Учебно-Научный Центр

Биоинформатика

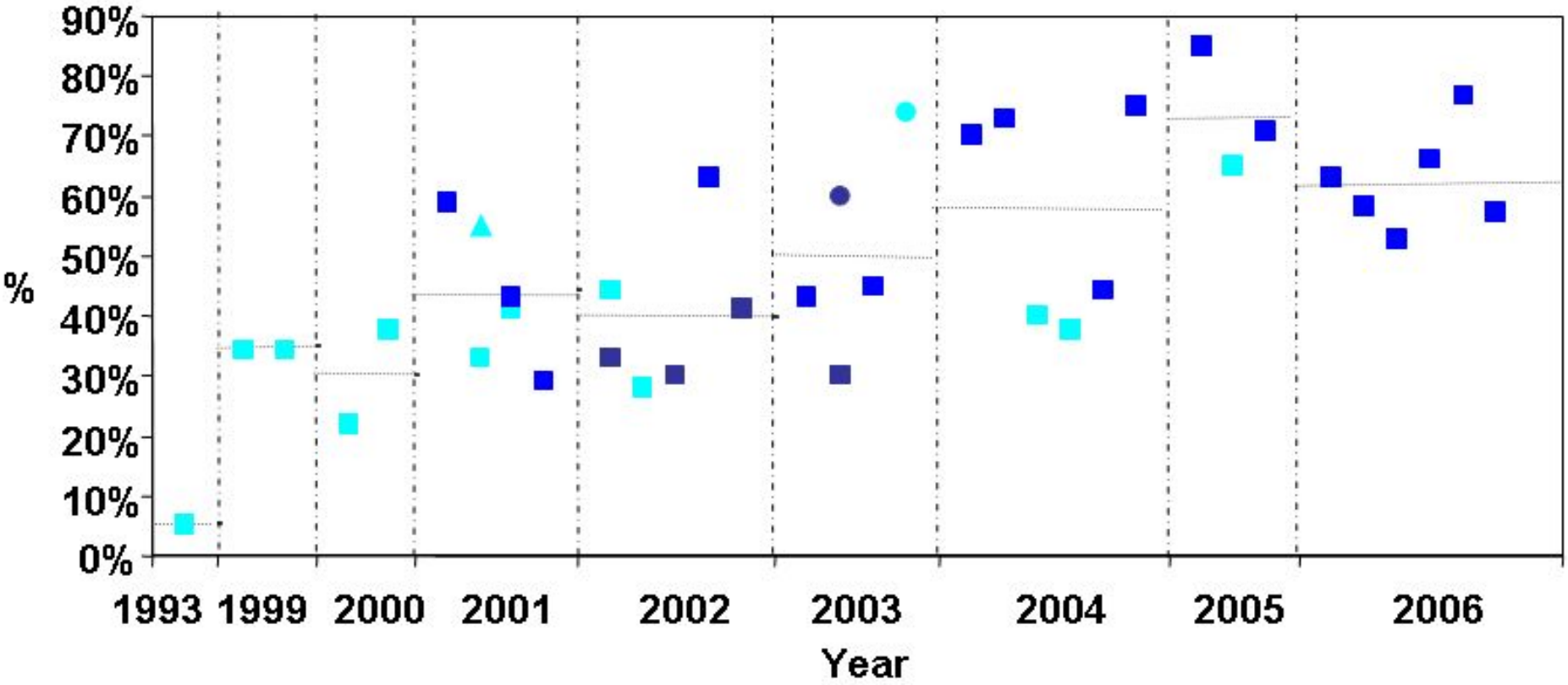
Альтернативный сплайсинг – детская площадка эволюции

М.Гельфанд

Институт проблем передачи информации им. А.А.Харкевича РАН

«Молекулярная и клеточная биология», 2 апреля 2007

Оценки % альтернативно сплайсируемых генов млекопитающих (по году публикации)



- Человек (выборка из генома)
- Человек (полные хромосомы)
- Мышь (выборка из генома)

- Все гены
- Только гены с ≥2 экзонами
- Гены с высоким покрытием EST

Значение альтернативного сплайсинга

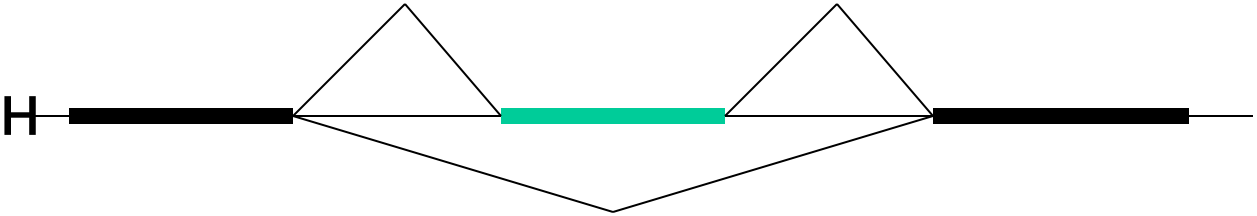
- **Функциональное:**
  - поддержание белкового разнообразия
    - Человек: ~30.000 генов, >100.000 белков
  - поддержание белкового **едино**образия
    - Например, секретируемые, мембранные и цитоплазматические изоформы
  - регуляция?
- **Эволюционное**

Общий план

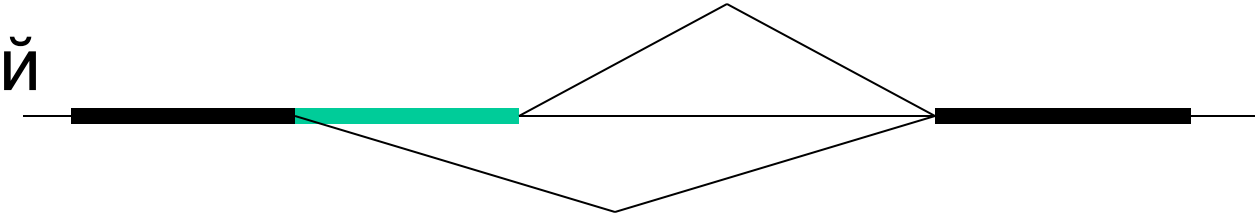
- **Эволюция альтернативной экзон-интронной структуры**
  - млекопитающие: человек, мышь, собака
  - двукрылые насекомые: *Drosophila melanogaster*, *D. pseudoobscura*, *Anopheles gambiae*
- **Скорость эволюции и тип отбора в постоянных и альтернативных областях**
  - Человек и мышь
  - *D. melanogaster* и *D. pseudoobscura*
  - человек-шимпанзе vs. человеческие SNP
- **Альтернативный сплайсинг и структура белка**

