

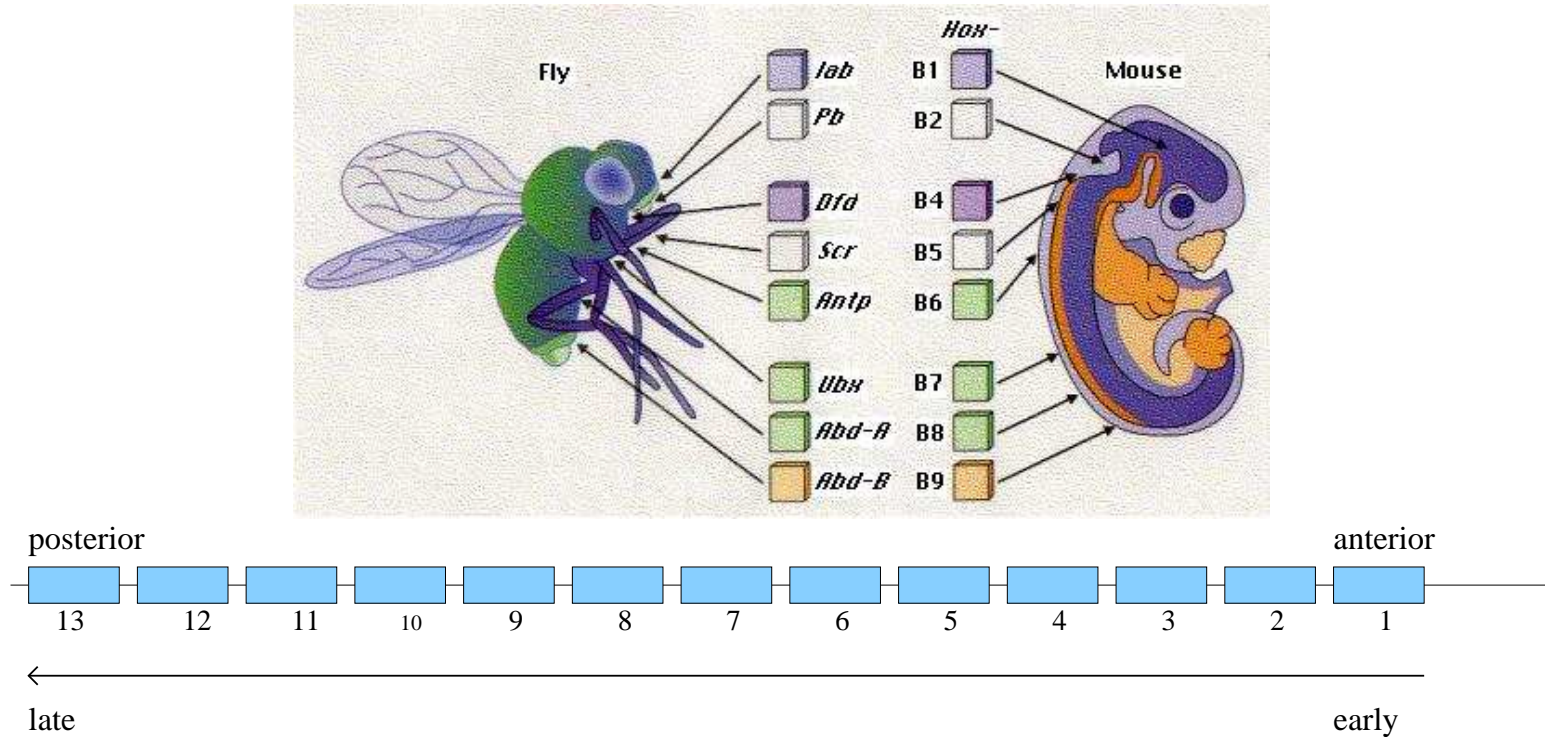
The Mystery of Hox Cluster Evolution

Claudia Fried

Bioinformatic, Institute of Informatic, Univ.Leipzig, Germany

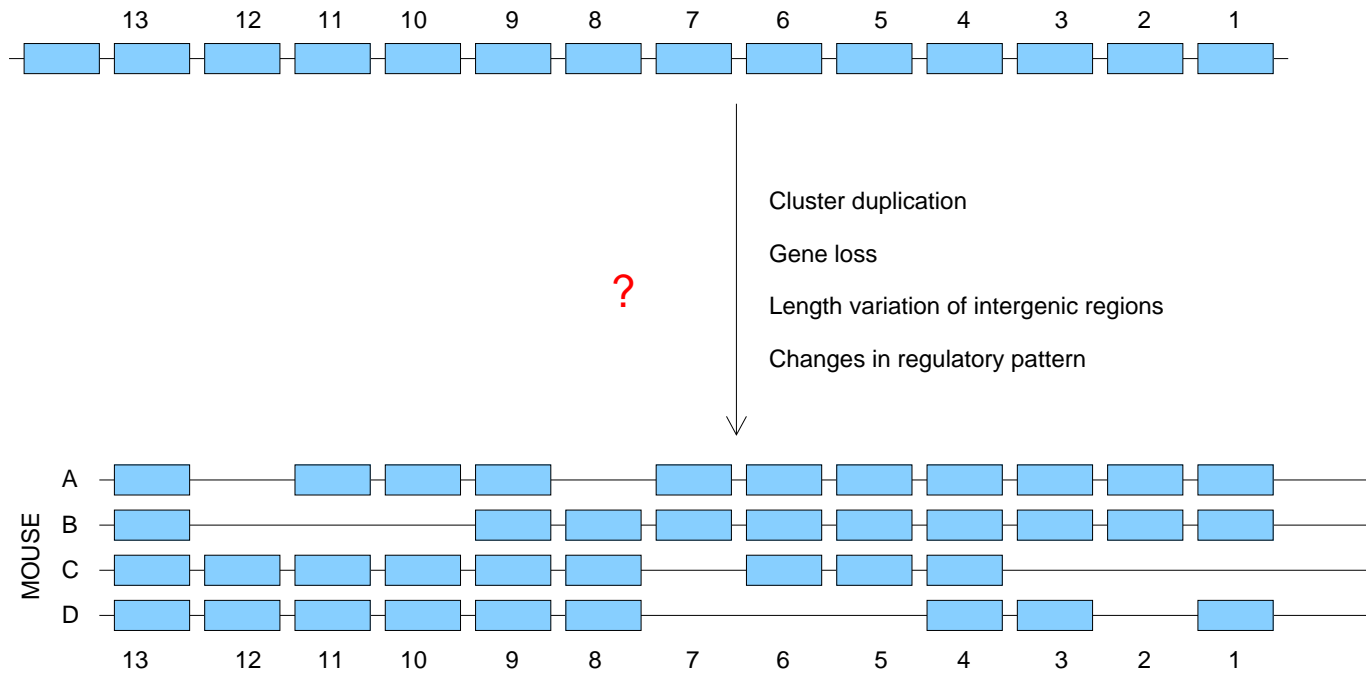
Bled, 24. Februar 2004

Hox Genes and their Function in Organisms

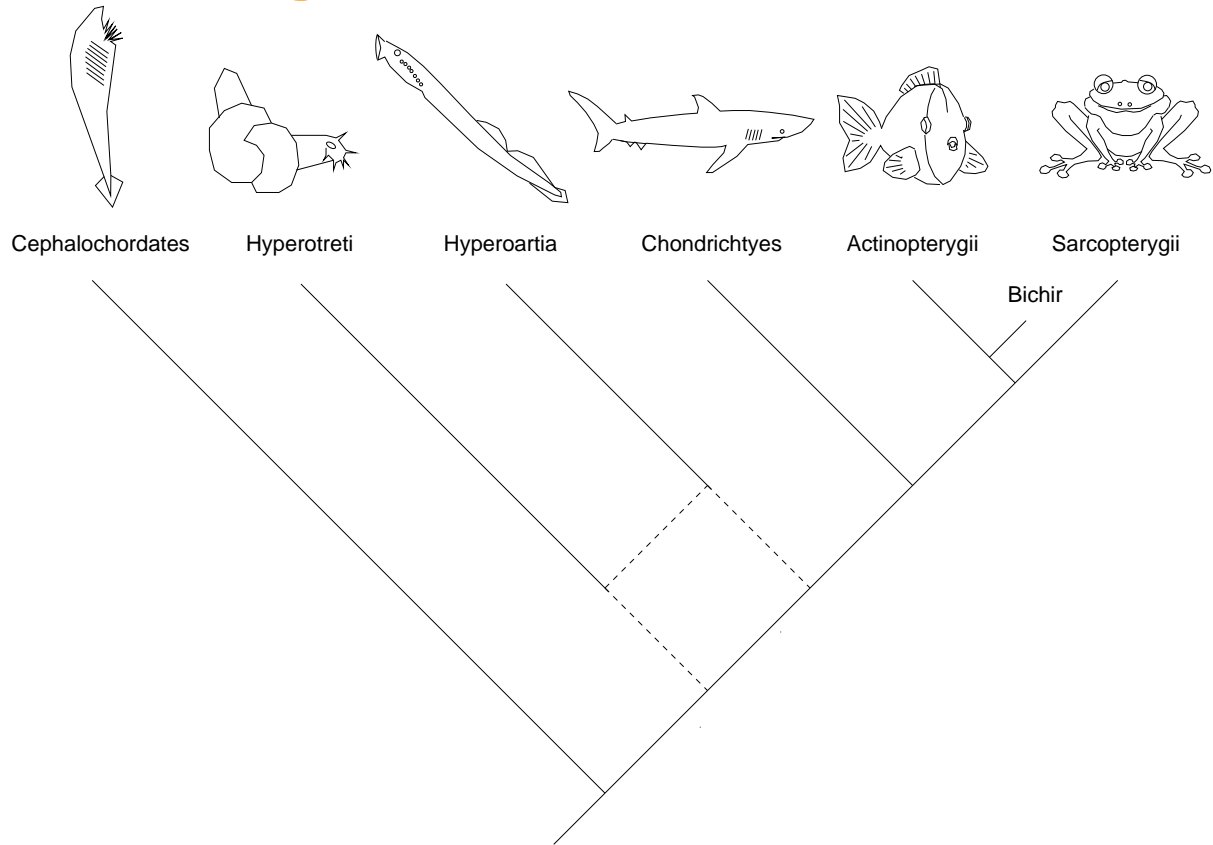


- hox genes encode a family of transcription factors operating in a regulatory cascade
- they are organized in clusters and the order of their expression matches with their sequential order
- consists of 13 paralogous genes that are related through genome duplication

