

В. Любецкий (лаб. 6), Л. Рубанов (лаб. 2)

**Массовый компьютерный поиск генов,
потеря/приобретение которых в эволюции может
лежать в основе потери/приобретения
фенотипического признака**

29 сентября 2022 года

Мы расскажем часть следующей статьи, поданной в биологический журнал:

A large-scale search for genes whose loss/emergence in evolution could underlie the loss/emergence of a specific phenotypic trait

**Vassily A. Lyubetsky¹, Lev I. Rubanov¹,
Oleg A. Zverkov¹, ..., Andrey G. Zaraisky²**

¹ Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute);

² Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences

Также в доклад включена часть наших результатов из других публикаций:

Korotkova D.D., Lyubetsky V.A., Rubanov L.I., Seliverstov A.V., Zverkov O.A.,, Zاراisky A.G. [Bioinformatics screening of genes specific for well-regenerating vertebrates reveals c-answer, a regulator of brain development and regeneration](#) // Cell Reports, 2019, V. 29, № 4, P. 1027–1040.e6. DOI: 10.1016/j.celrep.2019.09.038 [IF 8.1, WoS-Q1].

Rubanov L.I., Zاراisky A.G., Shilovsky G.A., Seliverstov A.V., Zverkov O.A., Lyubetsky V.A. [Screening for mouse genes lost in mammals with long lifespans](#) // BioData Mining, 2019, V. 12, ID 20. DOI: 10.1186/s13040-019-0208-x (IF2.7, WoS-Q1).

Общий взгляд: рассматриваются

очень далёкие уровни организации материи:

например, **уровень геномов** и

уровень фенотипических проявлений генома.

Хотим установить связь между событиями

на таких уровнях (во времени, в эволюции):

событие на одном уровне соответствует

какому событию на другом уровне?

1-й большой пример. **Виды:** позвоночные – от рыб и амфибий до плацентарных млекопитающих

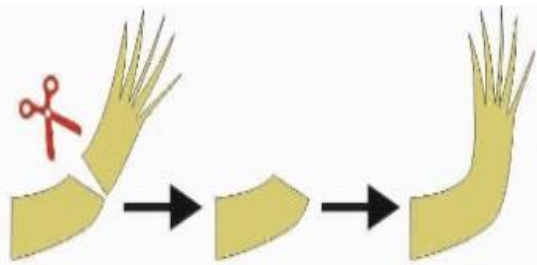
Фенотип-й признак: регенеративная потенция вида (например, отрастёт ли после ампутации у детёныша большой придаток тела).

Вопрос: какие гены отвечают за регенерацию?

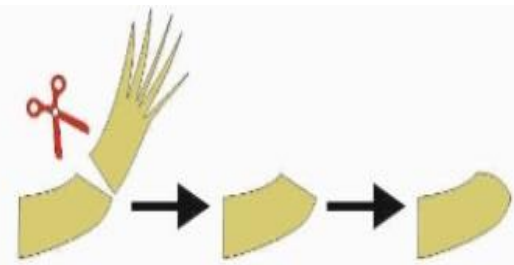
Нижние виды – те, что обладают данным признаком (здесь: рыбы и амфибии).

Верхние виды – те, что решительно не обладают этим признаком (здесь: плацентарные млекопитающие).

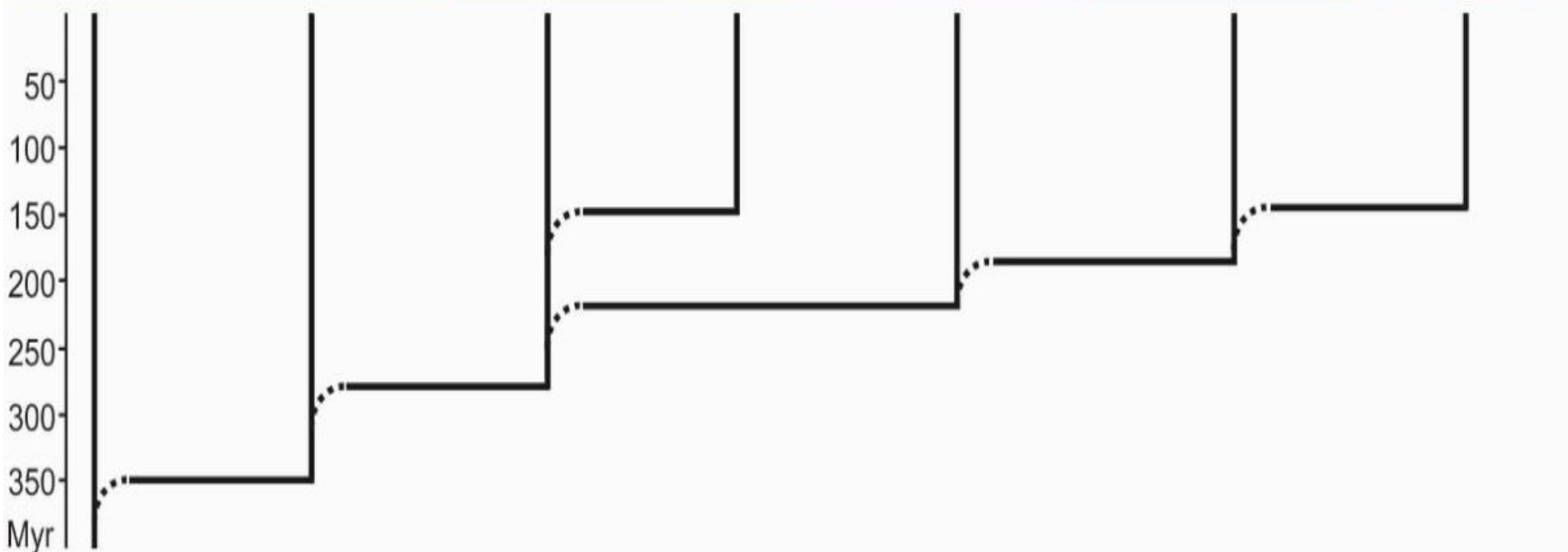
Средние виды – некоторые и в некоторой степени обладают этим признаком (= между нижними и верхними).



good regeneration



bad regeneration



Итак, проблема: **какие гены отвечают за регенеративную потенцию?**

Наш ответ: те, которые у нижних видов присутствуют, у верхних точно отсутствуют, а у средних есть или нет в некоторой степени.

Средние виды – это «серая зона»: например, некоторые ящерицы сохранили способность регенерировать свой хвост, хотя не полностью; детёныши примитивных млекопов (в частности, опоссума) могут регенерировать части конечности более эффективно, чем детёныши плацентарных млекопов.

Изучение перехода через средние виды поможет понять **механизм потери регенеративной потенции.**

По сравнению с нижними видами (амфибиями и рыбами)

рептилии, птицы, млеко

резко снижают регенеративную потенцию,

так что в первую очередь важно изучить ситуацию, когда

средних видов нет.

Тогда рыбы и амфибии – нижние,

а все, кто выше них, – верхние.

Чтобы найти такие гены, мы перебираем все белок-

кодирующие гены лягушки, которые присутствуют у

рыб и отсутствуют у рептилий, птиц, млеко-

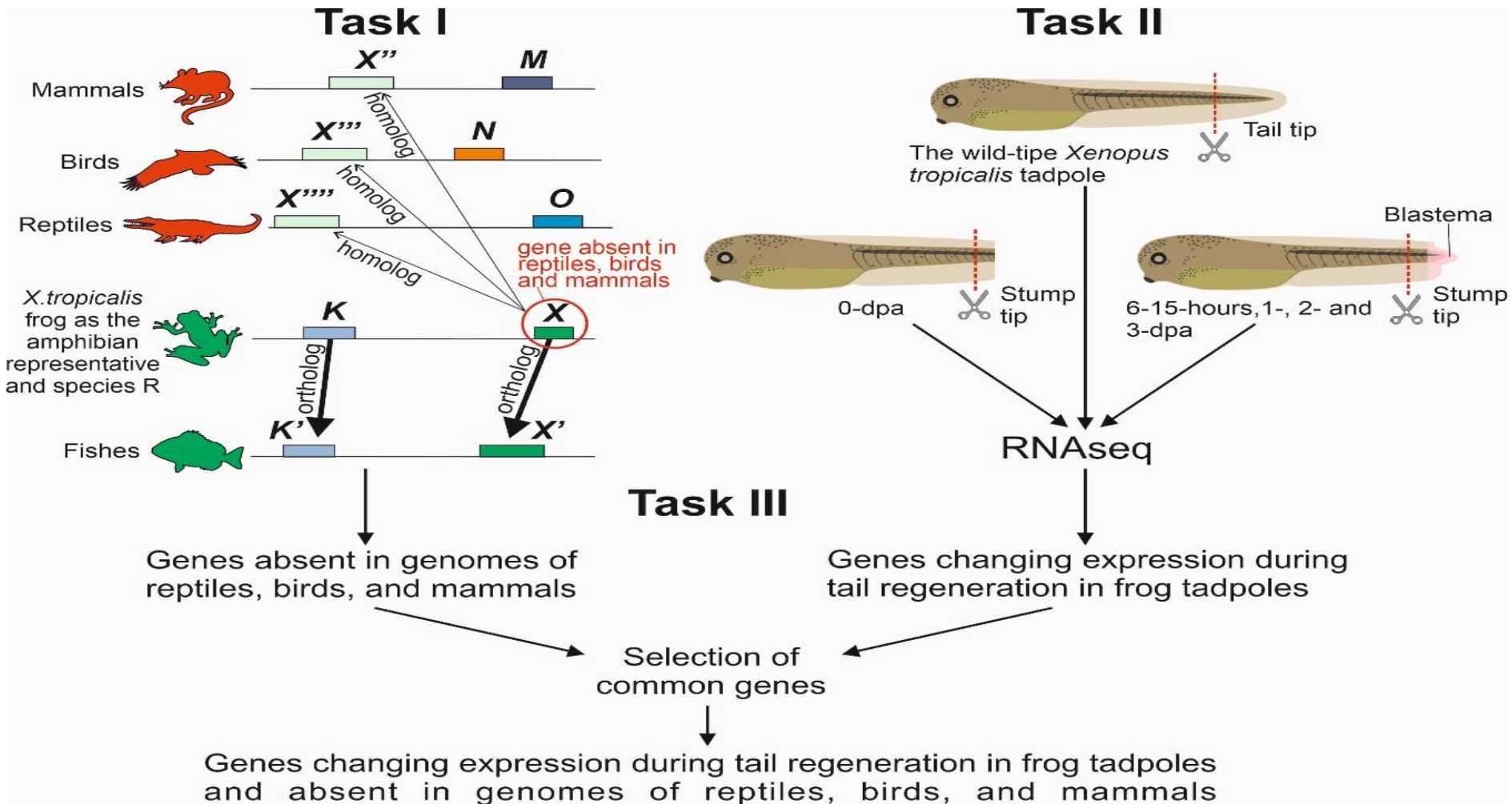
Идея нашего метода: для каждого гена **X** из **R** проверим: ген X

I) присутствует со свидетелями (=с локальной синтенией) у рыб

и отсутствует у верхних (там **X** только гомолог); и

II) у модельного вида **R** ген **X** значительно меняет уровень

экспрессии после ампутации хвоста.



Сначала обсудим **более простую задачу II**.

Нам были доступны 6 групп образцов (SRA project SRP091865, GEO project GSE88975) – число ридов (с длиной ~200) всех транскриптов после ампутации в WT и через 0, 6, 15, 24, 72 часов после ампутации.

Мы привязали риды к интересующим генам 10-й сборки генома *Xenopus tropicalis* (BWA-MEM aligner) и оценили уровень экспрессии каждого гена в образце относительно всех генов образца (DESeq2). Используя RNA-sequence analysis, нашли гены, которые **значимо меняют свою экспрессию**. Эти гены лягушки составили то, что мы называем **список-2**.

Итоговый список (=выход/результат нашего алгоритма) есть пересечение **списка-1** и **списка-2**. Это пересечение содержит гены лягушки (и дополнительную информацию).

Трудная часть алгоритма состояла в получении **списка-1**, который состоит из **генов лягушки, присутствующих в нижних видах и отсутствующих в верхних видах**.

