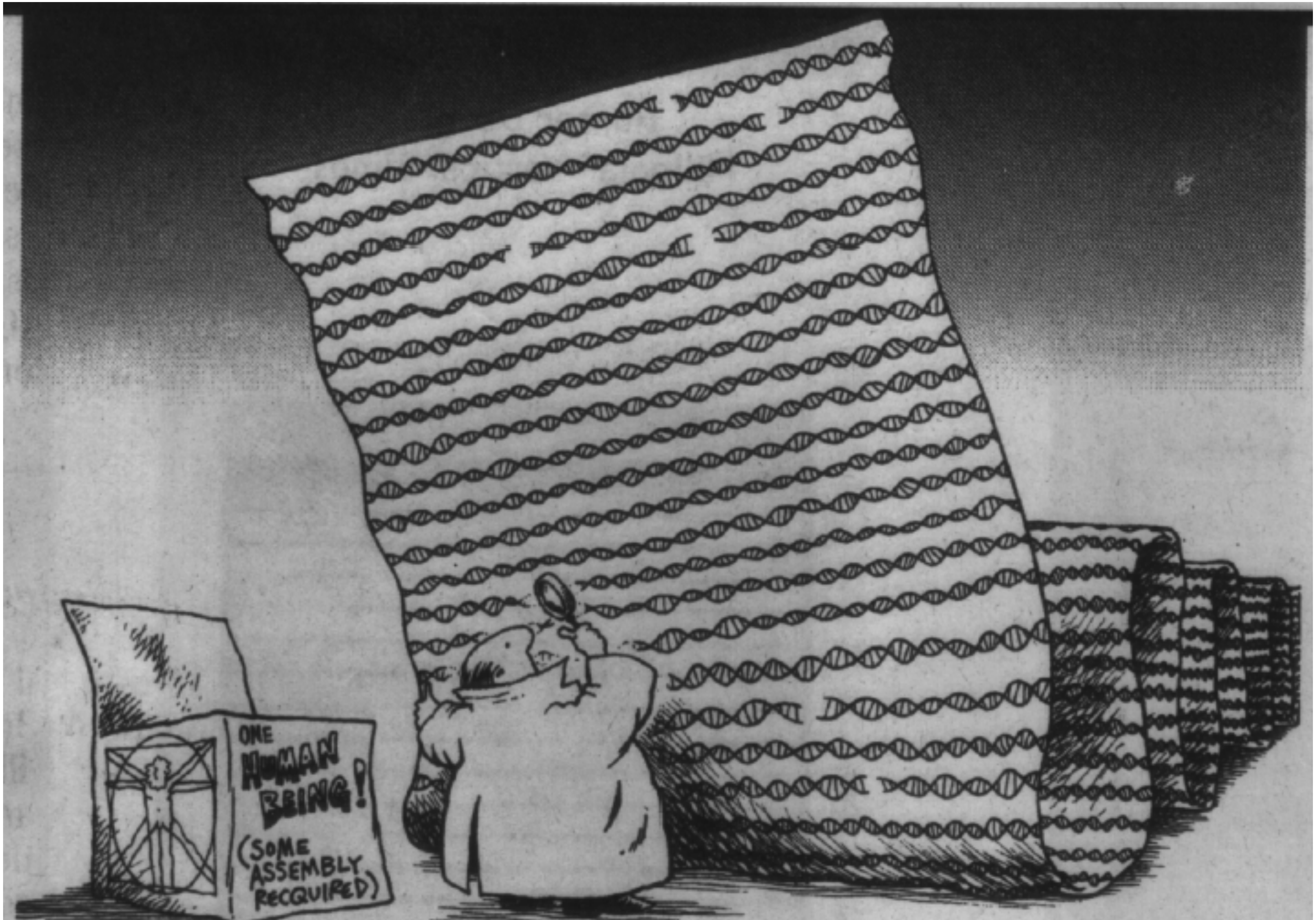
Axel Meyer
Universität Konstanz
Germany

Genome duplications during the evolution of vertebrates

Woods Hole Molecular Evolution Course 2011



BY AUTH FOR THE PHILADELPHIA INQUIRER

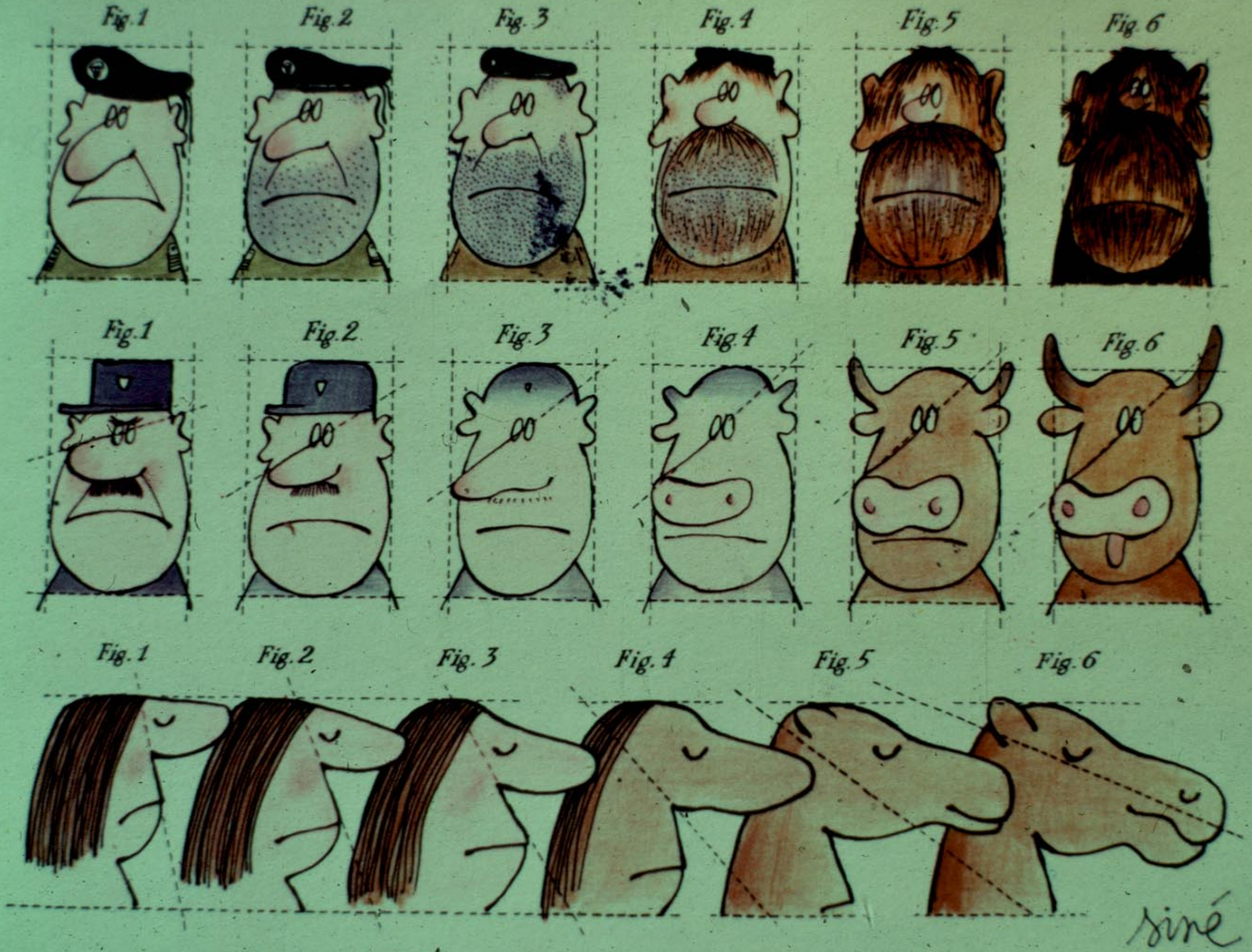
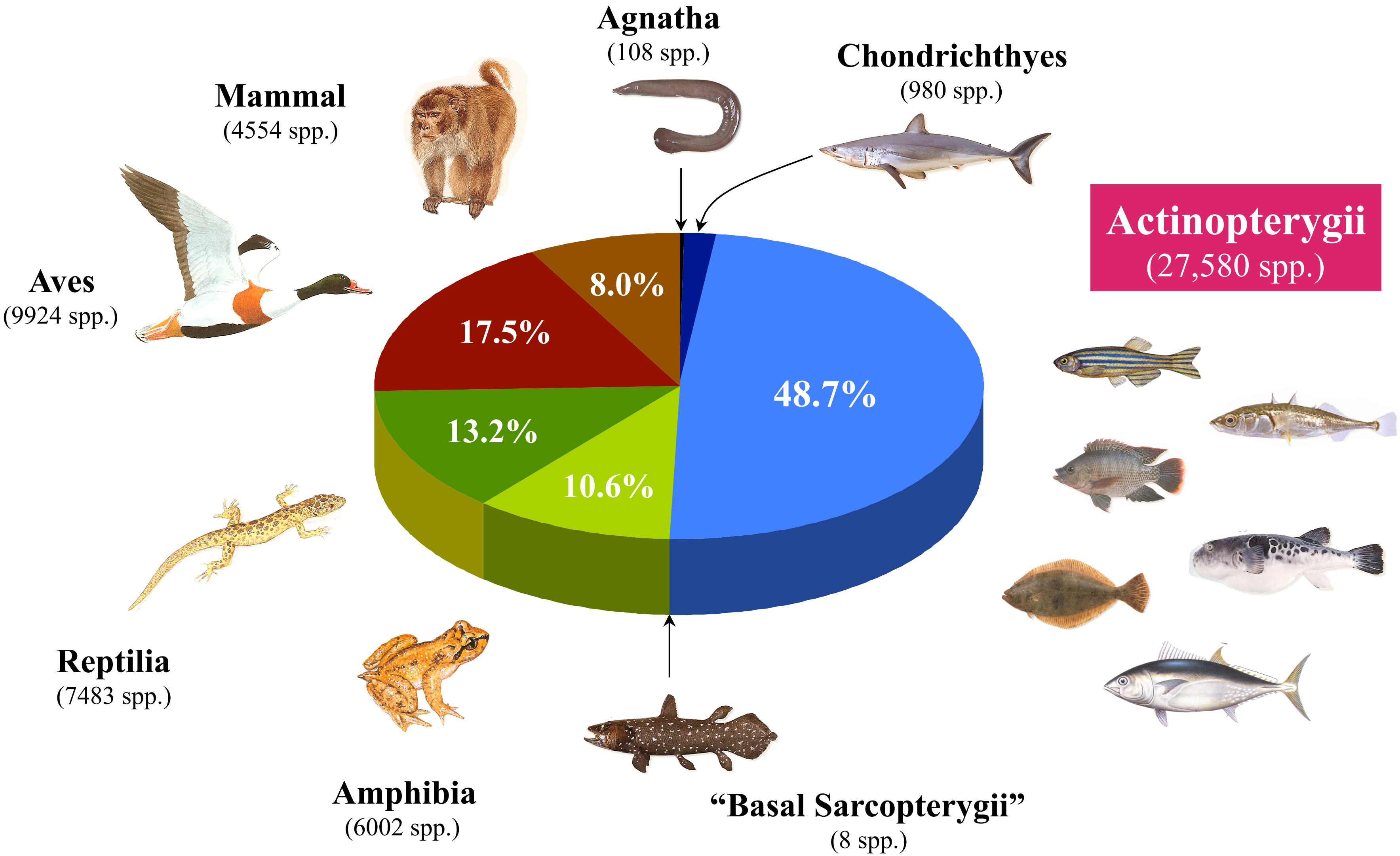


