

Молекулярная Зоология, весна 2011

лекция 1

**Введение в курс. Происхождение генного разнообразия:
дубликации, мобильные элементы, альтернативный
сплайсинг.**

Курс Молекулярная Зоология, или, что то же, «Системная зоология» (“System zoology”)

Мюге Николай Сергеевич

Институт Биологии Развития (ИБР РАН)

Институт Рыбного Хозяйства и Океанографии (ВНИРО)

- ▶ mugue@mail.ru
- ▶ 8(916)6877783
- ▶ <http://mugue.narod.ru/MolZool2011.html>



Пропасть между первым и вторым отделением

- ▶ Существует разрыв между современной молекулярной биологией и медициной, которые стремятся «разобрать по винтикам» живую клетку и организм с целью понять «механизм жизни». Основным объектом исследования стал человек, а также другие организмы, на которых можно смоделировать процессы человека в норме и патологии с целью понимания механизмов болезней и создания лекарств.
- ▶ С другой стороны, область «организменной биологии» (зоология, ботаника, экология и прочее «Второе отделение») стремятся реконструировать эволюцию живых организмов, их распределение в пространстве и связи (взаимодействие) организмов со средой и с другими организмами.



Пропасть начинает сужаться

- ▶ Главным образом благодаря прогрессу в области геномики и биологии развития, в настоящее время мы видим колоссальные изменения в подходах при изучении «традиционной» зоологии. Чтобы понять механизм «работы» организма человека, просто знание генома человека (опубликован в 2000г.) оказалось категорически недостаточно. Необходим стал «эволюционный подход» и количество «модельных организмов» стало стремительно расти, включая теперь не только мышь и обезьяну, но и большое количество беспозвоночных. Никого не удивишь теперь исследованием диабета на мухах. Изучение генов, участвующих в морфогенезе теперь начинается от трихоплакса и гидры. Отсеквенированы десятки полных геномов беспозвоночных, и эти данные находятся по большей части в свободном доступе что позволяет **зоологам не только изучать морфологию, но и пытаться понять как эта морфология создается на молекулярно генетическом уровне - как в процессе эмбриогенеза, так и в филогенетическом (эволюционном) аспекте.**
-



Молекулярная филогения все больше подменяет традиционную систематику

- ▶ Обилие и доступность данных по генам (как митохондриальным, так и ядерным) позволяет все с большей достоверностью реконструировать филогению.
- ▶ Однако объект (организм) должен быть первичным, все больше появляется работ, когда организм отсеквенировали, а определить забыли.
- ▶ С другой стороны, пренебрегать данными молекулярной филогении непрактично.
- ▶ **ВЫВОД: Хороший молекулярный систематик должен быть в первую очередь хорошим зоологом!**



Три ключевых вопроса биологии

- ▶ Блок проблем 1. Как из одной клетки образуется сложный организм? (эмбриология)
- ▶ Блок проблем 2. Как из одного вида образуется много других видов? (эволюция)
- ▶ Блок проблем 3. Как связано сознание (мышление, память, творчество) и структура мозга?
- ▶ Список крайне субъективен. Кто хочет его дополнить?
- ▶ На первые 2 вопроса будем пытаться отвечать.



Структура курса

- ▶ Лекция 1. Введение в курс. Происхождение генного разнообразия: дупликации, мобильные элементы, альтернативный сплайсинг.
 - ▶ Лекция 2. Гены видообразования. Генетические механизмы участвующие в формировании новых видов
 - ▶ Лекция 3. Геногеография (филогеография). Реконструкция расселений и формирования фаун путем сравнения генов.
 - ▶ Лекция 4. От генотипа к фенотипу: регуляция на уровне транскрипции и трансляции, белок-ДНК и белок-белок взаимодействие.
 - ▶ Лекция 5. Организация биохимических каскадов и их регуляция.
-



Структура курса (часть 2, Evo-Devo)

- ▶ Лекция 6 и 7. Происхождение многоклеточных. От двухслойных к трехслойным. Нох-гены и эволюция членистоногих. Формирование продольной и дорзовентральной осей.
- ▶ Лекция 8. От генотипа к фенотипу (и обратно).
- ▶ Лекция 9. Механизмы адаптаций на молекулярном уровне
- ▶ . Зоология в «пост-геномную» эру. Анализ экспрессии генов на микрочипах, Новые технологии в секвенировании - платформа 454, Solid и Illumina. Протеомика.
- ▶ Лекция 10. Заключительная. Обзор современного «инструментария» зоолога и что нас ждет в ближайшие 5-10 лет.



Уровни организации

- ▶ Гены (ДНК, Геном)



- ▶ Информационная РНК (Экспрессом)



- ▶ Белки (Протеомика)



- ▶ Белок-белковые и белок-ДНК взаимодействия



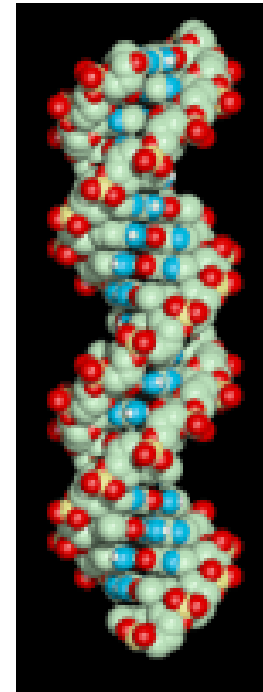
- ▶ Фенотип (морфология, физиология, гомеостаз, адаптации)



ДНК – самая сложная структура в живых организмах!

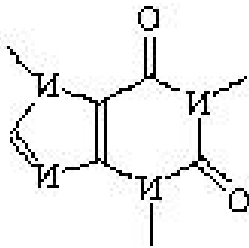


VS.

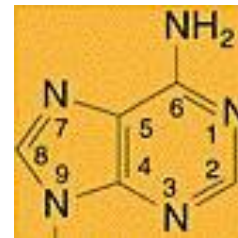


С другой стороны, ДНК по составу мало отличается от бутылки Кока-Колы!

Coke	DNA	Solubility
water	water	-
Sugar (sucrose)	Sugar (dexoyribose)	VERY High
Phosphate (PO ₄ ⁻ acid)	phosphate	moderate
caffeine	bases (A,T,C,G)	extremely low



caffeine



Adenine





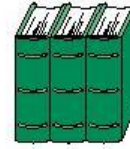
Phage λ
50 kb
2 pages



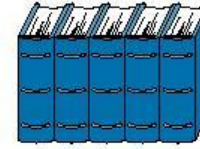
Escherichia coli
(bacteria)
4.7 Mb
200 pages



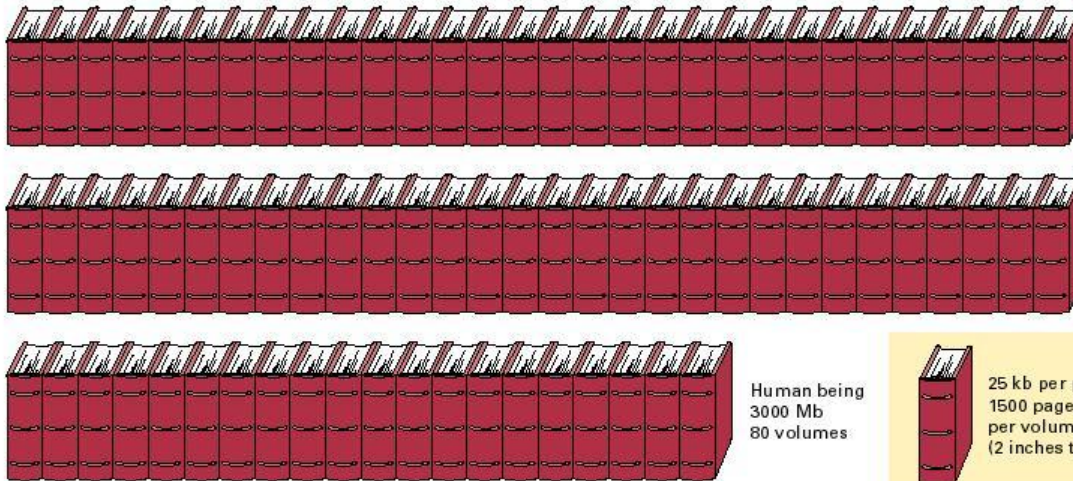
Saccharomyces cerevisiae
(yeast)
12.5 Mb
500 pages



Caenorhabditis elegans
(nematode)
Arabidopsis thaliana
(plant)
100 Mb
3 volumes



Drosophila melanogaster
(fruit fly)
165 Mb
5 volumes



Human being
3000 Mb
80 volumes



25 kb per page
1500 pages
per volume
(2 inches thick)

У мужчин на два тома меньше...

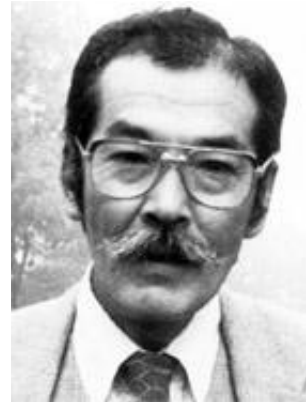
Но многие организмы имеют геном еще
больше!



- ▶ Может ли обезьяна, бьющая по клавишам компьютера, напечатать сонет Шекспира?

А может ли функциональный ген возникнуть из случайного набора нуклеотидов?

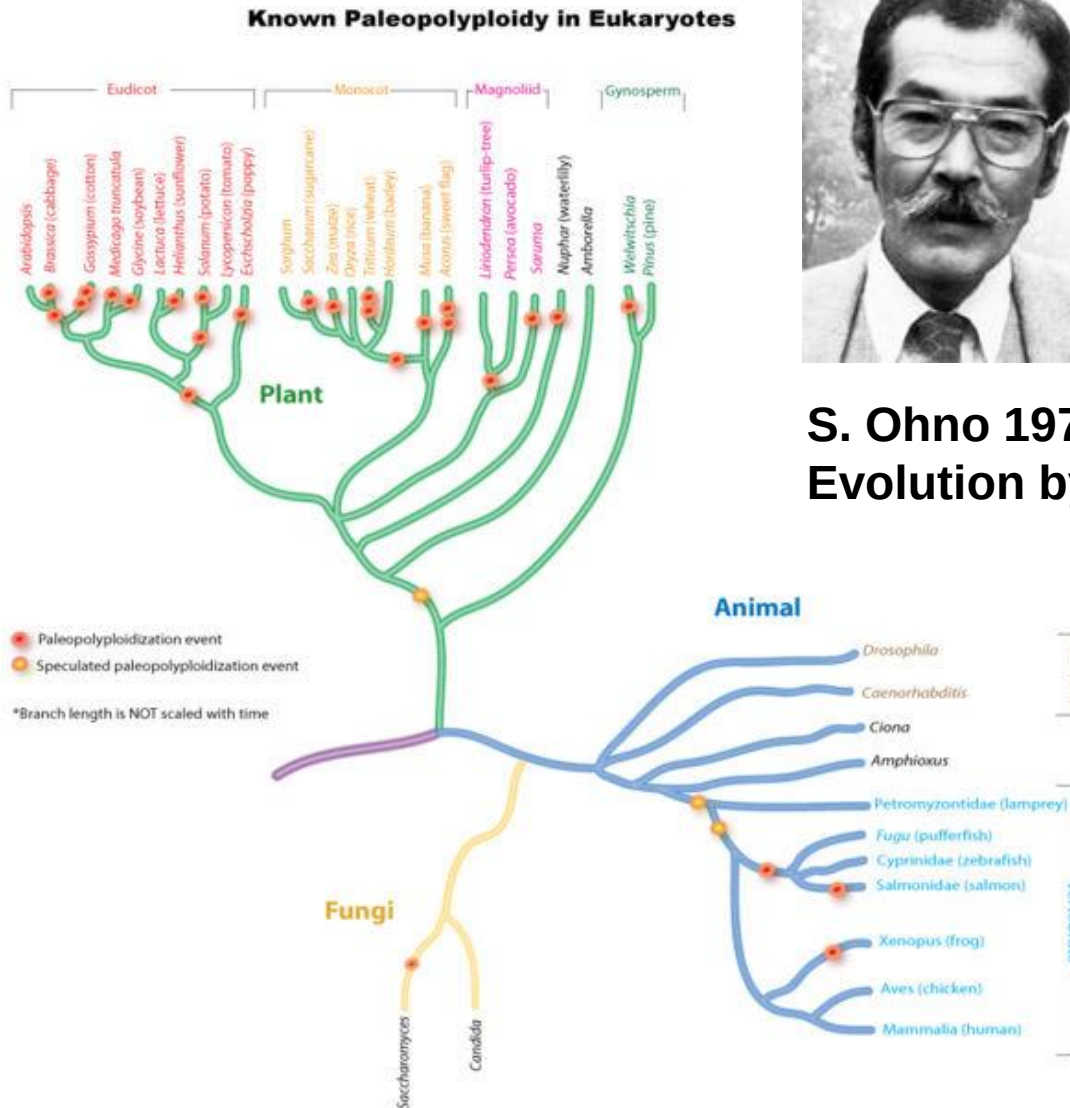
Роль дупликации генов и геномов. Полиплоидия у многоклеточных



Susumu Ohno

Susumu Ohno
Feb. 1, 1928 - Jan 13, 2000

S. Ohno 1970
Evolution by gene duplication



ТЕРМИНОЛОГИЯ

- ▶ Аутополиплоид – удвоение генома одного вида
 - ▶ Аллотетраплоид (Амфитетраплоид) –слияние геномов двух близких видов
 - ▶ Гомологичные последовательности называют *ортологичными*, если к их разделению привел акт видообразования: если ген существует у некоего вида, который дивергирует с образованием двух видов, то копии этого гена у дочерних видов называются ***ортологами***.
 - ▶ Гомологичные последовательности называют *паралогичными*, если к их разделению привело удвоение гена: если в пределах одного организма в результате хромосомной мутации произошло удвоение гена, то его копии называют ***паралогами***.
-



Странные свойства ЛДГ у вьюна из холодильника.

Держали вьюна (*Misgurnus fossilis*) в холодильнике и при комнатной температуре по 2-3 недели

Выделяли ЛДГ-А из мышц

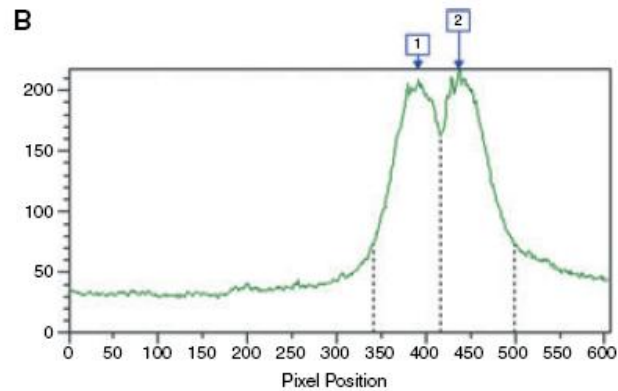
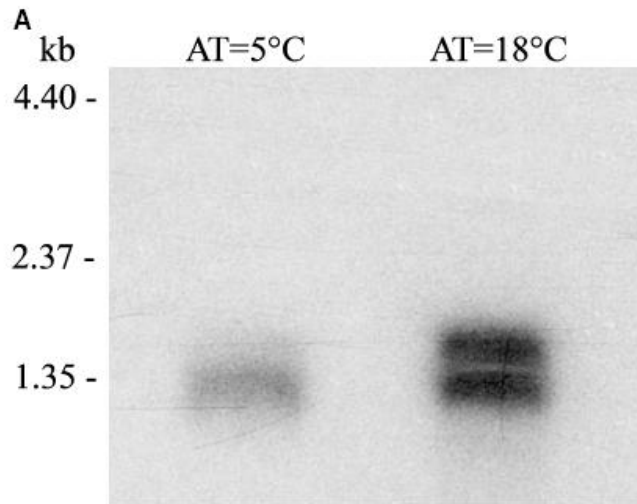
Различия между «теплой» и «холодной» ЛДГ:

Различная активность,

Термостабильность,

денатурация мочевиной и т.д.





Northern hybridization – две формы mRNA

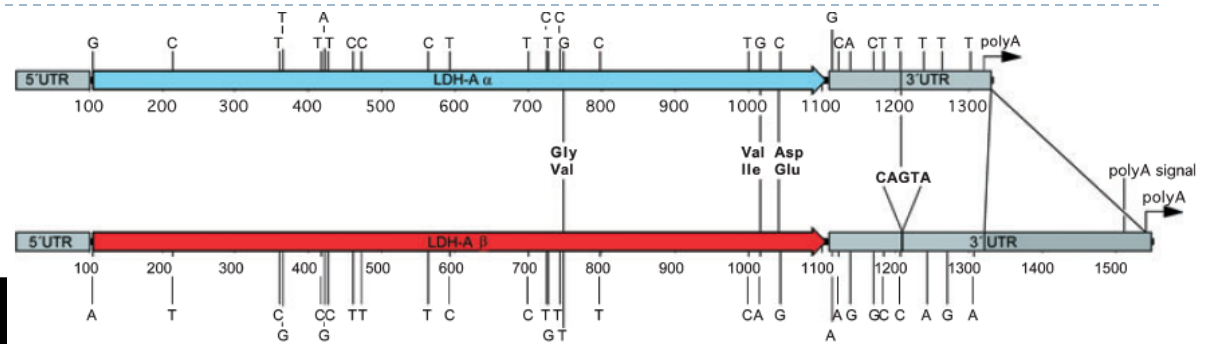


Fig. 2. Structure of the short ($mRNA_{ldh-a}^s = 1332$ bp) and long ($mRNA_{ldh-a}^l = 1550$ bp) forms of LDH-A mRNAs from skeletal muscle of weatherfish *Misgurnus fossilis*. Nucleotide mismatches are indicated outwards, whereas amino acid mismatches are indicated inwards.