Элементарные альтернативы

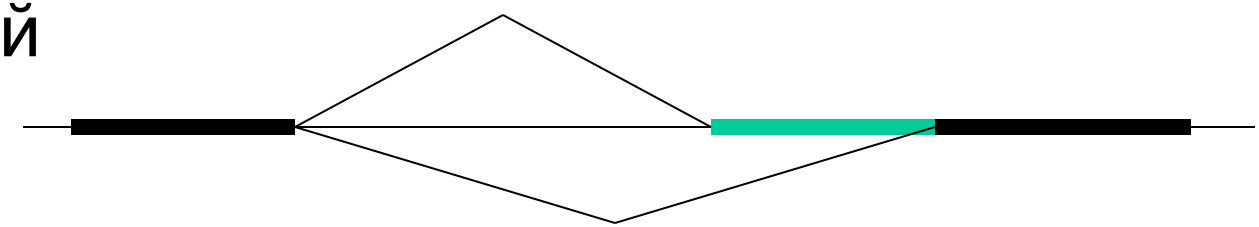
Кассетный экзон



Альтернативный донорный сайт



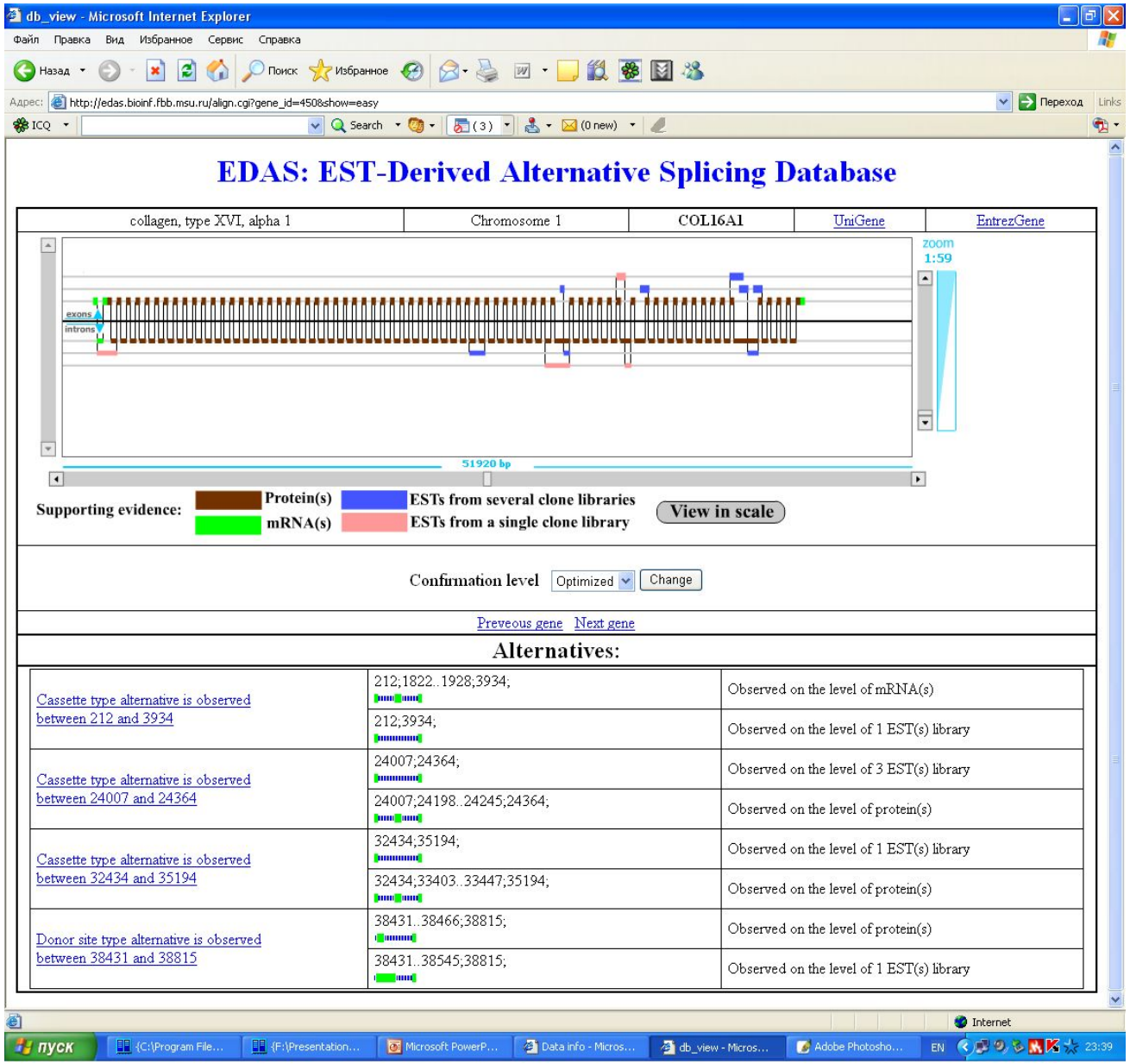
Альтернативный акцепторный сайт



И т.д.: взаимоисключающие экзоны, удержанные интроны, сложные альтернативы

EDAS: альтернативный сплайсинг генов человека

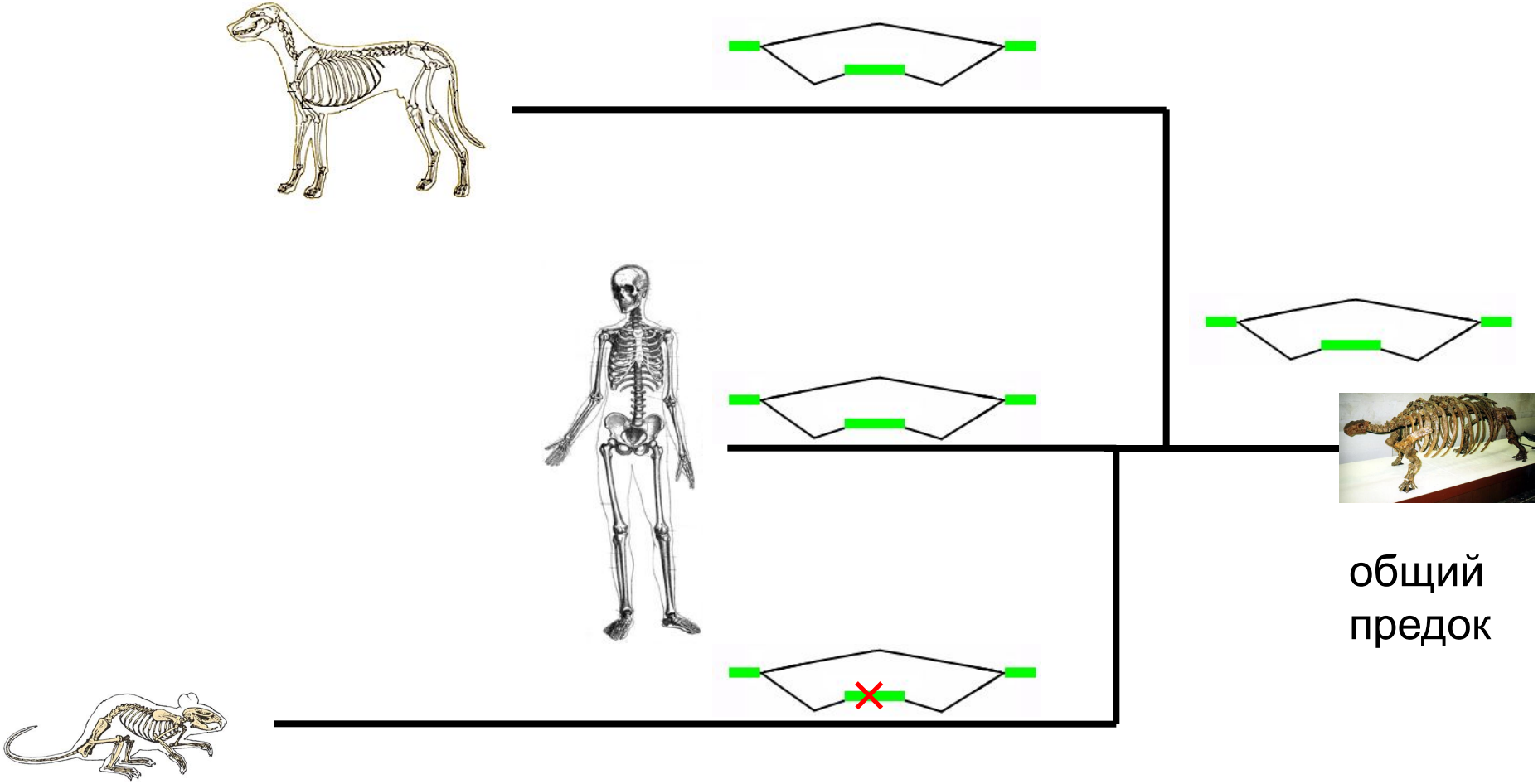
20809 генов;
114568 мРНК;
91835 белков;
51713 альтернатив,
из них 31746
элементарных