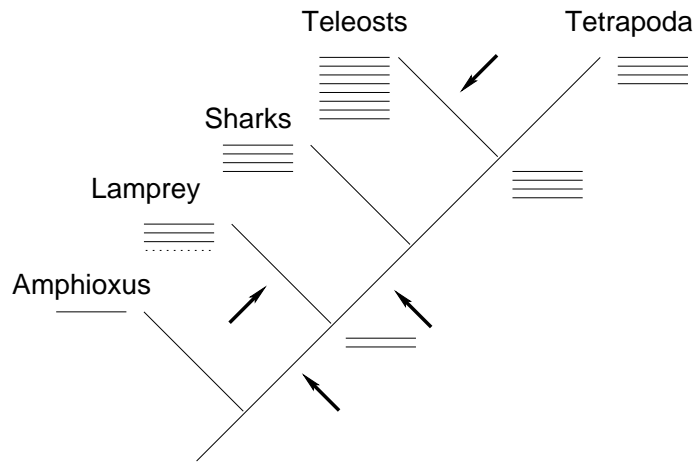
Hox Cluster Evolution



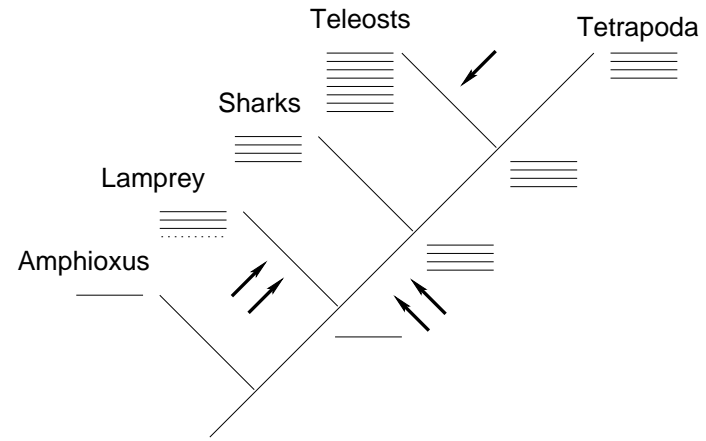
Phylogenetic Tree of Vertebrates



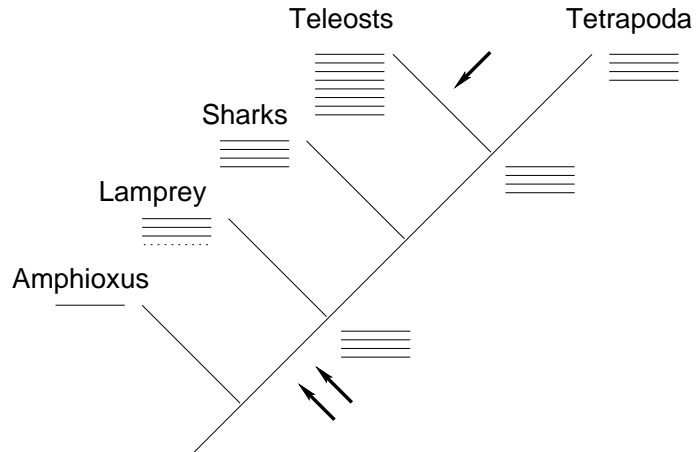
Hox-clusters in Lampreys



One independent duplication

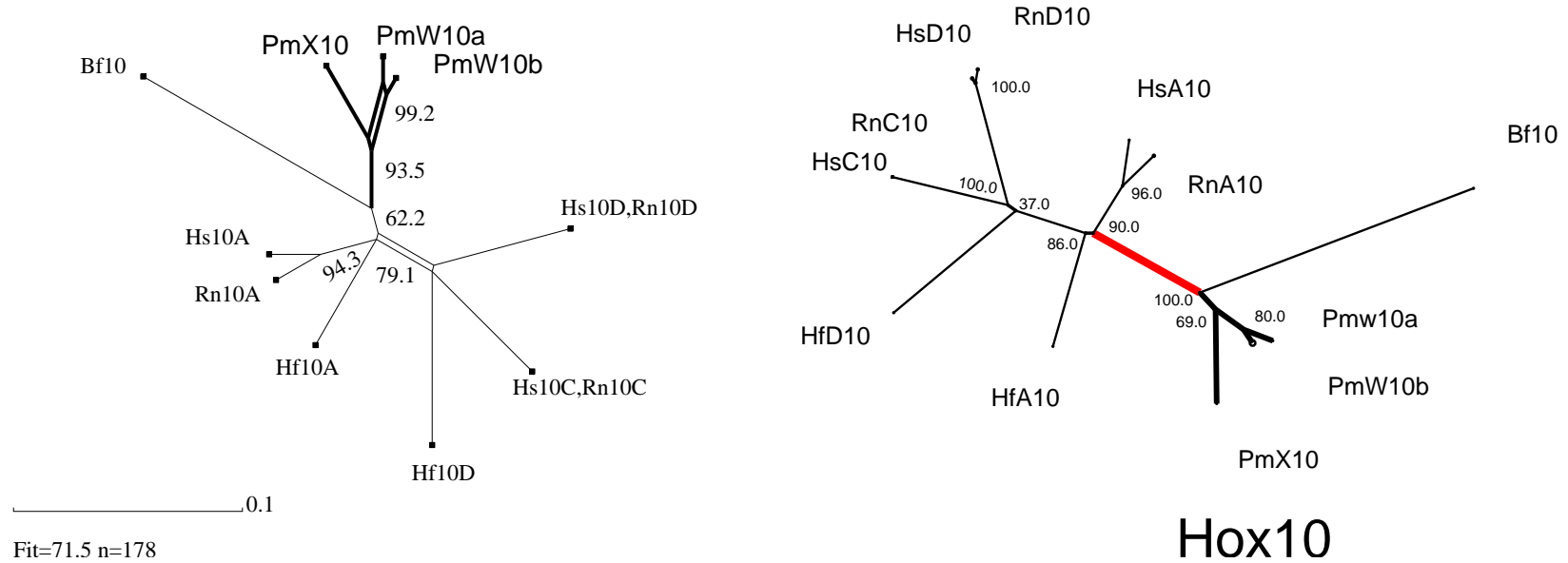


Two independent duplications



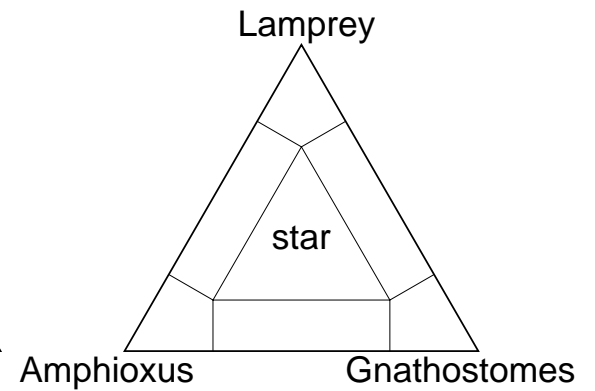
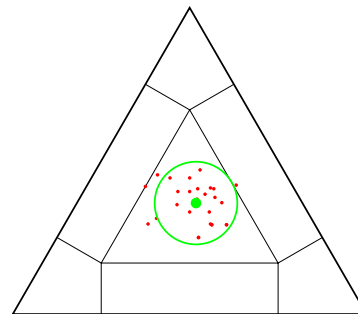
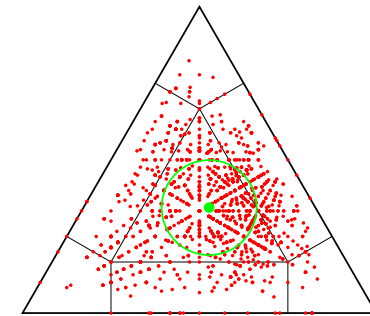
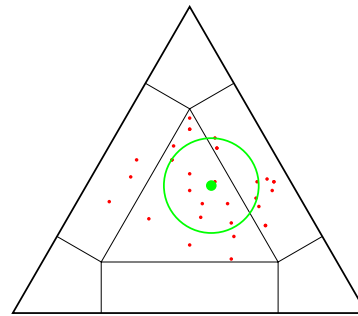
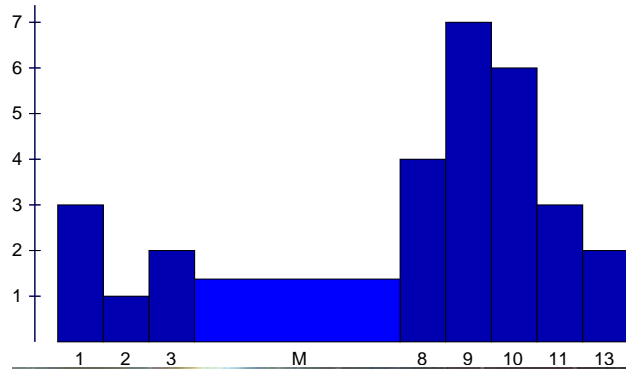