Здесь проблема: для б-к гена **X** лягушки (=геном **R**) и любого другого вида (=геном **A**) определить, **присутствует ли ген X в A, т.е. имеется ли в A ген Y «тот же самый» что X, т.е. \exists ли Y ОРТОЛОГИЧНЫЙ X.**

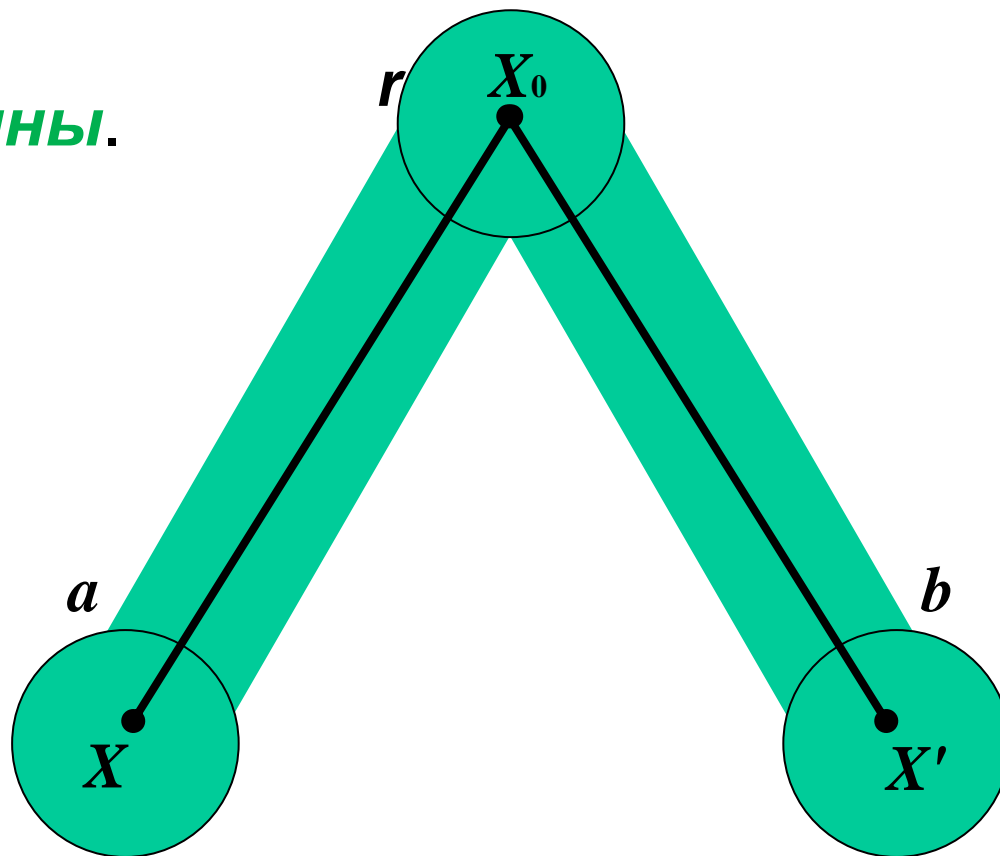
Но что значит ортологичен ??!

Вид r разделился на виды a и b (=видообразование). Ген

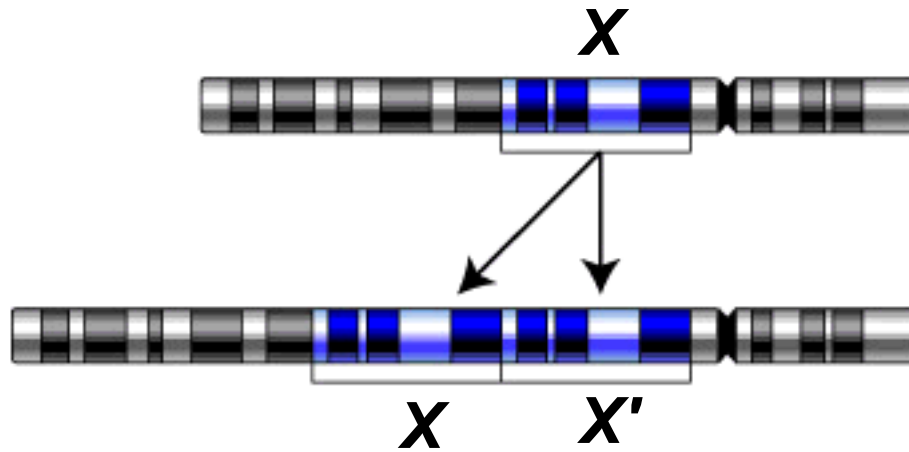
X_0 из r ушёл в вид a как X и в вид b как X' .

Какое-то эволюционное время X и X' одинаковы/похожи =
гомологичны по последовательности и по функции в
клетке.

Тогда X и X' ортологичны.



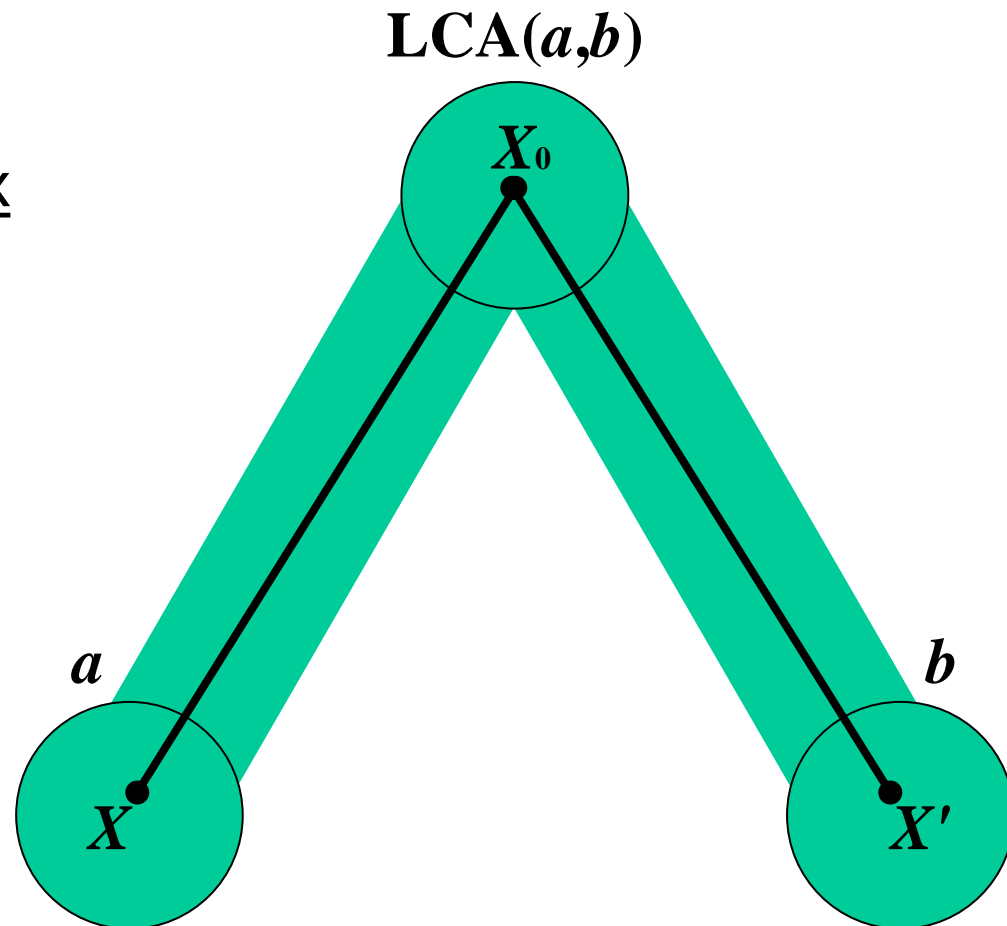
В виде **a** ген **X** дублировался на **X** и **X'**. Какое-то эволюционное время **X** и **X'** одинаковы/похожи = гомологичны по последовательности и возможно по функции в клетке. Тогда **X** и **X'** **паралогичны**.



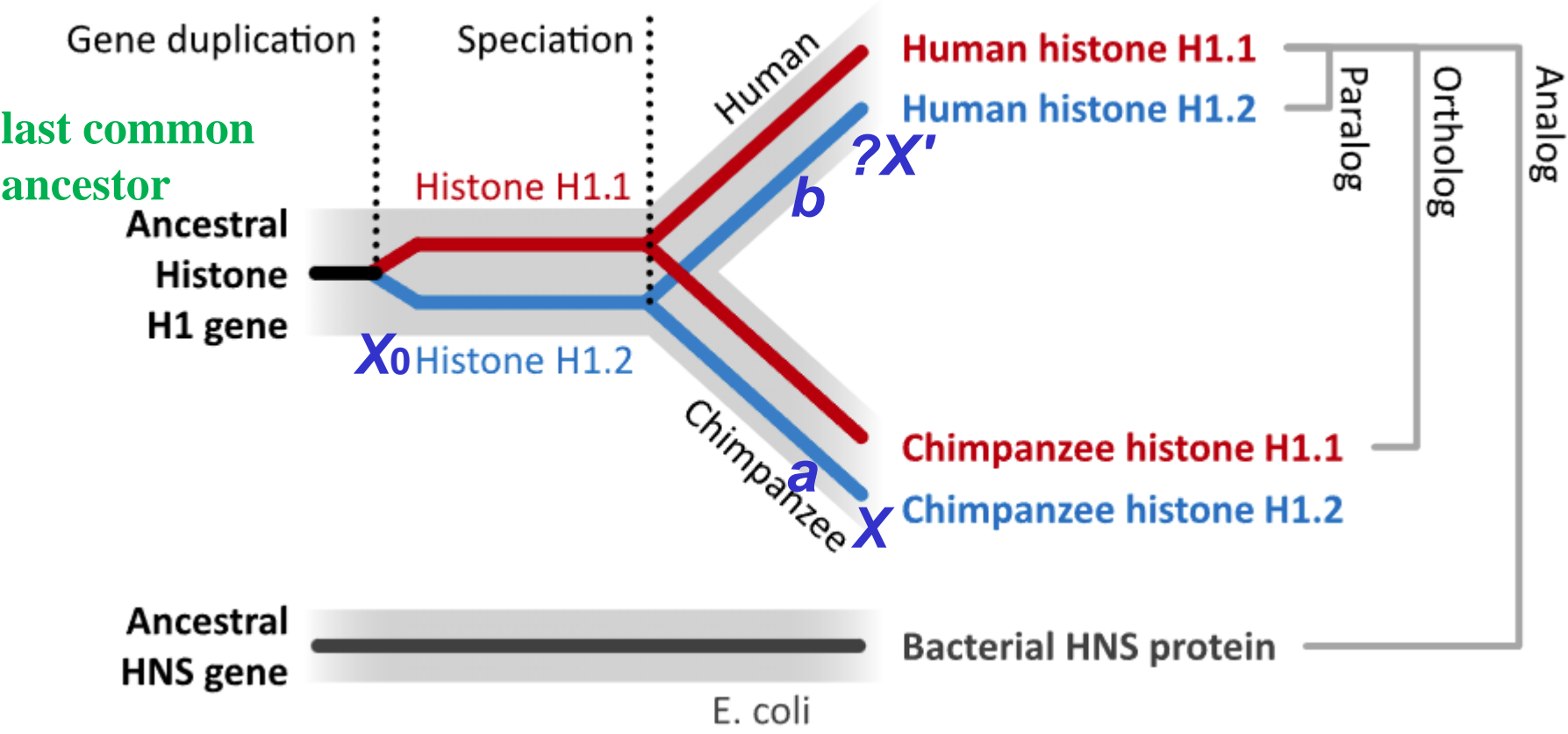
Ещё ортологи или паралоги могут теряться, а гены постоянно меняются во времени. Есть горизонтальный перенос и т.д.