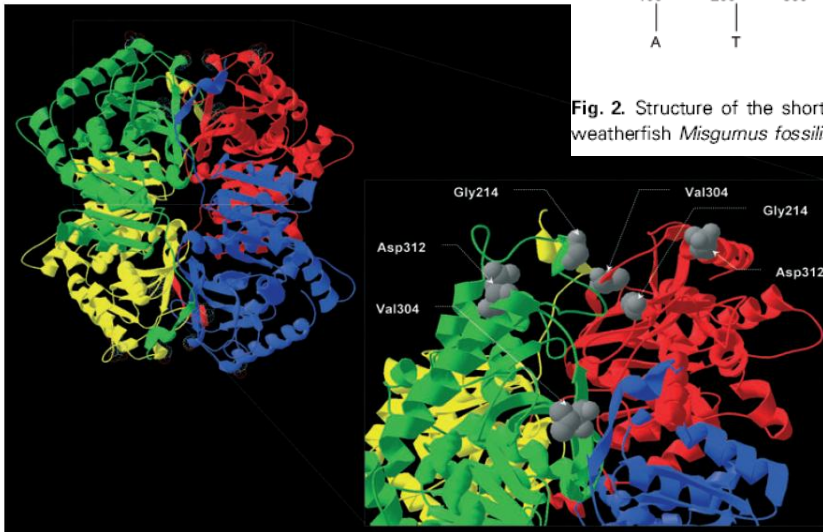
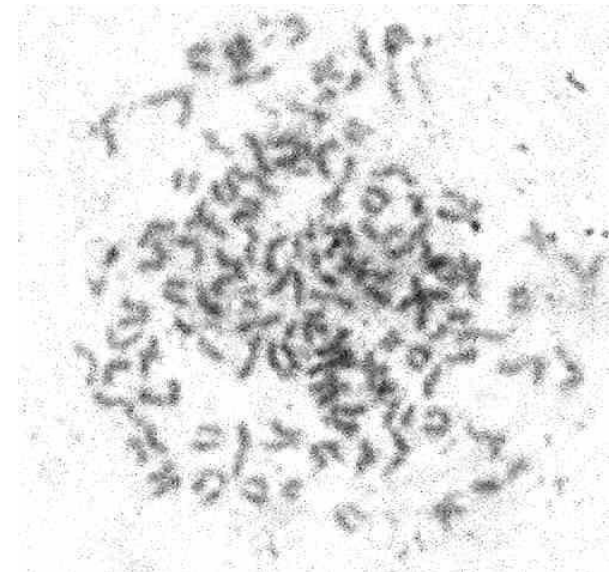
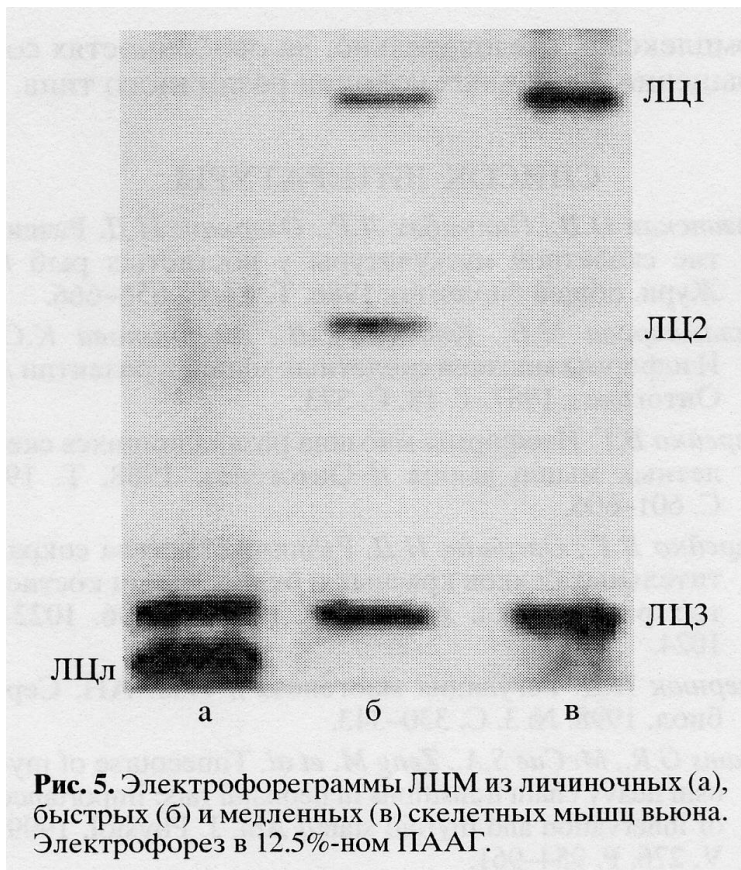
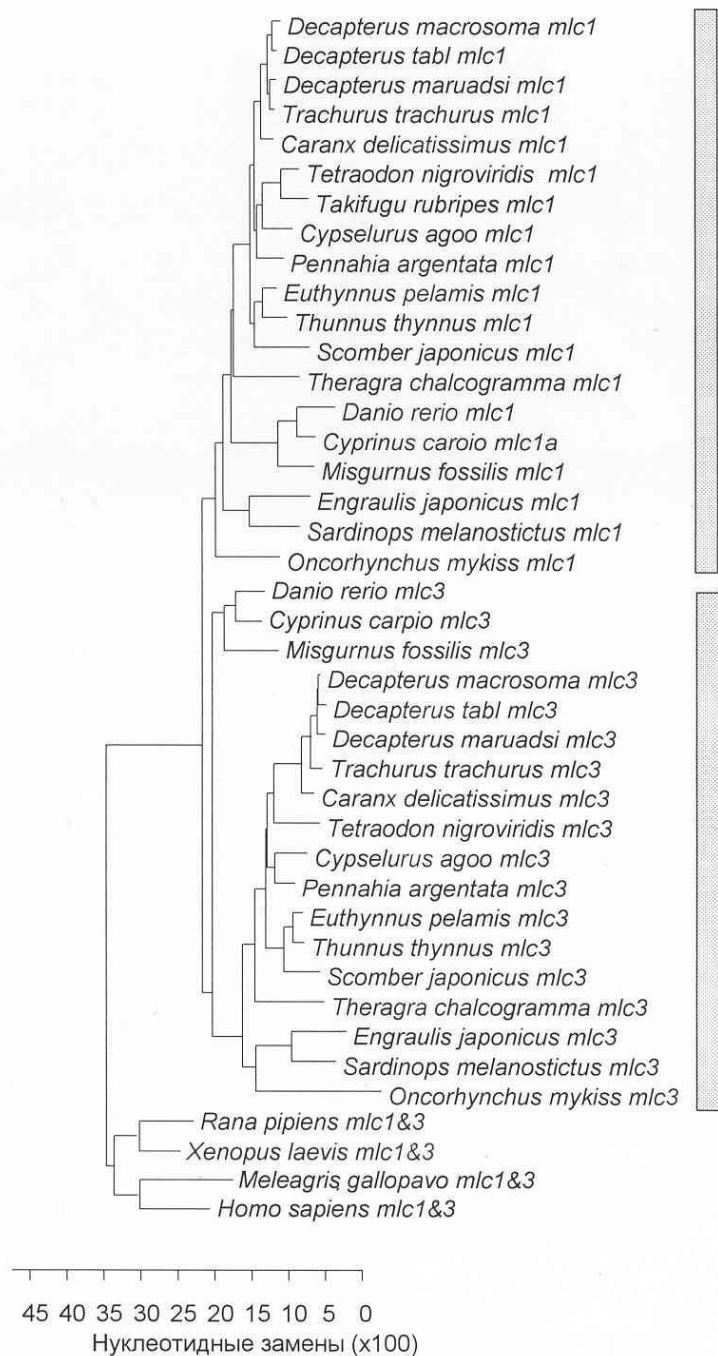


Fig. 3. Predicted quaternary structure of skeletal muscle LDH-A α_4 -homotetramer (LDH-A*) from weatherfish *Misgurnus fossilis* and close up view on the contact area between two neighbouring subunits. Each subunit is coloured and the corresponding mismatched amino acids are indicated.

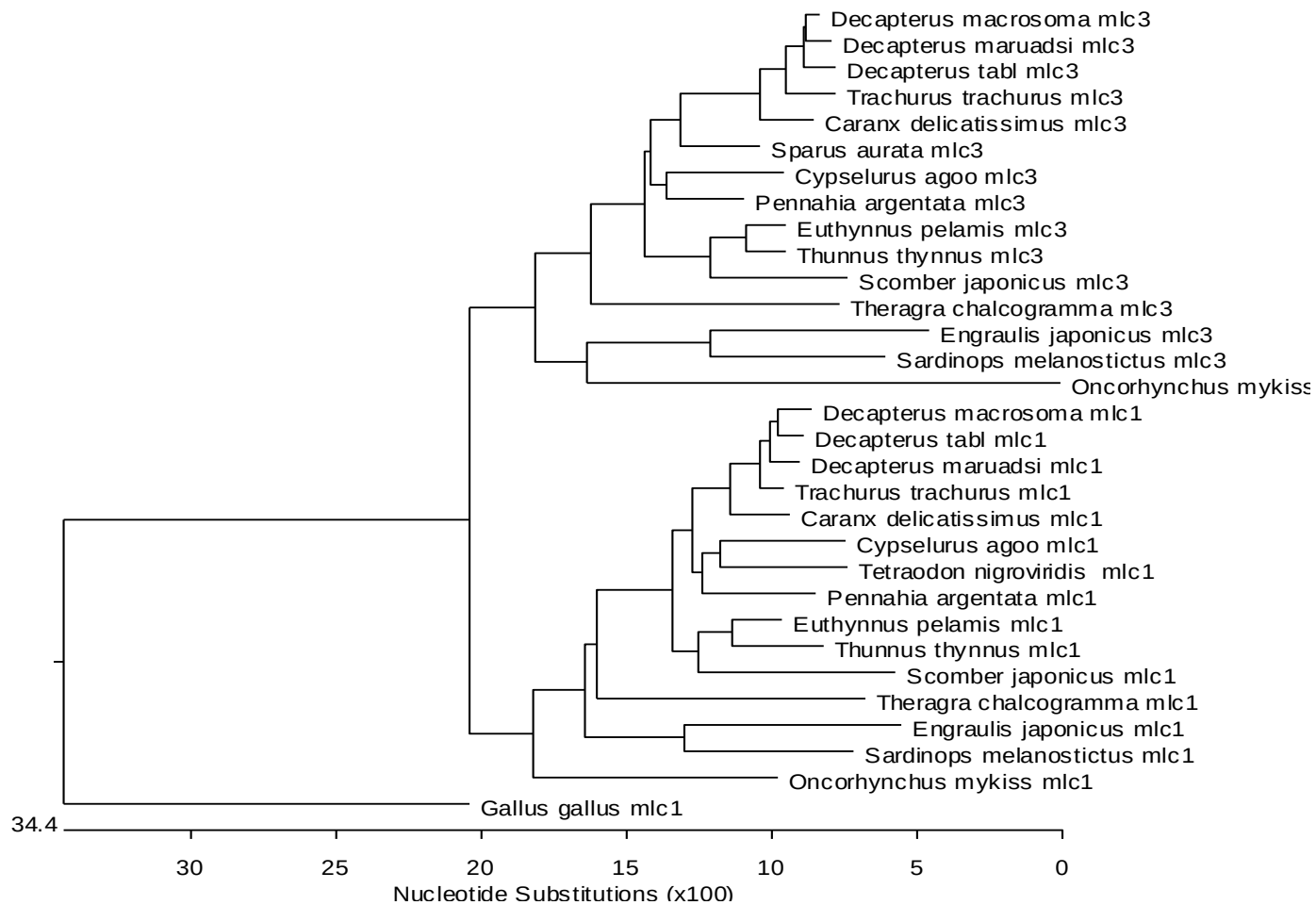


Пример 2. В поисках «эмбриональной» легкой цепи миозина





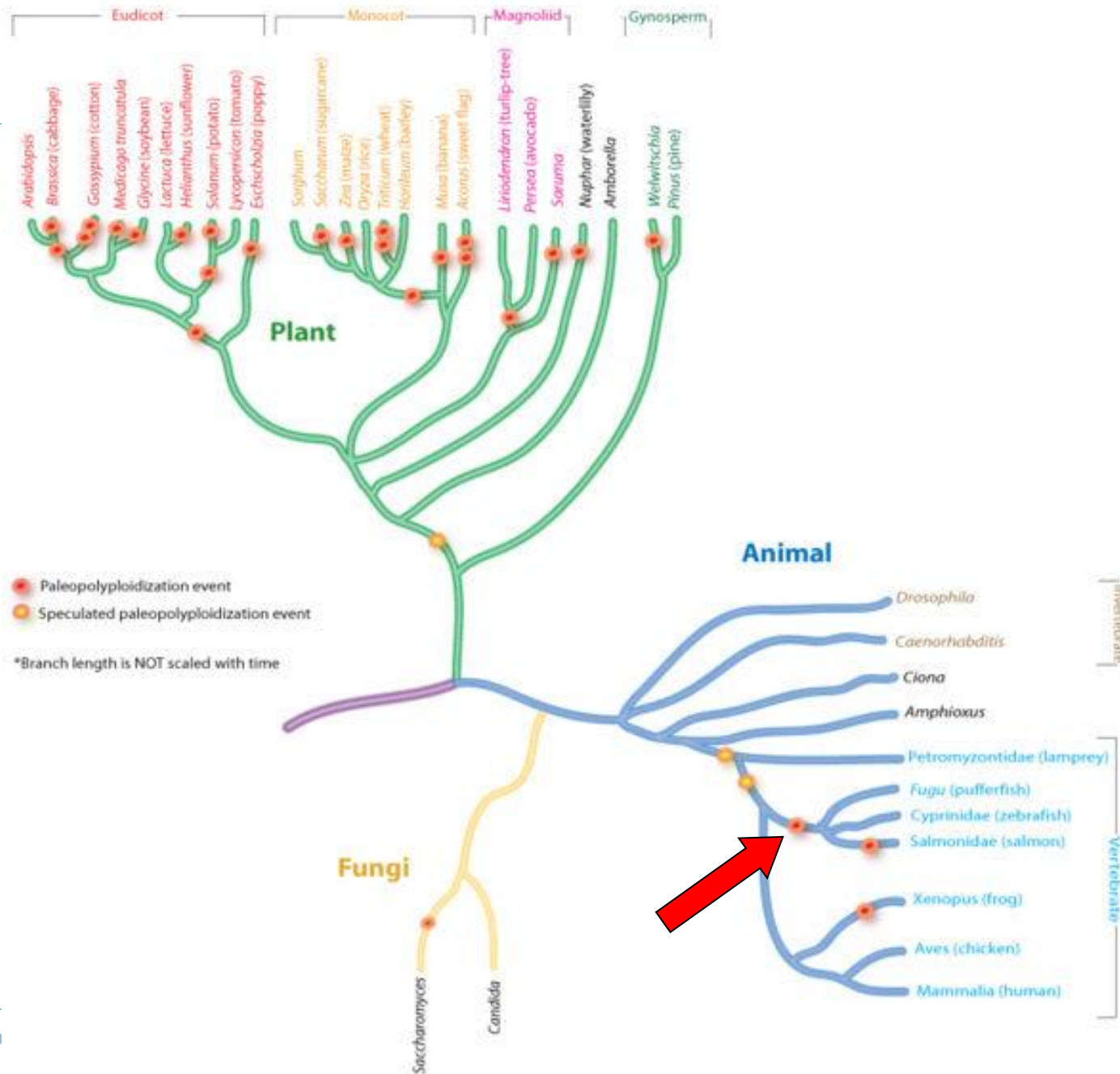
М-1 и ЛЦМ-3 стистых рыб



лцм3

лцм1

Known Paleopolyploidy in Eukaryotes



Genome Duplication, a Trait Shared by 22,000 Species of Ray-Finned Fish

John S. Taylor,^{1,2} Ingo Braasch,¹ Tancred Frickey,¹ Axel Meyer,^{1,4} and Yves Van de Peer³