No independent duplication

Hox Cluster Duplication of Lampreys



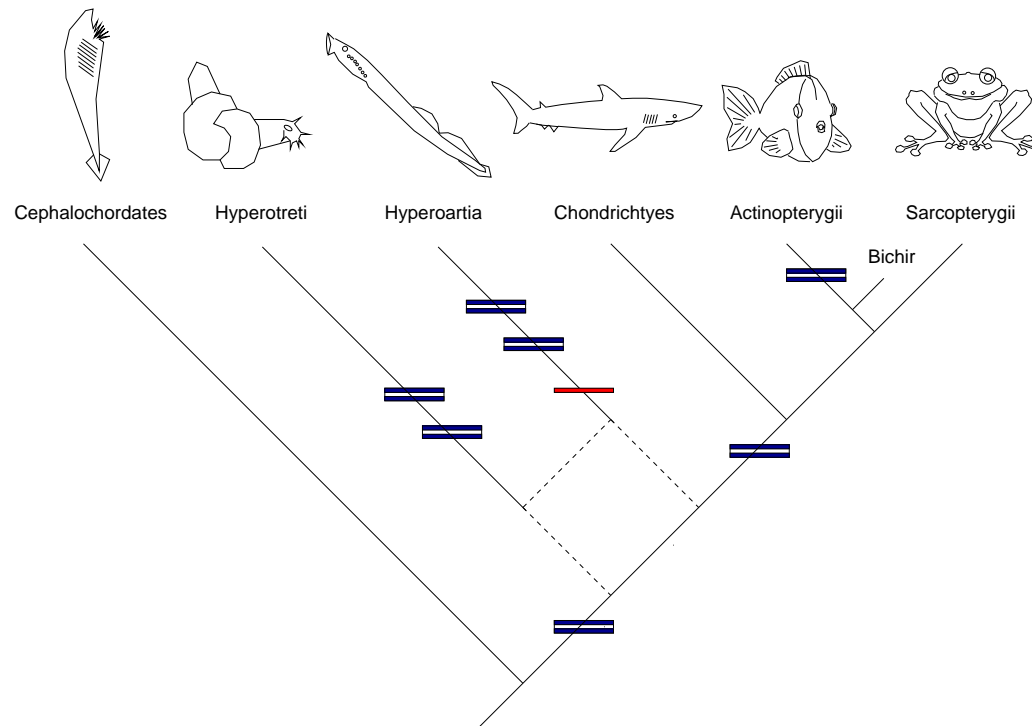
Phylogenetic relationship indicate independent duplication events

PCR Survey of a Hagfish



Weak signal for similarity of hagfish with gnathostome Hox sequences relative to Lamprey contradicts other phylogenetic signals

A Hypothesis for the Hox Cluster Evolution in Early Vertebrates



- Lamprey Hox genes are much more derived
- Probably almost complete resolution of redundancy after first duplication in Lampreys