Fig. 1
Fig. 2
Fig. 3
Fig. 4
Fig. 5
Fig. 6
Fig. 1
Fig. 2
Fig. 3
Fig. 4
Fig. 5
Fig. 6
Fig. 1
Fig. 2
Fig. 3
Fig. 4
Fig. 5
Fig. 6
siné


Cavalier King Charles
Treeing Walker
Coonhound
Pharaoh Hound
Sussex Spaniel
Puli
St. Bernard
Australian Terrier
German Shepherd Dog
Smooth Collie
Standard Poodle

Ray-finned fishes comprise about 50% of the total number of vertebrates



Agnatha
(108 spp.)

Chondrichthyes
(980 spp.)

Mammal
(4554 spp.)

Aves
(9924 spp.)

Actinopterygii
(27,580 spp.)

Reptilia
(7483 spp.)

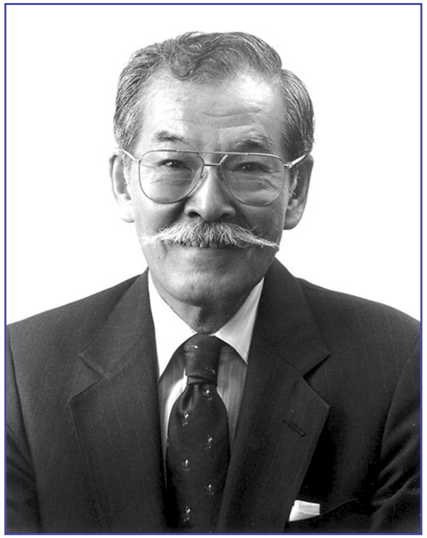
Amphibia
(6002 spp.)

“Basal Sarcopterygii”
(8 spp.)

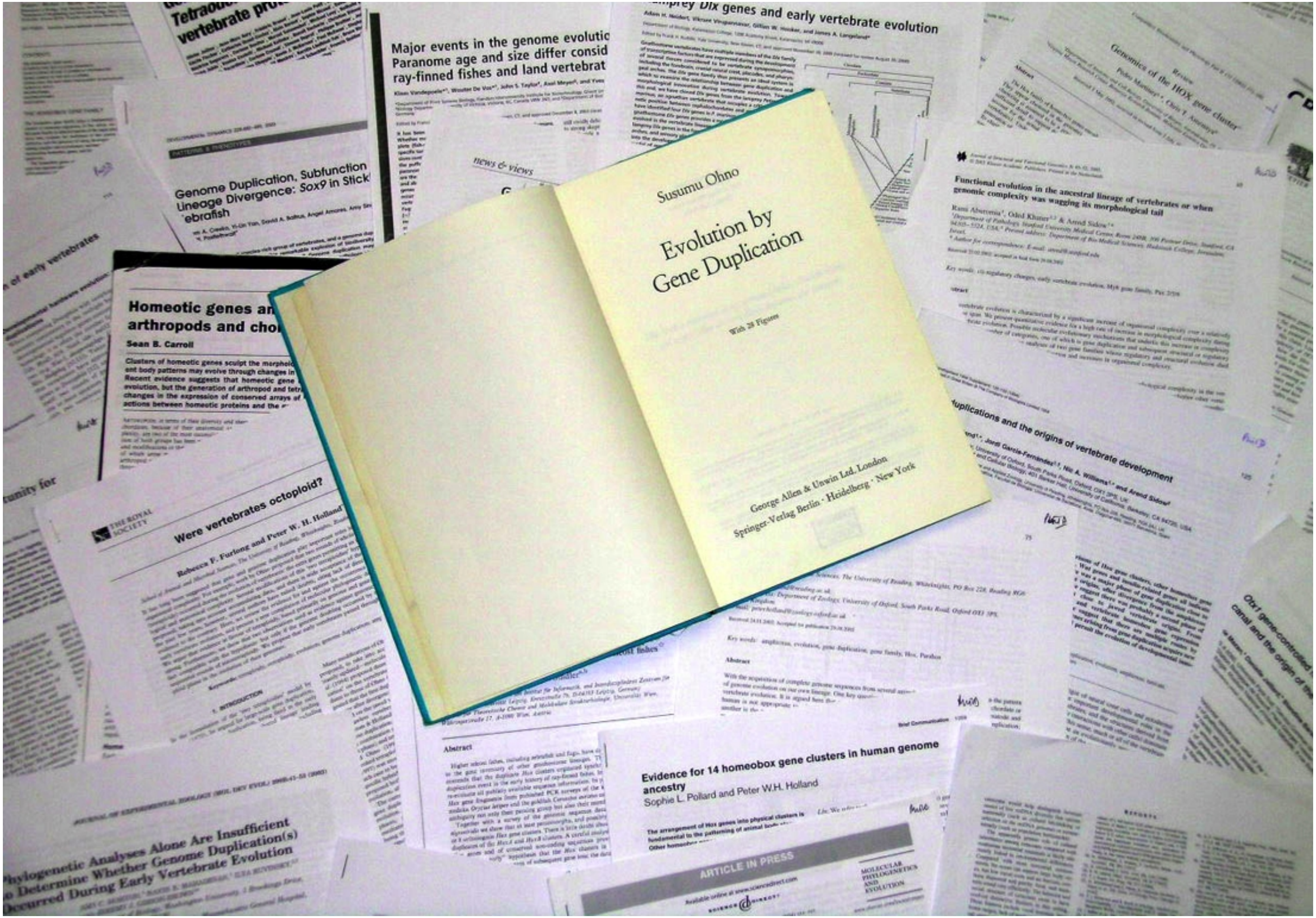
Leaps in evolution through gene and genome duplications