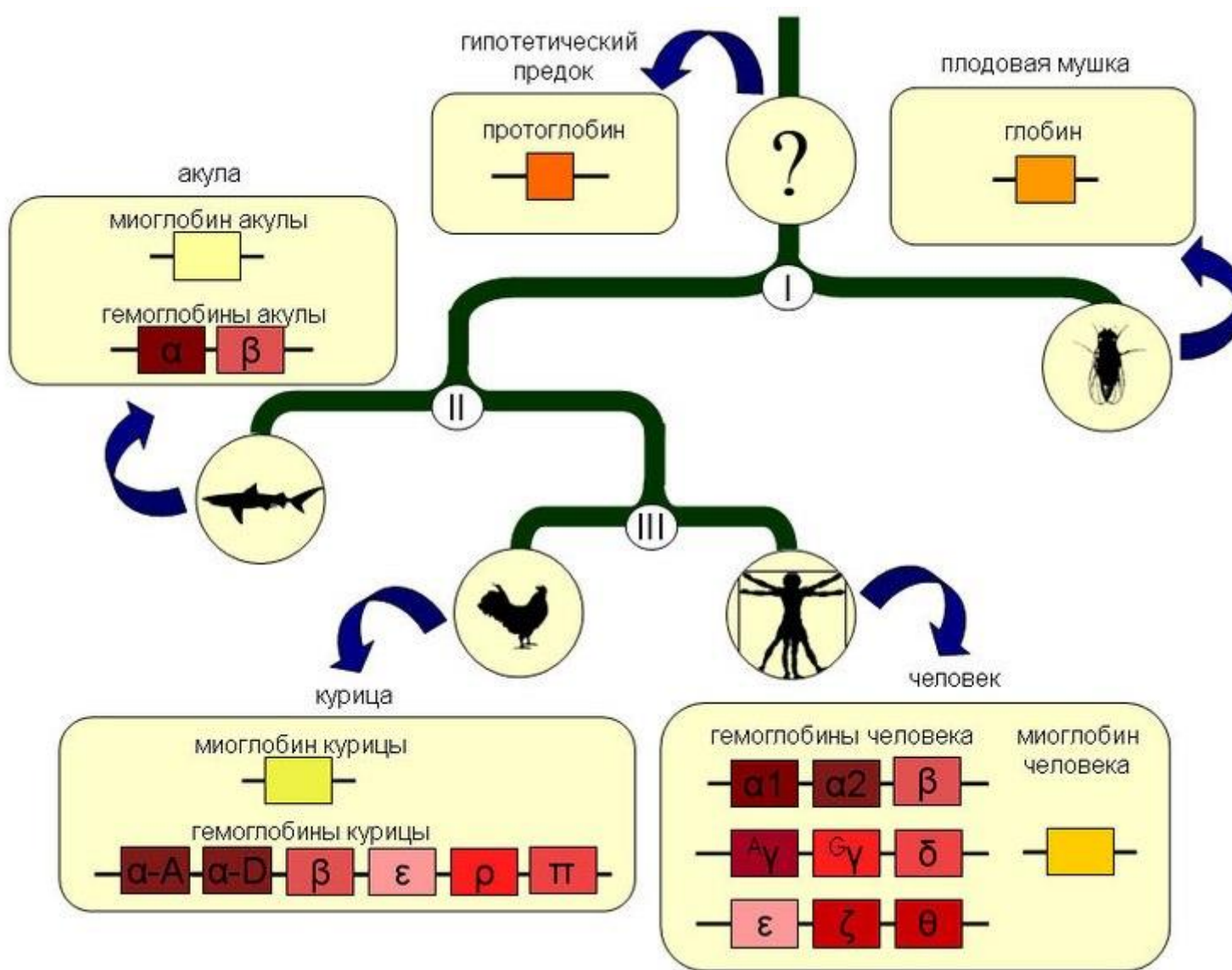
Ситуация ещё труднее: дан ген X в современном виде a ;
сохранился ли в другом современном виде b ген X' – **прямой**
потомок предкового гена X_0 , который в эволюции прямо
породил ген X в a .

Здесь $LCA(a,b)$
может быть далеко в предках
и по пути от предкового гена
 X_0 могло произойти
несколько паралогов!
Тогда X' **ортолог** для X .



Нахождение ортолога X' усложняется тем, что в эволюции видообразование происходит вместе с дупликацией гена/генома, потерей и горизонтальным переносом генов:





Ортологичные гены/белки -- гомологичные гены/белки, разошедшиеся в видообразовании, и также те, что передались от них в вертикальной эволюции. Но это **НЕ компьютерное определение!**

Итак, ортологичные гены – гены, которые у различных видов произошли видообразованием их *последнего* общего вида-предка (что зависит от таксона, в котором определяется общий предок рассматриваемых видов).

Есть разные способы определения ортологичности генов X и X' : деревья генов/белков, модели из марковских цепей и т.д. В БДх приводятся таблицы ортологов, часто неточные.

Отдельная трудность – большое число паралога X в a и X' в b . Из паралога часто рождается другая функция гена, и трудно решить, кто кому ортолог.

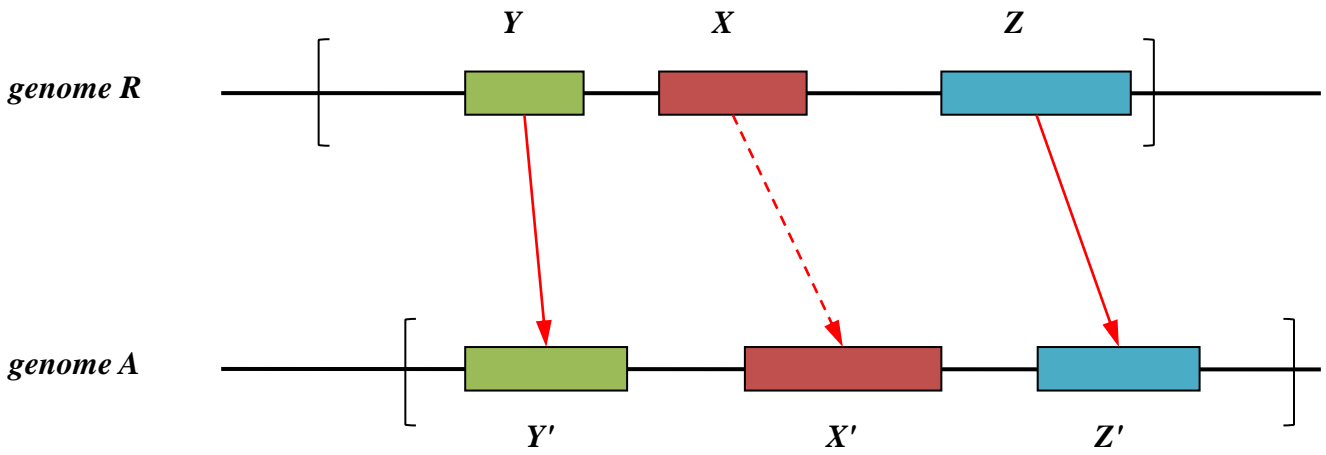
Мы **проверяем ортологичность генов X и X' следующим образом:**

выбираем **референсный** вид R (например, лягушку или другой вид); выбираем **нижнее** и **верхнее** множества.

Ищем все гены X из R , которые в нижних видах **присутствуют** и в верхних **отсутствуют**

в смысле следующих трёх условий (на **отношение «ортологичности»**).

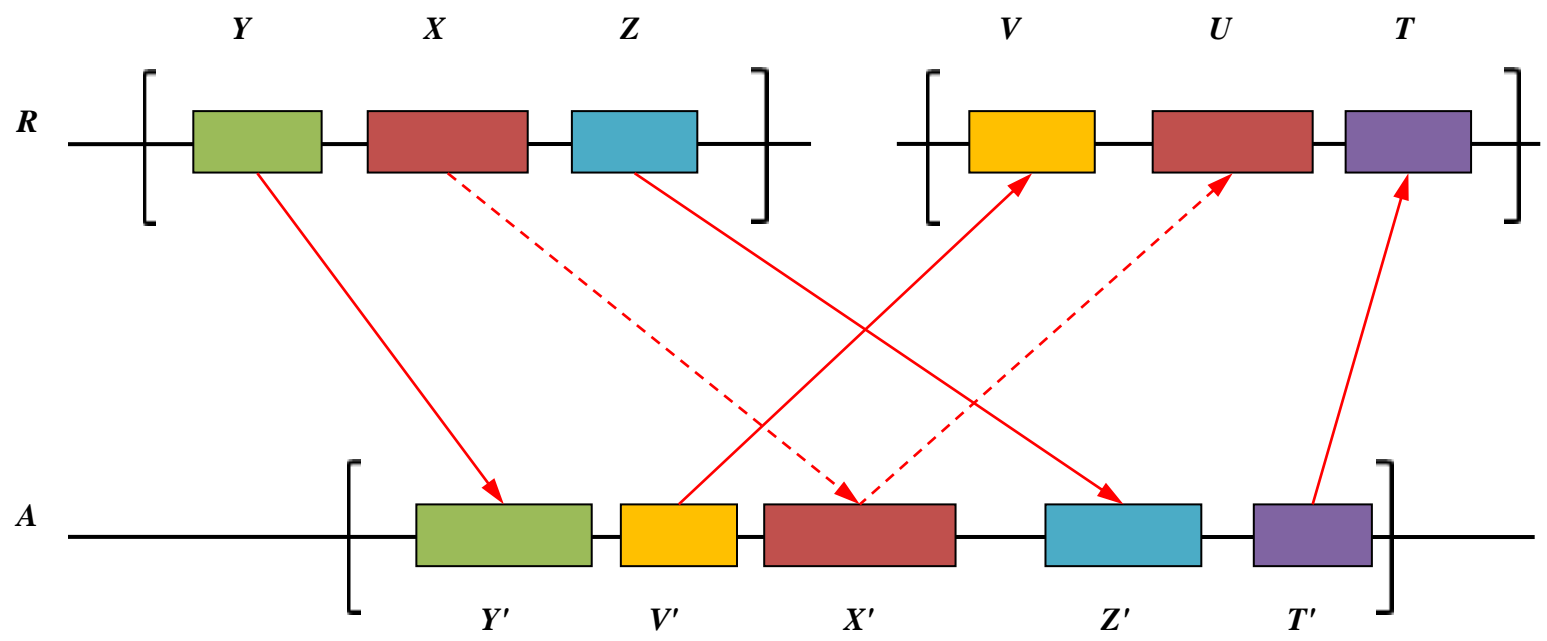
Первое условие: для каждого б-к гена X в R в каждом геноме A рассматриваем u лучших гомологов для X (если нет u штук гомологов, то сколько есть).



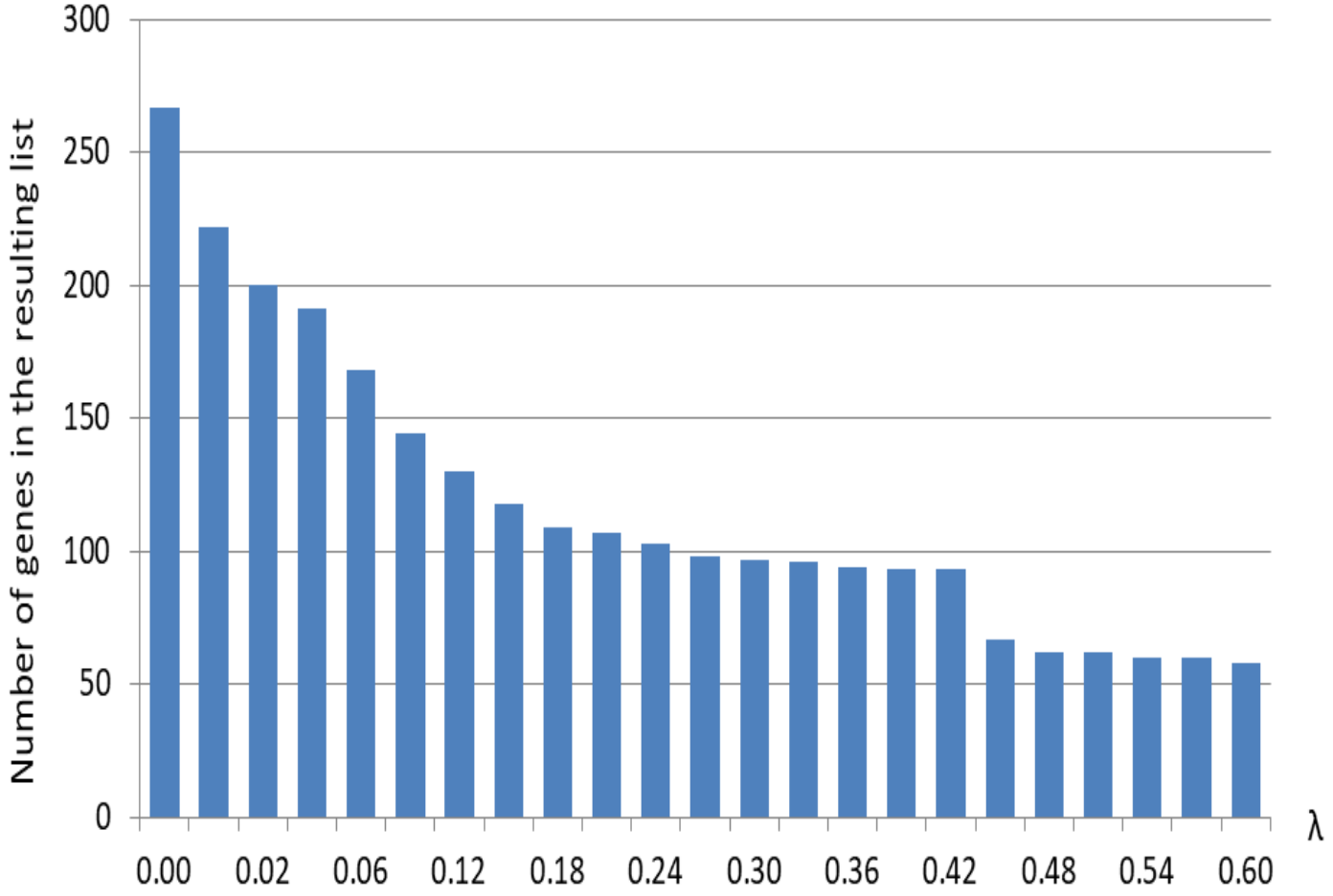
Второе условие:
X имеет локаль-
 ных свидетелей
 = локальная
 синтения.