Альтернативная экзон-интроная структура генов млекопитающих

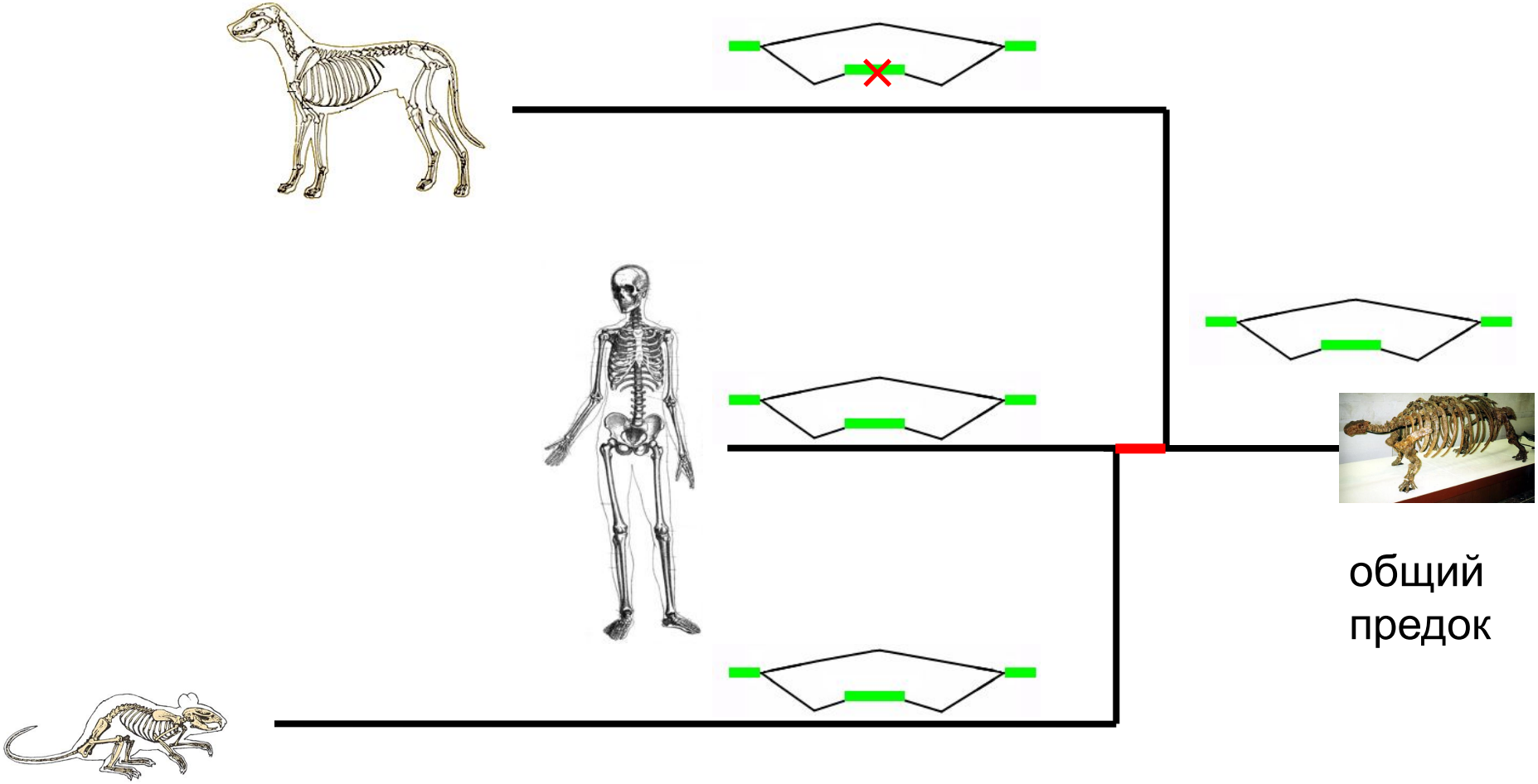
- Тройки ортологичных генов: человек-мышь-собака
- Следим за судьбой (консервативностью) альтернатив человека в геномах мыши и собаки

Потеря альтернативы в геноме мыши

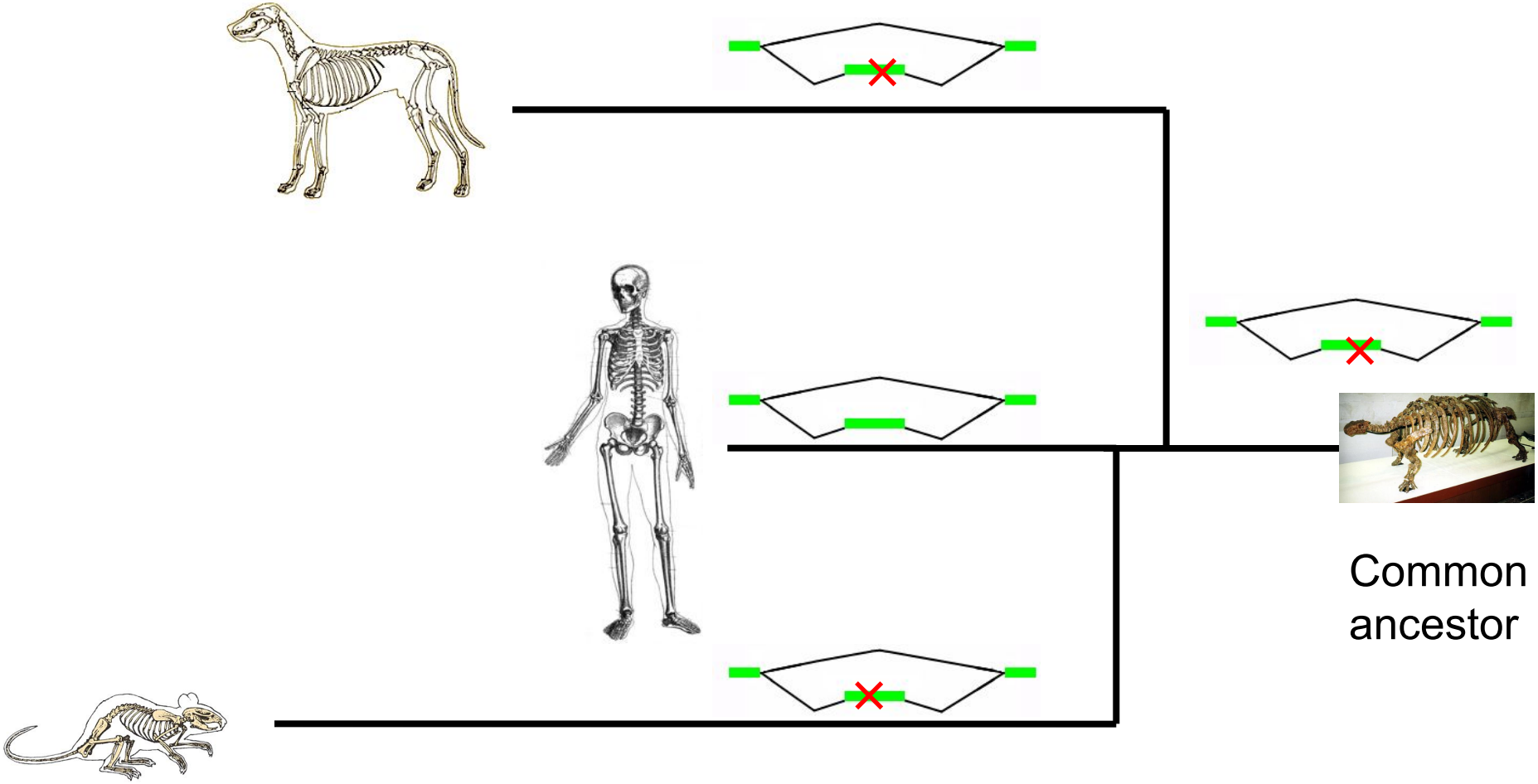


Потеря альтернативы в геноме собаки

(хотя теоретически возможно возникновение в общем предке приматов и грызунов)



Появление альтернативы в геноме человека (или ошибка сплайсинга, или экспериментальный шум)

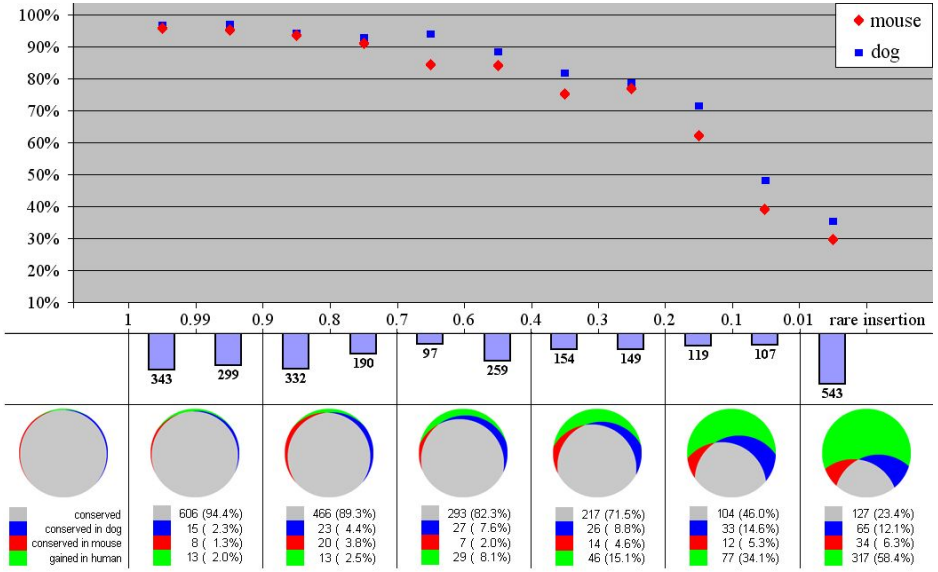


Альтернативы в генах человека, отсутствующие в генах мыши – реальны ли они?