Phylogenetic footprints

ATTAGATGGGT **GTCAGCA** TTATAGCAGCAA
AAGAAAGGCGA **ATCAGCA** TTTCAGCAGAAA
AAAAGTTGGGA **GTCAGCA** CTTTAAAAAAAAA
↑↑ ↑↑↑ ↑ ↑↑ ↑ ↑↑↑ ↑↑↑ ↑↑

- short
- gapless
- highly conserved
- mostly clustered

Phylogenetic Footprint

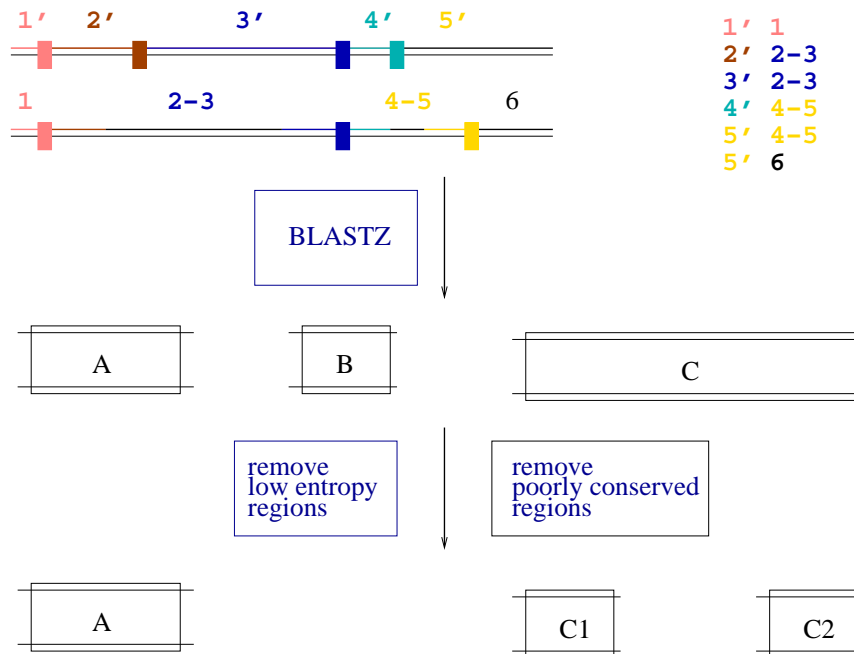
Gene regulatory regions in noncoding genomic sequences are subject to stabilizing selection and evolve slower than adjacent nonfunctional DNA.

Regulatory Elements

cis-regulatory sequence
enhancer elements

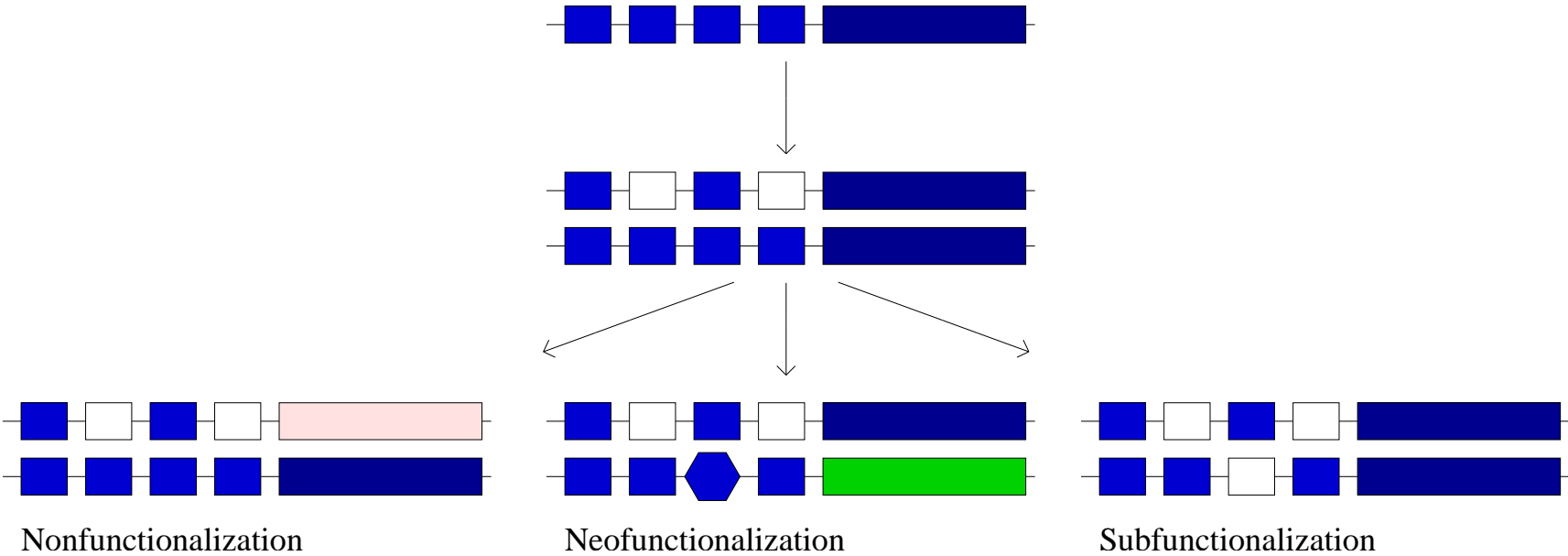
- 100-300bp long
- contain multiple short sequence-specific binding motifs for transcription factors
- carry out a regulatory function that is a subfunction of a complex regulatory pattern
- executes this subfunction when linked to a reporter gene
- may be separated from the target gene by several 100kb