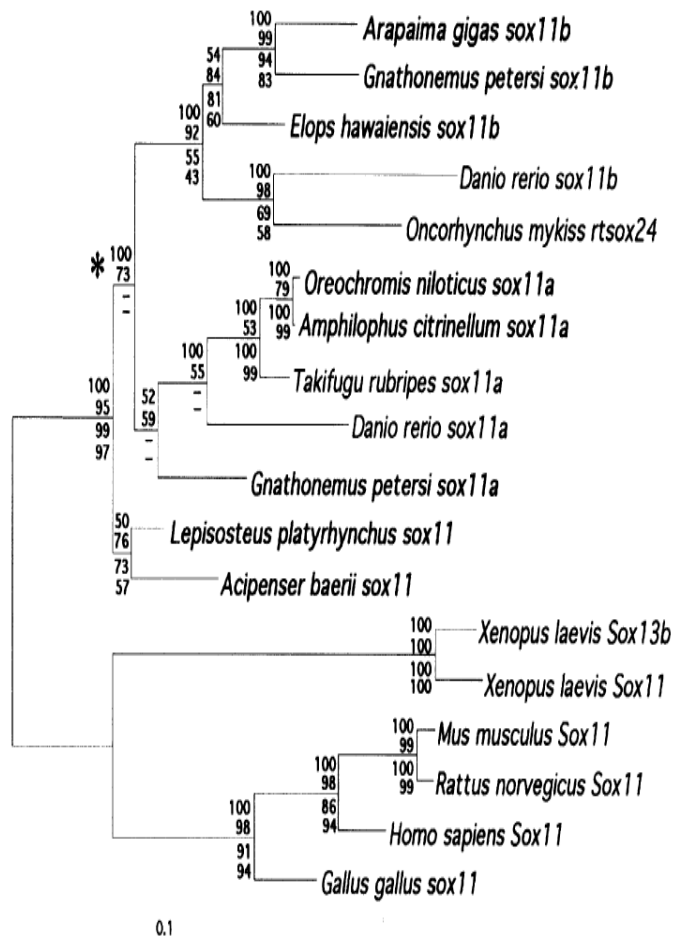
¹Department of Biology, University of Konstanz, D-78457, Konstanz, Germany; ²Biology Department, University of Victoria, Victoria, BC, V8W 3N5 Canada; ³Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology (VIB), Ghent University, K.L. Ledeganckstraat 35, B-9000 Ghent, Belgium

Table 1. Duplicated Fish Genes and Human Orthologs

Human Query Sequences	Each Column Lists Members of a Clade That is Orthologous to the Human Query Sequence			NJ	QP	AS
ATP1B1 (4502277)	Danio 9789577	Danio 11096273, Anguilla 1703468, Anguilla 7406523, Takifugu JGI22524	+	—	+	
ATP1B2 (4502279)	Danio 9789579	Danio 14150727, Takifugu JGI789	+	+	+	
ATP1B3 (4502278)	Danio 974774, Anguilla 7406521	Danio 9837579, Takifugu JGI9802	+	+	+	
BMP2 (4557369)	Danio 2149148	Danio 2804175, Takifugu JGI7838	+	+	+	
CYP19 (13904858)	Danio 12655890, Carassius 3913347, Oreochromis 4838530, Oreochromis 4838536, Takifugu JGI6225, Pimephales 14041612	Danio 12655892, Carassius 2662332, Oreochromis 3913346, Oreochromis 4838538, Takifugu JGI22275, Ictalurus 3913357, Oncorhynchus 228574, Oryzias 3913355, Dicentrarchus 14589321, Hippoglossus 13620178, Paralichthys 4239990	+	+	+	
DLL (10518497)	Danio 2809389, Takifugu JGI3940	Danio 1888392, Takifugu JGI18204	+	+	—	
DLX2 (4758168)	Danio 2842748	Danio 108243, Takifugu JGI119697	+	+	+	
DLX4 (4503343)	Danio 2842751	Danio 2842750	+	+	+	
EFNA5 (4503487)	Danio 2494365, Takifugu JGI4301	Danio 2462953, Takifugu JGI34618	—	—	—	
EN1 (7710119)	Danio 4322044, Takifugu JGI28510	Danio 417127, Takifugu JGI32850	—	—	—	
EN2 (11422302)	Danio 417128	Danio 417129, Takifugu JGI7515	—	—	—	
EPHB4 (4758290)	Danio 3163942, Takifugu JGI26074	Danio 3005901, Takifugu JGI17145	+	+	+	
FKH1 (4503735)	Danio 12004940	Danio 12004938, Takifugu JGI9390	—	—	—	
FKH5 (8134472)	Danio 2982347, Takifugu JGI20315	Danio 2982343, Takifugu JGI3282	—	—	+	
FLOT1 (5031699)	Danio 12751185, Carassius 2190561, Takifugu JGI8518	Danio 12751187, Carassius 12751189, Takifugu JGI3374	—	+	+	

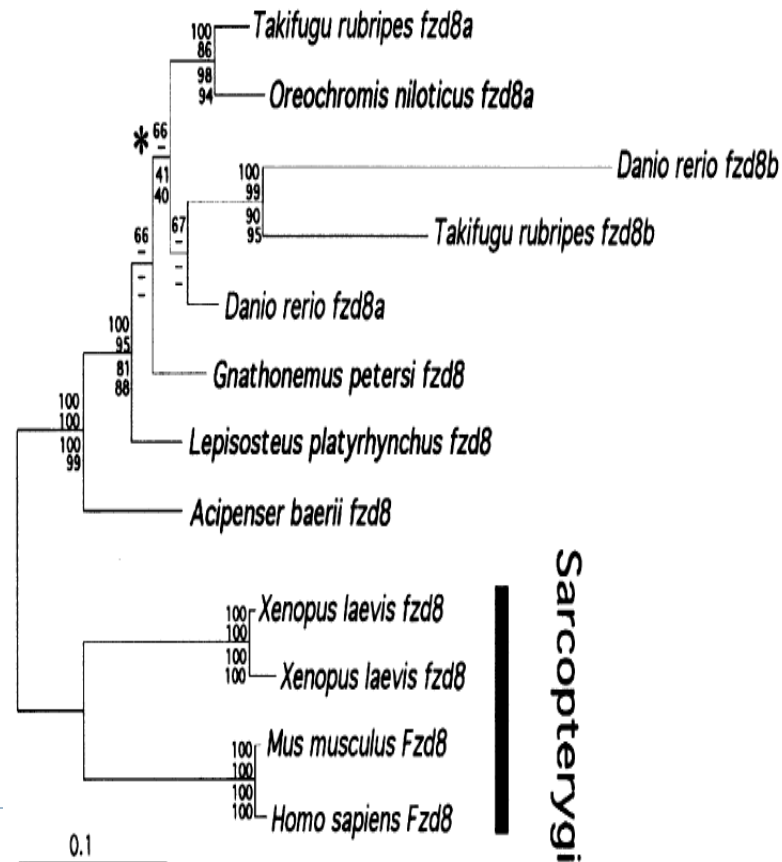
Phylogenetic Timing of the Fish-Specific Genome Duplication Correlates with the Diversification of Teleost Fish

Simone Hoegg, Henner Brinkmann,* John S. Taylor,† Axel Meyer



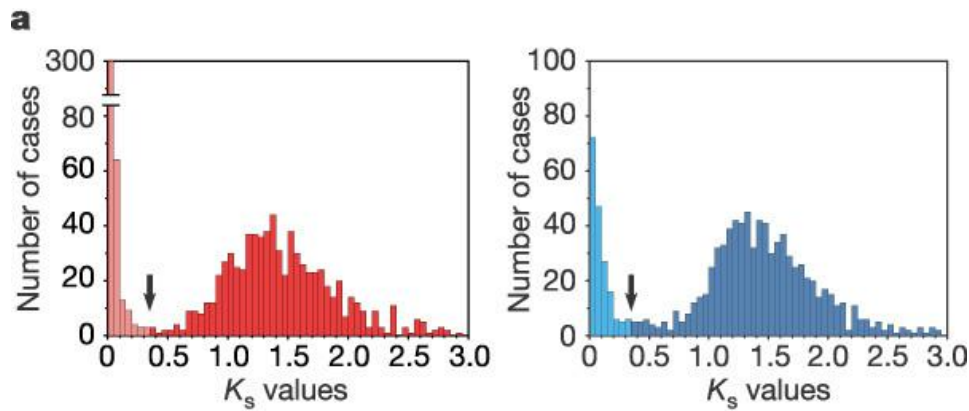
Actinopterygii

Sarcopterygii

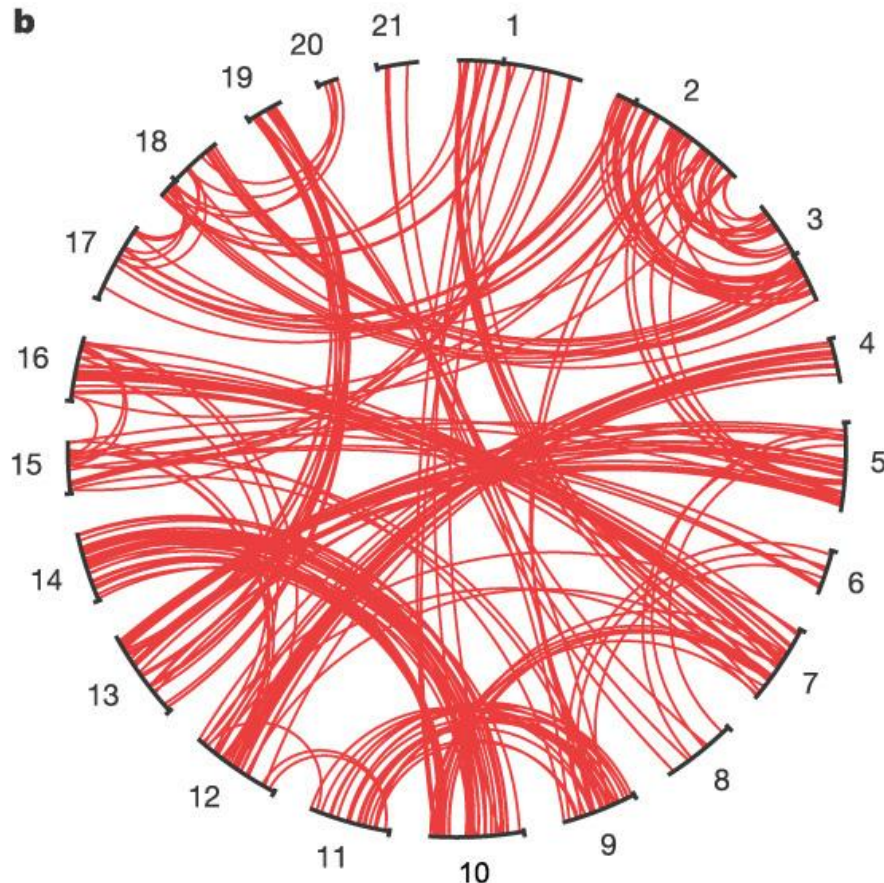


Actinopterygii

Sarcopterygii



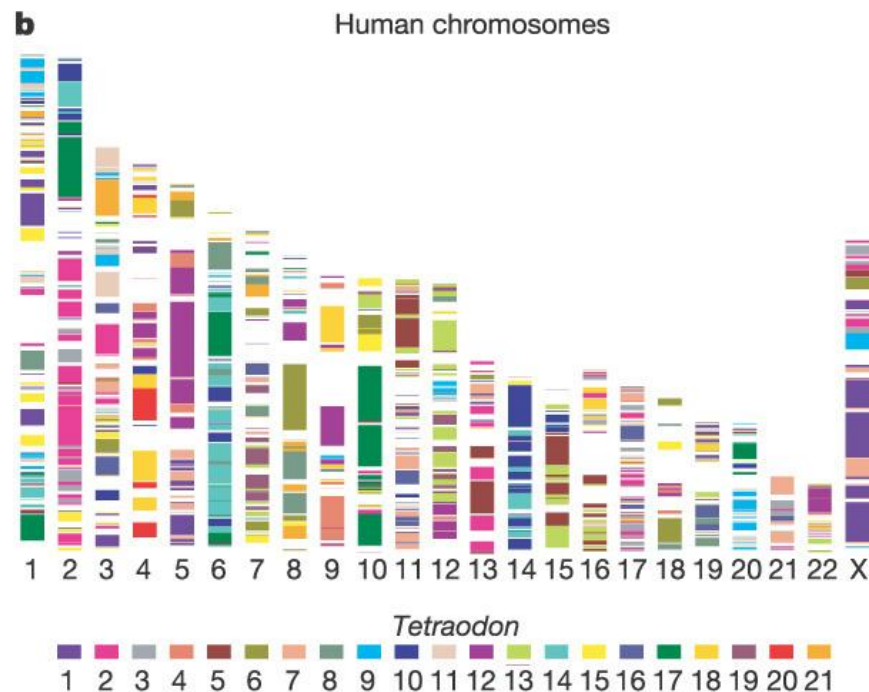
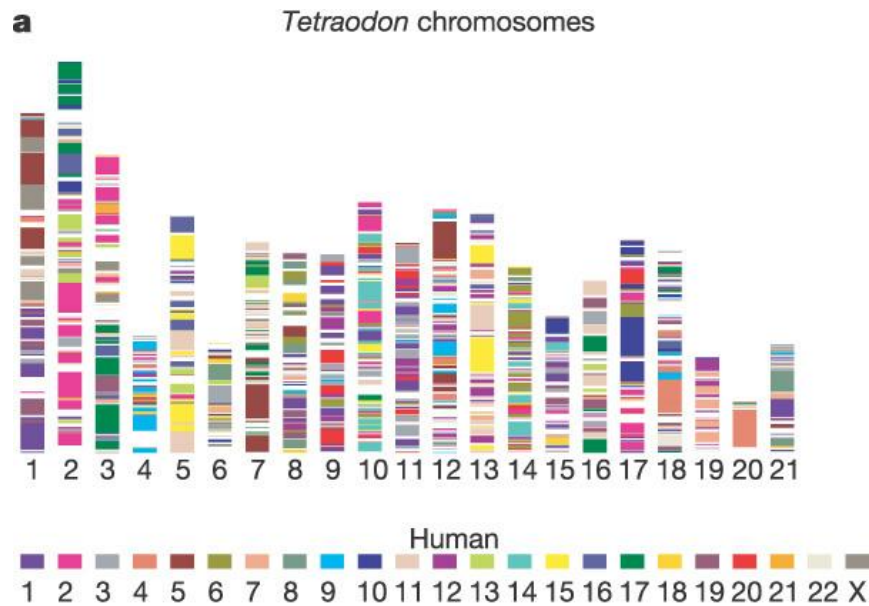
a, Distribution of K_s values of duplicated genes in *Tetraodon* (left) and *Takifugu* (right) genomes. Duplicated genes broadly belong to two categories, depending on their K_s value being below or higher than 0.35 substitutions per site since the divergence between the two puffer fish (arrows).



b, Global distribution of ancient duplicated genes ($K_s > 0.35$) in the *Tetraodon* genome. The 21 *Tetraodon* chromosomes are represented in a circle in numerical order and each line joins duplicated genes at their respective position on a given pair of chromosomes.

Jaillon et al., Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype

Nature **431**, 946-957 (21 October 2004) |



a, For each *Tetraodon* chromosome, coloured segments represent conserved synteny with a particular human chromosome. Synteny is defined as groups of two or more *Tetraodon* genes that possess an orthologue on the same human chromosome, irrespective of orientation or order. *Tetraodon* chromosomes are not in descending order by size because of unequal sequence coverage. The entire map includes 5,518 orthologues in 900 syntenic segments.

b, On the human genome the map is composed of 905 syntenic segments. See Supplementary Information for the synteny map between *Tetraodon* and mouse (Supplementary Fig. S11).