Третье условие:

$$(\exists U : W(U, X') - W(X, X') > \lambda) \Rightarrow X' \text{ не ортолог}$$



Важную роль играет подбор параметров и особенно параметра λ в третьем условии. Неожиданное и приятное свойство (в итоговом списке, средних видов нет): число выдаваемых генов из ≈ 22 тысяч б-к генов лягушки: **267** при $\lambda=0$ и **93** при $\lambda=0.42$. С ростом λ число выдаваемых генов **убывает!**



Следующие гены/белки найдены компьютером и вошли в **ИТОГОВЫЙ СПИСОК** (= пересечение списка-1 и списка-2):

- 1) **C-C motif chemokine 4 = cc/34** (LOC100493779); its expression increases approximately 20-fold upon regeneration;
- (2) **eotaxin-like** (LOC101733569); its expression increases more than in 8 times during regeneration;
- (3) **an unknown protein** containing two Cu/Zn-binding domains (LOC100495179), its expression increases more than 140-fold during regeneration;
- (4) **fibronectin-like protein** (LOC101733911), its expression increases by more than 60 times during regeneration;
- (5) **potential protein kinase**, an unknown protein containing two protein kinase domains (LOC100493453), its expression increases by more than 60 times during regeneration;
- (6) **potential regulator of Wnt-signaling**, Shisa-4 (LOC101734421); its expression increases in almost 16 times during regeneration.

**В том числе, алгоритм нашёл ранее неизвестный ген,
названный нами**

c-answer,

ответственный за регенерацию

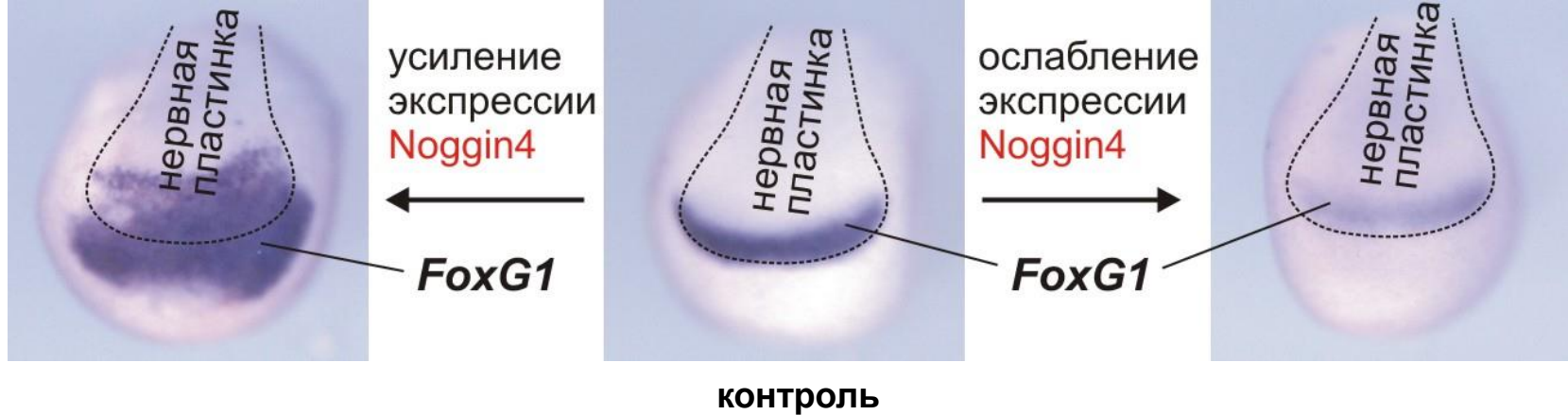
и за развитие конечного мозга!

**С найденными генами были поставлены «мокрые»
эксперименты (в лаборатории А. Зарайского).**

Эти опыты указали на тесную связь найденных генов с регенерацией у амфибий и даже предсказали механизмы этой связи.

Это предмет для отдельного доклада, включающего обсуждение экспериментальных методов, использованных в опытах.

Несколько картинок на тему этих опытов:



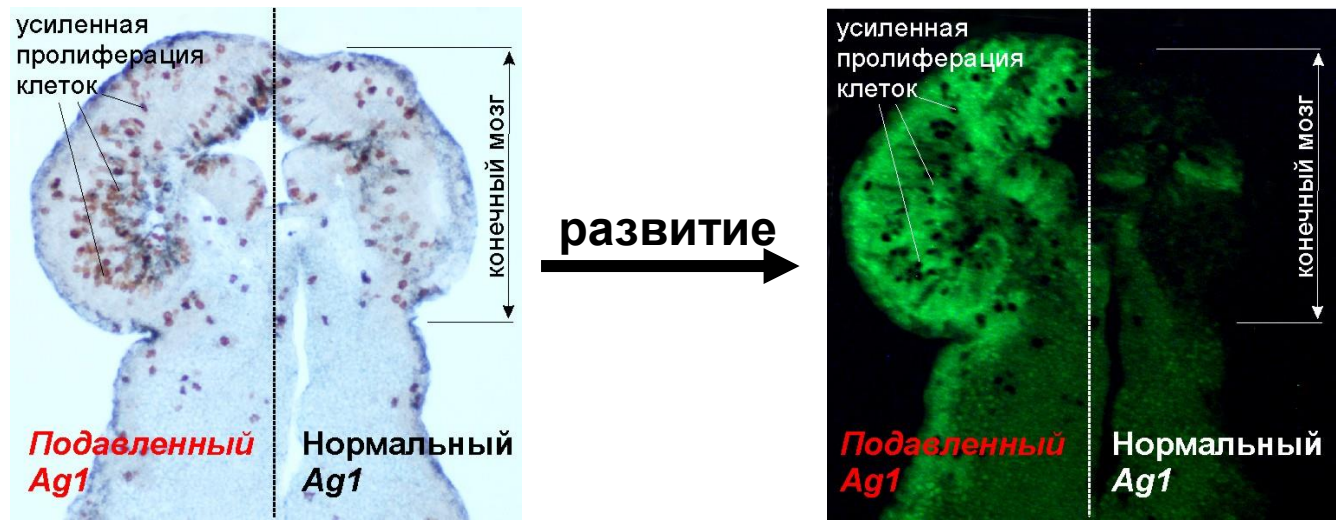
контроль

<Нервная пластинка – зачаток ЦНС у эмбрионов хорд. *Pallium* ~ кора головного мозга.> Наши гены *Ag1*, *Ras-dva1* и *Noggin4* вызывают экспрессию гена *FoxG1* (= развитие **вентральной** области конечного мозга). Одновременно **FoxG1 ингибирует** развитие **дорсальной** области конечного мозга.

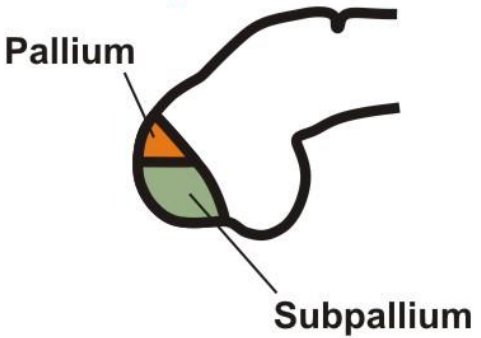


Можно предположить: исчезновение *Ag1*, *Ras-dva1* и *Noggin4* у предков высших позвоночных/млекопитающих привело к снижению уровня экспрессии *FoxG1*, что, стимулировало развитие дорсальной области конечного мозга (и в конечном счёте – коры).

Подавление генов *Ag1*, *Ras-dva1* или *Noggin4* у эмбриона шпорцевой лягушки вызывает подавление=ингибирование экспрессии гена *FoxG1* и стимулирует деление клеток, что приводит к увеличению размеров конечного мозга.