“Had evolution been entirely dependent upon natural selection, from a bacterium only numerous forms of bacteria would have emerged. The creation of metazoans, vertebrates and finally mammals from unicellular organisms would have been quite impossible, for such big leaps in evolution required the creation of new gene loci with previously non-existent functions”

Susumu Ohno 1970



Evolution by Gene Duplication

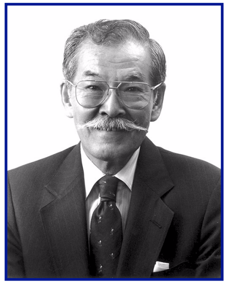

Susumu Ohno
Evolution by Gene Duplication
With 28 Figures
George Allen & Unwin Ltd. London
Springer-Verlag Berlin · Heidelberg · New York
Major events in the genome evolutic
Paranome age and size differ consid
ray-finned fishes and land vertebrat
Tetraodo
vertebrate pro
mprey Dlx genes and early vertebrate evolution
Genome Duplication, Subfunction
Lineage Divergence: Sox9 in Stick
ebrafish
news & views
Genomics of the HOX gene cluster
Functional evolution in the ancestral lineage of vertebrates or when genomic complexity was wagging its morphological tail
Homeotic genes an
arthropods and cho
Sean B. Carroll
of early vertebrates
Were vertebrates octoploid?
Rebecca F. Furlong and Peter W. H. Holland
THE ROYAL SOCIETY
uplications and the origins of vertebrate development
Evidence for 14 homeobox gene clusters in human genome ancestry
Sophie L. Pollard and Peter W.H. Holland
ARTICLE IN PRESS
MOLECULAR PHYLOGENETICS AND EVOLUTION
hylogenetic Analyses Alone Are Insufficient
o Determine Whether Genome Duplication(s)
ccurred During Early Vertebrate Evolution
Otx1 gene-controlled
canal and the origin of th
Abstract
REPORTS

Susumu Ohno

Book Reviews
Phylogenetic Mechanism Evolution by Gene Duplication. SUSUMU OHNO. Springer-Verlag, New York, 1970. xvi, 160 pp., illus. $16.50.

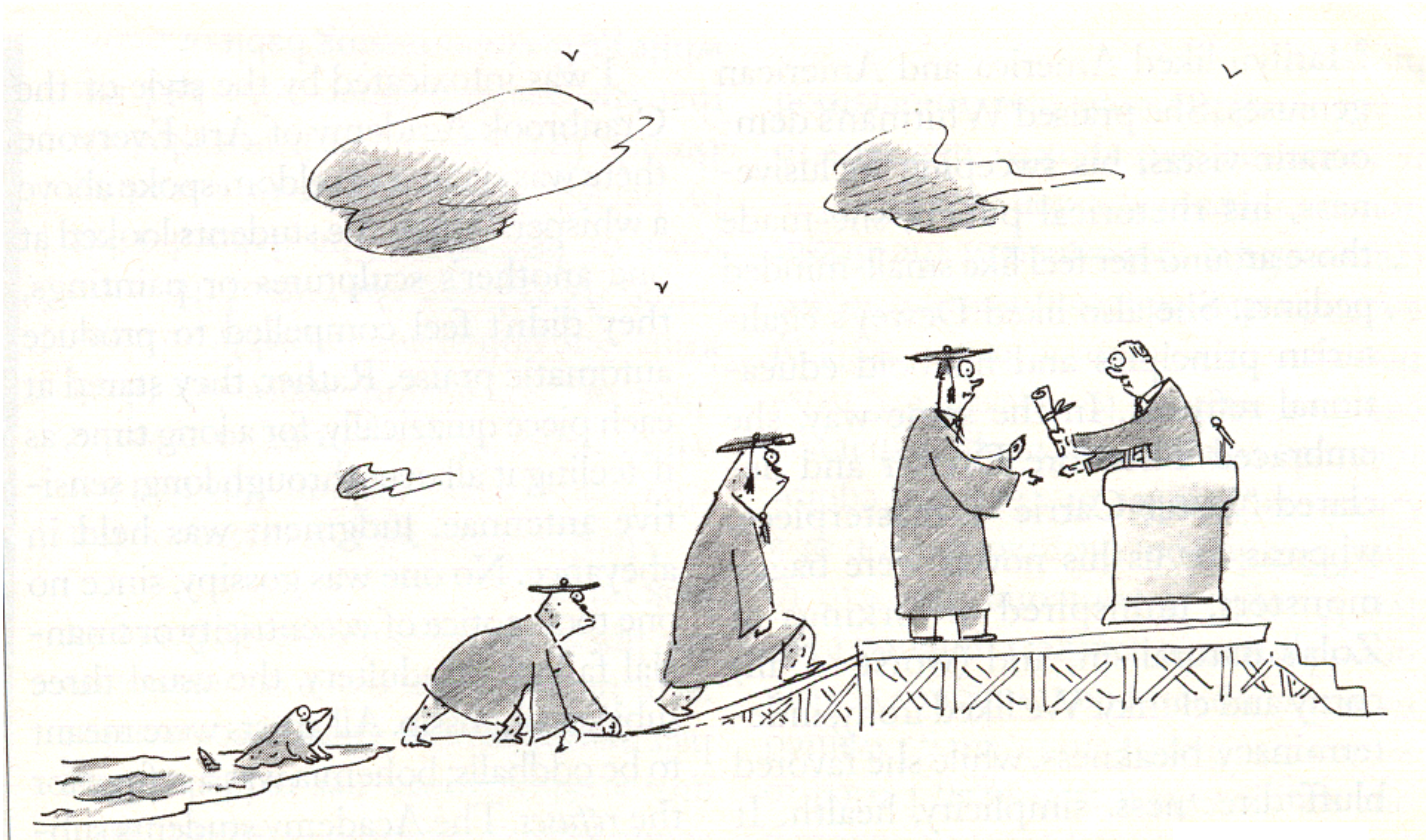
Blue Crossopterygian Man dominates the dust jacket. A forewarning? Ah yes-you are warned: you will be stimulated, cajoled, and, unless you have a more than passing acquaintance with at least one of the subject matters from which the author spins his argument, perhaps misled.

Science 175: 617-618 (1972) by Janice Spofford



"In this golden age of biology, a book faces the danger of becoming obsolete before its publication. It is my belief that in order to avoid early obsolescence, the author, judging on the basis of the scant evidence available, is obliged to anticipate future developments and paint a picture with broad strokes of his brush. This I have done rather freely in this book"