Jaillon et al., Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype

Nature **431**, 946-957 (21 October 2004) |

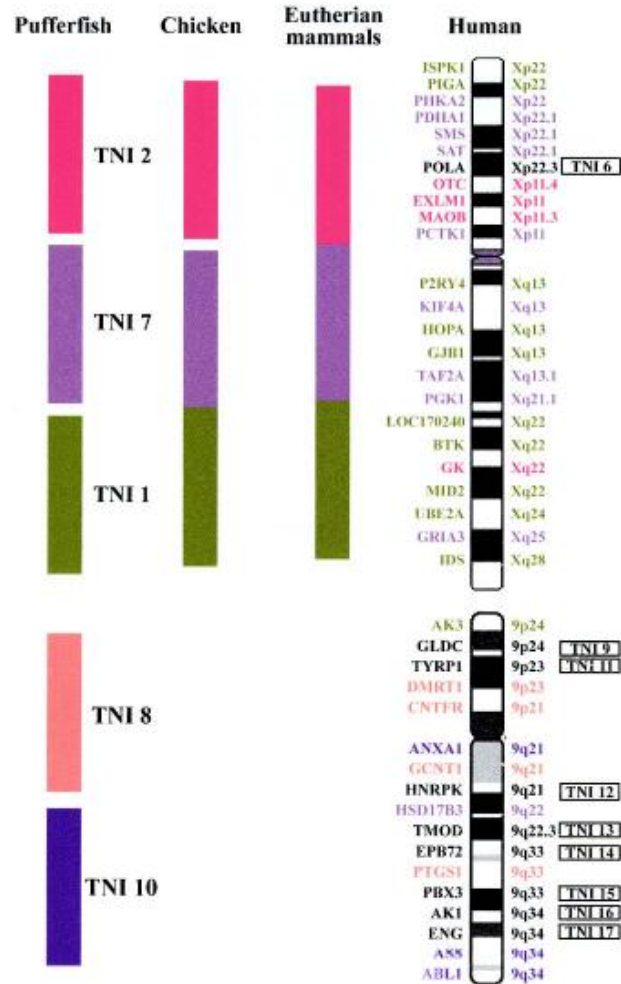


Figure 1 Syntenic relationships of human chromosomes X and 9 with the *Tetraodon* genome and an evolutionary model for the mammalian X. The top shows the common phylogenetic origin of pufferfish chromosomes TNI 1, TNI 2, and TNI 7 and the human X. At left, the three *Tetraodon* blocks of conserved synteny are depicted as green (TNI 1), red (TNI 2), and purple (TNI 7) bars. At right is shown the mosaic synteny of human X and the chromosomal localizations of the comparatively mapped genes. The gene order is based on genomic sequence data (<http://www.genome.ucsc.edu/cgi-bin/hgGateway> December 2001). The *POLA* gene, indicated in black, maps to a different *Tetraodon* chromosome (TNI 6) and, thus, does not belong to the delineated synteny groups. The two central schematic drawings represent hypothetical ancestral X chromosomes. Because human X orthologous genes from TNI 1 and TNI 7 are linked in chicken, we conclude that an ancestral X autosome contained these two sets of Xp and Xq genes even before the avian-mammalian split. Because the human X orthologs belonging to the TNI 2 block are autosomal in marsupials and monotremes, they represent a recent addition to the eutherian sex chromosomes. For simplicity, the three X syntenic *Tetraodon* chromosomes are depicted as blocks on the hypothetical X ancestors. However, this does not reflect the real gene order, which may differ significantly from both the order in the three *Tetraodon* chromosomes and in the human X. At bottom is shown a much lower degree of synteny conservation for human chromosome 9 in the pufferfish than for the X. Comparative mapping of human 9 orthologous genes reveals two syntenic blocks on TNI 8 (orange) and TNI 10 (blue). One gene each reside on TNI 1 (green) and TNI 7 (purple), respectively. The genes indicated in black all map to different TNI chromosomes.

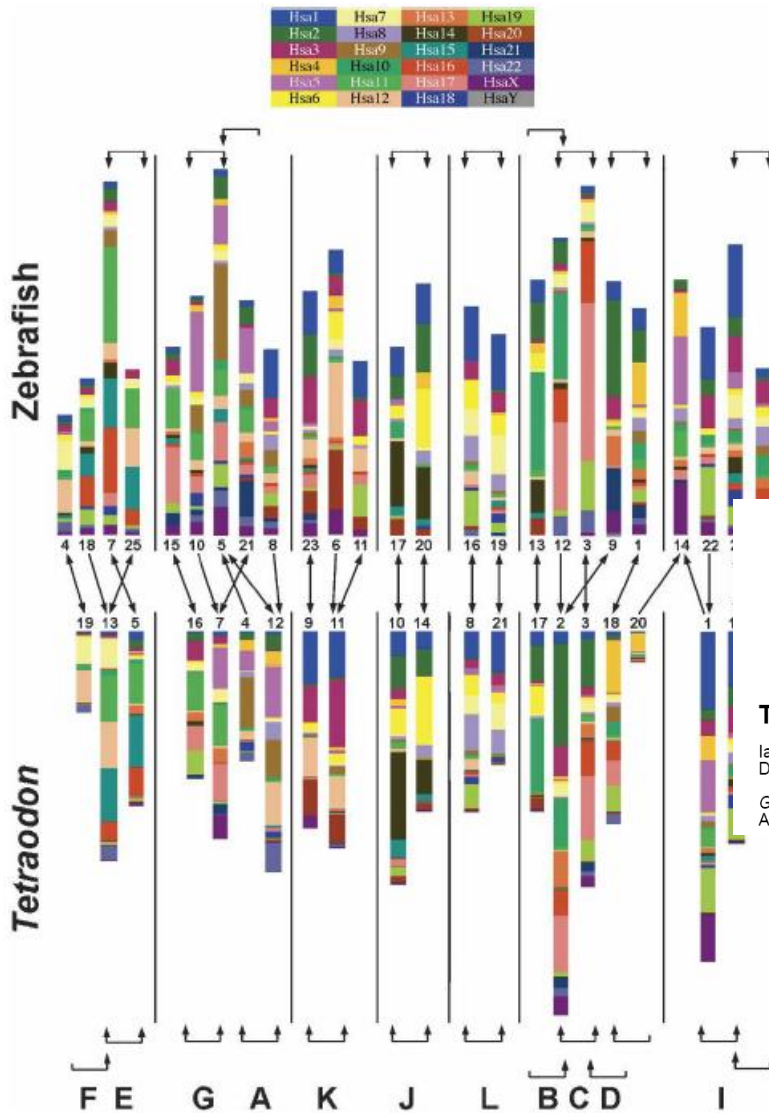


Figure 5. Reconstruction of ancestral vertebrate proto-chromosomes by comparison of zebrafish and *Tetraodon* gene maps. The 25 zebrafish

GENOME RESEARCH

The zebrafish gene map defines ancestral vertebrate chromosomes

Ian G. Woods, Catherine Wilson, Brian Friedlander, Patricia Chang, Daengnoy K. Reyes, Rebecca Nix, Peter D. Kelly, Felicia Chu, John H. Postlethwait and William S. Talbot

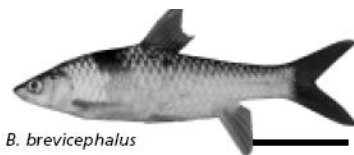
Genome Res. 2005 15: 1307-1314; originally published online Aug 18, 2005;
Access the most recent version at doi:[10.1101/gr.4134305](https://doi.org/10.1101/gr.4134305)

Полиплоидия и видообразование





B. acutirostris



B. brevicephalus



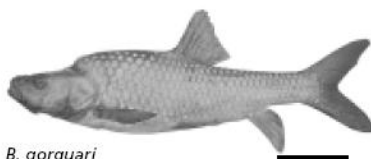
B. crassibarbis



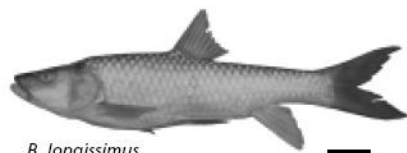
B. dainellii



B. gorgorensis



B. gorguari



B. longissimus



B. macropthalmus



B. megastoma



B. nedgia



B. platydorsus



B. surkis



B. truttiformis



B. tsanensis

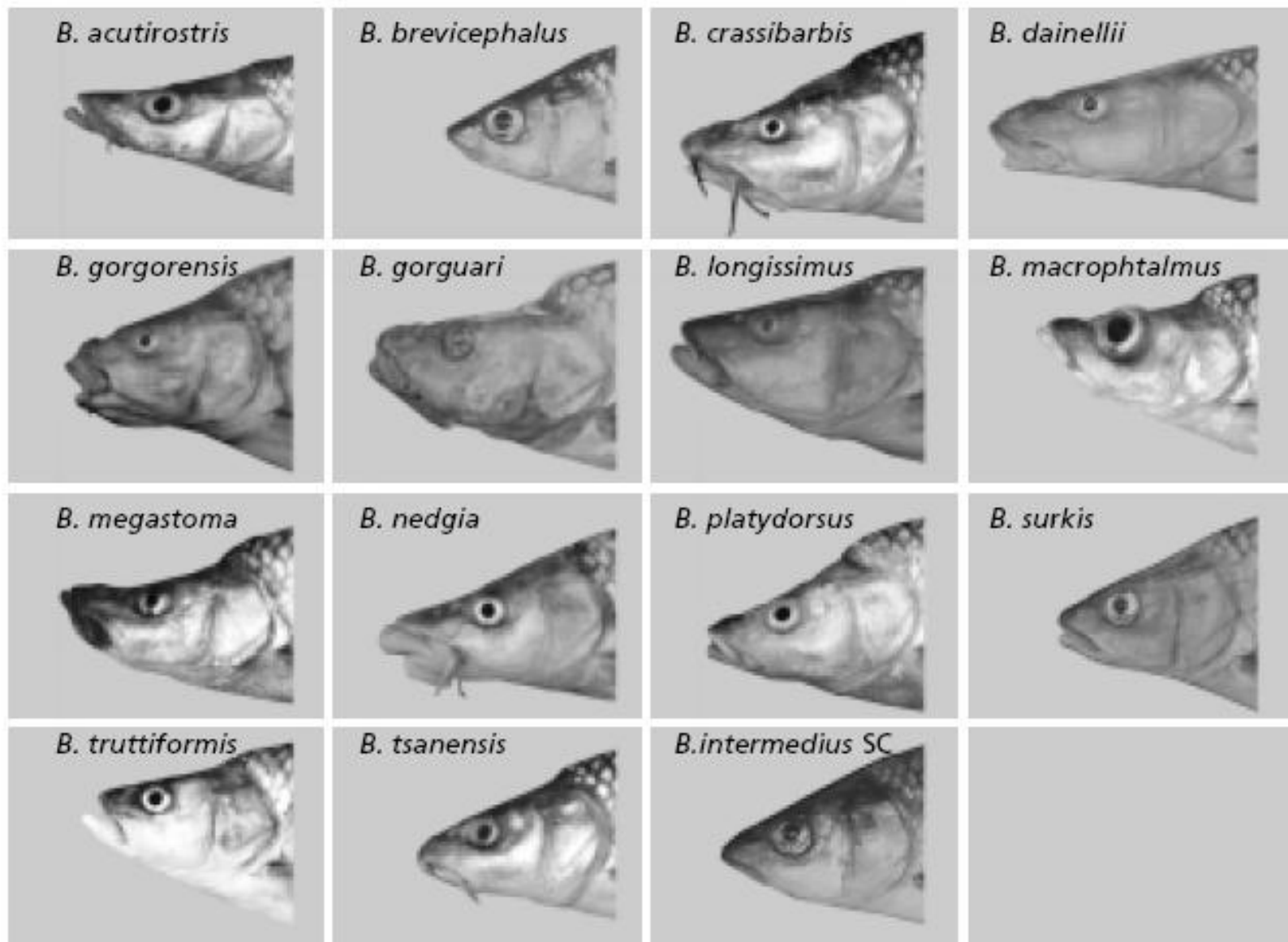


Figure 1. Heads of the 15 *Barbus* species of Lake Tana.

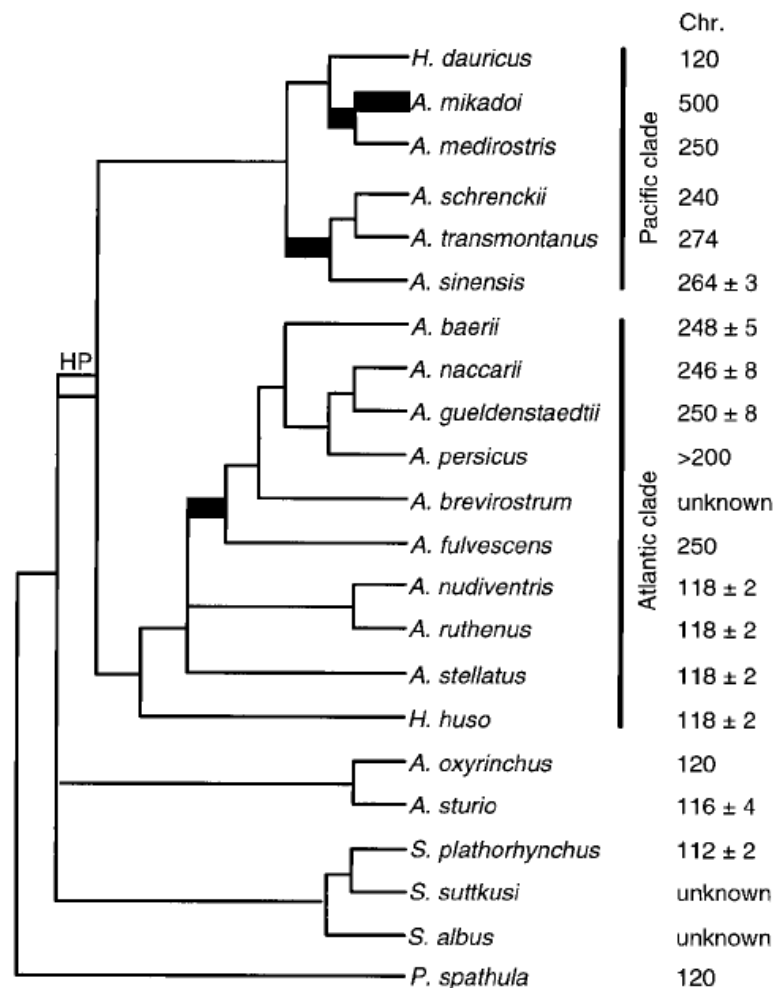


FIGURE 4.—Ploidy evolution in the Acipenseridae, including some related Acipenseriform taxa for comparison. The tree represents the topology, calculated using maximum parsimony, of the entire cytochrome-*b* gene. Polyploidization events are mapped on the tree using black boxes. The existence of heteroplasmy resulting from length variations of the mtD-loop is shown by a white box (data were taken from LUDWIG *et al.* 2000; *A. schrenckii*, *A. sinensis*, and *H. dauricus*; A. LUDWIG, unpublished data; ZHANG 1998). The number of chromosomes is reviewed in BIRSTEIN *et al.* (1997). See also Table 3.

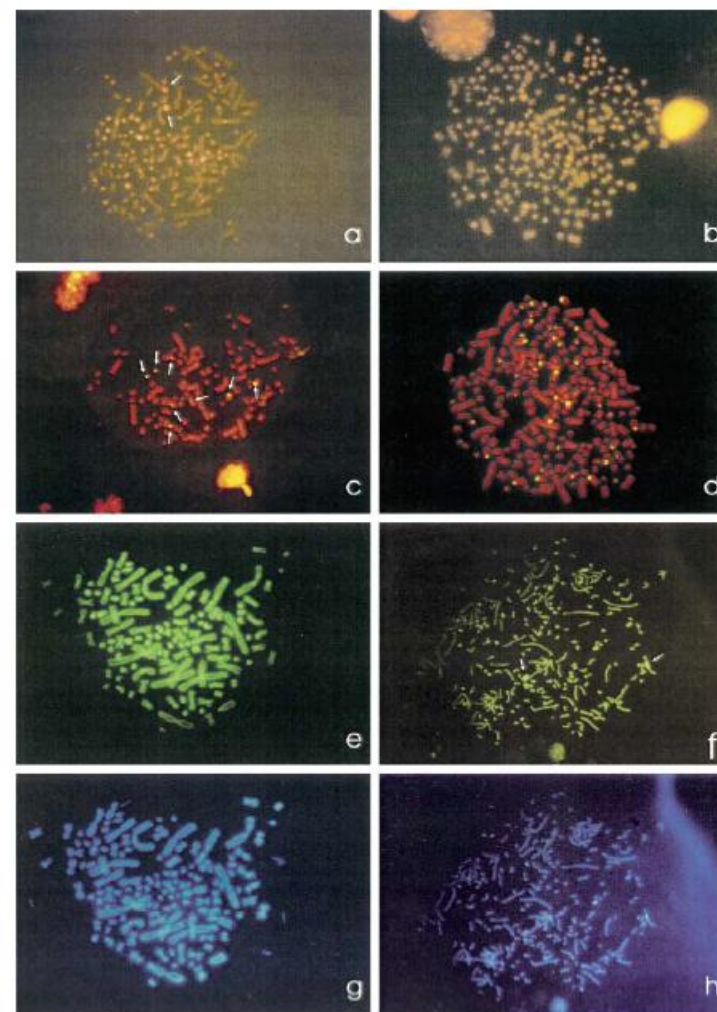
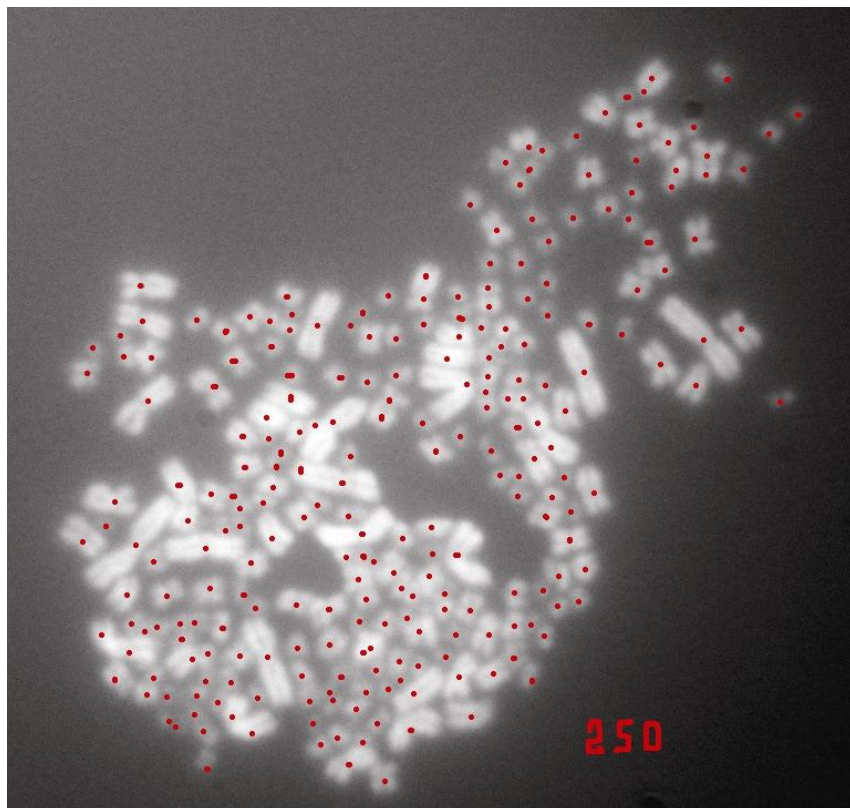
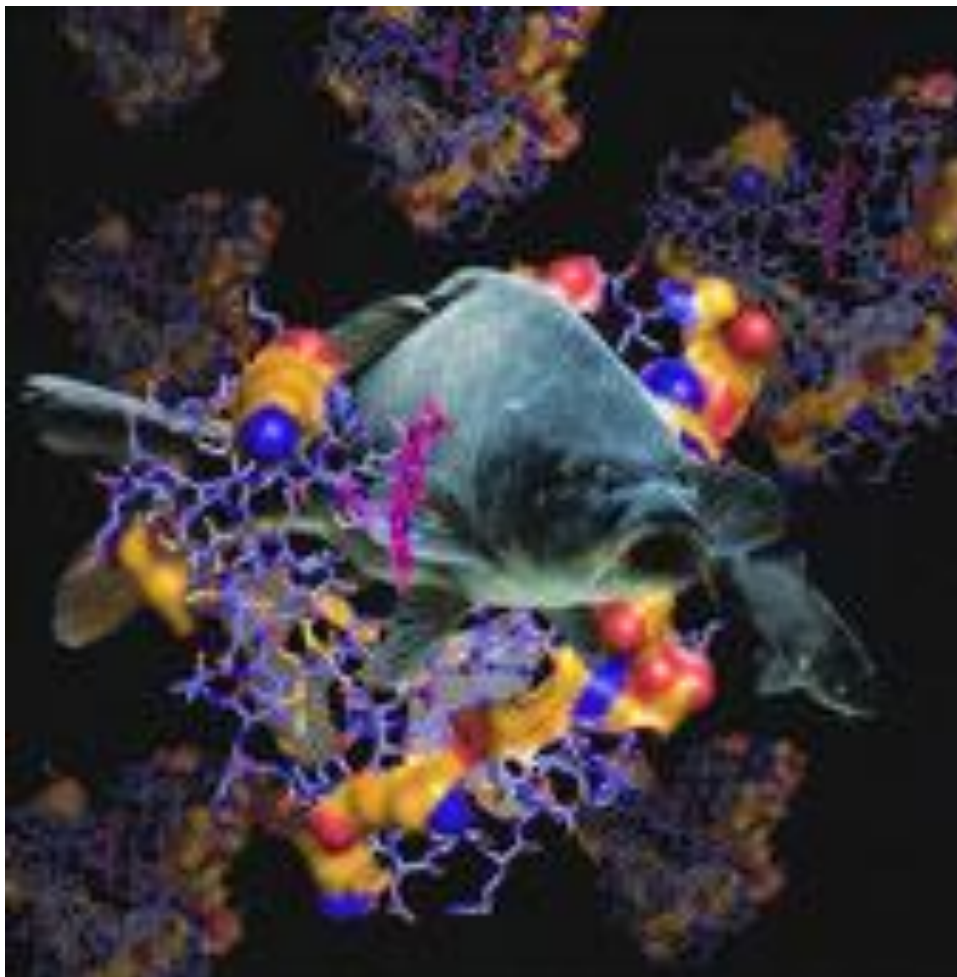


