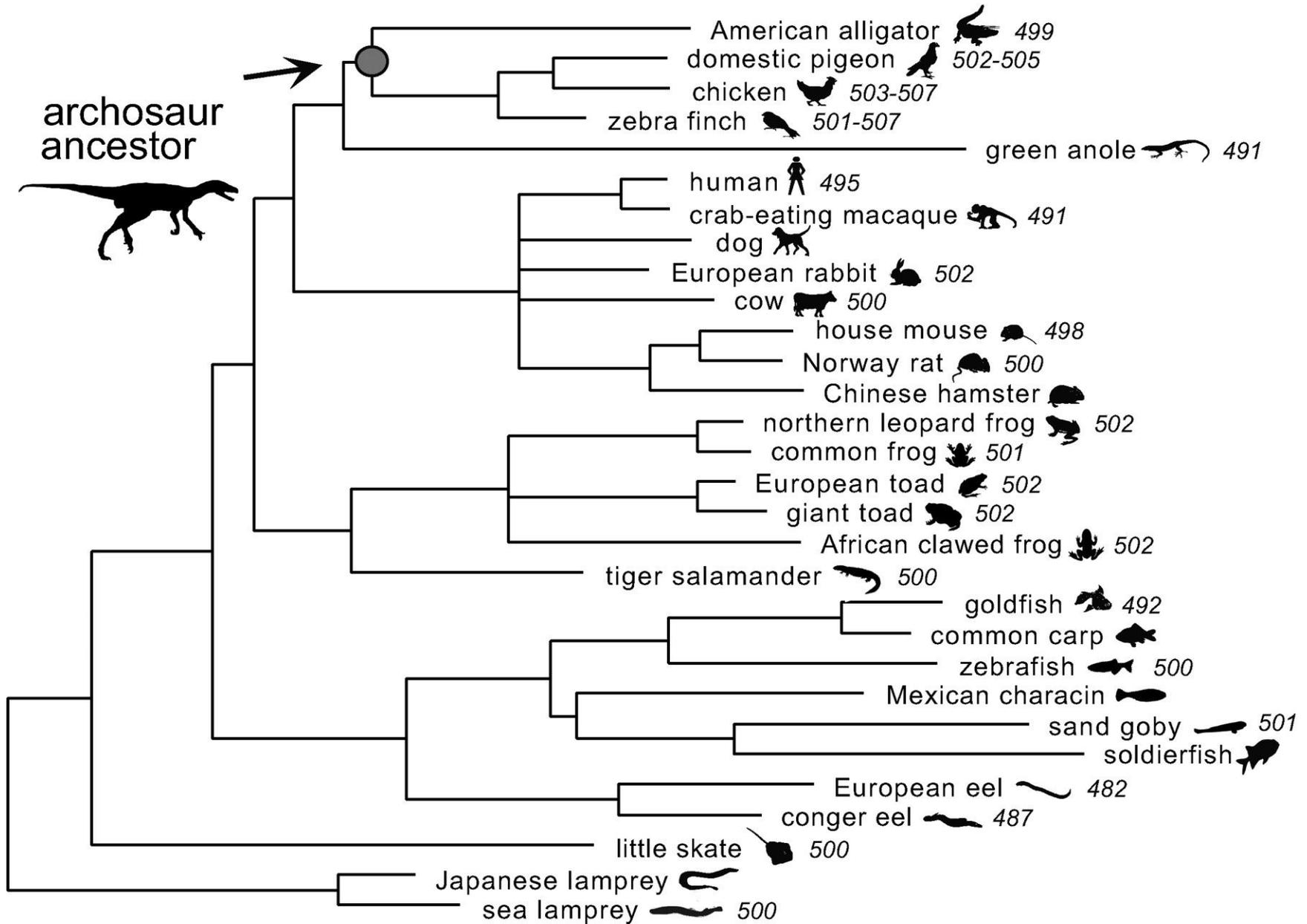
Belinda S. W. Chang, Karolina Jönsson,* Manija A. Kazmi,* Michael J. Donoghue,† and Thomas P. Sakmar**