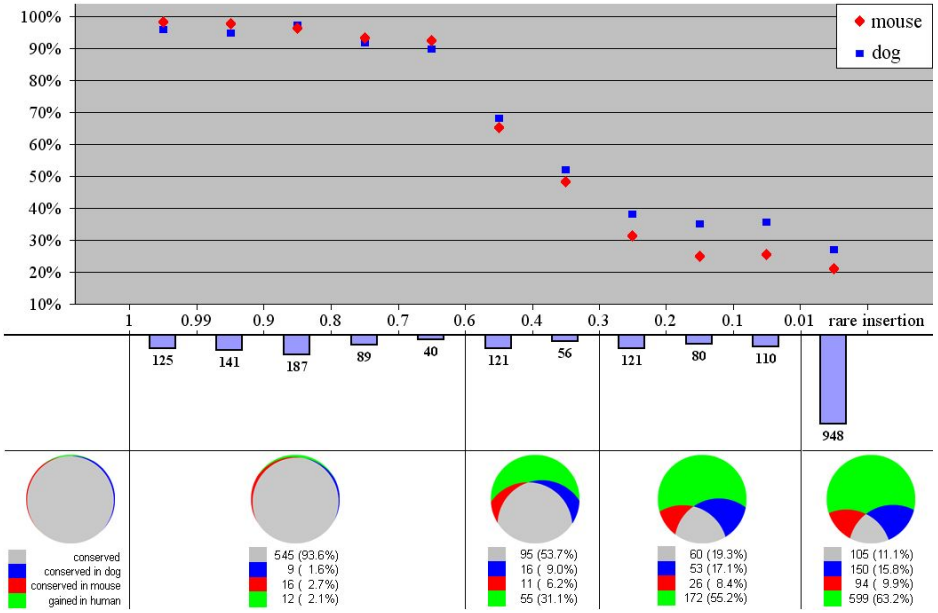
Добавим геном собаки

Наблюдения

нет сдвига рамки



сдвиг рамки



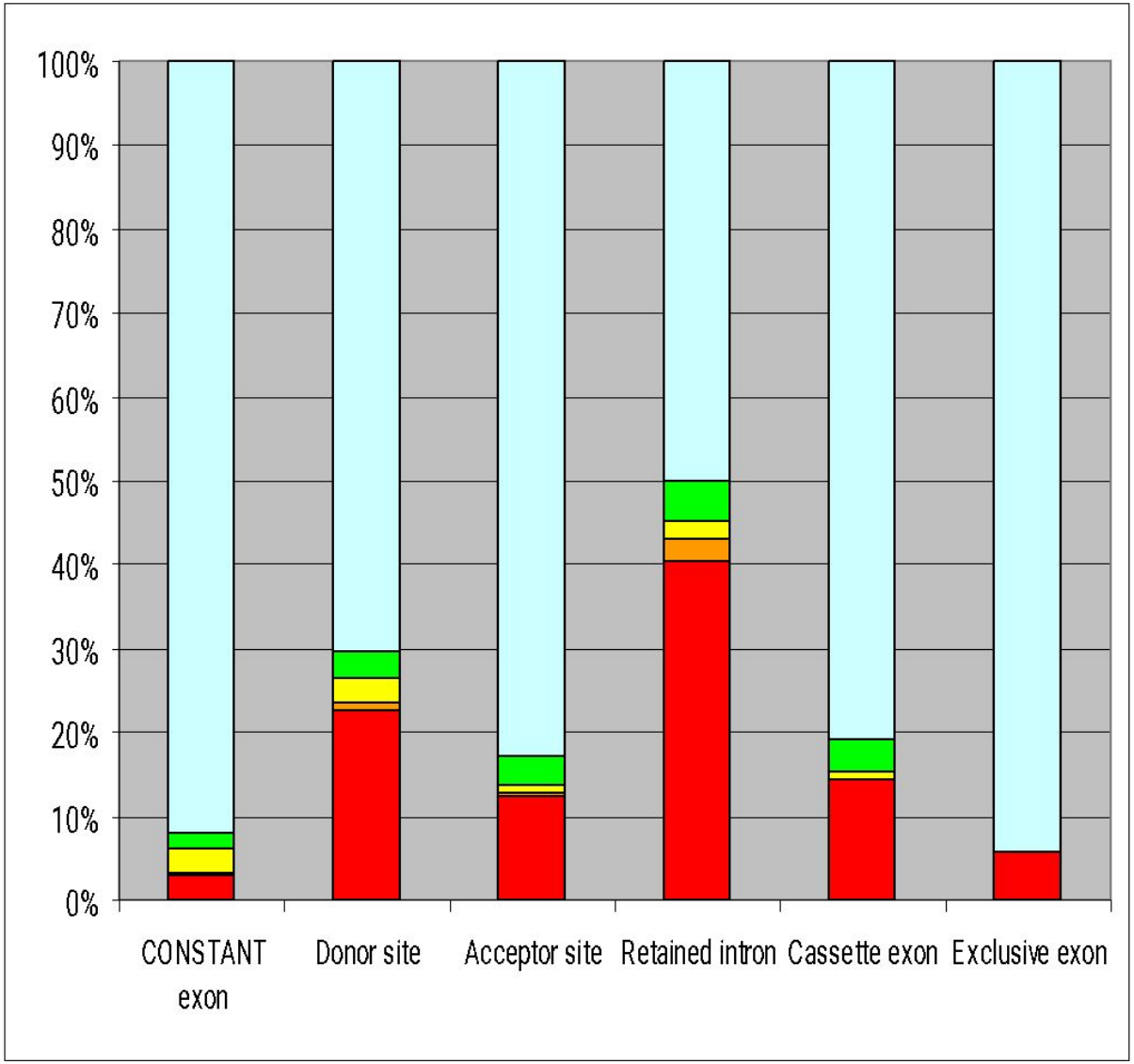
- Часто вставляемые экзоны консервативны независимо от того, сбивают ли они рамку
- Редко вставляемые экзоны менее консервативны, особенно сбивающие рамку
- Много геном-специфичных потерь
  - Больше в мыши, чем в собаке
  - Чаще для экзонов, сбивающих рамку
- Тем не менее, ~40% редко вставляемых экзонов консервативны хоть в одном экзоне

Консервативность белок-кодирующих областей в генах насекомых

Технически сложнее (сложности с выравниванием), но наблюдения те же: **альтернативные сегменты менее консервативны, чем константные**

	Константные сегменты	Альтер-нативные сегменты
D. melanogaster – *D. pseudoobscura*	97%	75-80%
D. melanogaster – *Anopheles gambiae*	77%	~45%

Консервативность элементарных альтернатив *D.melanogaster* в генах *D. pseudoobscura*



голубой – сохранились точно

зеленый – вставка интрона в D.ps. (или потеря в D.mel.)