Fig. 1. Metaphases of species of group A (on the left) and B (on the right), stained with different techniques. C-banding stained with propidium iodide: (a) *A. stellatus*, (b) *A. baerii*. Fluorescence *in situ* hybridization with HindIII satellite DNA: (c) *H. juuso*, (d) *A. baerii*. Chromomycin A_3 counterstained with methyl green: (e) *H. juuso*, (f) *A. guddenaeae*. DAPI staining: (g) and (h) respectively, the same metaphases in (e) and (f).

В поисках молодых паралогов



Recent Duplication of the Common Carp (*Cyprinus carpio* L.) Genome as Revealed by Analyses of Microsatellite Loci

Lior David,^{†} Shula Blum,[†] Marcus W. Feldman,^{*} Uri Lavi,[‡] and Jossi Hillel^{*}*

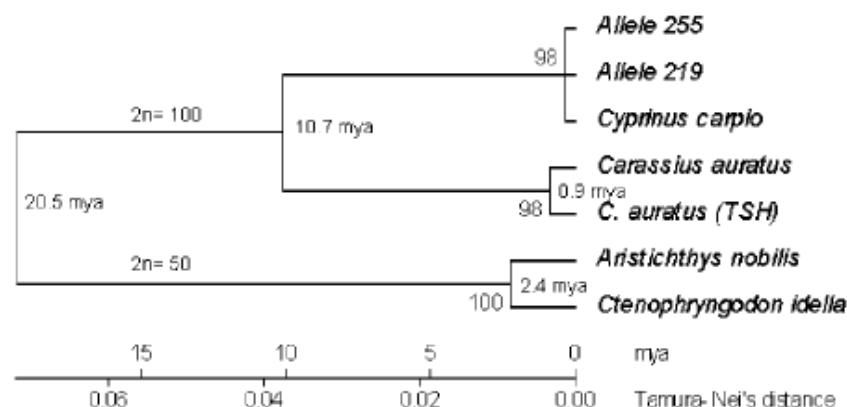


Table 1
Summary of Segregation Patterns of Microsatellite Loci in Families

Number	Type ^a	Primer Pair	Family I (<i>n</i> = 59 fish)	
1	1	<i>Koi53-54</i>	1 or 2 fragments in parents ^b	Du
2	1	<i>Koi35-36</i>	>2 fragments in 59 offspring ^d	Du
3	1	<i>CCA 17</i>	Duplicated ^c (differential fixation ^e)	Du
4	1	<i>Koi17-18</i>	1 or 2 fragments in parents ^b	>2
5	1	<i>Koi111-112</i>	>2 fragments 58 offspring ^d	>2
6	1	<i>Koi89-90</i>	Duplicated ^c	Du
7	1	<i>Koi49-50</i>	Duplicated ^c	>2
8	1	<i>Koi3-4</i>	>2 fragments in 5 offspring ^d	Du
9	1	<i>Koi115-116</i>	Duplicated ^c	>2
10	1	<i>Koi19-20</i>	Duplicated ^d	>2
11	1	<i>MFW 14</i>	No data	Du
12	1	<i>MFW 7</i>	>2 fragments in parents ^{b,f}	Du
13	1	<i>MFW 4</i>	>2 fragments in parents ^{b,f}	Du
14	1	<i>MFW 22</i>	No data	Du
15	1	<i>Koi55-56</i>	1 or 2 fragments in parents ^b	Du
16	1	<i>MFW 26</i>	>2 fragments in parents ^{b,f}	Du

FIG. 5.—Gene tree of some cyprinids. Phylogeny is based on the DNA sequence of the *thyrotropin* gene using the Tamura-Nei distance. The tree includes flanking sequences of two microsatellite alleles that are located down stream to the gene. Numbers outside the nodes are bootstrap values (%), and those inside nodes are estimates for divergence time (MYA). Above the family branches are chromosome numbers (2n).

CarpBASE - Microsoft Internet Explorer

Файл Правка Вид Избранное Сервис Справка

Адрес: <http://legr.liv.ac.uk/carpbase/index.htm> Переход

Назад Поиск Избранное

Ссылки bioinf biosoft conferences e-shopping friends fun google gulag home info news sturgeon travel Каспий library

Google Go Bookmarks 79 blocked Check AutoLink AutoFill Send to Settings



Laboratory for Environmental Gene Regulation
Data Centre

Homepage	Data Resources	Software	BLAST	ExprAlign	EST-Ferret	Register
Search Databases	carpBase	squirrelBase	roachBase	zfBase	troutBase	stickleBase

carpBASE

Environmental Transcriptomic Project is funded by Natural Environment Research Council (NERC). The aim of Carp EST analysis is to provide the Carp ESTs with gene annotations to aid the interpretation of microarray expression data. We have sequenced approximately 13,000 high-quality Carp ESTs from 14 cDNA libraries. All these ESTs have been analysed by the [EST-ferret](#) package. Further details of this project can be found at [Carp environmental transcriptomics project](#) and [our paper on PNAS \(PDF format\)](#).

The CarpBASE contains the updated result from Carp ESTs data analysis. You can browse from data set tables by using [CarpBase-Browser](#) or search information by using [CarpBase-Search](#). For the [Carp cDNA Library Resources](#), please see the [Library Resources](#). To see how we built the CarpBase, please refer to [EST-ferret](#) and [ExprAlign](#).

For more details on our projects, please visit our [LEGR Home](#).

[SEARCH the databases](#)
[BROWSE carpBASE](#)
[cDNA LIBRARY](#)
[LEGR HOME](#)

Интернет

Direct generation of SNP markers for carp paralogs

LEGR Annotation Data Search

[LEGR HOME](#)

[HELP](#)

Choose database		carpBASE 5.0 ▾
Clone ID		Example: 17o10 or 01b07
By cluster info	Maingroup ID	Example: 341 or 971 or 357
	Subgroup ID	1785-1 Example: 341-1
By BLAST info	Against database:	ALL Databases ▾
	Max # of BLAST hits	Best hit only ▾
	Best BLAST hit ID, entry or acc #	Example: P05065 or ALFA_RAT
	Best BLAST hit description	Example: Aldolase A
EC number		Example: 4.1.2.13
Gene Ontology Term/ID		Example: Aldolase or 4332

Search

or

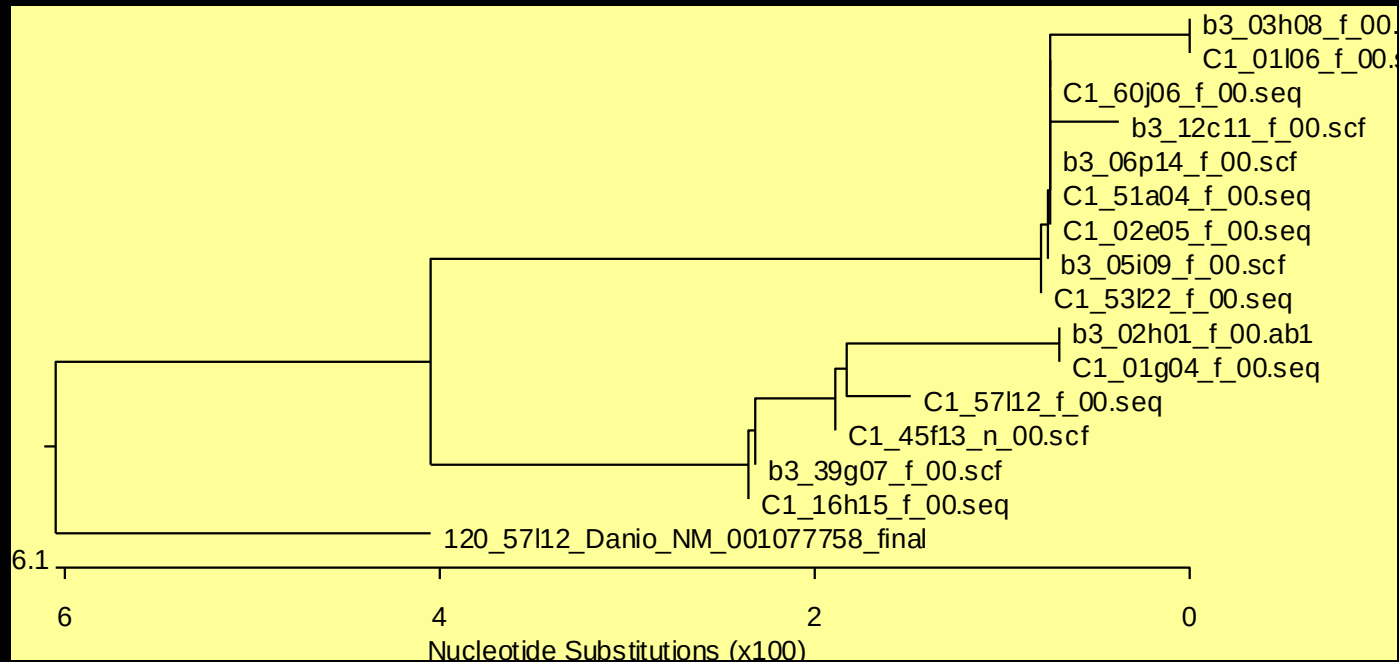
Reset



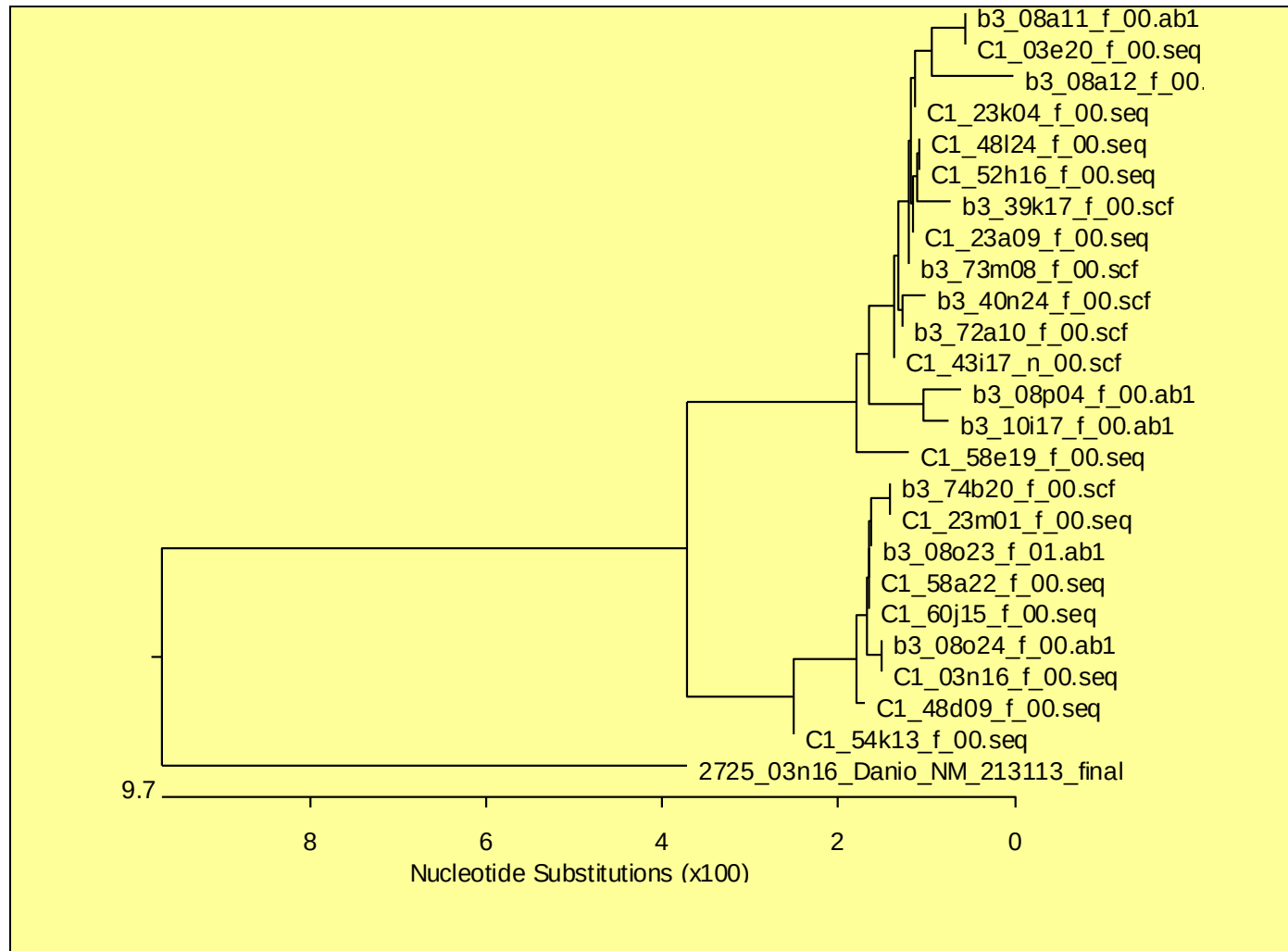
Search Results: There are 1 matches

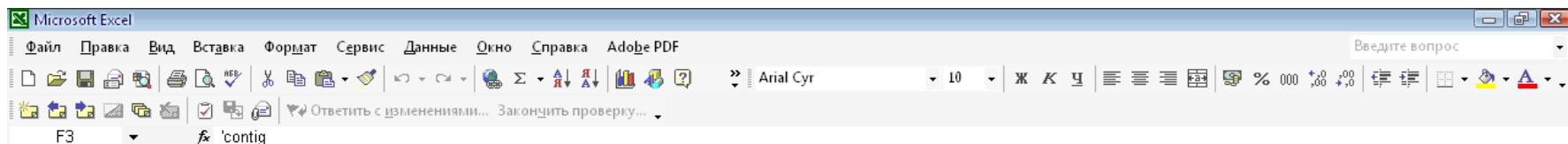
Cluster information	BLAST Searches	Other functional annotation																								
Sub-group ID: 1785-1 (Consensus) Cluster Tpye: contig Consensus Length: 840 bp 63 clone members 46b19 (sequence) 41i02 (sequence) 46m06 (sequence) 41b06 (sequence) 46b23 (sequence) 25f21 (sequence) 44e10 (sequence) 41j20 (sequence) 45g12 (sequence) 44k17 (sequence) 45o03 (sequence) 33k20 (sequence) 45c20 (sequence) 45h04 (sequence) 45c06 (sequence) 44j03 (sequence) 46p24 (sequence) 47o14 (sequence) 44k01 (sequence)	SwissProt <table><tr><th>DB id</th><th>DB entry</th><th>Descriptions</th><th>E-value</th><th>Identity</th><th>Alignment length</th></tr></table> RefSeq <table><tr><th>DB id</th><th>DB entry</th><th>Descriptions</th><th>E-value</th><th>Identity</th><th>Alignment length</th></tr></table> nr <table><tr><th>DB id</th><th>DB entry</th><th>Descriptions</th><th>E-value</th><th>Identity</th><th>Alignment length</th></tr></table> Zabrafish protein <table><tr><th>DB id</th><th>DB entry</th><th>Descriptions</th><th>E-value</th><th>Identity</th><th>Alignment length</th></tr></table>	DB id	DB entry	Descriptions	E-value	Identity	Alignment length	DB id	DB entry	Descriptions	E-value	Identity	Alignment length	DB id	DB entry	Descriptions	E-value	Identity	Alignment length	DB id	DB entry	Descriptions	E-value	Identity	Alignment length	Gene Ontolgoy Cellular Component: extracellular space Molecular Function: cytokine activity Biological Process: Enzyme Annotation
DB id	DB entry	Descriptions	E-value	Identity	Alignment length																					
DB id	DB entry	Descriptions	E-value	Identity	Alignment length																					
DB id	DB entry	Descriptions	E-value	Identity	Alignment length																					
DB id	DB entry	Descriptions	E-value	Identity	Alignment length																					