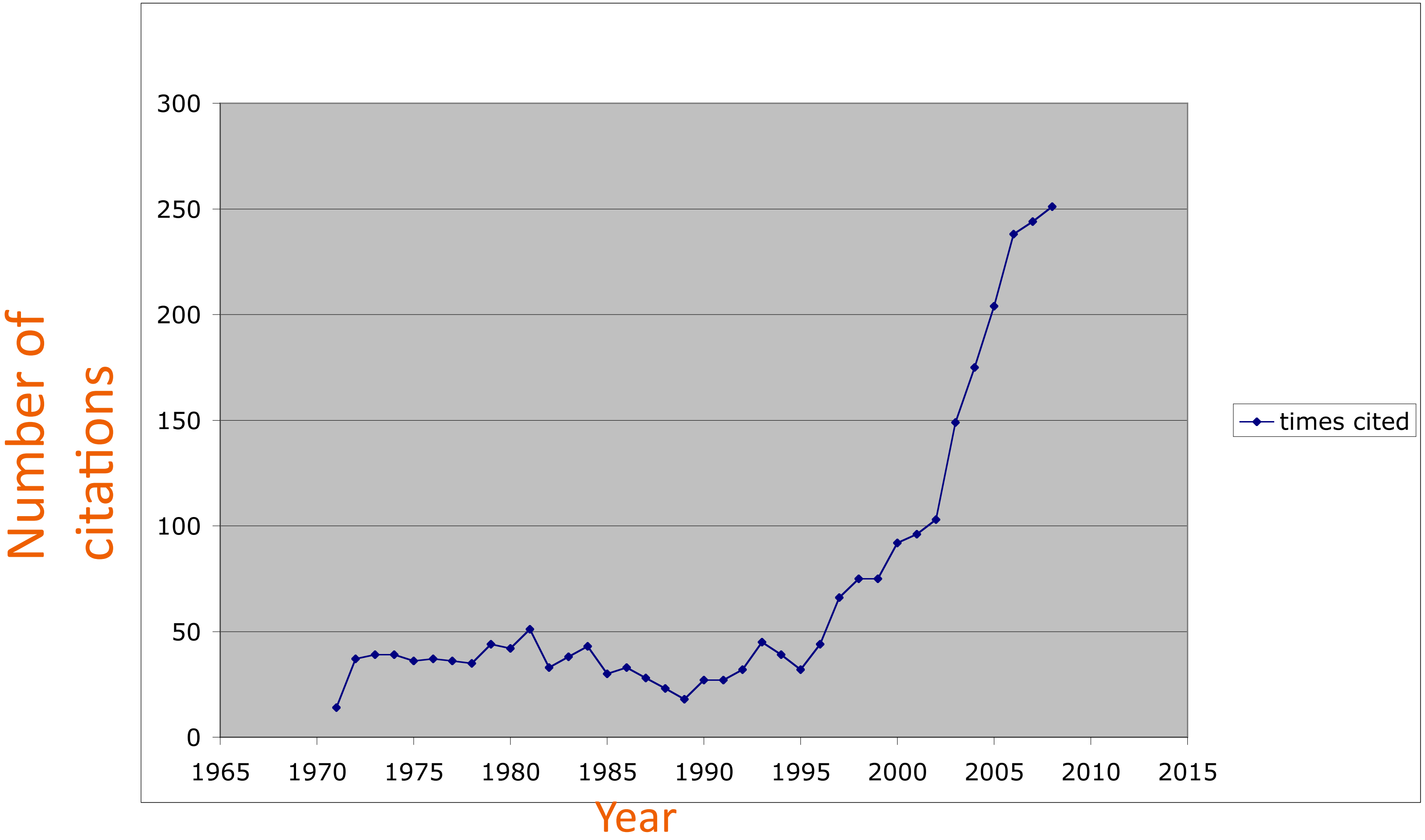
1970 - Evolution by gene duplication



“natural selection merely modified while redundancy created”

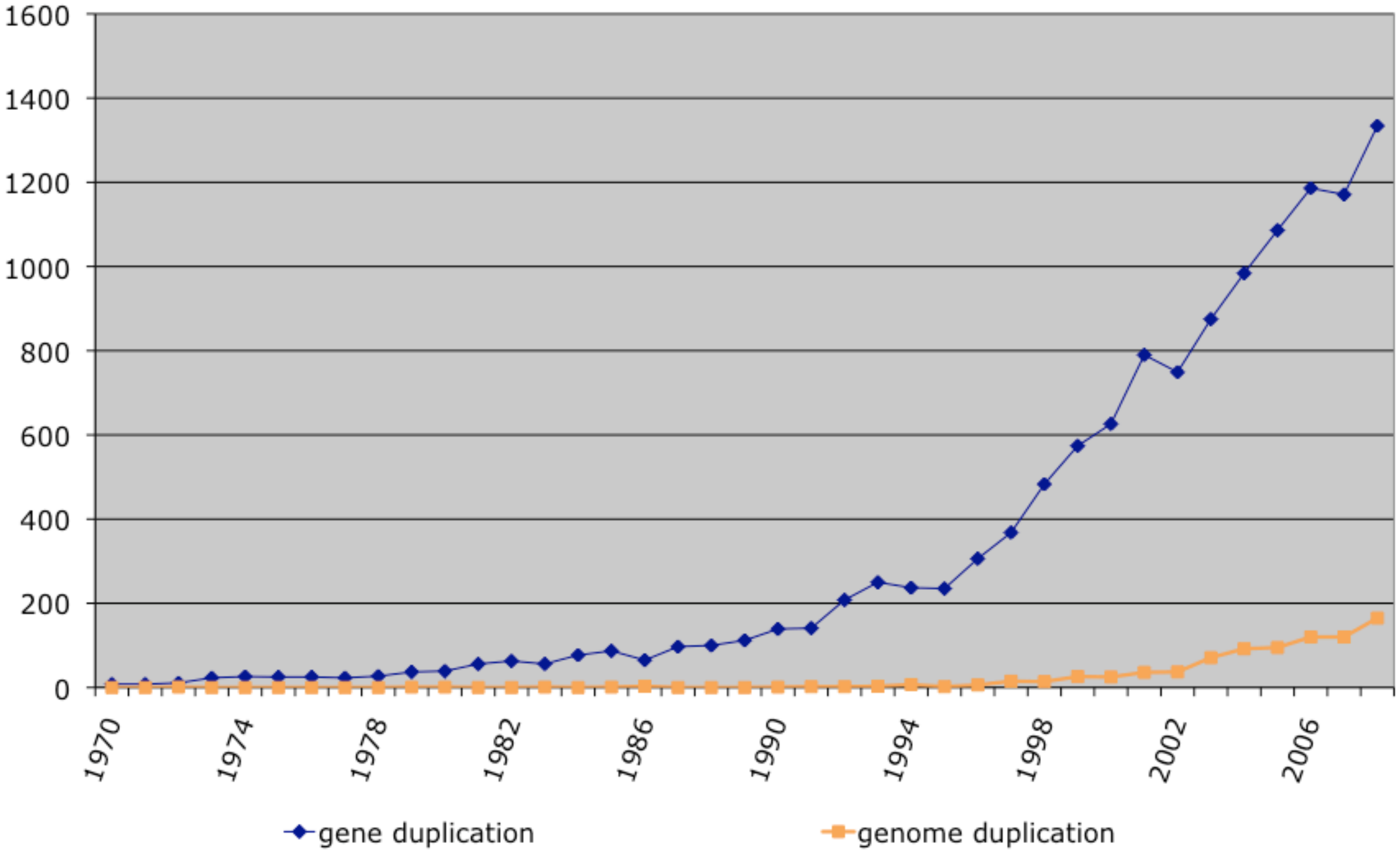
Susumu Ohno (1970)

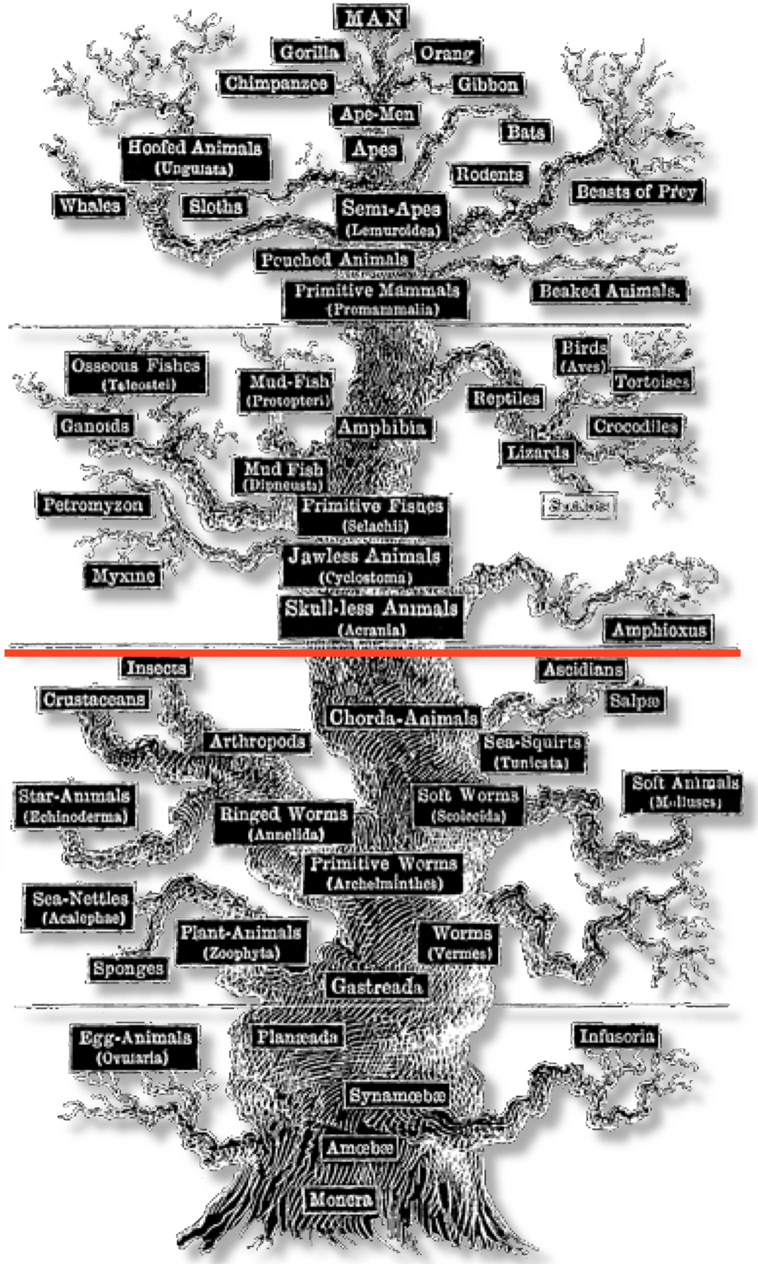
Number of citations of Susumu Ohno's 1970 book "Evolution by gene duplication"