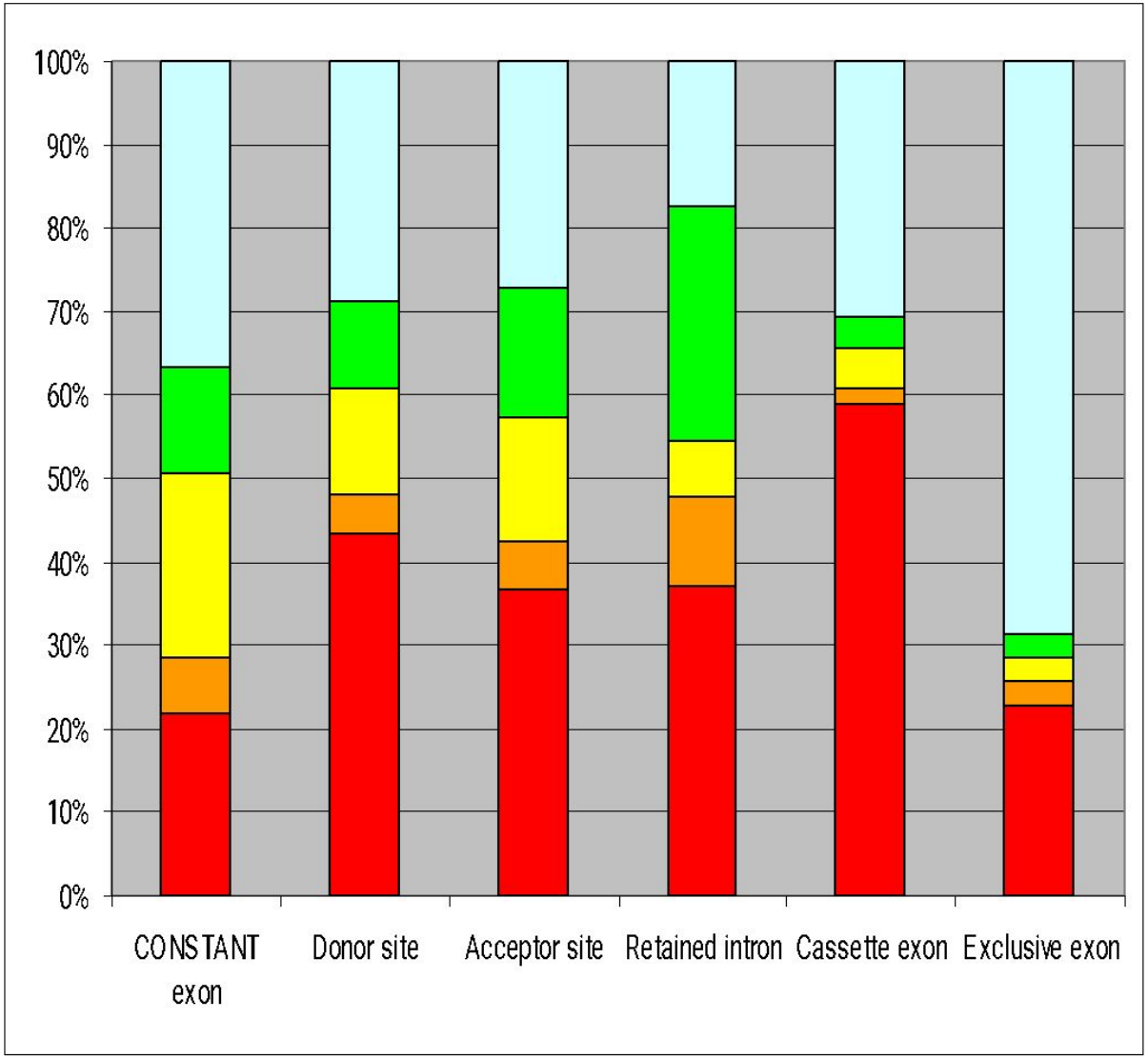
желтый – вставка интрона в D.mel.(млм потеря в D.ps.)

оранжевый – множественные вставки/потери

красный – неконсервативно

- Удержанные интроны наименее консервативны **(все ли они функциональны?)**
- Взаимоисключащие экзоны столь же консервативны, как конститутивные

Консервативность элементарных альтернатив *D.melanogaster* в генах *Anopheles gambiae*



голубой – сохранились точно

зеленый – вставка интрона в Anopheles (или потеря в Drosophila)

желтый – вставка интрона в Drosophlia (млм потеря в Anopheles)

оранжевый – множественные вставки/потери

красный – неконсервативно

- ~30% вставок у дрозофилы, ~10% вставок у комара (вообще, у него примерно 3 интрона на ген, а у дрозофилы примерно 4)
- Почти нет вставок интронов во взаимоисключающие экзоны

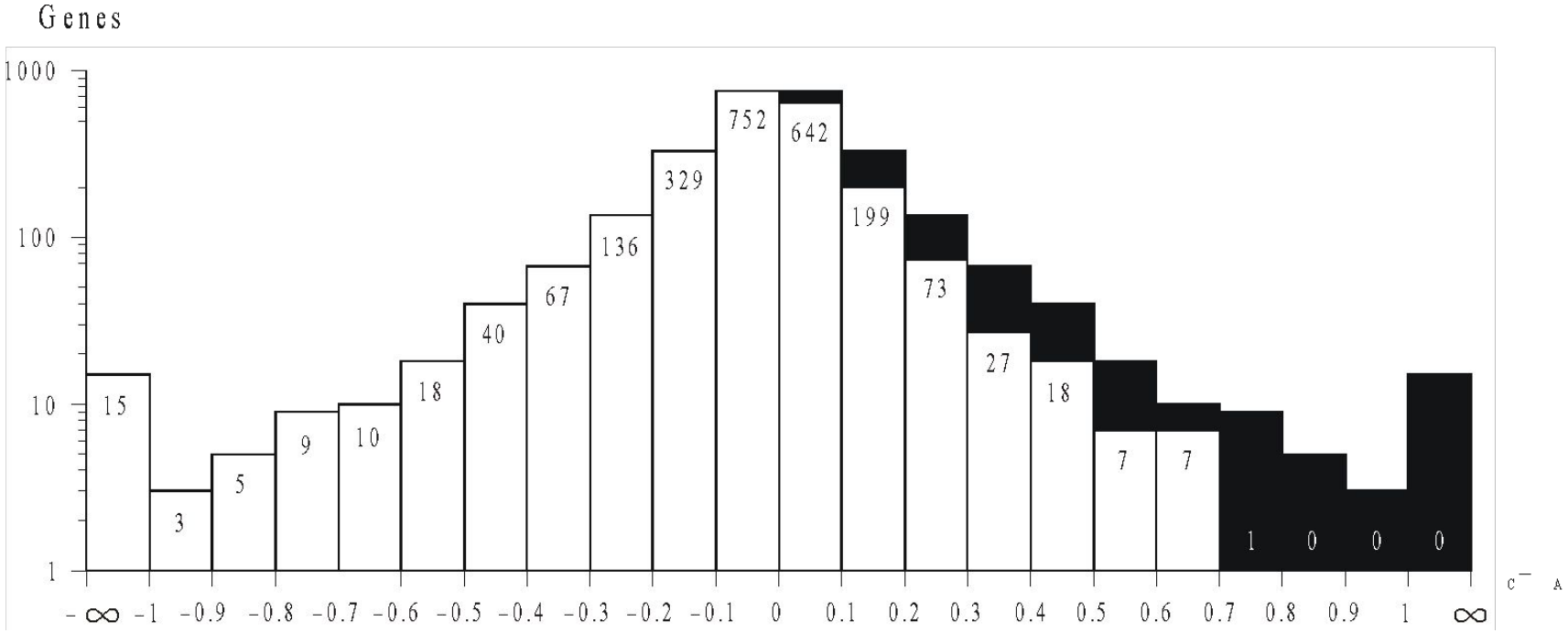
Скорость эволюции и тип отбора в альтернативных и константных областях

- Пары ортологичных генов
  - человек и мышь
  - *D. melanogaster* и *D. pseudoobscura*

- Отношение скорости синонимичных и несинонимичных замен (**d_n/d_s**):

  большая доля несинонимичных замен (изменяющих аминокислоту)
  => слабый стабилизирующий отбор или отбор на изменчивость
  (в такой постановке неразличимы)

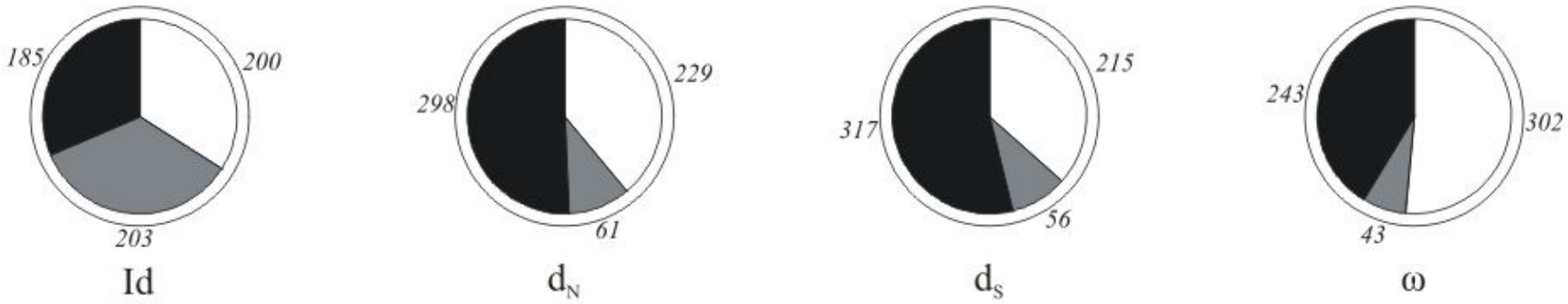
Гены человека и мыши: несимметричная гистограмма $d_n/d_s(const)–d_n/d_s(alt)$



Черный: «тень» левой части.