Cluster 120 -serine (or cysteine) proteinase inhibitor



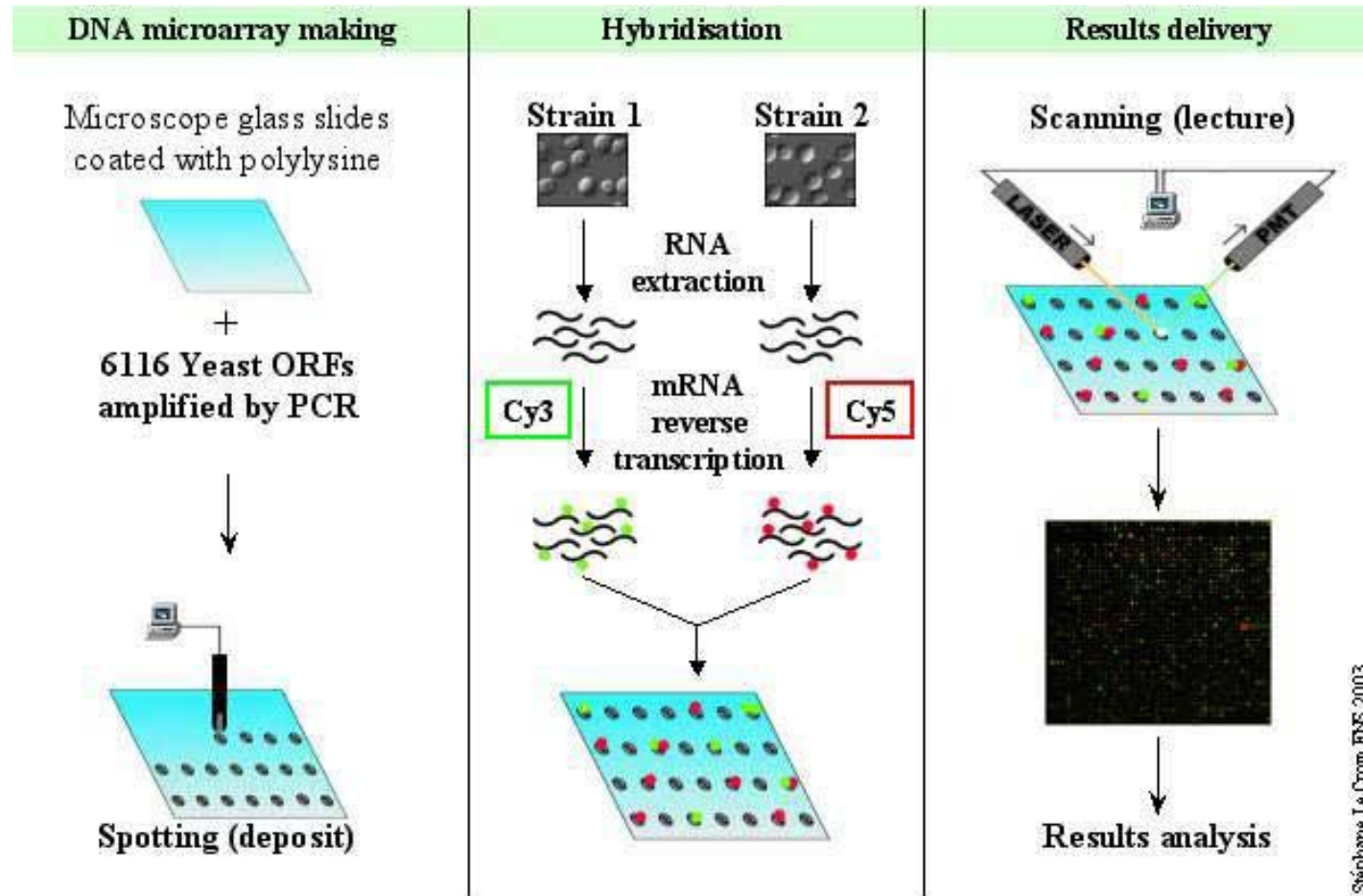
Cluster 2725 - ribosomal protein L26





11.03.03_genome_subgroup_table_sortedover.xls																
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Row_num	Maingroup	Subgroup	Subgroup_id		Cluster_ty	Contig_singlet	Conser	Subgr	Subgroup_members	DB_BLAS	Bit_score	E-value	Identif	Gene_name	
2	2485	1708	11	1708-11		contig	Contig11	1544	754	Cc_03a21_f_00.seq Cc_3806_n_00.scf	nr	105	4,00E-21	50%	1	mtDNA - NOT hypothetical 18K pro
3	4188	2592	4	2592-4		contig	Contig4	1274	359	Cc_57c18_f_00.seq Cc_57c08_f_00.seq	swissprot	284	4,00E-76	99%	1	Hemoglobin alpha chain
4	2480	1708	6	1708-6		contig	Contig6	1110	201	Cc_40e24_n_00.scf Cc_37c23_n_00.scf	nr	N/A	N/A	N/A	1	Unclassifiable EST
5	3906	2439	1	2439-1	2.2 (??)	contig	Contig1	771	174	Cc_51k01_f_00.seq Cc_17m03_f_00.se	swissprot	296	4,00E-80	96%	1	Hemoglobin beta chain
6	2846	1844	1	1844-1	??	contig	Contig1	1280	108	Cc_54o15_f_00.seq Cc_52b02_f_00.seq	swissprot	480	e-135	94%	2	40S ribosomal protein S2
7	2709	1798	2	1798-2		contig	Contig2	706	102	Cc_33j07_n_00.scf Cc_10j01_f_00.seq	nr	N/A	N/A	N/A	1	Unclassifiable EST
8	3639	2302	2	2302-2	3.3	contig	Contig2	702	89	Cc_59f03_f_00.seq Cc_60c02_f_00.seq	swissprot	230	2,00E-60	97%	1	40S ribosomal protein S20
9	3770	2360	4	2360-4	3.0	contig	Contig4	492	83	Cc_53o08_f_00.seq Cc_49f13_f_00.seq	nr	N/A	N/A	N/A	1	Unclassifiable EST
10	3741	2346	1	2346-1		contig	Contig1	1403	73	Cc_57k20_f_00.seq Cc_51o05_f_00.seq	swissprot	396	e-110	76%	2	Trypsin III precursor (EC 3.4.21.4) (
11	3572	2270	1	2270-1	~2.0	contig	Contig1	832	71	Cc_57a06_f_00.seq Cc_53k19_f_00.se	nr	N/A	N/A	N/A	1	Unclassifiable EST
12	4248	2619	2	2619-2	??	contig	Contig2	1292	68	Cc_49d01_f_00.seq Cc_49i14_f_00.seq	swissprot	124	8,00E-28	42%	1	Pentraxin fusion protein precursor
13	2677	1785	1	1785-1		contig	Contig1	840	63	Cc_46b19_n_00.scf Cc_41i02_n_00.scf	swissprot	140	3,00E-33	98%	5	Granulin-1
14	3574	2272	1	2272-1		contig	Contig1	793	56	Cc_48e18_f_00.seq Cc_54i11_f_00.seq	swissprot	132	1,00E-30	90%	5	Proto-oncogene tyrosine-protein kir
15	1058	760	1	760-1	??	contig	Contig1	1029	56	Cc_14j02_f_00.seq Cc_57a15_f_00.seq	swissprot	394	e-109	72%	2	Chymotrypsin A precursor (EC 3.4.
16	2975	1908	1	1908-1		contig	Contig1	1395	54	Cc_45o09_n_00.scf Cc_34g07_n_00.sc	nr	N/A	N/A	N/A	1	Unclassifiable EST
17	3642	2304	1	2304-1	1.1	contig	Contig1	1119	50	Cc_52j04_f_00.seq Cc_55j05_f_00.seq	swissprot	352	1,00E-96	92%	1	60S ribosomal protein L9
18	2736	1809	2	1809-2	~1.4	contig	Contig2	1127	48	Cc_46j19_n_00.scf Cc_34e12_n_00.scf	zebra.pt	81.6	9,00E-16	66%	5	hypothetical protein LOC449557 [D
19	2847	1844	2	1844-2	??	contig	Contig2	963	48	Cc_18i04_f_00.seq Cc_34n03_n_00.scf	swissprot	471	e-132	98%	2	40S ribosomal protein S2
20	3691	2326	1	2326-1	??	contig	Contig1	515	48	Cc_48g11_f_00.seq Cc_51p12_f_00.se	swissprot	203	2,00E-52	99%	1	60S ribosomal protein L34
21	4050	2525	3	2525-3		contig	Contig3	651	48	Cc_52a13_f_00.seq Cc_50i14_f_00.seq	nr	N/A	N/A	N/A	1	Unclassifiable EST
22	2491	1708	17	1708-17		contig	Contig17	2199	47	Cc_44e08_n_00.scf Cc_38o12_n_00.sc	nr	302	3,00E-80	91%	1	hypothetical 18K protein - goldfish i
23	2676	1784	1	1784-1		contig	Contig1	1068	47	Cc_31d07_n_00.scf Cc_37d11_n_00.sc	swissprot	127	7,00E-29	42%	1	Natlectin precursor
24	78	56	1	56-1	3.8	contig	Contig1	1093	47	Cc_02e13_f_00.seq Cc_02i04_f_00.seq	swissprot	130	5,00E-30	32%	2	Apolipoprotein A-I-2 precursor (Apo
25	3628	2300	1	2300-1	3.9	contig	Contig1	942	45	Cc_49i18_f_00.seq Cc_60a23_f_00.se	swissprot	340	4,00E-93	68%	2	60S ribosomal protein L6 (TAX-resp
26	3753	2353	3	2353-3	~1.3 (3.8)	contig	Contig3	793	43	Cc_57d02_f_00.seq Cc_44o02_n_00.sc	swissprot	174	3,00E-43	88%	5	40S ribosomal protein S11
27	3605	2287	1	2287-1		contig	Contig1	1117	42	Cc_55g13_f_00.seq Cc_56f08_f_00.seq	swissprot	387	e-107	80%	2	Trypsin II precursor (EC 3.4.21.4) (F
28	2481	1708	7	1708-7		contig	Contig7	1163	41	Cc_45h10_n_00.scf Cc_34m24_n_00.sc	nr	N/A	N/A	N/A	1	Unclassifiable EST
29	3563	2265	3	2265-3		contig	Contig3	771	38	Cc_48n03_f_00.seq Cc_53i06_f_00.seq	swissprot	162	7,00E-40	60%	1	Fatty acid-binding protein, brain (B-
30	3596	2282	2	2282-2	??	contig	Contig2	677	37	Cc_58f07_f_00.seq Cc_59k05_f_00.se	nr	N/A	N/A	N/A	1	Unclassifiable EST
31	4062	2531	1	2531-1	2.9	contig	Contig1	733	37	Cc_55e13_f_00.seq Cc_54o18_f_00.se	swissprot	151	2,00E-36	100%	7	Ubiquitin
32	103	73	1	73-1		contig	Contig1	1162	37	Cc_52g07_f_00.seq Cc_50h09_f_00.se	swissprot	285	1,00E-76	58%	2	Elastase-1 precursor (EC 3.4.21.36
33	2774	1829	3	1829-3	??	contig	Contig3	1003	36	Cc_50p03_f_00.seq Cc_50n19_f_00.se	swissprot	317	2,00E-86	96%	1	Translationally-controlled tumor pro
34	2790	1835	1	1835-1		contig	Contig1	1093	36	Cc_41g17_n_00.scf Cc_53k05_f_00.se	swissprot	352	8,00E-97	67%	2	Chymotrypsin B (EC 3.4.21.1) [Cor
35	2350	1637	3	1637-3		contig	Contig3	1109	33	Cc_44c02_n_00.scf Cc_61a11_f_00.se	swissprot	463	e-130	75%	5	40S ribosomal protein SA (p40) (34
36	3589	2278	1	2278-1	1.8	contig	Contig1	715	33	Cc_49h05_f_00.seq Cc_54p08_f_00.se	swissprot	348	6,00E-96	96%	1	60S ribosomal protein L11
37	3873	2425	2	2425-2		contig	Contig2	1841	33	Cc_50o08_f_00.seq Cc_48h08_f_00.se	swissprot	759	0.0	100%	5	Actin, cytoplasmic L1 (Beta-actin)
38	4112	2556	2	2556-2	4.8	contig	Contig2	944	33	Cc_50g18_f_00.seq Cc_58g08_f_00.se	swissprot	440	e-123	94%	2	60S ribosomal protein L7a (Surfeit I
39	4417	2700	1	2700-1	9.2	contig	Contig1	813	33	Cc_03d14_f_00.seq Cc_49f24_f_00.se	zebra.pt	218	4,00E-57	73%	1	PREDICTED: similar to 14 kDa apo
40	4210	2598	1	2598-1	2.3	contig	Contig1	426	30	Cc_57o12_f_00.seq Cc_48k05_f_00.se	nr	N/A	N/A	N/A	1	Unclassifiable EST
41	3426	2183	3	2183-3		contig	Contig3	1152	29	Cc_47i03_n_00.scf Cc_47i02_n_00.scf	i swissprot	223	1,00E-57	49%	2	High choriolytic enzyme 1 precursor
42	3871	2424	1	2424-1	3.1 - 3.5	contig	Contig1	1419	29	Cc_04e16_f_00.seq Cc_51d18_f_00.se	swissprot	511	e-144	58%	5	Carboxypeptidase B precursor (EC
43	4003	2501	1	2501-1		contig	Contig1	802	29	Cc_54e19_f_00.seq Cc_54n22_f_00.se	swissprot	134	4,00E-31	93%	5	Transmembrane protein 47 (Transm
44	4125	2562	2	2562-2		contig	Contig2	788	29	Cc_58g24_f_00.seq Cc_50k11_f_00.se	swissprot	195	1,00E-49	60%	1	39S ribosomal protein L43, mitochr
45	3386	2160	2	2160-2	4.9	contig	Contig2	1229	28	Cc_11i06_f_00.seq Cc_31b24_n_00.scf	swissprot	134	5,00E-31	47%	1	Lysozyme C precursor (EC 3.2.1.1)
46	3710	2335	1	2335-1	3.1	contig	Contig1	955	28	Cc_51j09_f_00.seq Cc_59f19_f_00.se	swissprot	387	e-107	96%	2	40S ribosomal protein S5
47	3866	2420	2	2420-2	3.1	contig	Contig2	1374	28	Cc_53g20_f_00.seq Cc_50i12_f_00.se	swissprot	263	1,00E-69	99%	1	40S ribosomal protein S18