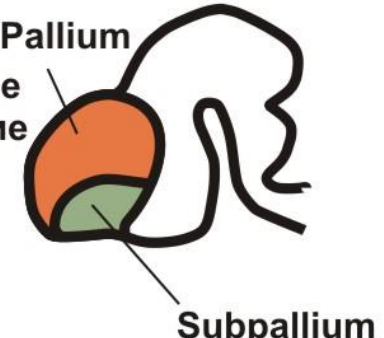
Примитивный конечный мозг рыб и амфибий



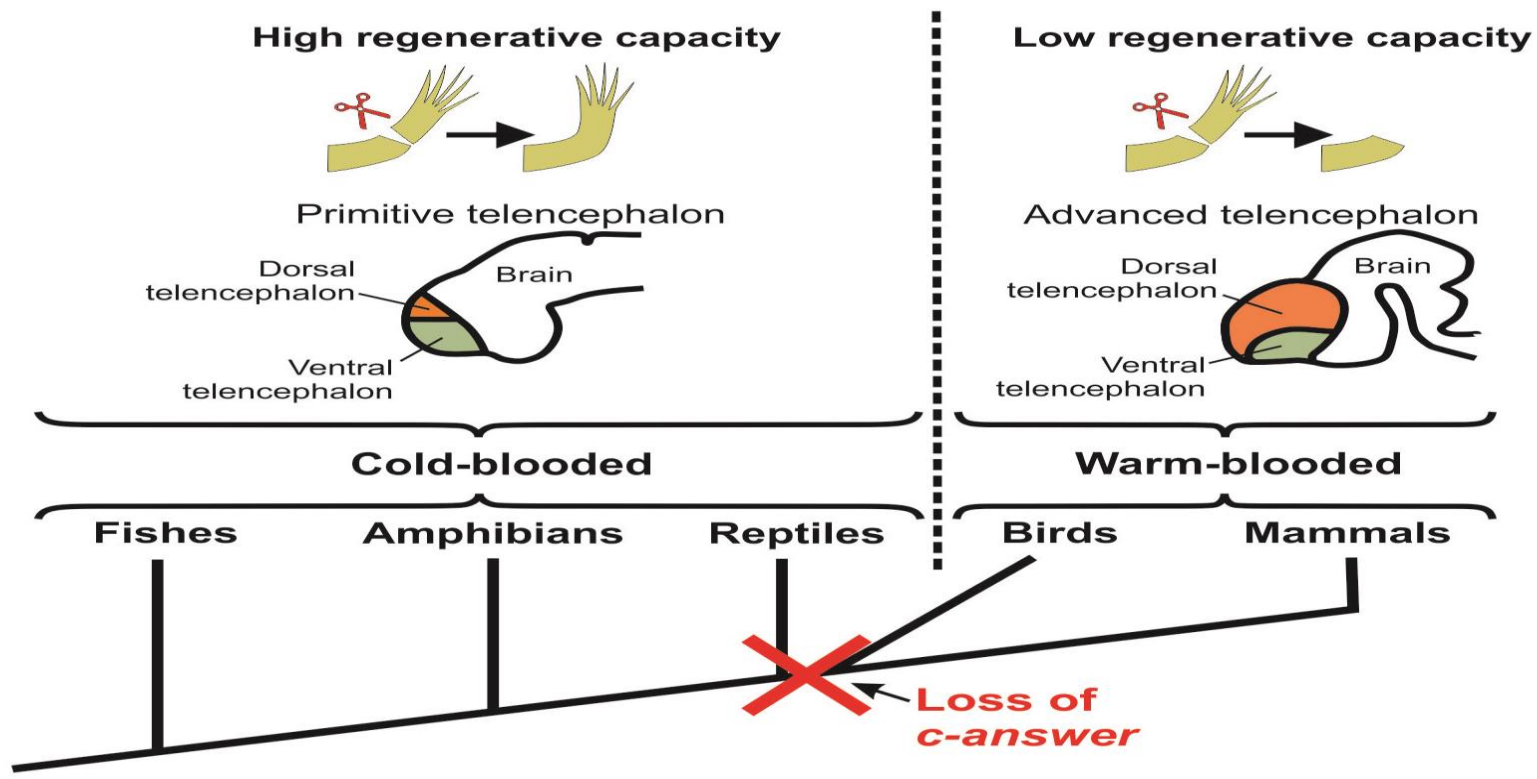
Потеря *Ag1*, *Ras-dva1* и *Noggin4*

интенсивное размножение клеток

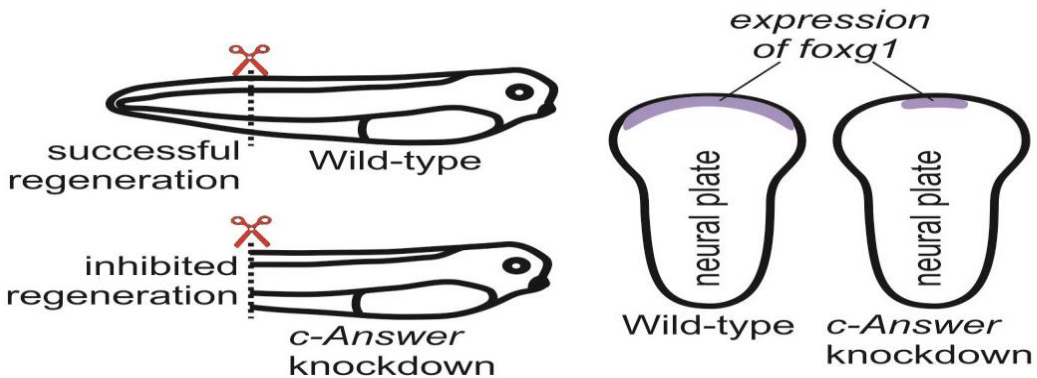
Развитый конечный мозг высших позвоночных



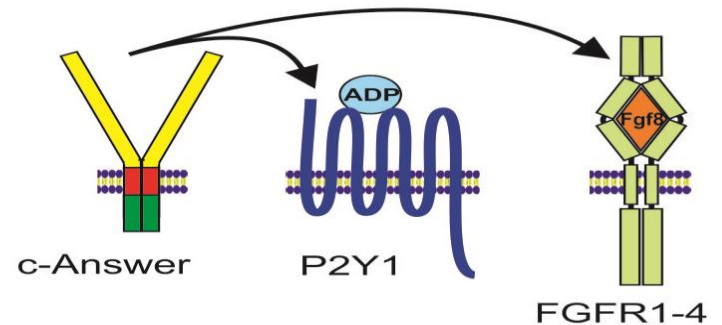
Loss of *c-answer* in evolution coincided a decrease of regenerative capacity and an increase of the dorsal telencephalon in warm-blooded



Knockdown of *c-answer* in frog inhibits regeneration and expression of the ventral telencephalon regulator *foxg1*



c-Answer promotes receptors' activity



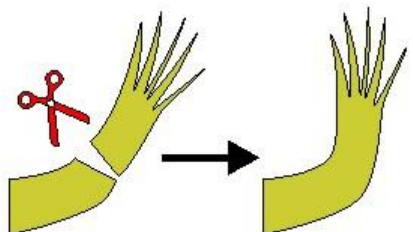
Роль генов *Ag1*, *Ras-dva1* и *Noggin4* в регенерации?

Высокая регенерационная способность рыб и амфибий

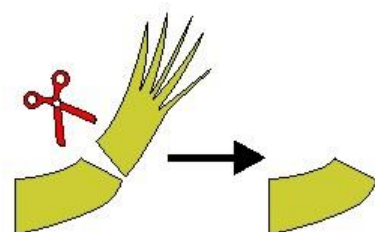
=анамнии

Низкая регенерационная способность рептилий, птиц и млекопитающих

=амниоты



ЭВОЛЮЦИЯ →



<i>Ag1</i>	■	■	НЕТ	НЕТ	НЕТ
<i>Ras-dva1</i>	■	■	■	НЕТ	НЕТ
<i>Noggin4</i>	■	■	■	■	НЕТ



Рыбы Амфибии Рептилии Птицы Млекопитающие

Исчезновение *Ag1*

Исчезновение *Ras-dva1*, и *Noggin4*

Используется

и другое,

трёхвидовое

условие: R –

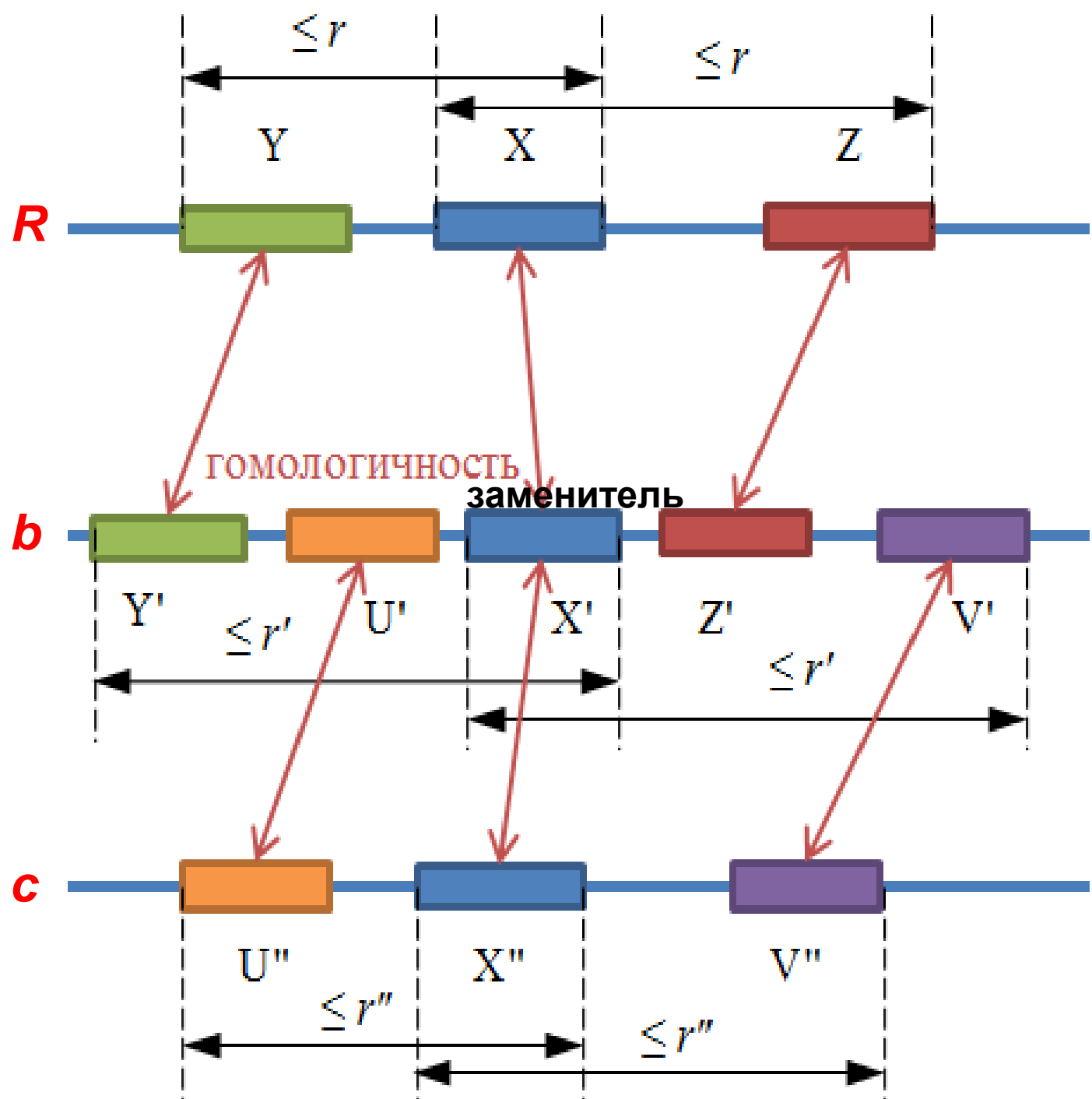
лягушка,

b – рыба,

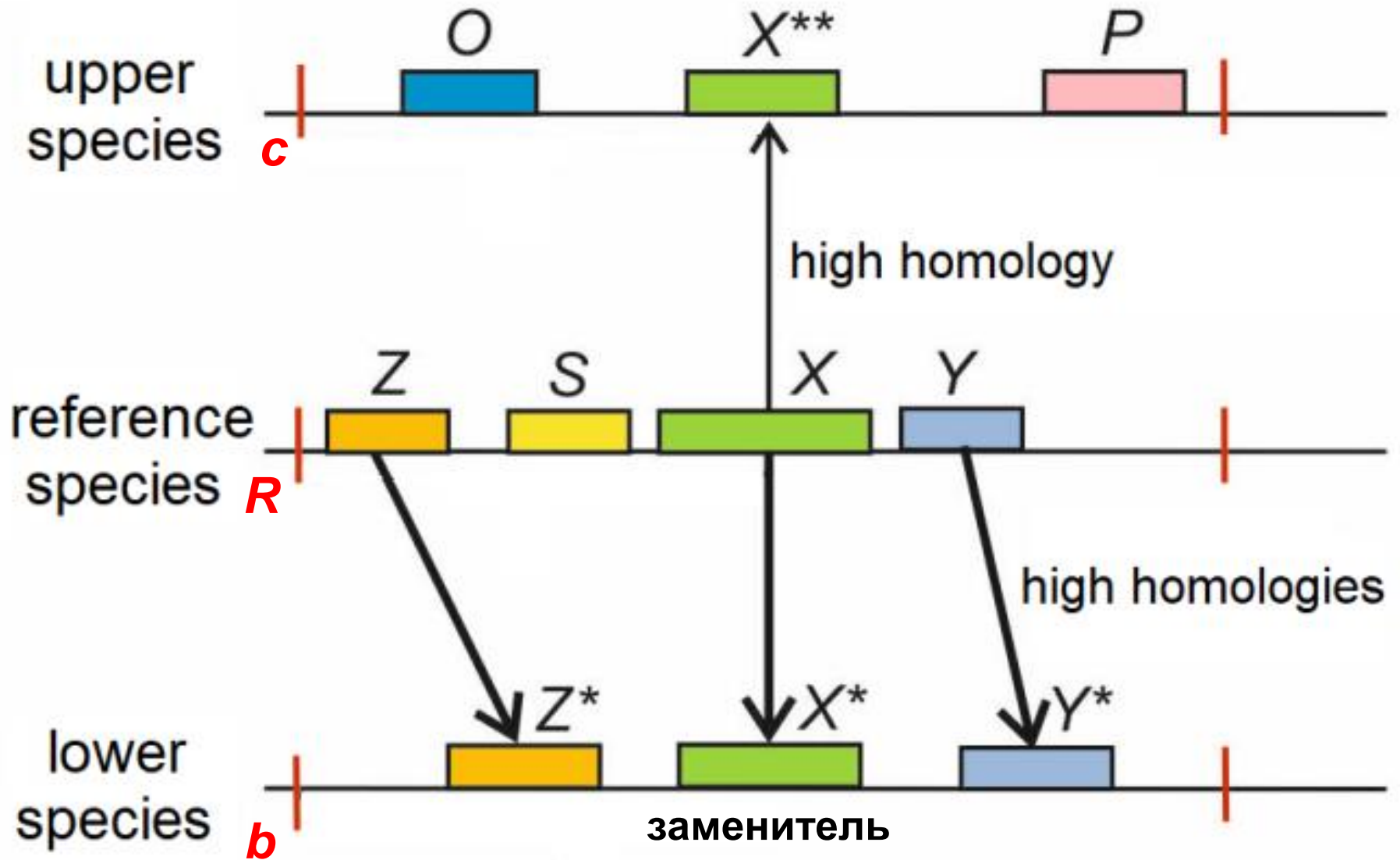
c – нотто;

здесь

X не потерян:



Трёхвидовое условие: здесь X потерян:



Использование трёхвидового условия для поиска
потерянных генов:

ген X представлен в референсном виде R (лягушке), но не представлен со свидетелями «вверх»; зато X представлен в рыбе b геном X'' – «заменителем» X ; а X'' представлен со свидетелями «вверх» в c (например, в человеке или морской свинке).