The tracker approach

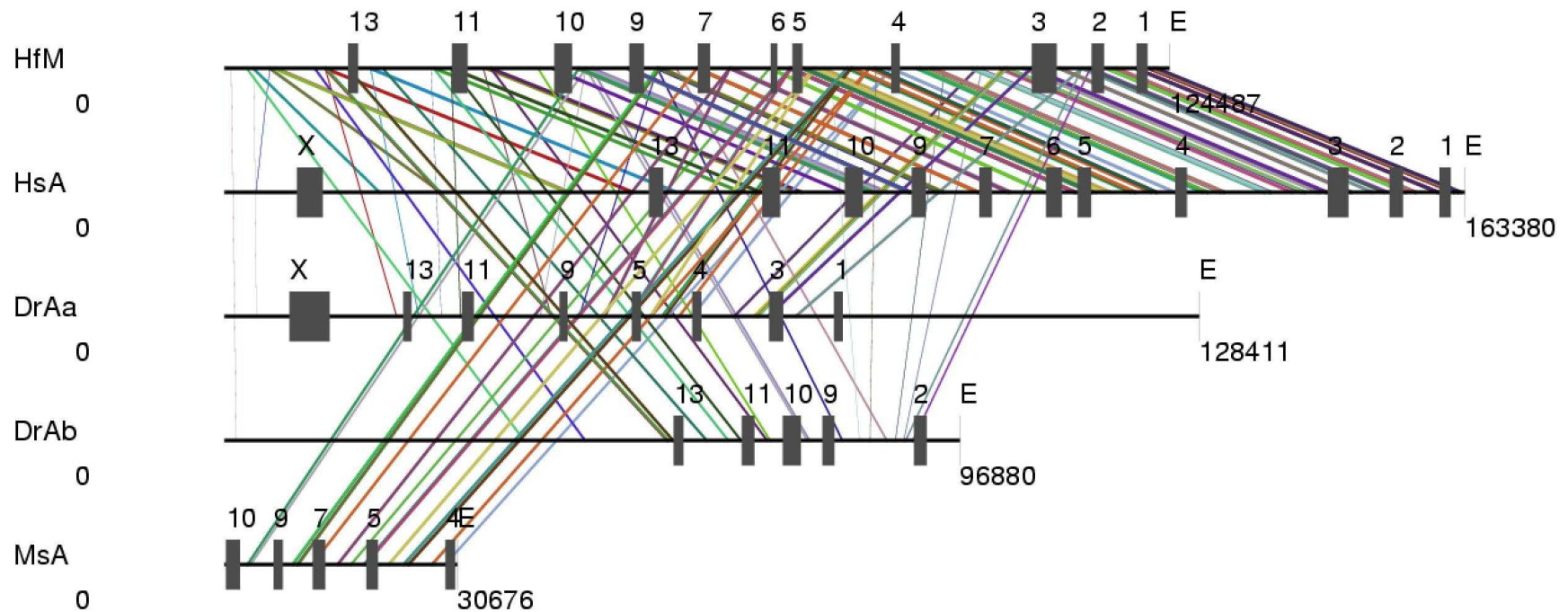


Footprint	HsA	HsB	HsC	HsD
109			6159 39	36668 43
110			6674 32	40204 32
111		46526 30	27862 30	
112	58930 58	46757 77		
113	59976 49			99324 49
114		49778 37		93768 37
115		50076 58	28353 54	
116		51297 30		100243 30
117		54151 23	32951 23	
118		59861 19		100473 19
119			33681 47	101401 47
120			33734 44	102148 44
121		60068 27	37179 27	
122		60997 42	37449 42	
123			38192 27	104539 27
124			52761 42	114312 42
125			52861 48	114411 47
126		70060 29	57886 29	
127		72845 31	45157 31	
128		78573 60		107279 70
129		79503 30	59852 30	
130		80399 48	38601 48	
131		80751 137	15959 138	
132		80786 46	16115 43	
133		80887 28	16197 28	
134		88113 93	15965 96	
135		88235 21	16089 21	
136		88312 62	16163 62	
137		92264 35		96351 35
138		94387 49	60658 49	
139		97213 18	39684 18	
140		98365 34	60894 34	
141		99301 152	15960 155	
142		99498 67	16157 68	
143		100192 20	50350 20	
144		100374 144	15959 145	

Footprint Evolution after Gen Duplication



Regulatory pattern of HoxA cluster



Shark has much more in common with **human** than other **fish species**

Teleosts share only a few regulatory regions

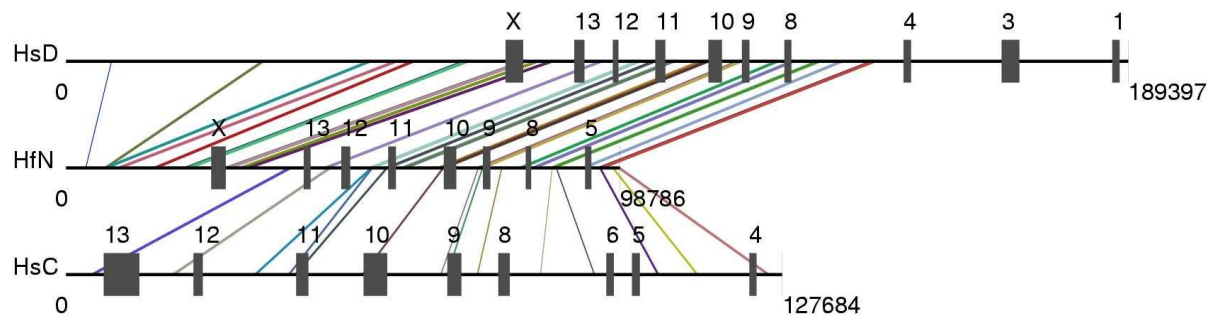
Few regulatory sequences are shared between all taxa