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

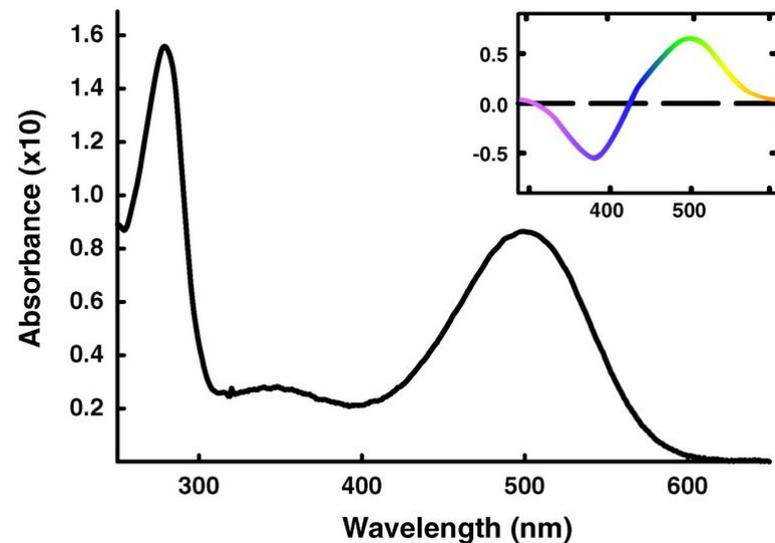
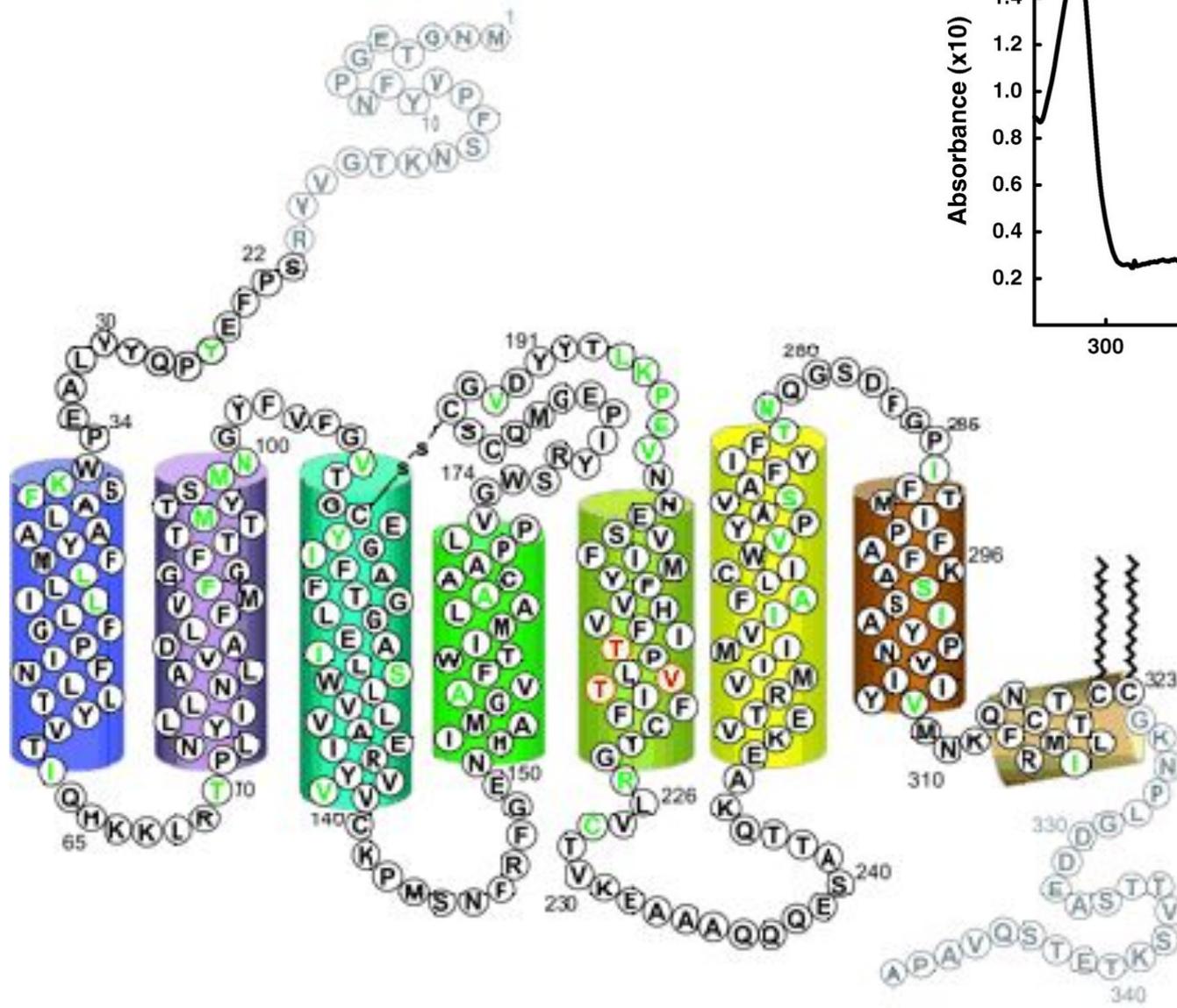
Mol. Biol. Evol. 19(9):1483–1489. 2002