Принцип создания и работы кДНК микрочипа

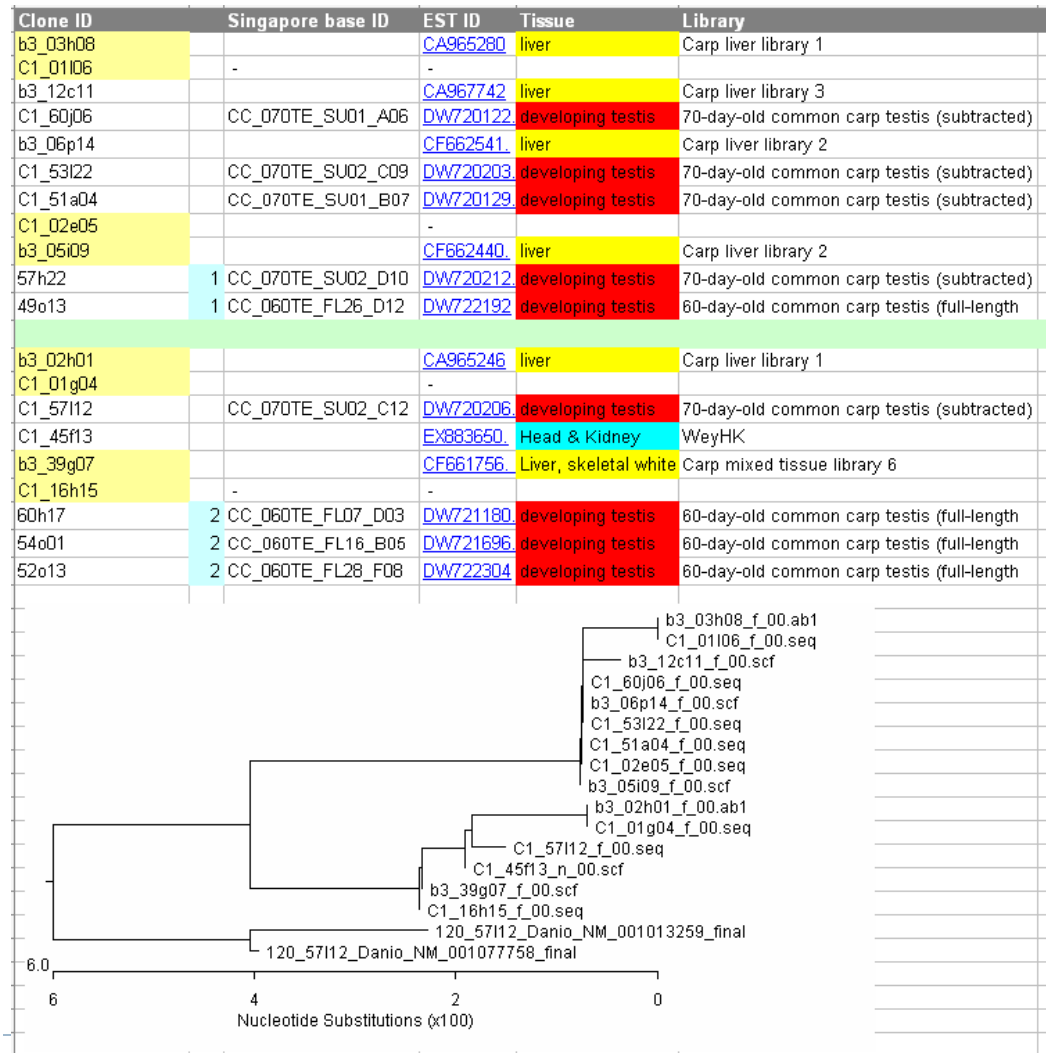


Возможные пути изменения паралогов после дупликации:

- ▶ Оба паралога продолжают выполнять одинаковую функцию
- ▶ Один из паралогов перестает экспрессироваться (становиться псевдогеном)
- ▶ Паралоги приобретают тканеспецифичность экспрессии
- ▶ Паралоги начинают выполнять разную функцию в одной ткани.



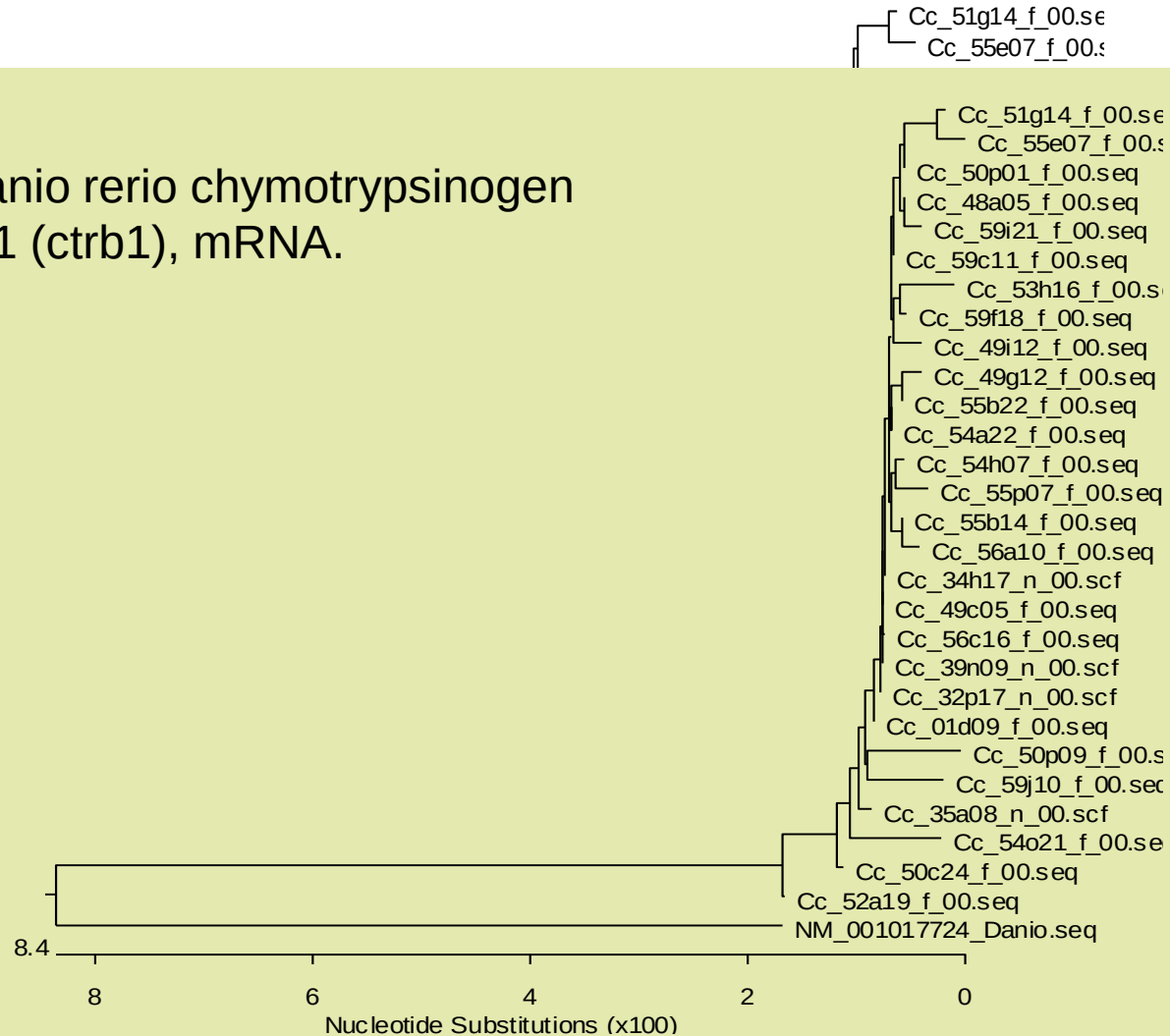
Оба паралога продолжают выполнять одинаковую функцию



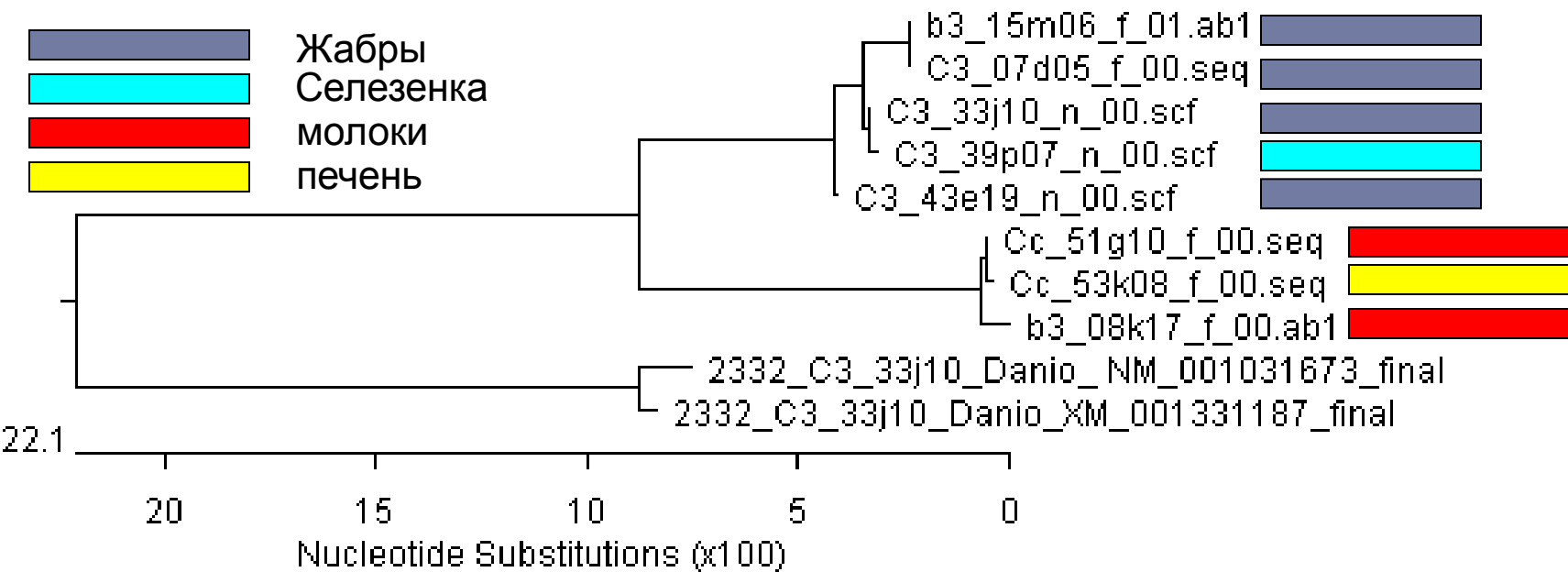
Cluster 120-1 Alpha-1-antitrypsin homolog precursor

Один из паралогов перестает экспрессироваться (становиться псевдогеном)

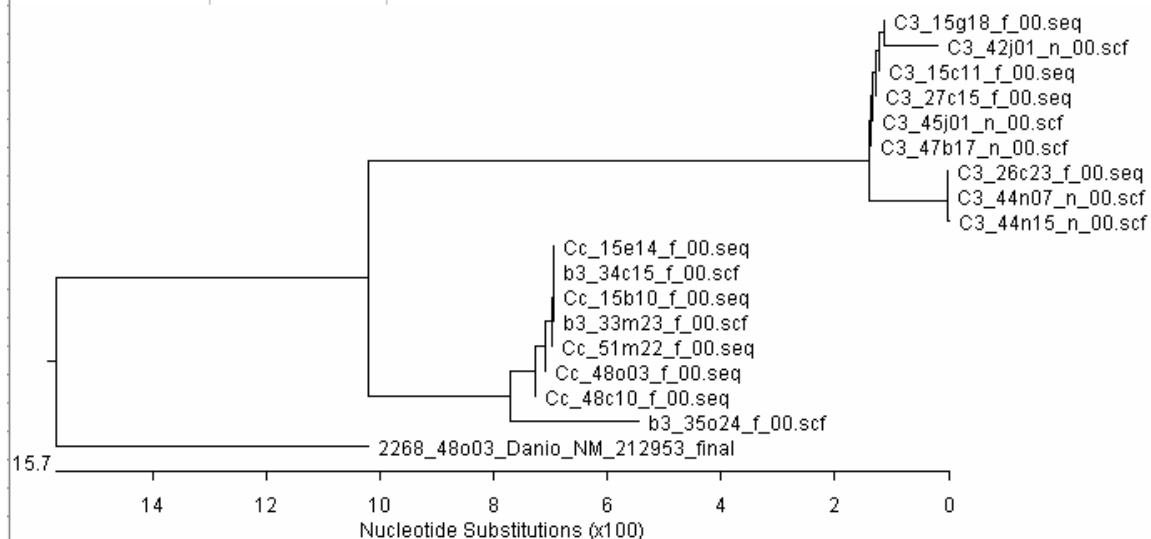
Danio rerio chymotrypsinogen
B1 (ctrb1), mRNA.



Паралоги приобретают тканеспецифичную экспрессию



Clone ID	Tissue	Library
C3_15g18	-	
C3_42j01	Gill	cDNA library prepared by Daryl
C3_15c11	-	
C3_27c15	-	
C3_45j01	Head Kidney	cDNA library prepared by Daryl
C3_47b17	Head Kidney	cDNA library prepared by Daryl
C3_26c23	-	
C3_44n07	Head Kidney	cDNA library prepared by Daryl
C3_44n15	Head Kidney	cDNA library prepared by Daryl
		
Cc_15e14	-	
b3_34c15	Muscle	Serially subtracted cDNA library prepared from muscle of warm, cold and hypoxia challenge
Cc_15b10	-	
b3_33m23	Liver	Serially subtracted cDNA library prepared from liver of warm, cold and hypoxia challenge
Cc_51m22	Brain	cDNA library prepared by Caroline
Cc_48o03	developing testis	80-day-old common carp testis (full-length)
Cc_48c10	developing testis	80-day-old common carp testis (full-length)
b3_35o24	Liver	Serially subtracted cDNA library prepared from liver of warm, cold and hypoxia challenge



Cluster 2268 : 40S ribosomal protein S29

Мобильные элементы генома

- ▶ Инсерционные элементы, например, IS1603
- ▶ Транспозоны, например, Tn5
- ▶ Профаги - латентная форма умеренных бактериофагов, например, транспозонподобного фага Mu (мю)
- ▶ Плазмиды - эписомы, например, половой фактор кишечной палочки (F-плазмида)



Транспозоны

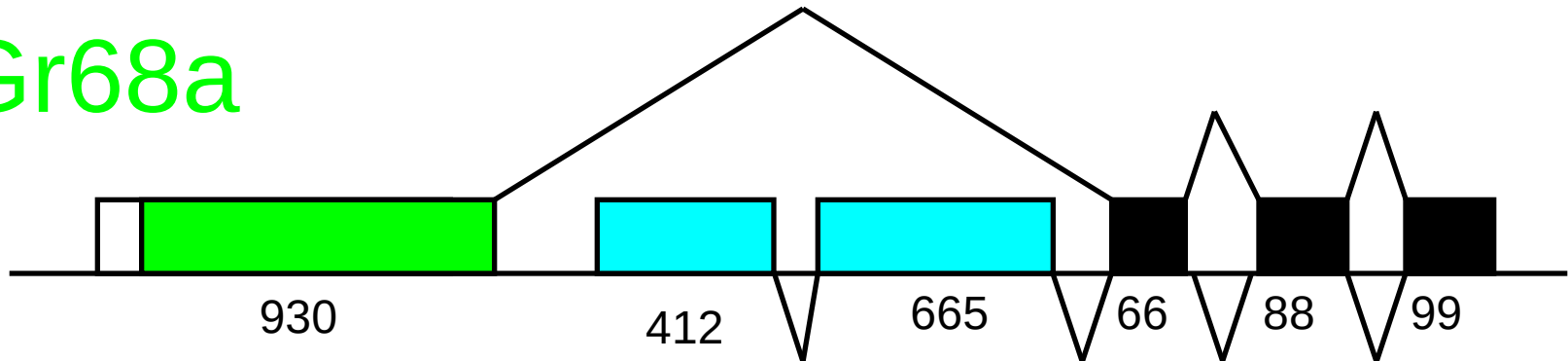
- ▶ **н** — последовательность ДНК, способная перемещаться внутри генома в результате процесса, называемого транспозицией. Транспозоны — один из классов мобильных элементов генома которые, встраиваясь в геном, могут вызывать мутации, в том числе и такие значительные как хромосомные перестройки.
- ▶ 1. Ретротранспозоны, которые перемещаются по геному путём обратной транскрипции с их РНК;
- ▶ 2. ДНК-транспозоны, перемещается путём прямого вырезания и вставки с использованием кодируемого транспозоном фермента транспозазы.
- ▶ 3 LINE-ы и SINE-ы