Since Ohno's 1970 book
"Evolution by gene duplication"

number of papers on gene and genome duplications



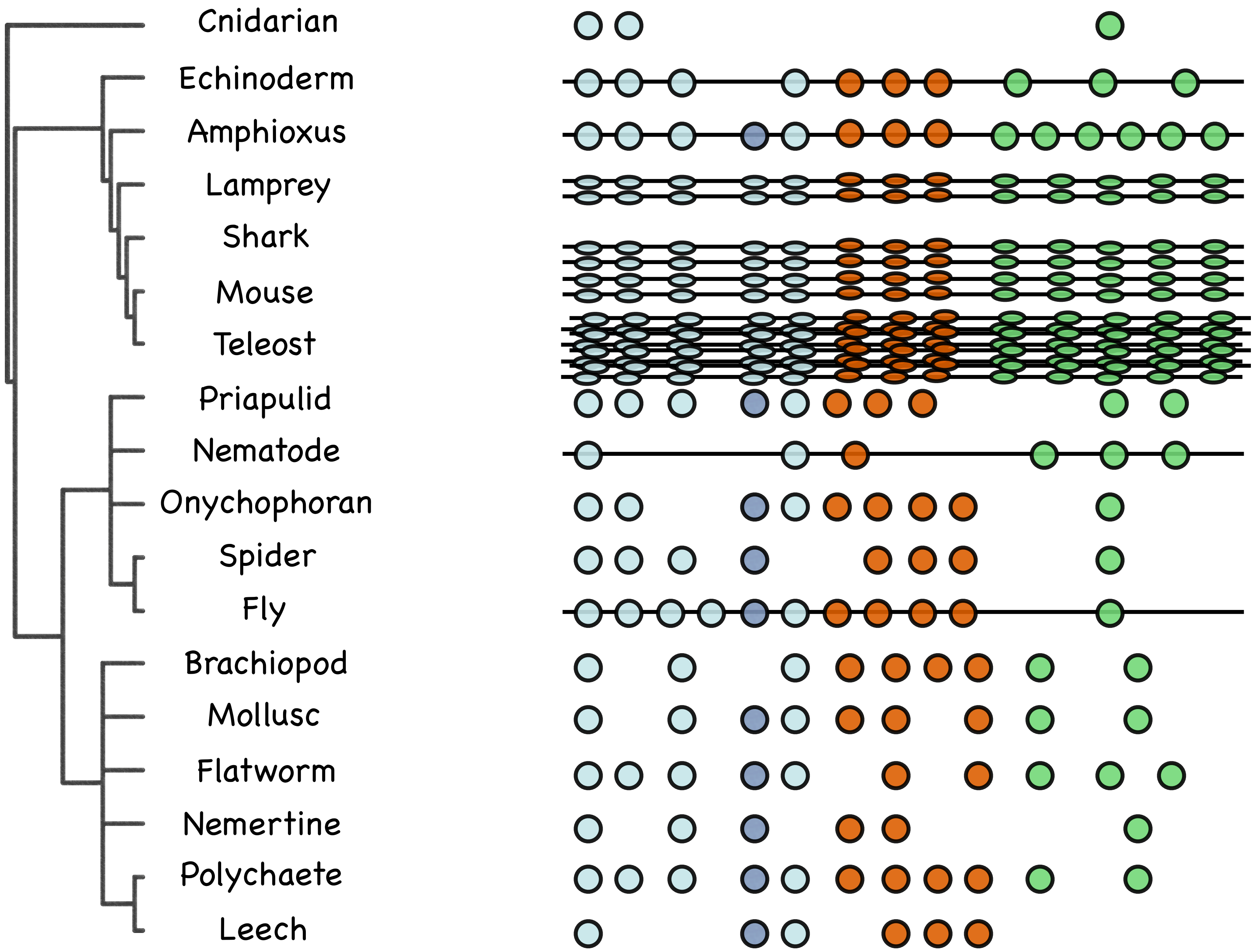


Mammals

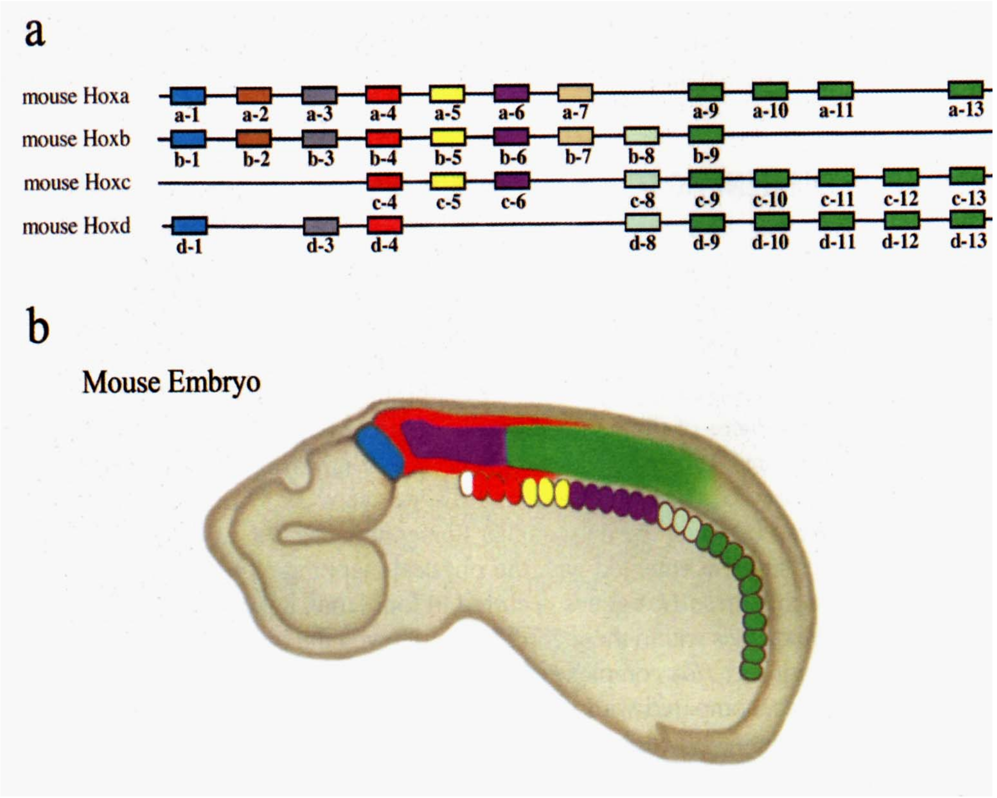
Vertebrates

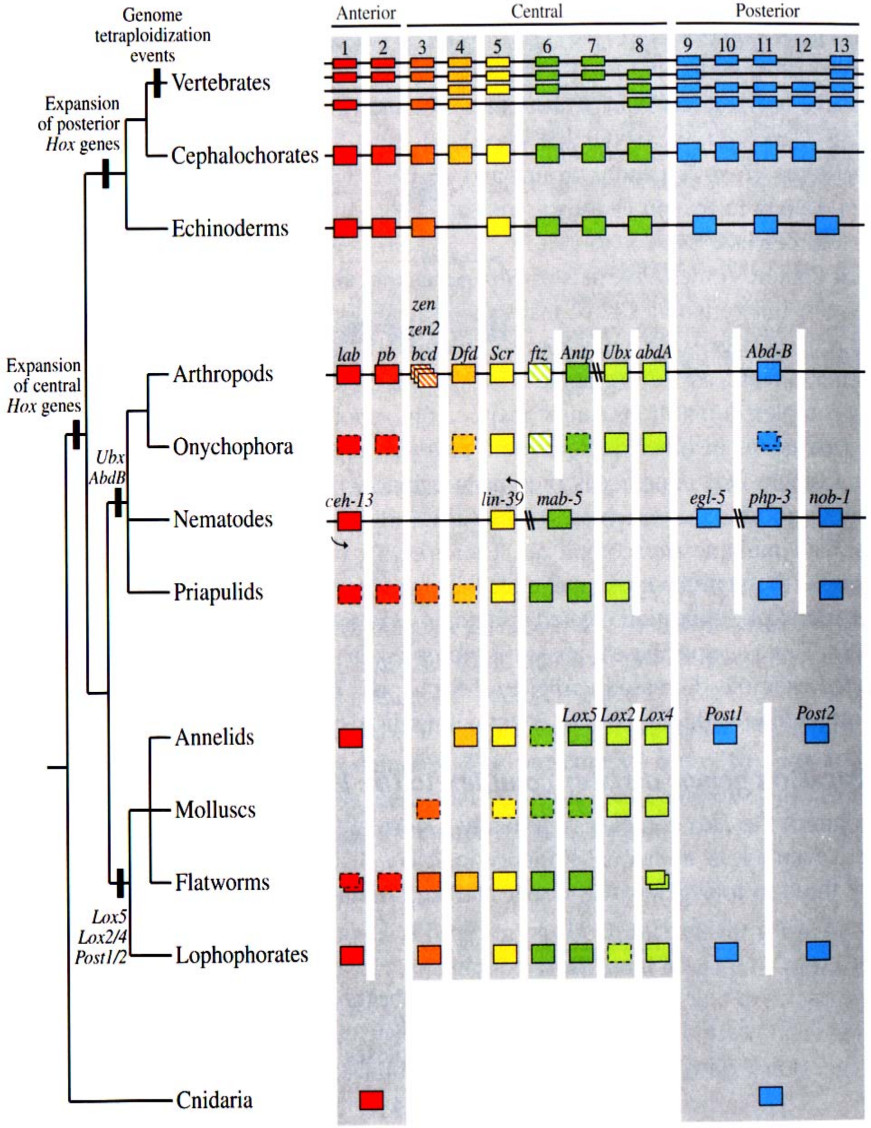
Invertebrate Intestinal animals

Primitive animals


Cnidarian
Echinoderm
Amphioxus
Lamprey
Shark
Mouse
Teleost
Priapulid
Nematode
Onychophoran
Spider
Fly
Brachiopod
Mollusc
Flatworm
Nemertine
Polychaete
Leech

Hox gene cluster evolution in animals





Carroll *et al*. (2001) “From DNA to Diversity”

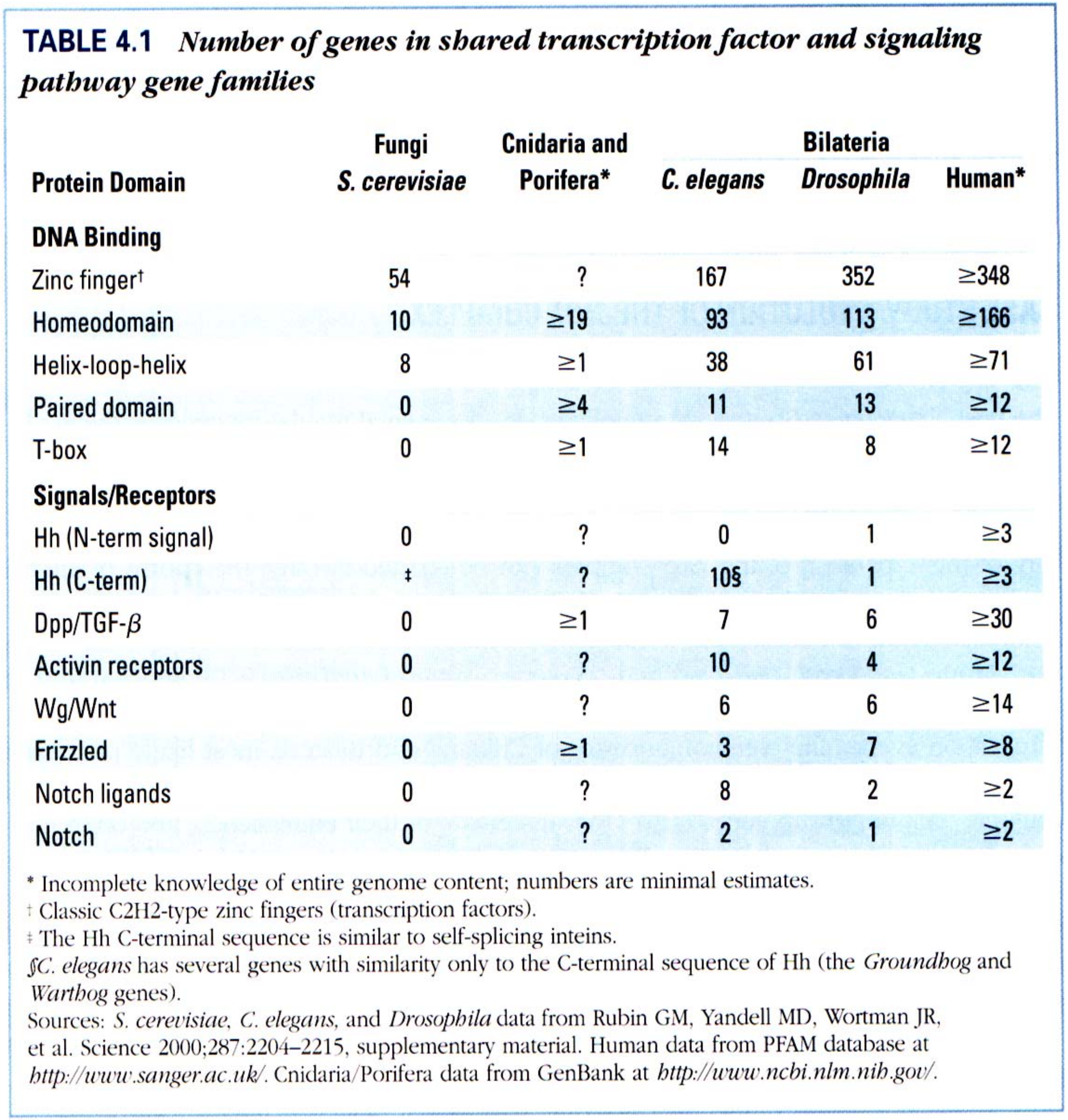


TABLE 4.1 ***Number of genes in shared transcription factor and signaling pathway gene families***

	Fungi	Cnidaria and	Bilateria		
Protein Domain	***S. cerevisiae***	**Porifera***	***C. elegans***	***Drosophila***	**Human***
DNA Binding					
Zinc finger[†]	54	?	167	352	≥348
Homeodomain	10	≥19	93	113	≥166
Helix-loop-helix	8	≥1	38	61	≥71
Paired domain	0	≥4	11	13	≥12
T-box	0	≥1	14	8	≥12
Signals/Receptors					
Hh (N-term signal)	0	?	0	1	≥3
Hh (C-term)	‡	?	10§	1	≥3
Dpp/TGF-β	0	≥1	7	6	≥30
Activin receptors	0	?	10	4	≥12
Wg/Wnt	0	?	6	6	≥14
Frizzled	0	≥1	3	7	≥8
Notch ligands	0	?	8	2	≥2
Notch	0	?	2	1	≥2

* Incomplete knowledge of entire genome content; numbers are minimal estimates.

† Classic C2H2-type zinc fingers (transcription factors).

‡ The Hh C-terminal sequence is similar to self-splicing inteins.