В большей части генов $d_n/d_s(alt) > d_n/d_s(const)$, особенно в области больших значений. Т.е. в альтернативных областях слабее стабилизирующий отбор

Гены *Drosophila*: слабее отбор в альтернативных областях?



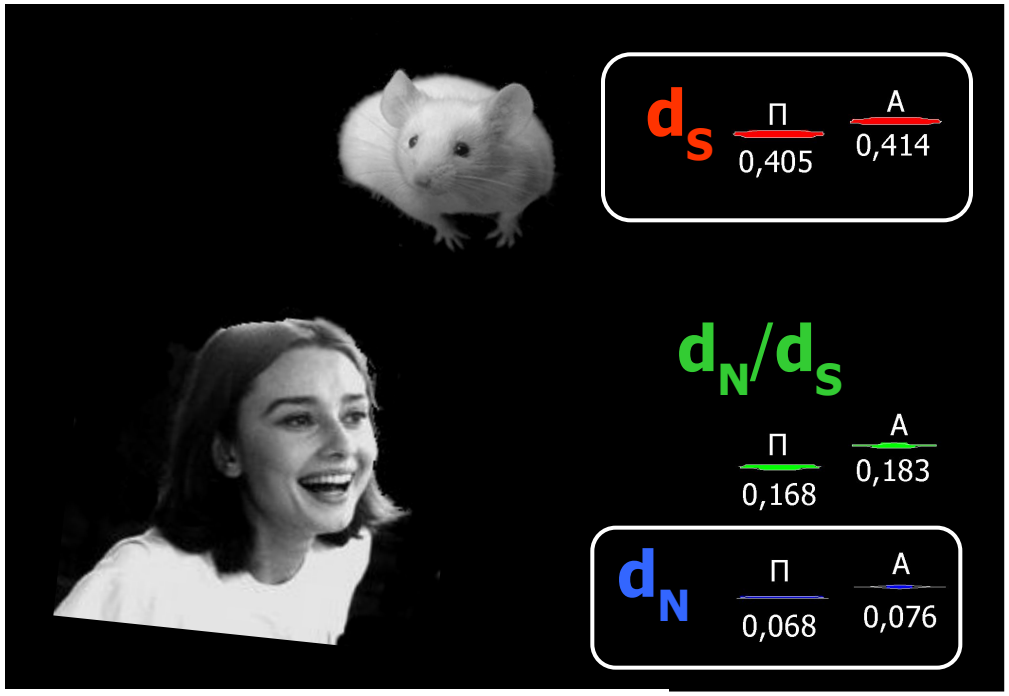
Больше замен в альт. обл.

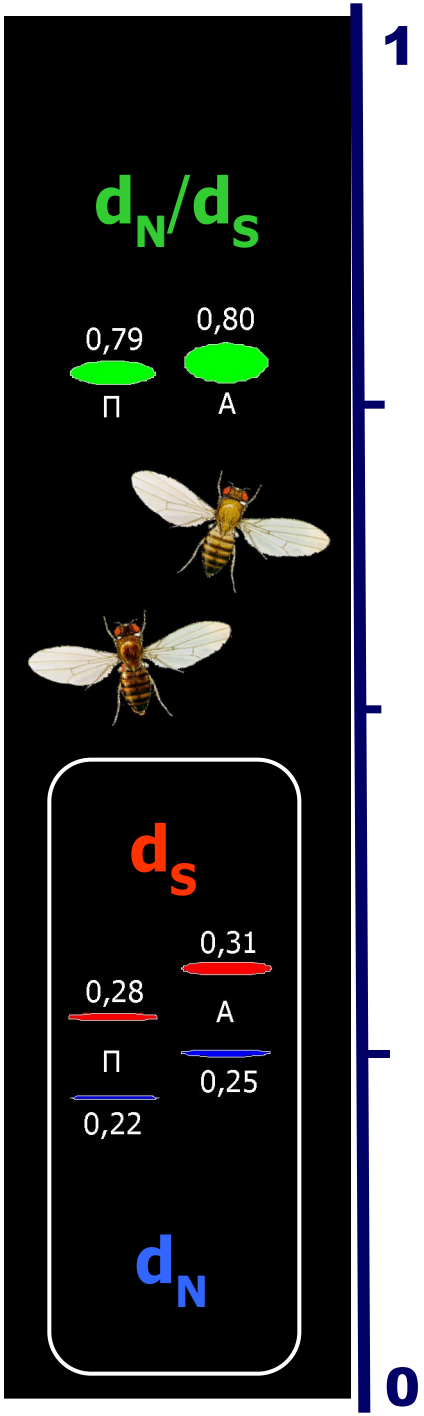
Равный уровень замен

Больше замен в конст. обл.

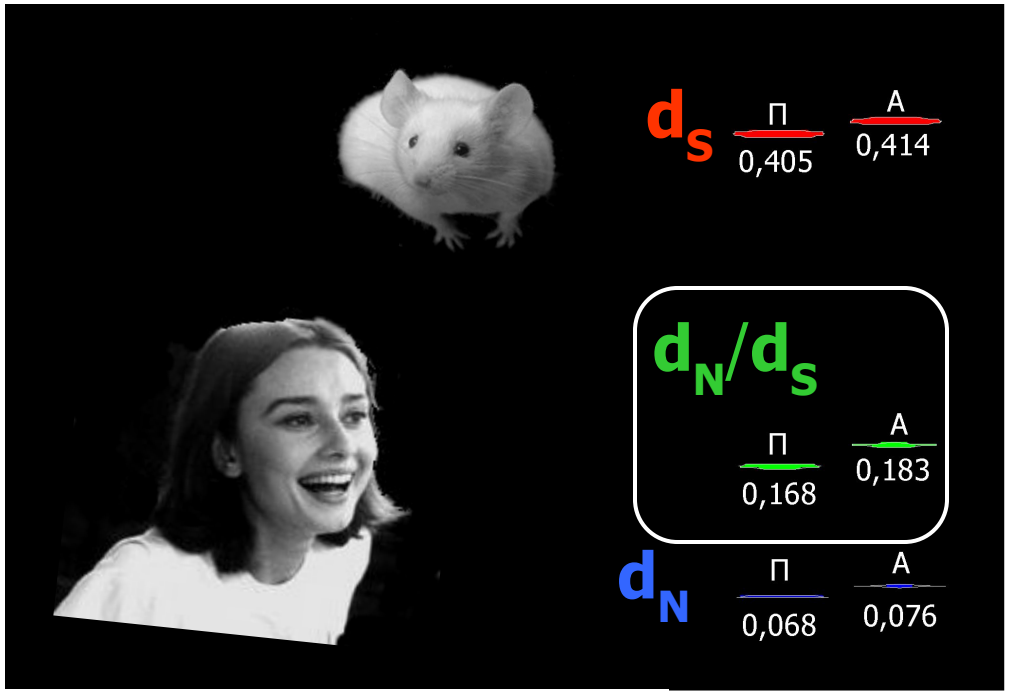
В большинстве генов, уровни и синонимичных, и несинонимичных замен выше в альтернативных областях по сравнению с конститутивными