Пример – на след слайде, где в роли c выступает морская свинка. Такой ген X также не потерян у млекоов.

Cavia porcellus
scaffold 30

ENSCPOG00000011826
ENSCPOP00000018492
FZD9

...174
...151
SERPINE1

...1746
...1577
ACHE

нет синтении с
млекопитающими

3.0E-59

BBH
0

Xenopus tropicalis
GL174277.1

ENSXETG00000023966
ENSXETP00000051708
sfrpx

...23967
...51709
LRTOMT

...23968
...51710
folr1

3.0E-23

BBH
0

BBH
9.0E-104

Gasterosteus aculeatus
group1

ENSGACG00000007608
ENSGACP00000010090
Fzd9b = заменитель ***sfrpx***

...7161
...9482
serpini1

...7230
...9613
BCHE

BBH
4.0E-100

...7832
...10380
LRTOMT

...7809
...10359
FOLR1

Трёхвидовое условие через рыбу

Ген лягушки ENSXETG00000023966 (*sfrpx*) **потерян** у птиц и плацентарных млекопитающих, но имеет «заменителя» – ген *fzd9b* у рыбы *Gasterosteus aculeatus* (трёхиглой колюшки), который **представлен как FZD9** у морской свинки *Cavia porcellus*. Ген не потерян у млекопитающих.

В геноме **лягушки** показан локус гена *sfrpx*:



Ген *sfrpx* также сохранился у **черепахи** (рептилии) *Pelodiscus sinensis* и у других рыб. (*sfrpx* содержит богатый цистеином внеклеточный домен С6 box, типичный для рецептора MuSK.)

Ген *sfrpx* не потерян = алгоритм не отбирает его.

Итак, наш метод = алгоритм = програм состоят в следующем:

**выбирается признак! и по нему выбираются
нижние, средние и верхние множества.**

Например: **хотим найти** гены, связанные с

**1) ослаблением регенеративной потенции и
развитием конечного мозга.**

Тогда **нижнее и верхнее множества:**

анамнии (амфибии и рыбы) против **амниот** (рептилии, птицы,
млекопитающие);

2) развитием теплокровности.

Тогда **нижнее и верхнее множества:** **холоднокровные**
против **теплокровных** (млекопитающие и птицы).

3) развитие плаценты.

Тогда **нижнее и верхнее множества:**

ранние млеки (однопроходные, сумчатые...) против
плацентарных (=высших млекоов).

Наш метод на другом материале,

2-й большой пример:

найти гены, утраченные у видов с

высокой видовой продолжительностью жизни (ПЖ),

<у видов с неотенией (= виды большую часть жизни

сохраняющие свойства раннего возраста,

например, аксолотль, голый землекоп), и т.д.>

Виды: приматы (отряд плацентарных млекопитающих:

обезьяны – сухоносые приматы, Haplorhini;

полуобезьяны – низшие, мокроносые приматы,

Strepsirrhini) и **грызунообразные, зайцевые.**

1-й компьютерный опыт =счёт:

референсный вид R – мышь.

Нами выполнен поиск генов X мыши, которые

сохранились с учётом синтении у **грызунов и кролика**

с малой ПЖ – **нижние виды**,

но **потеряны** у **голого землекопа, дамарского**

пескороя с высокой ПЖ – **верхние виды**.

Алгоритм нашёл потерянные гены мыши: все они **присутствуют**, где стоит знак **+**, и **отсутствуют**, где знак **-**.
Список генов не приводится. Указана ПЖ видов в годах.

Вид – « грызуны »	ПЖ	
серая крыса	3,8	+
сирийский хомячок	3,9	+
домовая мышь	4	+
европейский кролик	9	+
дамарский пескорой	15,5	-
голый землекоп	32	-

2-й компьютерный опыт =счёт:

референсный вид *R* – галаго Гарнетта.

Нами выполнен поиск генов **X** мыши, которые

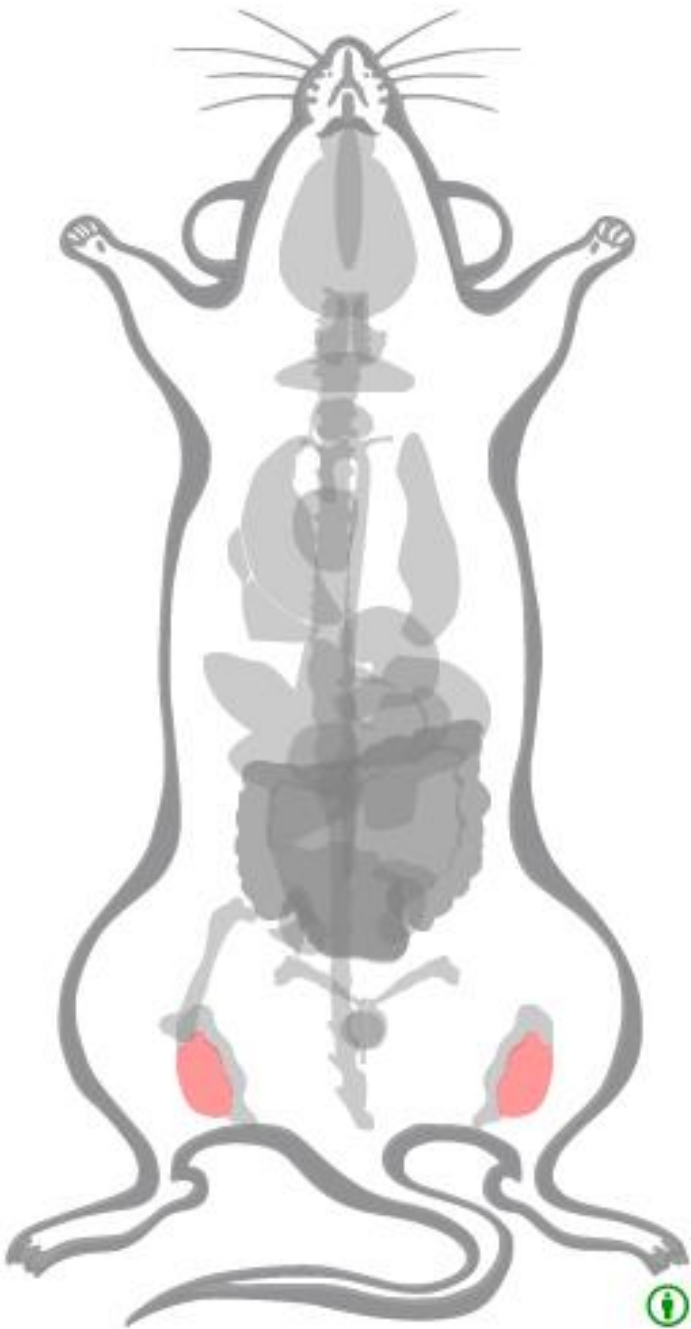
сохранились с учётом синтении у **приматов** с малой

ПЖ – **нижние виды**,

но **потеряны** у **человека, человекообразных** и

капуцина с высокой ПЖ – **верхние виды**.

Вид – приматы	ПЖ	
галаго Гарнетта	15,7	+
филиппинский долгопят	16	+
серый мышиный лемур	18,2	+
обыкновенная игрунка	22,8	+
белощёкий хохлатый гиббон	44,1	-
белоплечий капуцин	54,8	-
бонобо	55	
суматранский орангутан	58	-
шимпанзе	59,4	-
западная равнинная горилла	60,1	-
человек разумный	122,5	-



Найденные гены **X** мыши
(потерянные у видов с **высокой**
ПЖ) в наибольшей степени
экспрессируются в семенниках.

Эти гены могут быть опасны, если соответствующие белки накапливаются в течение длительного времени.

Но: онкологические и нейродегенеративные заболевания развиваются достаточно медленно. Поэтому такие гены не оказывают существенного влияния на молодые организмы и на виды, которые «по их природе» живут недолго.

А **при высокой продолжительности жизни** такие гены вредны и потому **теряются**.

Найденные нами гены могут быть примерами **плейотропных генов**.

(антагонистический) **Плейотропный ген** – полезен в молодом возрасте, когда половой отбор сильнее влияет на вид, но вреден позднее, когда размножения почти нет.

Плейотропные гены могут быть связаны с процессом старения. Они поддерживаются в популяции благодаря их позитивному эффекту в молодости (на **размножение – поэтому нужны в семенниках**),

несмотря на их негативный эффект в позднем возрасте; так как до него доживает незначительная часть популяции и там размножение менее интенсивно (Williams 1957).

Ещё пример найденного гена – Wap: он кодирует whey acidic protein (WAP). Этот белок в молозиве коров (=секрет молочной железы млекопитающих, вырабатываемый в конце беременности и после родов). WAP в небольшом количестве присутствует в молоке мышей, но в большом в молоке кроликов. WAP ингибирует рост бактерий *Staphylococcus aureus*, а также ингибирует пролиферацию mammary epithelial.

Потеря этого гена у видов с относительно высокой ПЖ может быть связана с **увеличением продолжительности кормления детёнышей молоком**. //Конец доклада.

Пример с белками-глобинами (ссылки по RefSeq, NCBI):

xgb (x globin) [496620](#) потерян у плацентарных;

но есть у рыб, утконоса, анолиса, дракона и двух черепах;

gby (globin y, cytoglobin) [100494515](#) потерян у плацентарных;

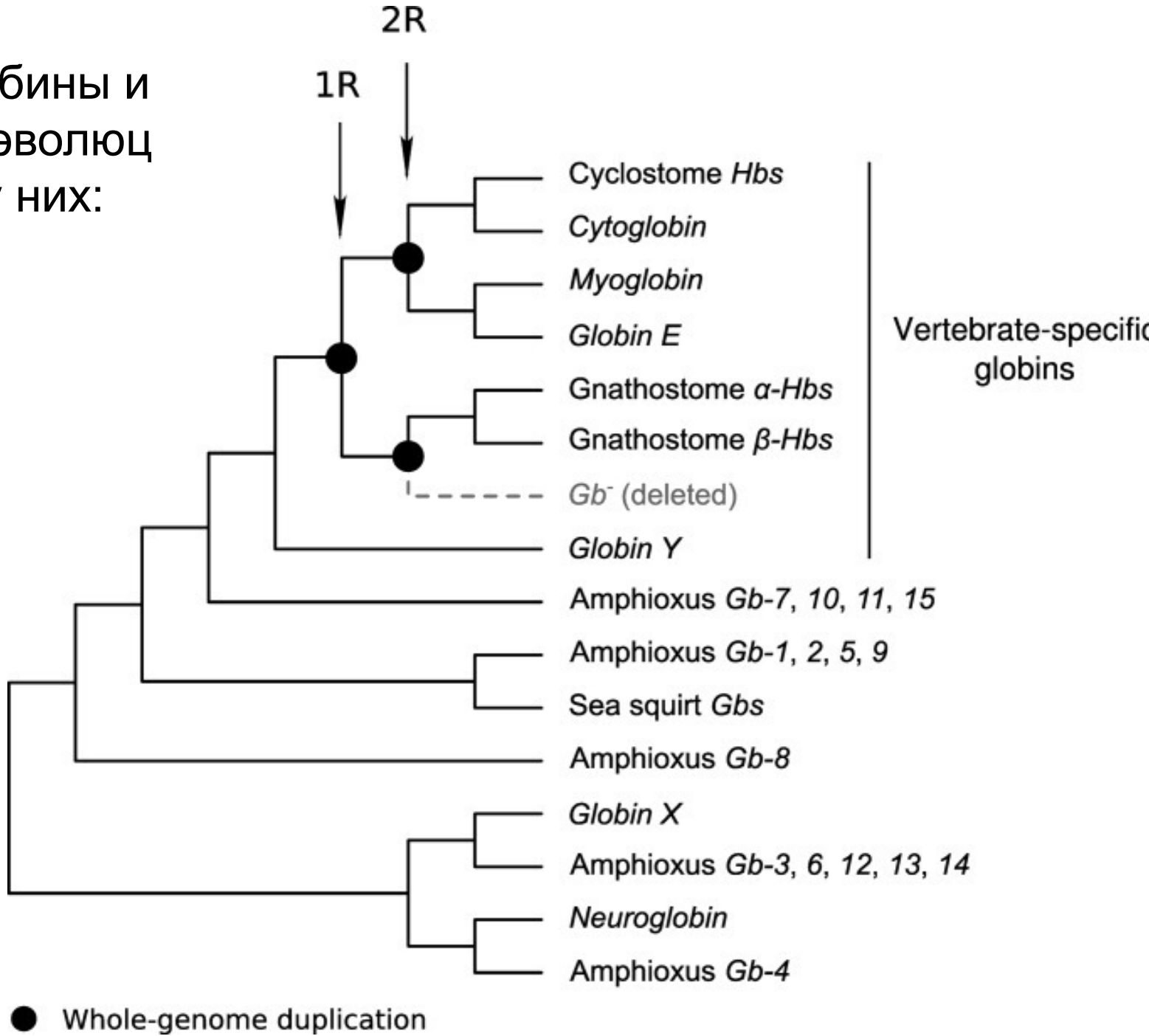
но есть у всех рыб, рептилий (кроме крокодила) и утконоса.