§*C. elegans* has several genes with similarity only to the C-terminal sequence of Hh (the *Groundhog* and *Warthog* genes).

Sources: *S. cerevisiae*, *C. elegans*, and *Drosophila* data from Rubin GM, Yandell MD, Wortman JR, et al. Science 2000;287:2204–2215, supplementary material. Human data from PFAM database at *http://www.sanger.ac.uk/*. Cnidaria/Porifera data from GenBank at *http://www.ncbi.nlm.nih.gov/*.

Carroll *et al*. (2001) “From DNA to Diversity”

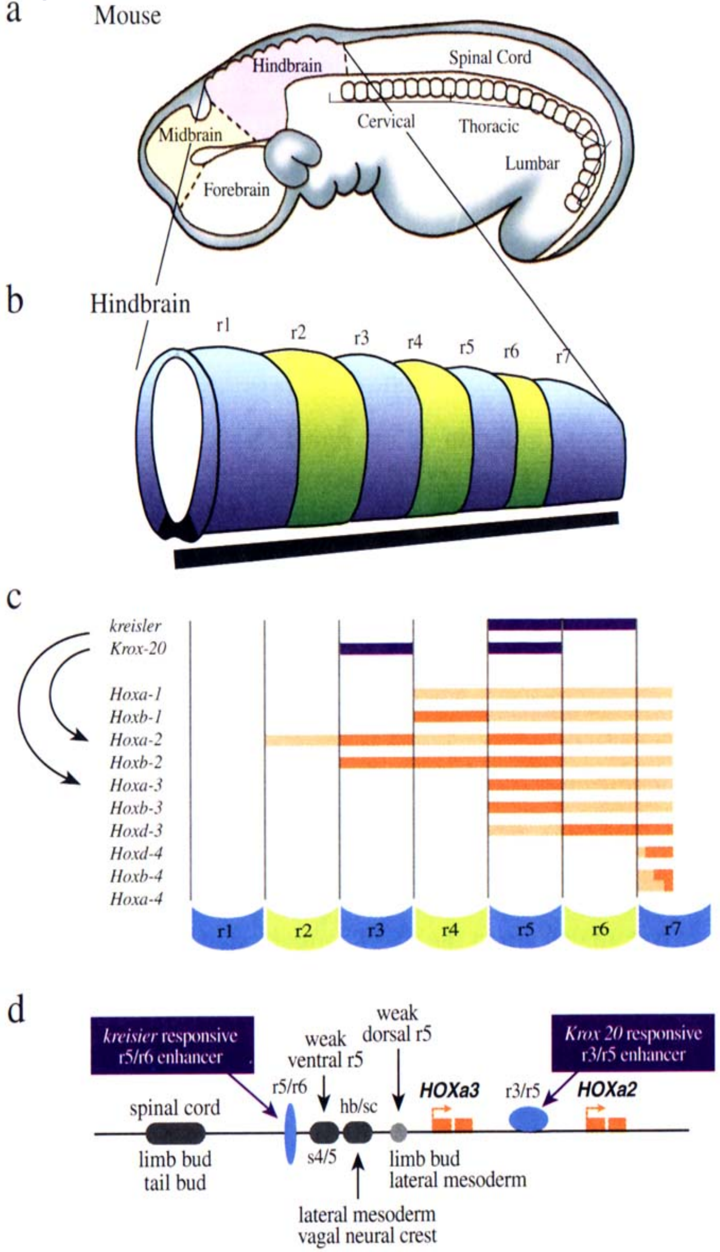
The current dogma

“Comparisons of developmental gene regulation between morphologically divergent animals, analyses of intraspecific variation, and the response of organisms and genes to selection all support the claim that regulatory DNA is the predominant source of the genetic diversity that underlies morphological variation and evolution.”

Sean Carroll, *Cell* 2000 (pp. 577–580)

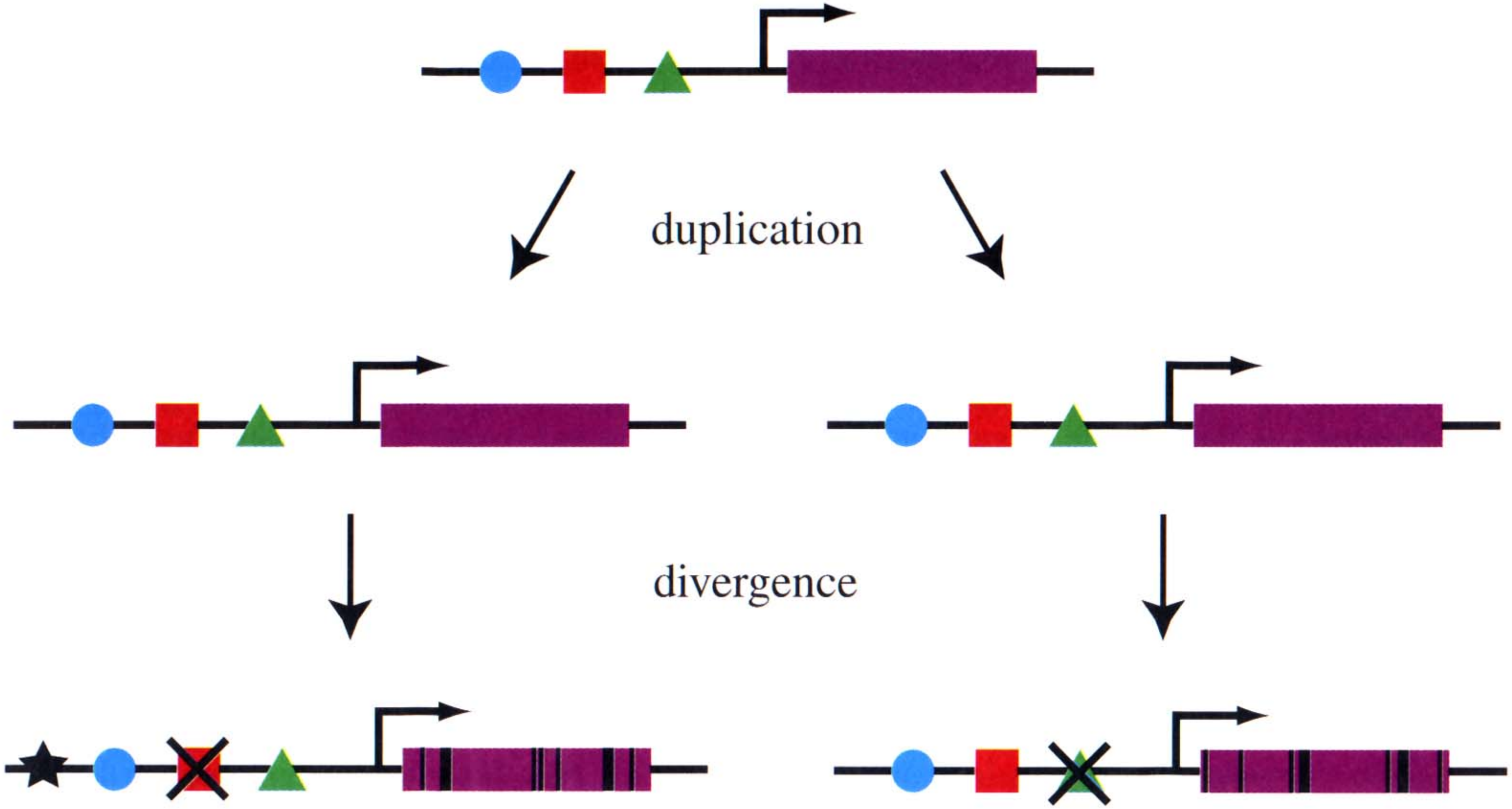
But see debate between Hoekstra & Coyne *contra* Wray

Regulation of *Hox* gene expression in the vertebrate hindbrain



Carroll et al. (2001) “From DNA to Diversity”