Identity of the Shark HoxN Cluster

percentages of protein sequence identities

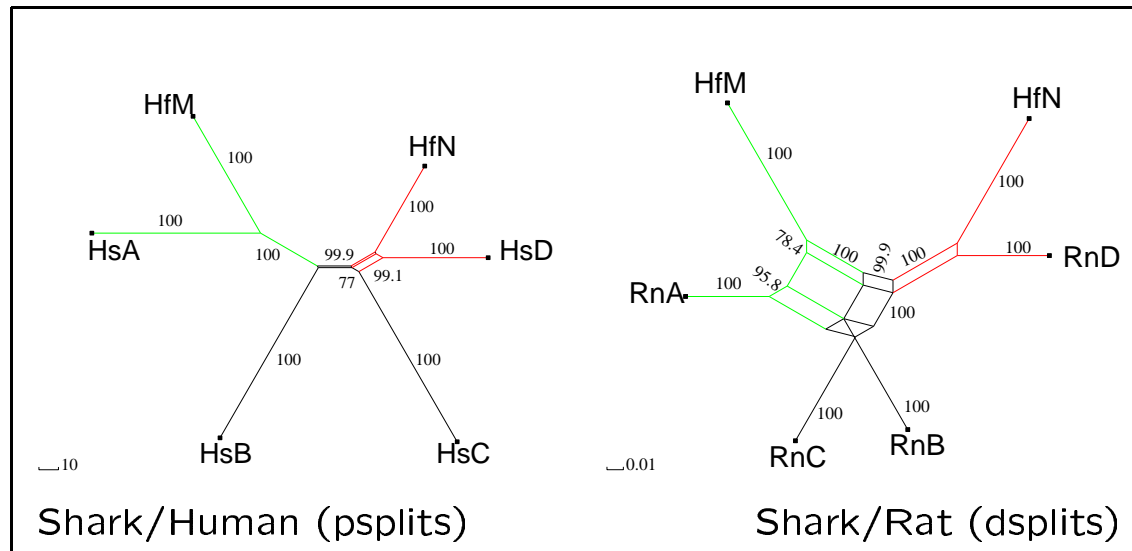
Cluster	evx	13	12	11	10	9	8	5
A	—	70	—	57	—	63	—	53
B	—	—	—	—	—	—	68	48
C	—	—	48	54	—	63	69	44
D	81	68	48	57	69	61	71	—

phylogenetic footprint cliques



Identity of the Shark HoxN Cluster

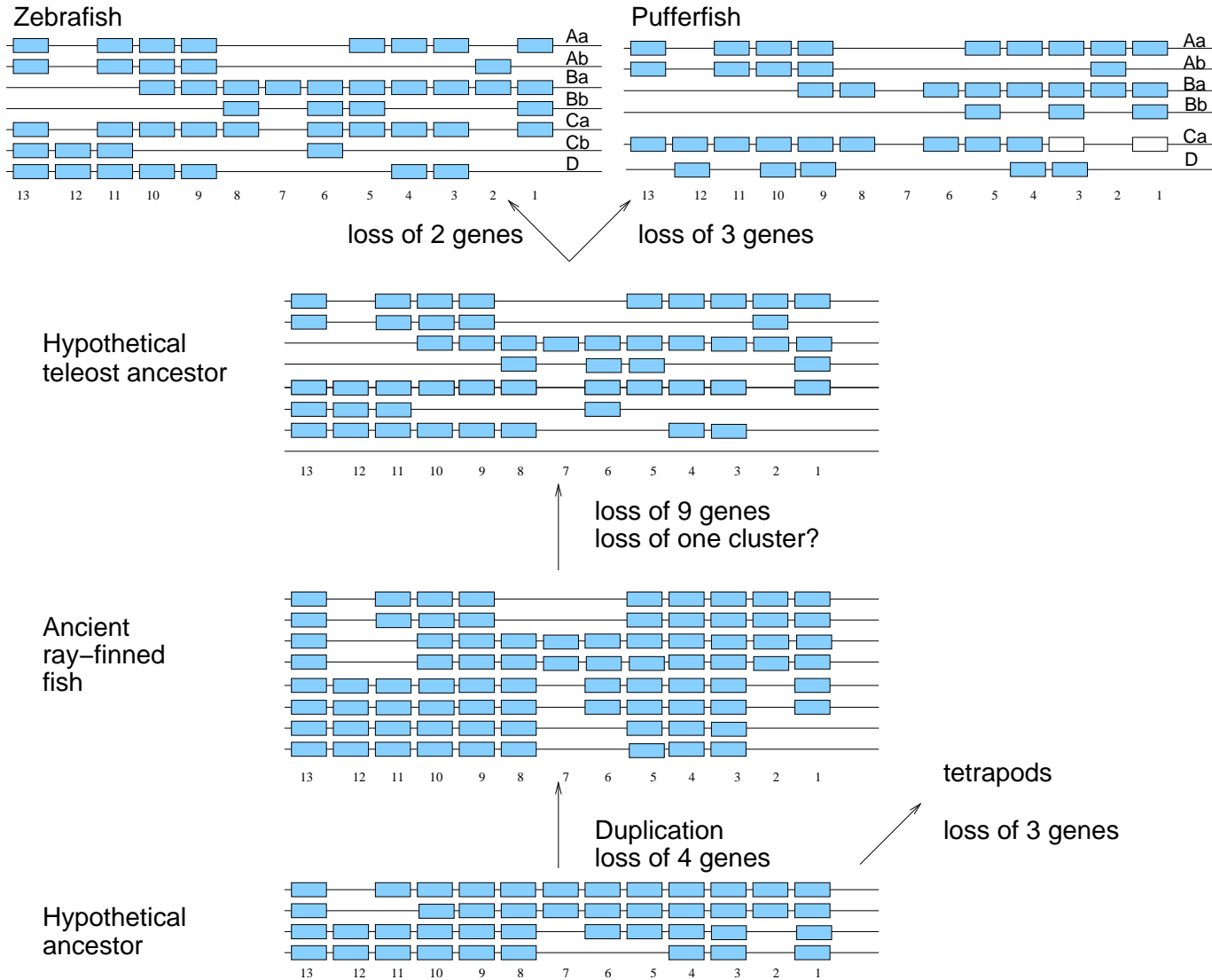
	Shark <i>HoxM</i>				Shark <i>HoxN</i>			
	<i>HoxA</i>	<i>HoxB</i>	<i>HoxC</i>	<i>HoxD</i>	<i>HoxA</i>	<i>HoxB</i>	<i>HoxC</i>	<i>HoxD</i>
	<i>evx</i> to <i>hox-4</i>							
Cliques	47	21	13	9	15	10	20	25
Length	3847	1905	646	1065	1728	961	1148	1995