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

Таксономическое дерево



Реконструированный белок и его свойства

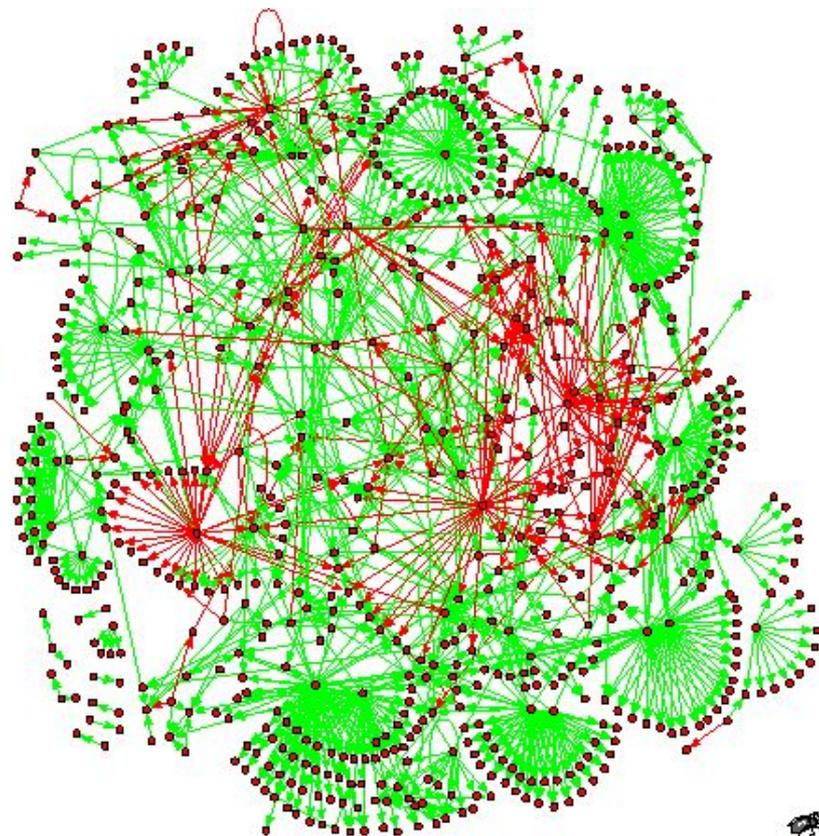
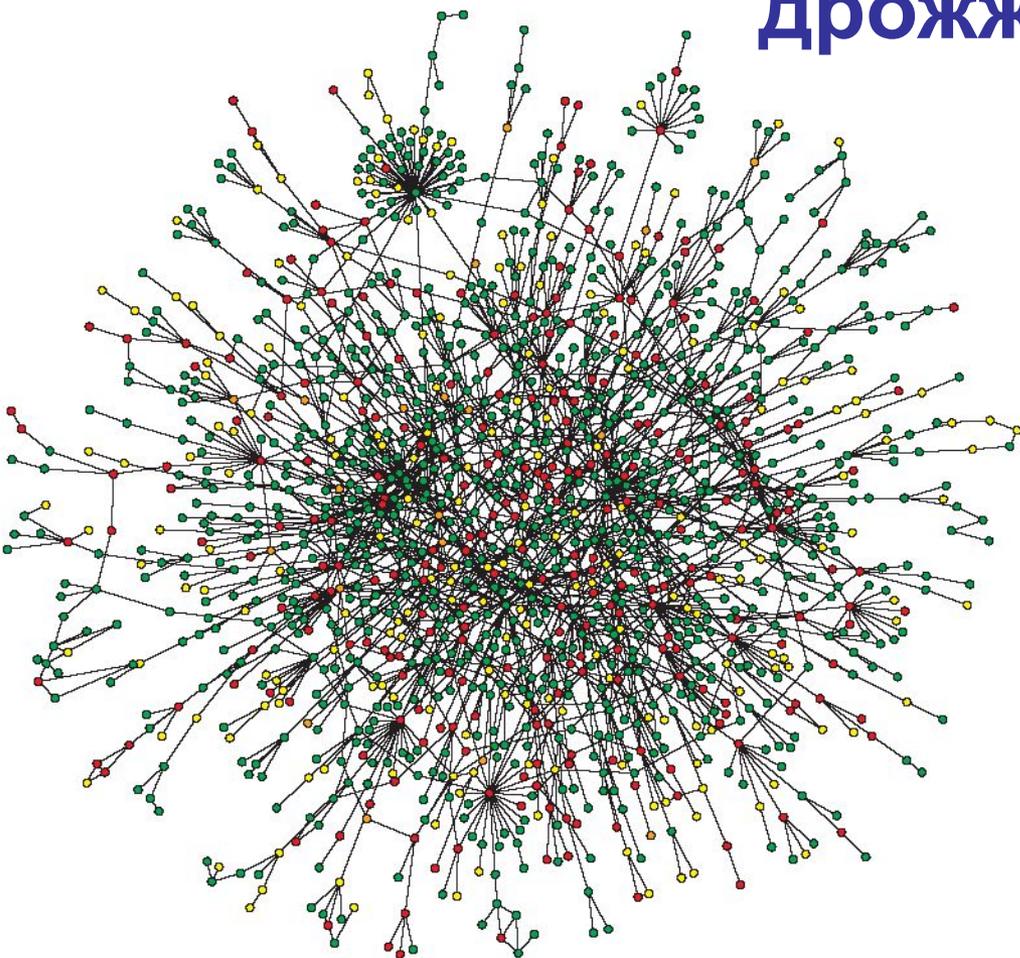