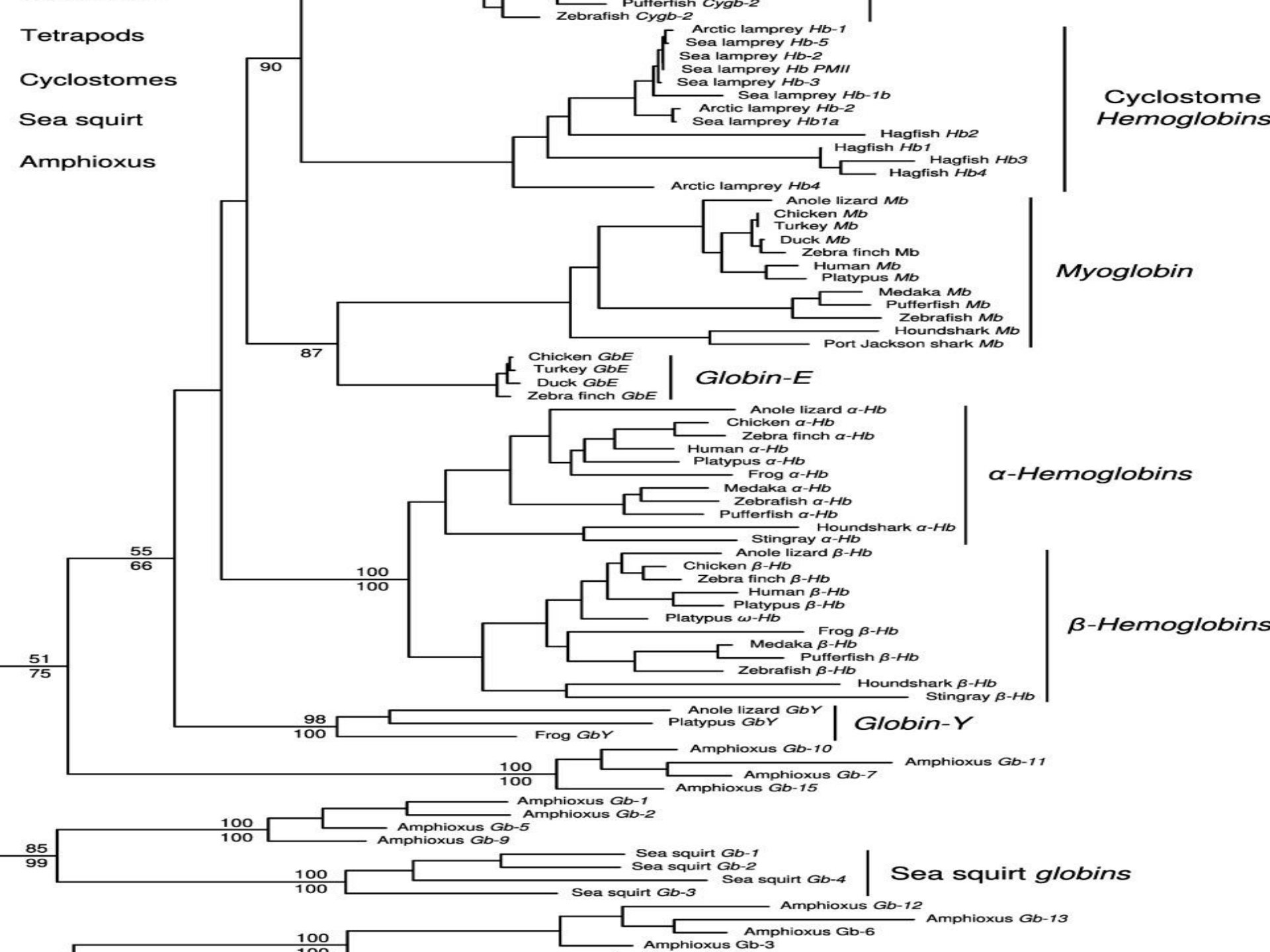
Ниже эволюция семейства глобинов, где R означает

полногеномные дупликации, Gbs – глобин,

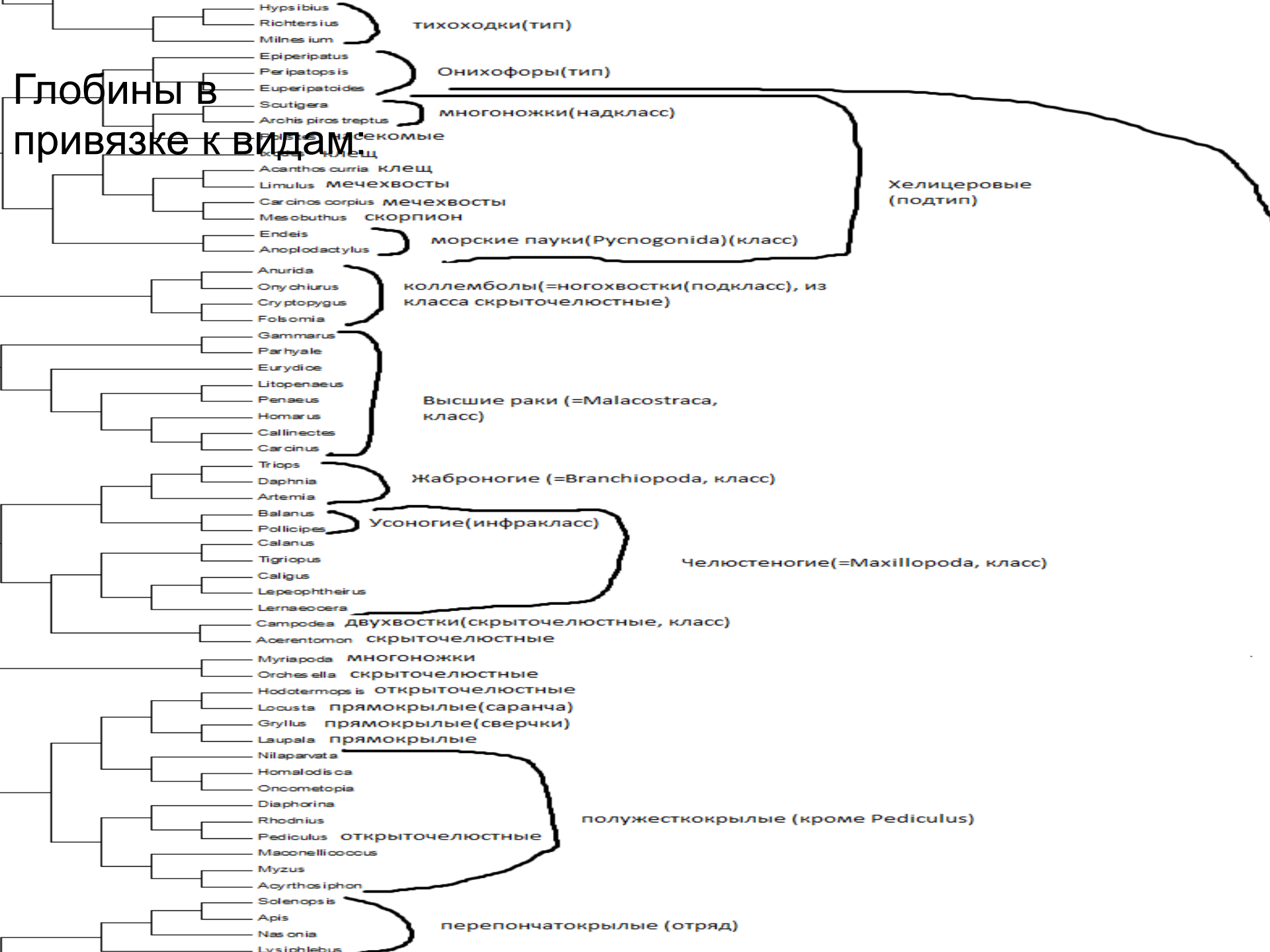
Hb(s) – α - и β -гемоглобины у челюстноротых.

Белки-глобины и
большие эволюц
события у них:





Глобины в привязке к видам



**В эмбриональном развитии
мозга происходит:
слева направо**

**Передний
мозг**

**Конечный
мозг**

**Промежуточный
мозг**

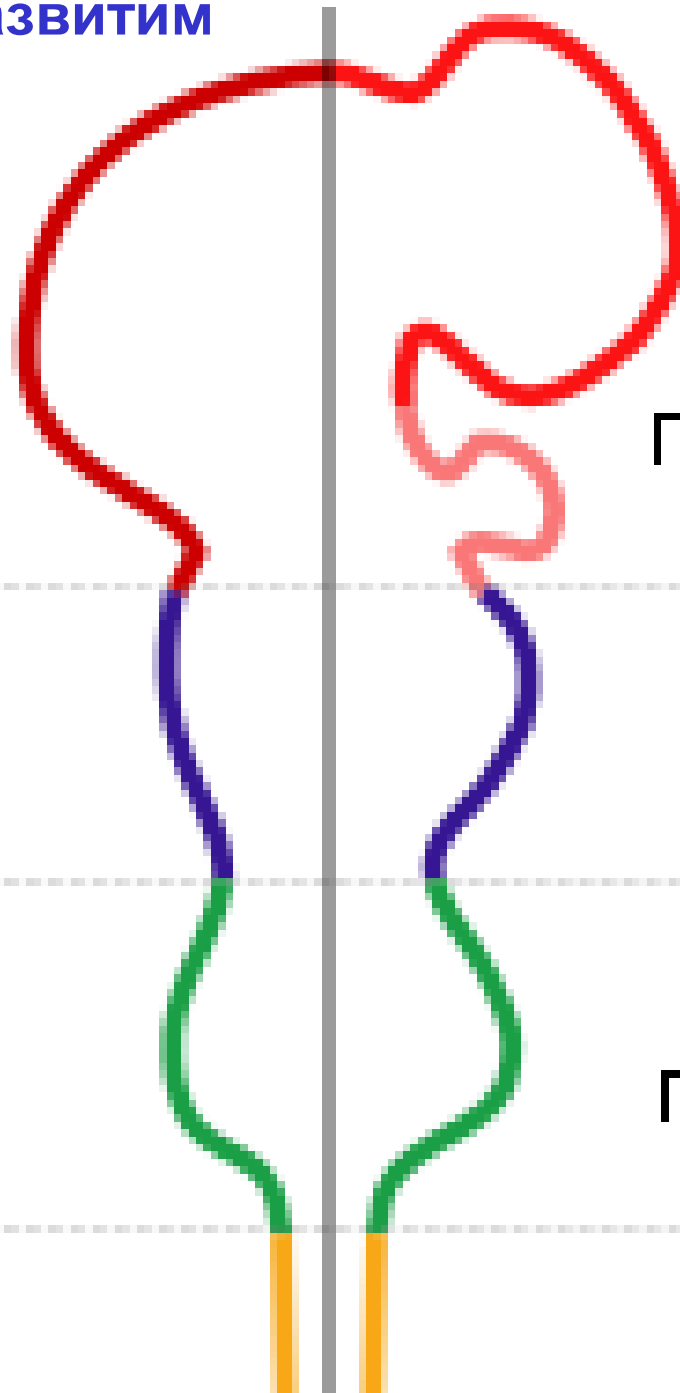
**Средний
мозг**

**Ромбовидный
мозг**

**Задний
мозг**

**Продолговатый
мозг**

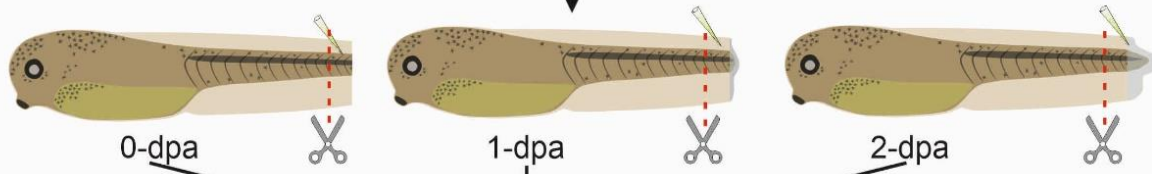
Спинной мозг



The wild-type *Xenopus laevis* tadpole stage 40

ген *ccl34*

Tail amputation ↓ *vivo*-MO injection



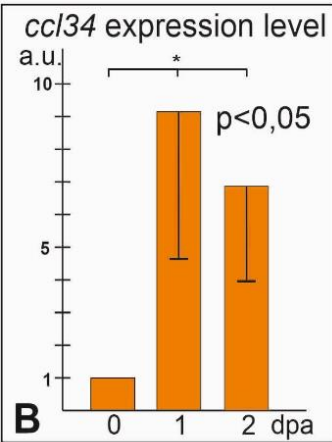
0-dpa

1-dpa

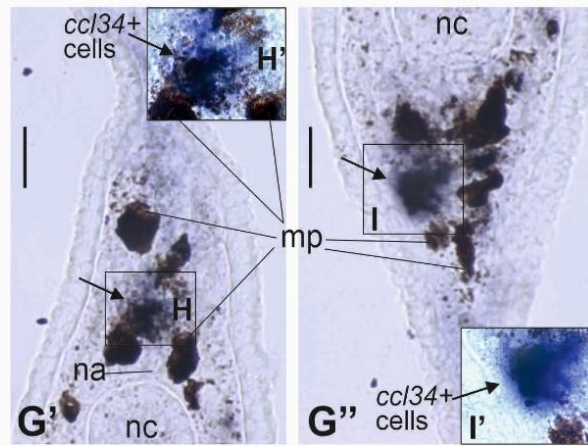
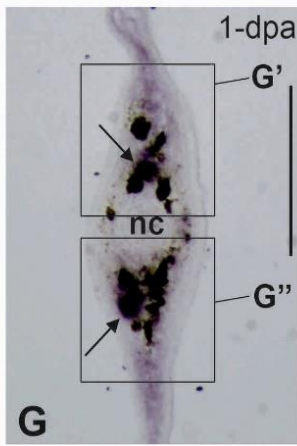
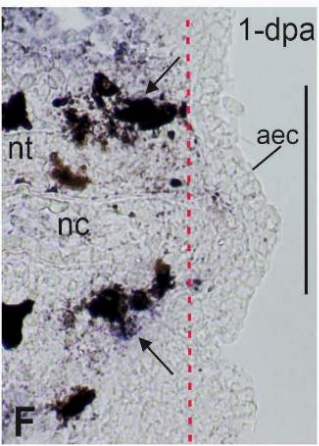
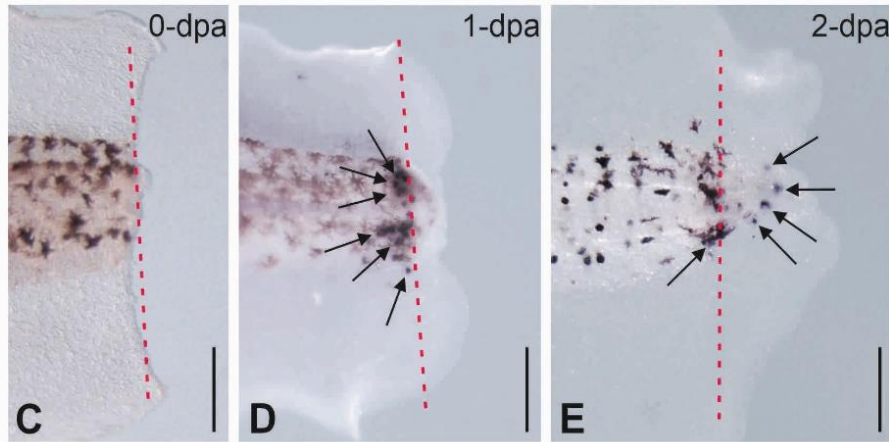
2-dpa

Samples collection for qRT-PCR and *in situ* hybridization

A



B



В. Результат qRT-PCR анализа экспрессии в клетках кончика ампутированного обрубка хвоста.

С–Е. Аналогично *in situ* hybridization. Боковой вид хвостового обрубка. Dorsal частью вверх. Уровень ампутации показан красными точками. Стрелки указывают на компактные группы *ccl34* экспрессирующих клеток.

A. Schema of experiments.

F–G''. Экспрессия в parasagittal (F) and transverse (G–G'') сечения хвостового обрубка hybridized in whole-mount at 1 dpa.

Шкалы: C–G 0.2 mm, G'–G'' 0.05 mm.

The effect of *ccl34* downregulation by injections of *ccl34* vivo MO on the regeneration of the amputated *Xenopus laevis* tadpole tail.

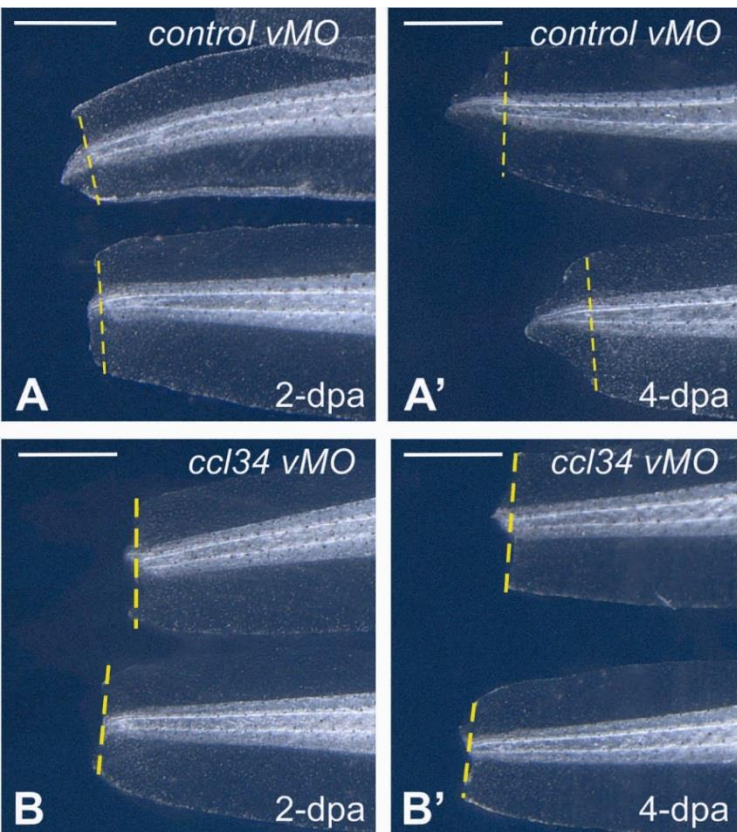
A–B'. Tail stumps injected with either control vivo MO (A–A') or *ccl34* vivo MO (B–B') at 2 and 4 dpa. The level of amputation is shown by yellow line.

C. Percent distribution of three types of the regenerating tail tips (normal regeneration, partial and no regeneration) among the 4 dpa tails injected with the control and *ccl34* vivo MO.

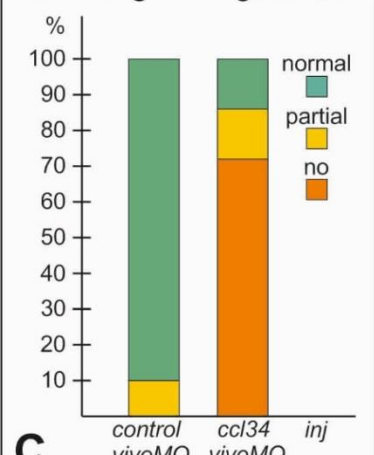
C'. Comparison on 4 dpa of the length of the regenerated tail tip in tails injected with the control and *ccl34* vivo MO.

C'. Comparison on 4 dpa of the length of the regenerated tail tip in tails injected with the control and *ccl34* vivo MO.

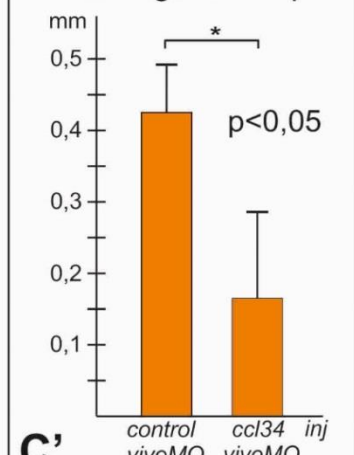
Шкала 0.5 mm.



Percentage of regenerates



Tail length on 2-dpa



since, without its activity, the blastema formation appears to be strongly suppressed. Besides demonstrating proof of principle of the validity of our screening, this finding has a self-own important meaning. First of all, this is stipulated by *ccl34* belonging to chemokine signaling proteins that regulate the migration of a huge number of cell types, in particular, cells involved in blastema formation during regeneration in cold-blooded and warm-blooded animals. During 1st and 2nd days after tail tip amputation, several events of cell migration take place, which could be associated with the enhancement of *ccl34* expression observed in this early period of tail regeneration. Thus, shortly after amputation, there is an intensive migration towards the wound surface of the innate immune system cells responsible for the initial inflammation and further clearing of damaged tissues [Mescher et al. 2013]. Simultaneously, mesenchymal cells, notochord precursors, and myogenic cells begin to accumulate in regions near the amputation plane, adjacent to the edge of the amputated notochord, thus forming a bulk of cells for further proliferation and the regeneration blastema formation [Sugiura et al. 2004]. Notably, the region of accumulation of all these cells near the notochord stump tip approximately corresponds to the location of cell clusters expressing *ccl34* (fig. 6B). Furthermore, the expression pattern of *ccl34* at 1 dpa very resembles that of interleukine-11, which was shown to induce and maintain progenitors of the blastema cell lines (compare fig. 6B in our work with fig. 1d in [Tsujioka et al. 2017]). Thus, one may conclude the expression of *ccl34* overlaps with territory in which the progenitors of the blastema cells are located at this stage. Given these overlaps, as well as the strong inhibitory effect of *ccl34* downregulation on blastema formation, it would be tempting to speculate that this chemokine may regulate at least some of the above cell migration events preceding blastema formation.

Secondly, the potential interest of *ccl34* as the regeneration regulator is due to the fact that according to Atlas of single-cell transcriptomics of the regenerating *Xenopus laevis* tail, *ccl34.S* is specifically expressed in two myeloid cell lines [Aztekin et al. 2019]. These myeloid cell lines, named by authors inflammatory and reparative, respectively, have been shown to play pivotal roles during tail regeneration in *Xenopus laevis* tadpoles [Aztekin et al. 2020]. The authors of this work have supposed that cells of the frog myeloid lines could correspond to M1 and M2 macrophage lines of mammals and known as cells producing different chemokines and thus regulating migration of other cells, including not belonging only to the immune system. In turn, the migration of macrophages themselves is also regulated by chemokine gradients [Arango Duque and Descoteaux 2014; Lacy 2015; Xuan et al. 2015]. Therefore, one may suppose that the cell clusters in which we observed the expression of *ccl34* may be just the macrophage-like cells. Thus, release of Ccl34 could be a part of the mechanisms by which these myeloid cells regulate regeneration.

It is obvious, however, that despite their plausibility all these assumptions should be subjects of careful testing in future studies. At the same time, one can already now argue that the very fact of the loss of such a critical regulator of regeneration as *ccl34* in reptile ancestors further strengthens our hypothesis that the complete loss of some genes could be one of the reasons for the decrease in regenerative potencies in evolution.



Время вдоль пути к корню убывает.