Joint work with Frank Ruddle, Chris Amemiya

HoxN of shark is most closely related to mammalian *HoxD*

Hox Clusters in Teleost Fishes

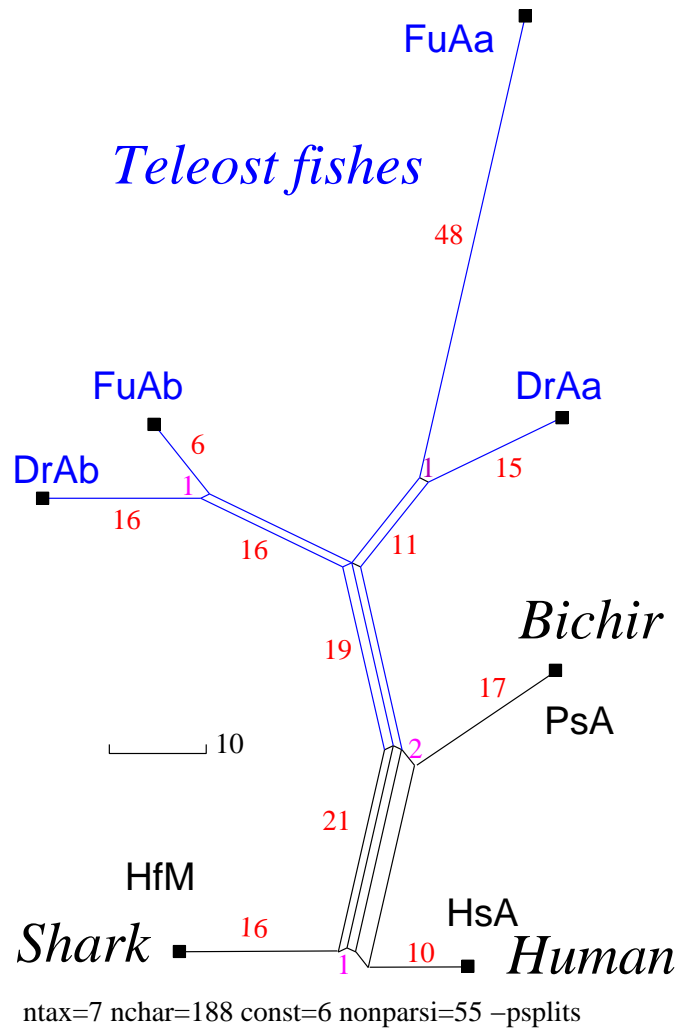


Bichir HoxA cluster



Joint work with C-h. Chiu *et al.*

Genomic evolution of a basal actinopterygian



- Higher teleosts underwent an additional Hox cluster duplication
- No duplication in bichir
- Significant teleost-specific footprints already exist in the bichir

What happens to footprints after duplication?

A Model For Footprint Evolution:

- (1) footprints are lost with their associated gene
- (2) stochastic resolutions for retained duplicates
- (3) footprints are lost if regulating gene is lost
- (4) non-structural contributions

$$r(F) = \underbrace{r(G)}_{(1)} \times \underbrace{\left[\frac{1}{2}P_1 + (1 - P_1) \right]}_{(2)} \times \underbrace{(1 - dQ)}_{(3)} \times \underbrace{(1 - \alpha)}_{(4)}$$

$r(G)$...	retention rate of genes in focus cluster
P_1	...	fraction of genes with 1st order paralogs
Q	...	fraction of extinct genes in entire network
d	...	fraction of genes from which a gene is regulated (Hox case: $d \approx 1$)

The parameter α measures *binding site turnover* and *adaptive effects*.

α-values of Teleost HoxA Cluster

Cluster	#genes	$P(1^{st})$	$r(F)/r(G)$		α
			data	structural	
<i>DrHoxAa</i>	7	0.43	0.49	0.69	0.29
<i>DrHoxAb</i>	5	0.60	0.51	0.62	0.18
<i>DrHoxA</i>	12	0.50	0.49	0.66	0.26
<i>TrHoxAa</i>	9	0.56	0.45	0.58	0.22
<i>TrHoxAb</i>	5	1.00	0.21	0.40	0.48
<i>TrHoxA</i>	14	0.71	0.37	0.52	0.29

Increased adaptive evolution in the Fugu HoxAb cluster?

Overview