Не только тексты

Можно использовать данные, которые порождаются другими типами массовых экспериментов

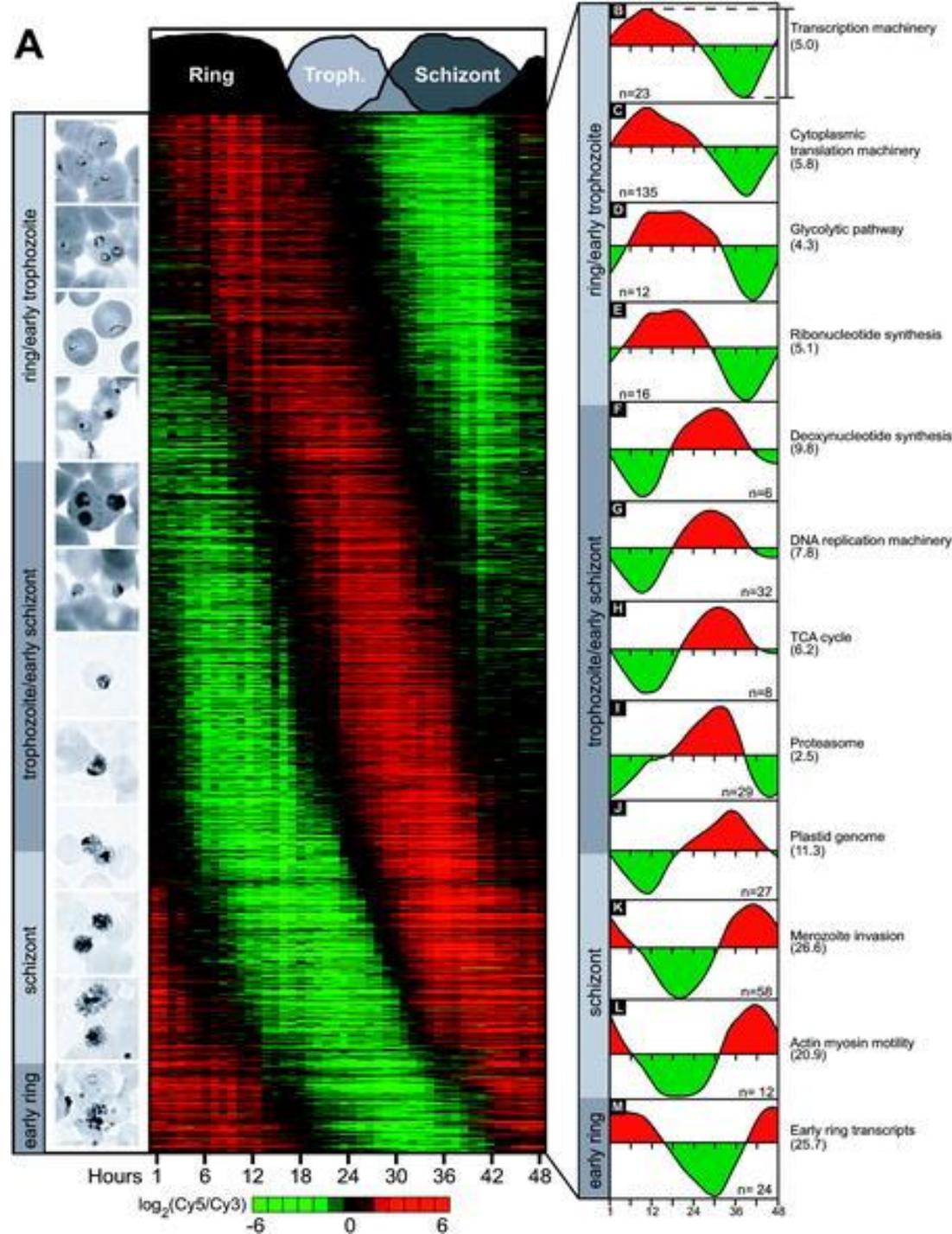
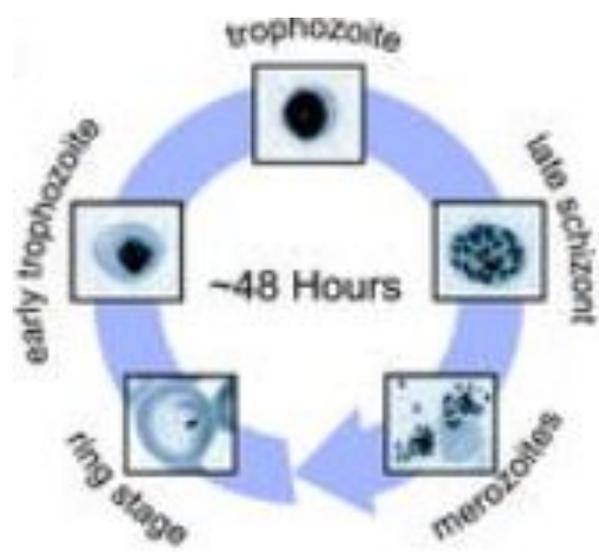
- Уровень экспрессии:
 - Концентрации мРНК
 - Концентрации белков
 - Время жизни мРНК и белков
- Взаимодействия:
 - Белок-ДНКовые
 - Белок-белковые
- Структура генома
 - Метилирование ДНК
 - Положение и модификация нуклеосом
- Функционально-генетические
 - Летальность мутаций
 - Фенотип
 - Синтетические 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 Bozdech¹, Manuel Llinás¹, Brian Lee Pulliam¹, Edith D. Wong¹, Jingchun Zhu², Joseph L. DeRisi^{1*}

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^a and William Martin^b

Trends in Genetics

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