Пример ретротранспонированного гена – Gr68a

Alternative splicing of Gr68a and Gr32a in D. virilis group

Gr68a

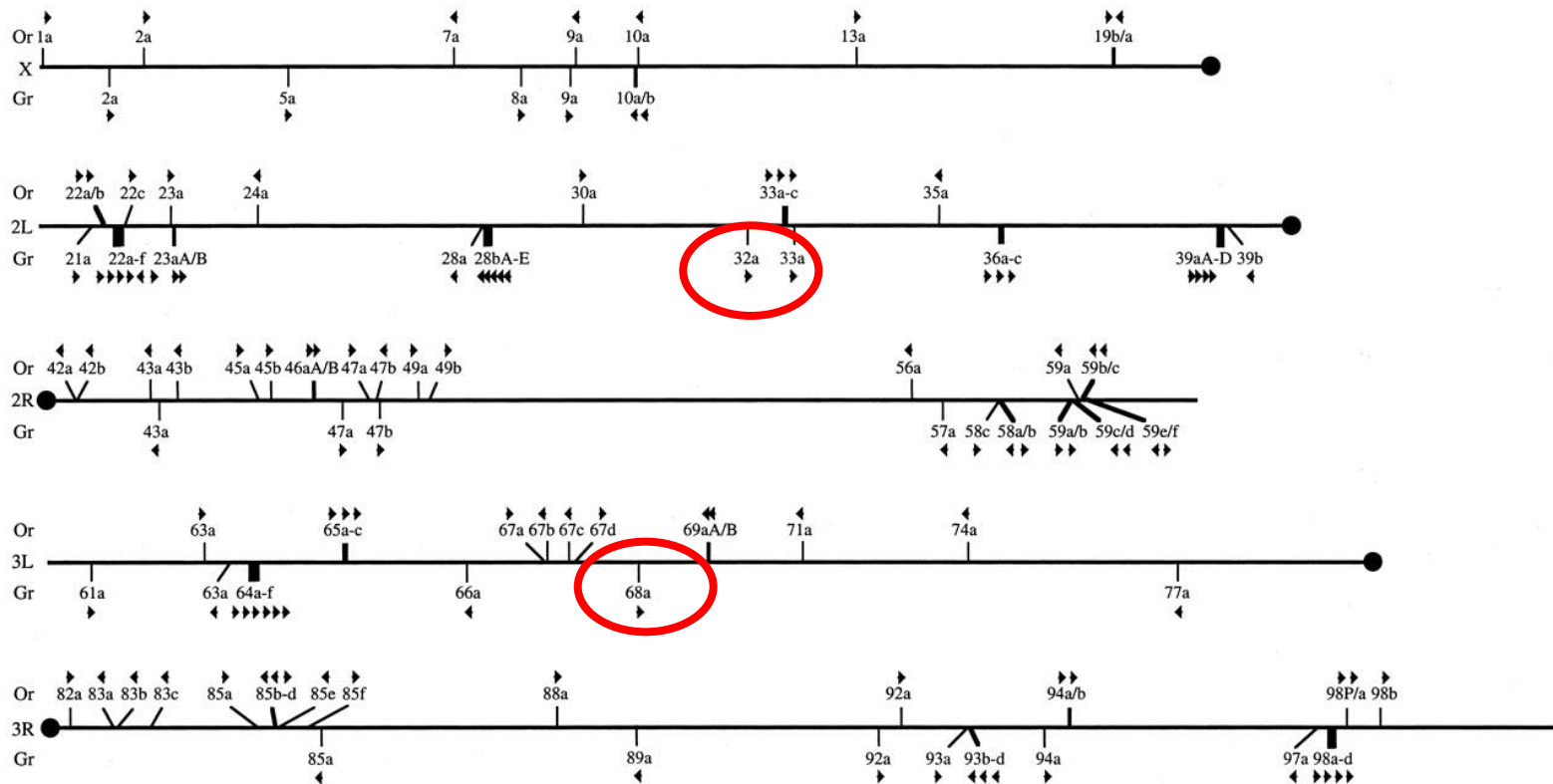


Gr32a

D. virilis.scaffold_12963

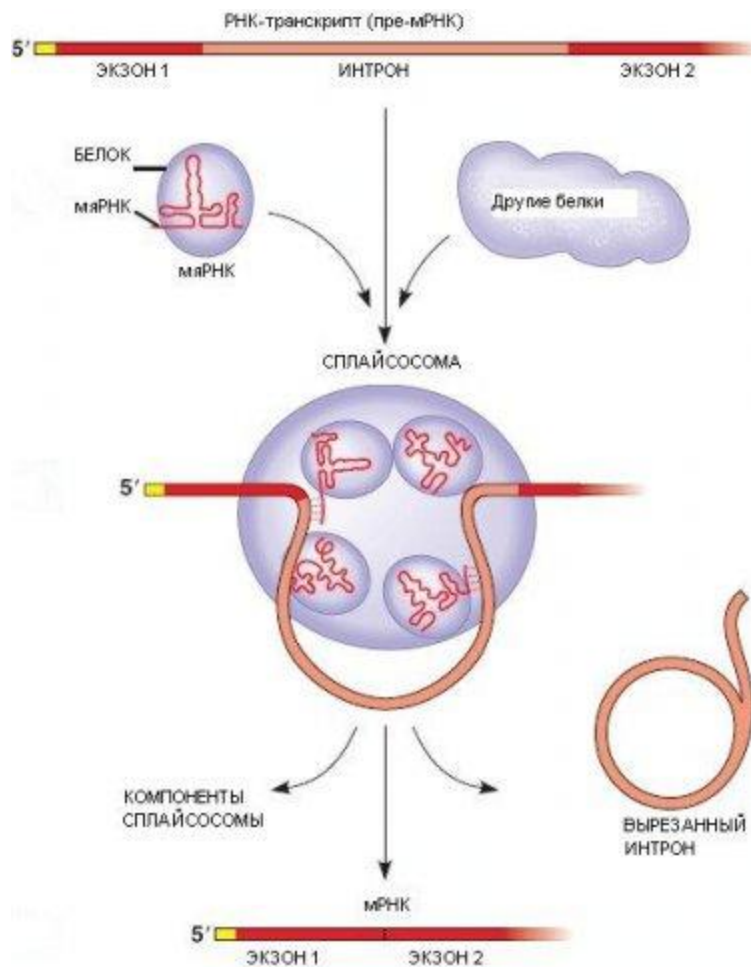
pos. 7917243 - 7920736

Genomic locations of the Or and Gr genes in *D. melanogaster*



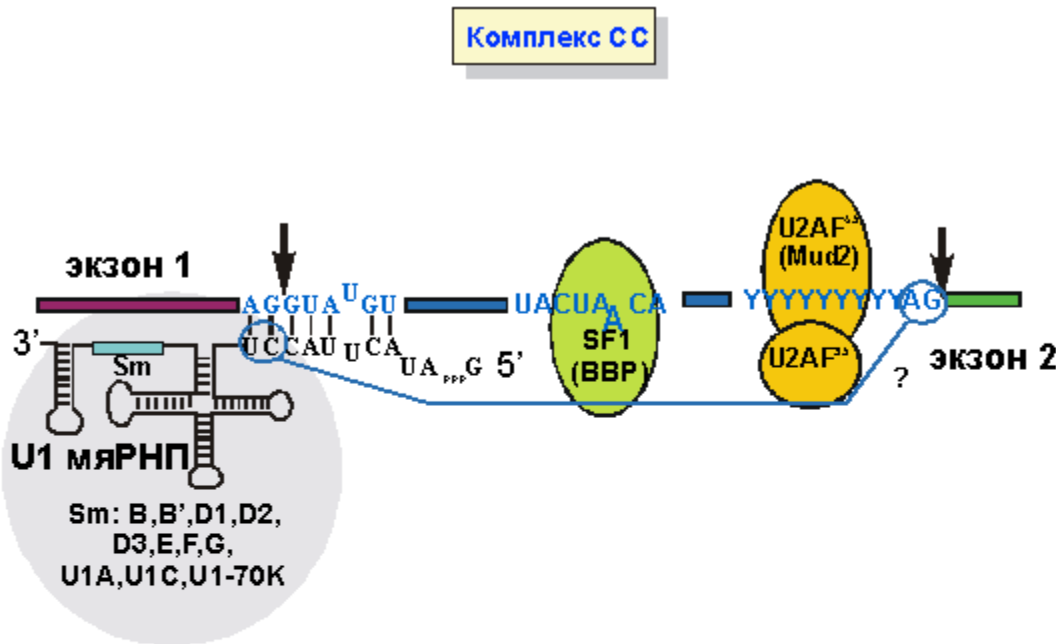
Robertson, Hugh M. et al. (2003) Proc. Natl. Acad. Sci. USA 100, 14537-14542

Механизм сплайсинга



Сплайсинг — это процесс вырезания внутренних участков пре-мРНК, называемых интронами. Бывает несколько разновидностей сплайсинга. Наиболее распространенный тип — это вырезание интронов с помощью малых ядерных РНП (мяРНП). Эти небольшие РНК, связанные с белками имеют ряд особенностей. Их 5'-конец закрыт кепом, причем в отличие от мРНК остаток гуанозина на 5'-конце не моно-, а триметилирован (по N7 и дважды по N2). Кроме того, мяРНК как правило содержат Sm последовательность, к которой присоединяется набор из семи Sm белков. Эти маркеры мяРНК нужны для направленного транспорта их в ядро и прикрепления к фосфорилированному CTD РНК полимеразы II.

Распознавание сайта сплайсинга



- ▶ Сначала U1 мяРНП узнает 5'-место сплайсинга, а возможно и 3'-место сплайсинга. Место разветвления узнается белком SF1, а полипиримидиновая последовательность, предшествующая 3'SS - гетеродимером U2AF. При этом образуется комплекс СС (commitment complex) – комплекс определяющий вырезание данного интрона.

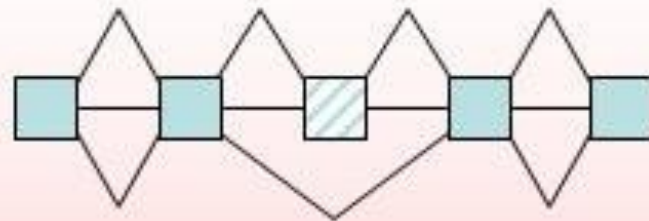
Конформационные перестройки

Связывание U1 мяРНП:
SR белки
(Prp39p, Prp40p?)

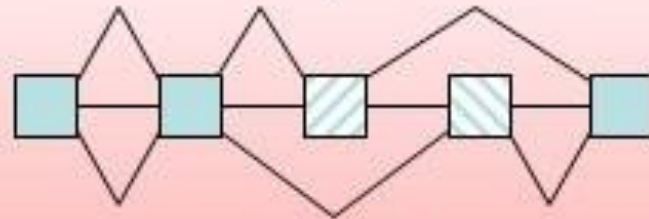
mRNA Splicing.flv



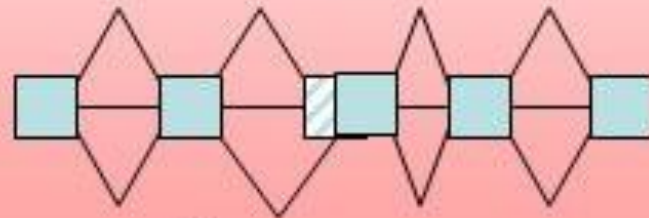
Типы альтернативного сплайсинга



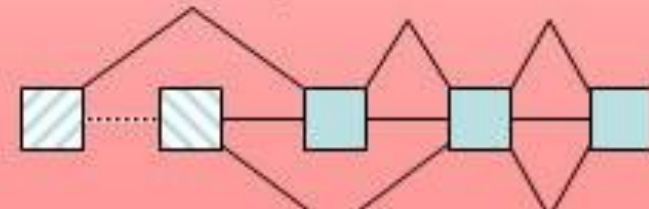
кассетный



взаимоисключающий

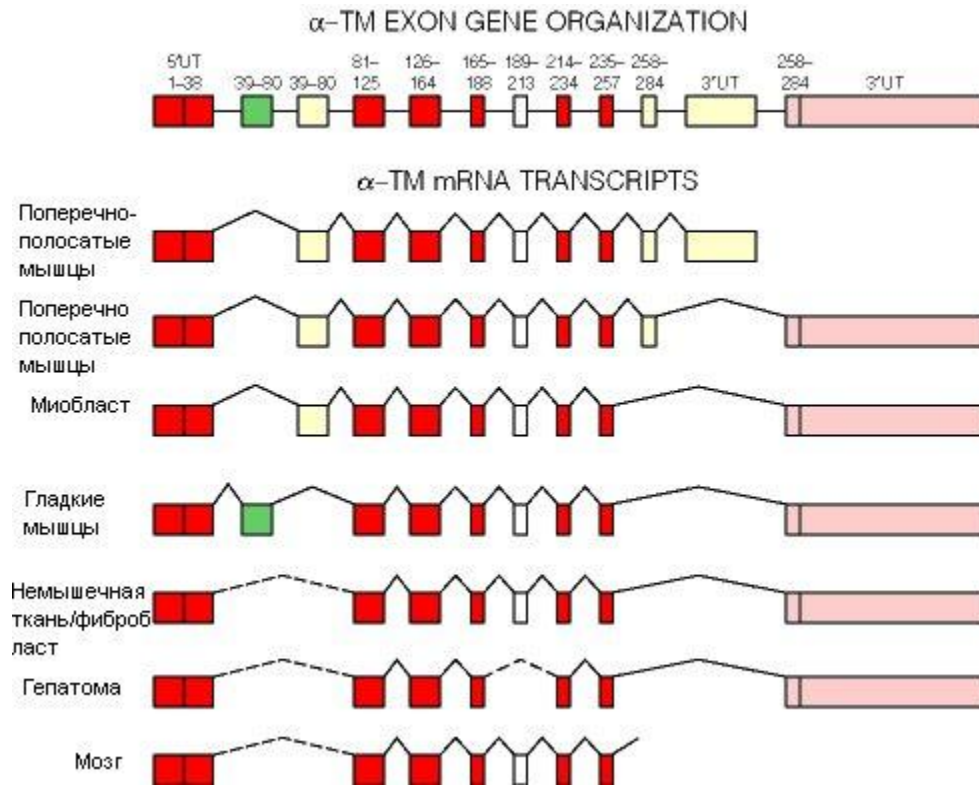


с внутренним акцепторным сайтом



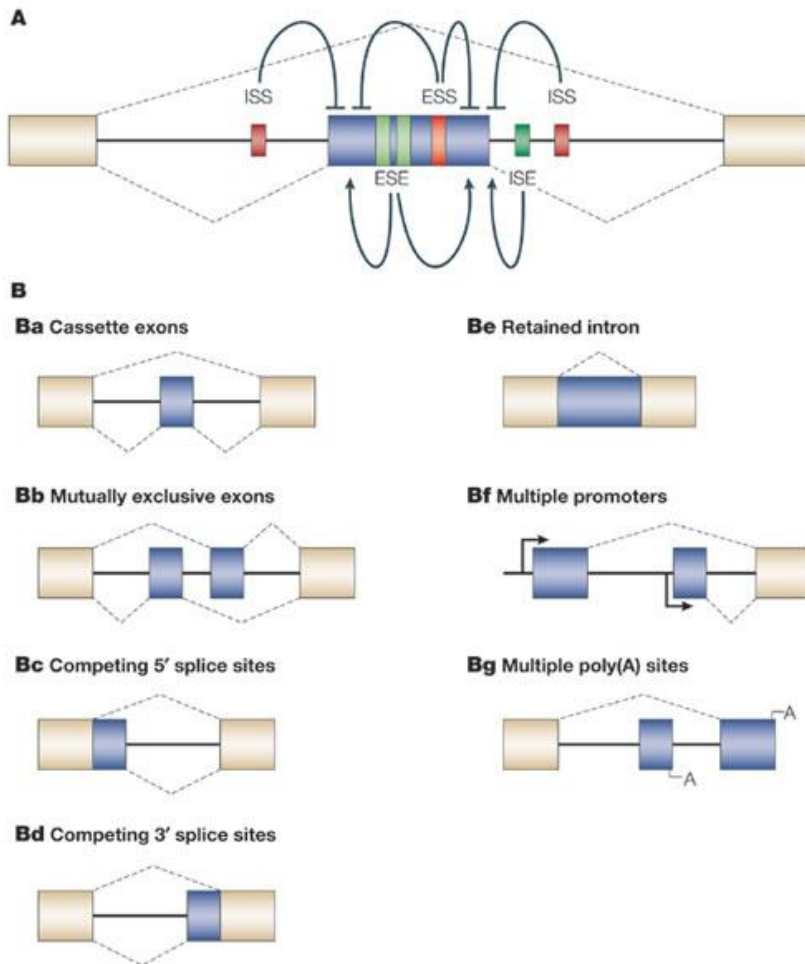
альтернативные промоторы

Альтернативный сплайсинг



- ▶ Альтернативный сплайсинг позволяет индивидуальным генам продуцировать множественные изоформы белков - играя тем самым центральную роль в генерации сложных протеомов.

cellular codes альтернативного сплайсинга



Регуляция сплайсинга как правило осуществляется с помощью модуляции ранних ступеней сборки сплайсосомы. Цис-элементы представляют собой энхансеры и сайленсеры сплайсинга, которые могут быть локализованы или в экзонах или в интронах и которые связывают активаторные и репрессорные белки. Иногда присутствие или отсутствие одиночного регулятора достаточно, чтобы предопределить пути альтернативного сплайсинга.

Комбинации наиболее широко распространённых факторов вовлекаются в выбор путей сплайсинга. Это привело к концепции 'cellular codes', которые состояются за счёт определенных комбинаций регуляторных факторов.

Альтернативный сплайсинг – полигон для апробации новых генов без потери старых (М.С.Гельфанд etc.)



- ▶ Сайты альтернативного сплайсинга чаще возникают или утрачиваются чем сайты конститутивного сплайсинга.
- ▶ Альтернативные геномы эволюционируют быстрее (по числу и синонимичных, и несинонимичных мутаций)
- ▶ Сравнение шимпанзе и человека показало наличие движущего отбора в альтернативных экзонах. ($K_n/K_s > 1$)