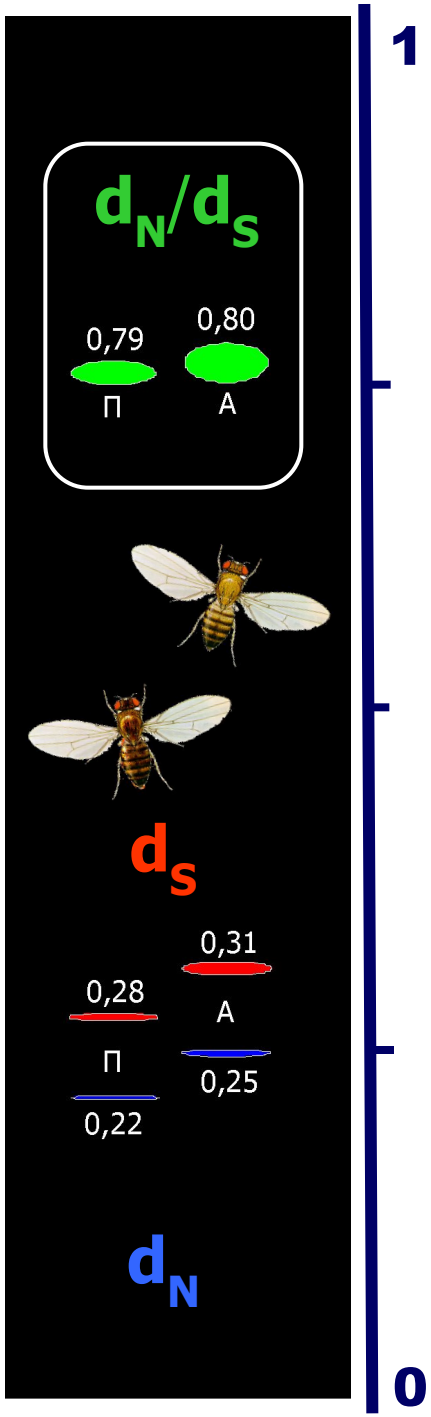
Альтернативные области изменяются быстрее, чем константные



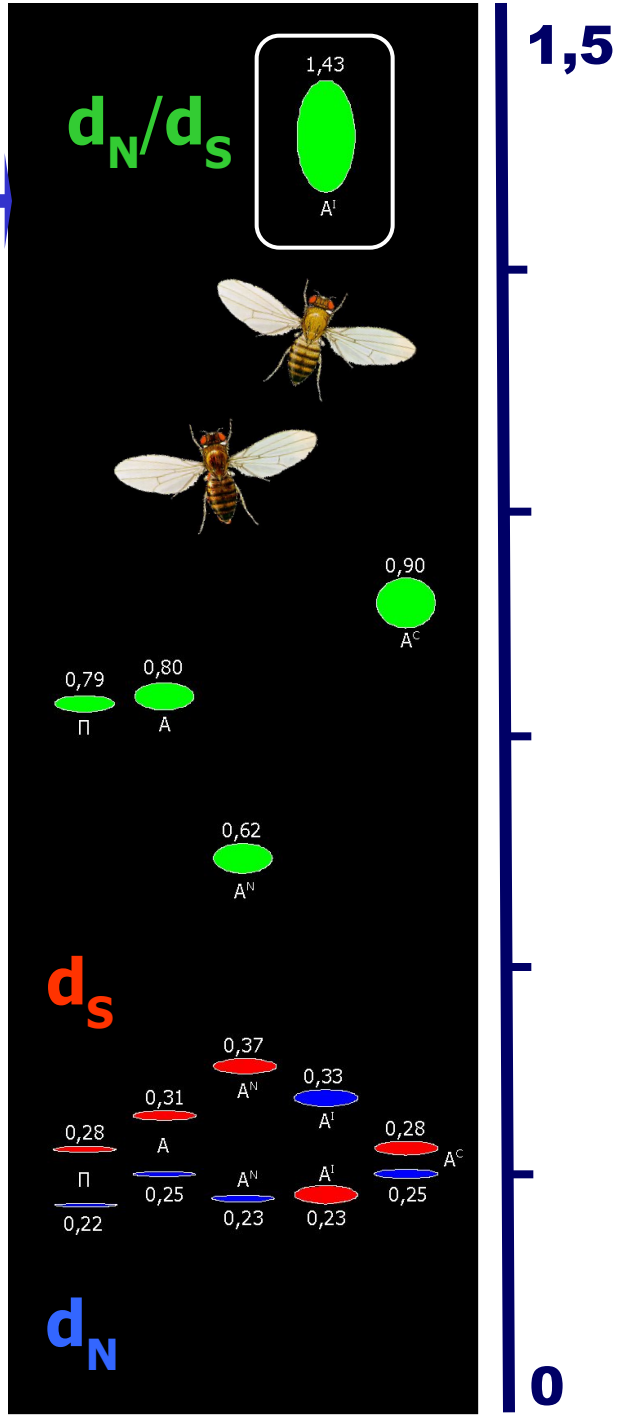


Ослабление стабилизирующего отбора в альтернативных областях





$$d_N(A^I)/d_S(A^I) = 1{,}43$$



Положительный отбор во внутренних альтернативных областях генов *Drosophila*

Тест МакДональза-Крейтмана: положительный отбор в минорных альтернативных областях

- Сравнение различий между генами человека и шимпанзе со SNP человека
- Экзоны, консервативные в генах мыши и/или собаки
- Гены с ≥60 ESTs
- Оценка значимости – тест Фишера

	Pn/Ps (SNP)	Kn/Ks (геномы)	разница	значим.
Конст.	0.72	0.62	– 0.10	0
Мажор.	0.78	0.65	– 0.13	0.5%
Минор.	**1.41**	**1.89**	**+ 0.48**	0.1%

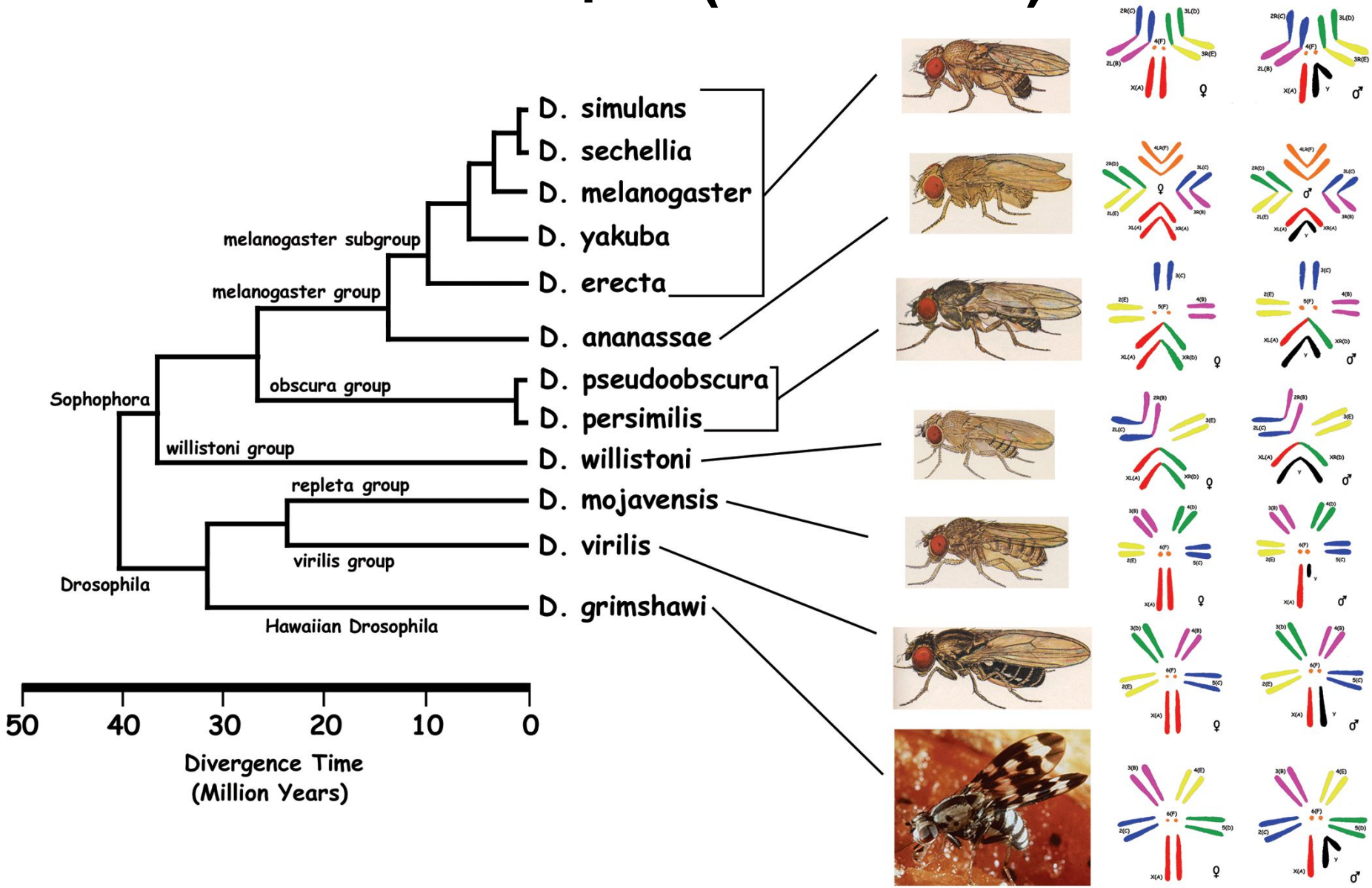
Минорные изоформы:

- Больше несинонимичных SNP: **Pn(alt_minor)=.12% >> Pn(const)=.06%**
- Больше несинонимичн. замен: **Kn(alt_minor)=.91% >> Kn(const)=.37%**
- **Положительный отбор** (не просто ослабление стабилизирующего): **α = 1 – (Pa/Ps) / (Ka/Ks) ~ 25% позиций**
- Аналогичные результаты для всех консервативных альтернатив или для всех генов с высоким покрытием EST

Попытка синтеза

- Альтернативный сплайсинг часто видоспецифичен
- *«молодые» альтернативные изоформы часто тканеспецифичны*
- … но все же функционален
- В альтернативных областях снижен уровень стабилизирующего отбора
  - избыток несинонимичных замен (по сравнению с синонимичными)
- В альтернативных областях действует положительный отбор
  - избыток несинонимичных замен (по сравнению с SNP)
- Альтернативный сплайсинг часто тасует белковые домены
- **Таким образом, альтернативный сплайсинг – это способ попробовать новые формы белков, не жертвуя старыми**

Планы: много дрозофил; млекопитающие (ENCODE)



Что делать?

- Оценить не только скорость потерь альтернатив, но и скорость приобретений (отличая молодые изоформы от ошибок сплайсинга)
- Следить за:
  - функциональность: транслируемые / нарушающие рамку
  - уровень: частые / редкие изоформы
  - паттерн тканеспецифичности (?)
  - Тип альтернативы: N-концевая / внутренняя / C-концевая

- Альтернативный сплайсинг в одном геноме / конститутивный – в другом (данные с микрочипов)

- Эволюция регуляции альтернативного сплайсинга
- Ошибки сплайсинга и мутации:
  удержаные интроны, пропущенные экзоны, скрытые сайты

Благодарности

- Обсуждения
  - Евгений Кунин (NCBI)
  - Игорь Рогозин (NCBI)
  - Всеволод Макеев (ГосНИИГенетика)
  - Дмитрий Петров (Stanford)
  - Дмитрий Фришман (GSF, TUM)
- Данные
  - King Jordan (NCBI)
- Поддержка
  - РАН (программа «Молекулярная и клеточная биология»)
  - РФФИ
  - Howard Hughes Medical Institute
  - INTAS

Авторы

- **Андрей Миронов** (МГУ, ИППИ)
- **Рамиль Нуртдинов** (МГУ)
  – человек/мышь/собака
- **Дмитрий Малько** (ГосНИИГенетика)
  – дрозофилы/комар
- **Екатерина Ермакова** (ИППИ)
  – Kn/Ks
- **Василий Раменский** (ИМБ)
  – SNP
- **Ирена Артамонова** (ИОГен)
  – человек/мышь, график из обзора
- **Алексей Неверов** (ГосНИИГенетика)
  – функциональность изоформ

Основные публикации

10,1 D.B.Malko, V.J.Makeev, **A.A.Mironov**, **M.S.Gelfand** (2006) Evolution of the exon-intron structure and alternative splicing in fruit flies and malarial mosquito genomes. Genome Research. 16: 505-509 .

4,1 **E.O.Ermakova**, **R.N.Nurtdinov**, **M.S.Gelfand** (2006) Faster evolutionary rate in alternatively spliced regions of mammalian genes. BMC Genomics. 7: 84.

0,36 **Е.О.Ермакова**, Д.Б.Малько, **М.С.Гельфанд** (2006) Эволюционные отличия альтернативных и постоянных белок-кодирующих участков альтернативно сплайсируемых генов Drosophila. Биофизика. 51: 581-588.

0,36 **Р.Н.Нуртдинов**, **А.Д.Неверов**, Д.Б.Малько, И.А.Космодемьянский, **Е.О. Ермакова**, В.Е.Раменский, **А.А.Миронов**, **М.С.Гельфанд** (2006) EDAS – база данных альтернативно сплайсированных генов человека. Биофизика. 51: 589–592.

книга **A.D.Neverov**, **A.A.Mironov**, **M.S.Gelfand** (2006) Similarity-based gene recognition. Handbook of Computational Molecular Biology (Chapman & Hall/CRC), Chapter 2.

книга **M.S.Gelfand** (2006) Computational analysis of alternative splicing. Handbook of Computational Molecular Biology (Chapman & Hall/CRC), Chapter 16.

В работе:

Nurtdinov et al. Conserved and species-specific alternative splicing in mammalian genomes

Ramensky et al. Positive selection in alternatively spliced regions of human genes

Публикации в других областях

Системная биология

10,2 V.Spirin, **M.S.Gelfand**, **A.A.Mironov**, L.A.Mirny (2006) Metabolic network in the evolutionary context: multi-scale structure and modularity. Proceedings of the National Academy of Sciences of USA. 103: 8774-8779.

0,89 А.В.Любецкая, Л.И.Рубанов, **М.С.Гельфанд** (2006) Потоковая модель метаболизма аминокислот кишечной палочки. Биохимия. 71: 1544-1549.

Сравнительная геномика и эволюция регуляторных систем

5,9 C.Yang, **D.A.Rodionov**, X.Li, **O.N.Laikova**, **M.S.Gelfand**, O.Zagnitko, M.F.Romine, A.Y.Obraztsova, K.H.Nealson, A.L.Osterman (2006) Comparative genomics and experimental characterization of N-acetylglucosamine utilization pathway of Shewanella oneidensis. Journal of Biological Chemistry. 281: 29872-29885.

4,2 **D.A.Rodionov**, P.Hebbeln, **M.S.Gelfand**, T.Eitinger (2006) Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters: Evidence for a novel group of ATP-Binding cassette transporters. Journal of Bacteriology. 188: 317-327.

2,1 **D.A.Rodionov**, **M.S.Gelfand** (2006) Computational identification of BioR, transcriptional regulator of biotin metabolism in alpha-proteobacteria, and its binding signal. FEMS Microbiology Letters. 255: 102-107.

0,44 Н.А.Дорощук, **М.С.Гельфанд**, **Д.А.Родионов** (2006) Регуляция метаболизма азота у грамположительных бактерий. Молекулярная биология. 40: 919-926.

2,2 **A.E.Kazakov**, **E.A.Permina**, **O.V.Kalinina**, **M.S.Gelfand** (2006) Comparative genomics of regulation of heavy metal resistance in Eubacteria. BMC Microbiology. 6: 49.

пока нет **D.A.Rodionov**, **M.S.Gelfand**, J.D.Todd, A.R.J.Curson, A.W.B.Johnston (2006) Comparative reconstruction of transcriptional network controlling iron and manganese homeostasis in alpha-proteobacteria. PLoS Computational Biology. 2: e163.

нет **G.Y.Kovaleva**, G.A.Bazykin, M.Brudno, **M.S.Gelfand** (2006) Comparative genomics of transcriptional regulation in yeasts and its application to identification of a candidate alpha-isopropylmalate transporter. Journal of Bioinformatics and Computational Biology. 4: 981-998.

Структуры и функции белков

4,7 **O.V.Kalinina**, **M.S.Gelfand** (2006) Amino acid residues that determine functional specificity of NADP- and NAD-dependent isocitrate and isopropylmalate dehydrgenases. Proteins. 64: 1001-1009.

нет N.S.Sadovskaya, R.A.Sutormin, **M.S.Gelfand** (2006) Recognition of transmembrane segments in proteins: review and consistency-based benchmarking of internet servers. Journal of Bioinformatics and Computational Biology. 4: 1033-1056.

Обзоры

9,6 **M.S.Gelfand** (2006) Evolution of transcriptional regulation networks in microbial genomes. Current Opinion in Structural Biology. 16: 420-429.

0,44 **М.С.Гельфанд** (2006) Цис-регуляторные структуры РНК бактерий. Молекулярная биология. 40: 609-619.