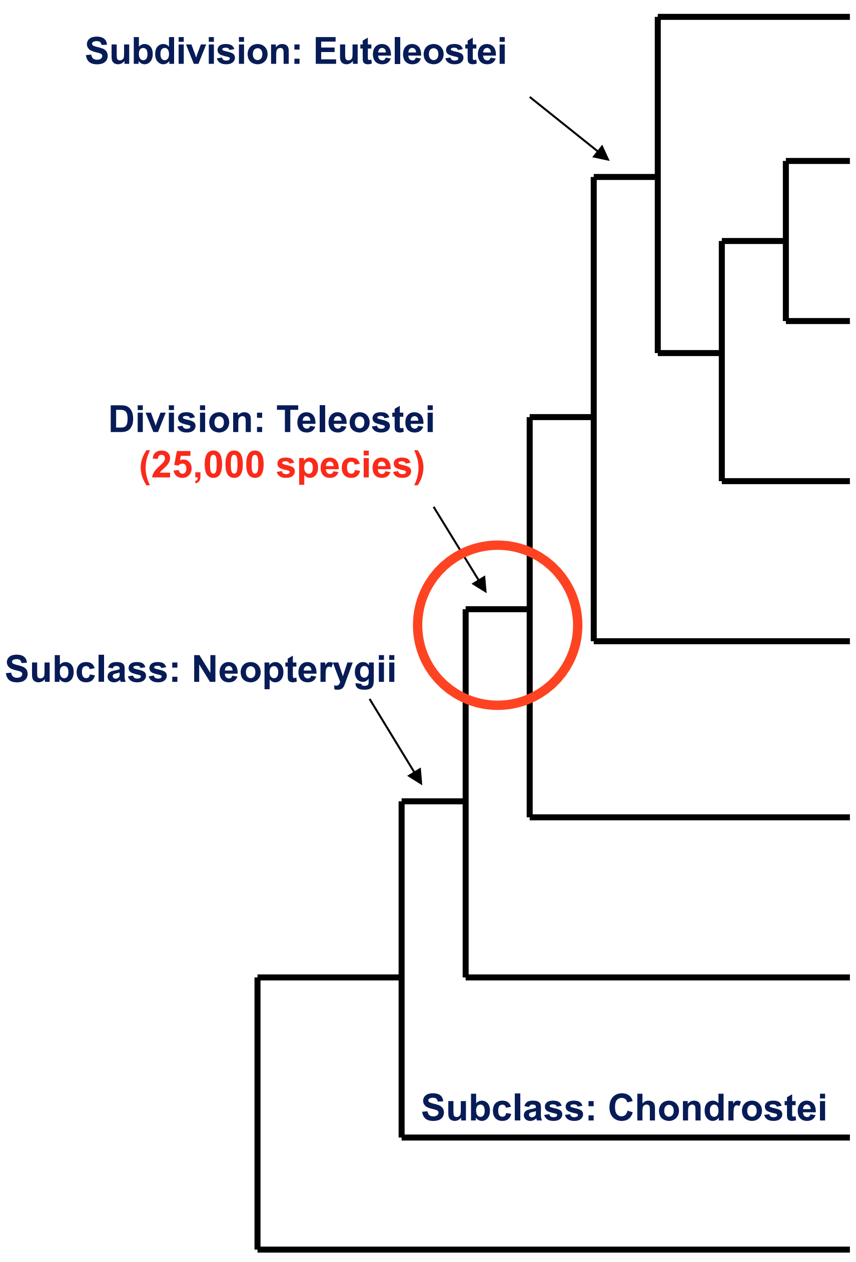
One mechanism of divergence in gene function - combining gene duplication and regulatory evolution



Force et al. (1999) *Genetics*

Carroll et al. (2001) “From DNA to Diversity”

Evolution of the fish genome - phylogenetic timing of the FSGD



Subdivision: Euteleostei

Division: Teleostei
(25,000 species)

Subclass: Neopterygii

Subclass: Chondrostei

Zebrafish

Pufferfish

Cichlid

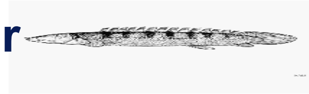
Medaka

Eel

Arapaima

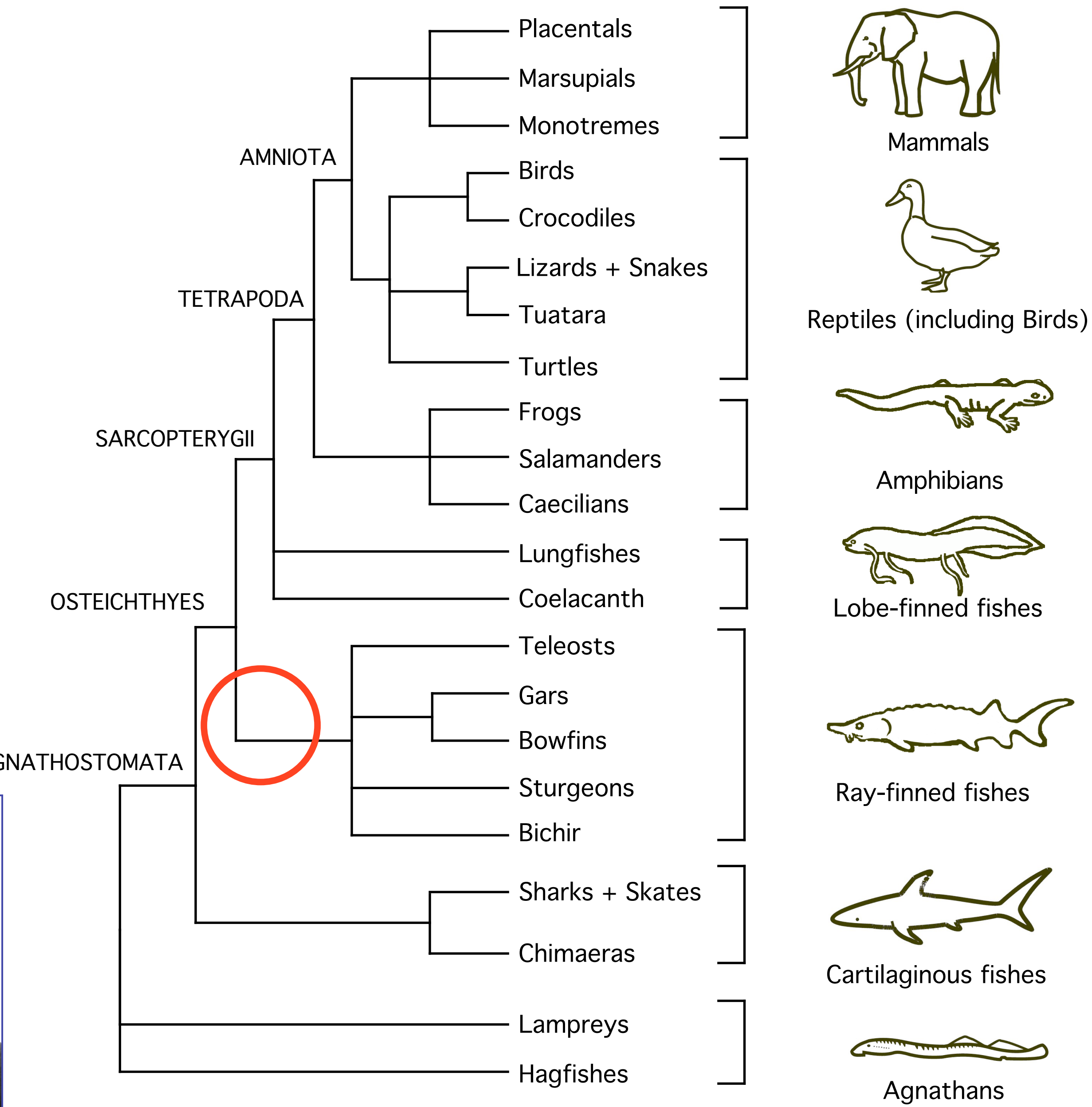
Gar

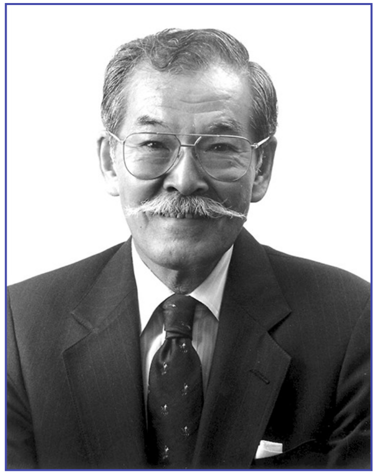
Bichir



Land vertebrates

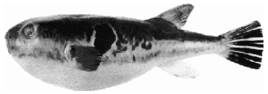
Hoegg, Brinkmann & Meyer Journal of Molecular Evolution (2004)

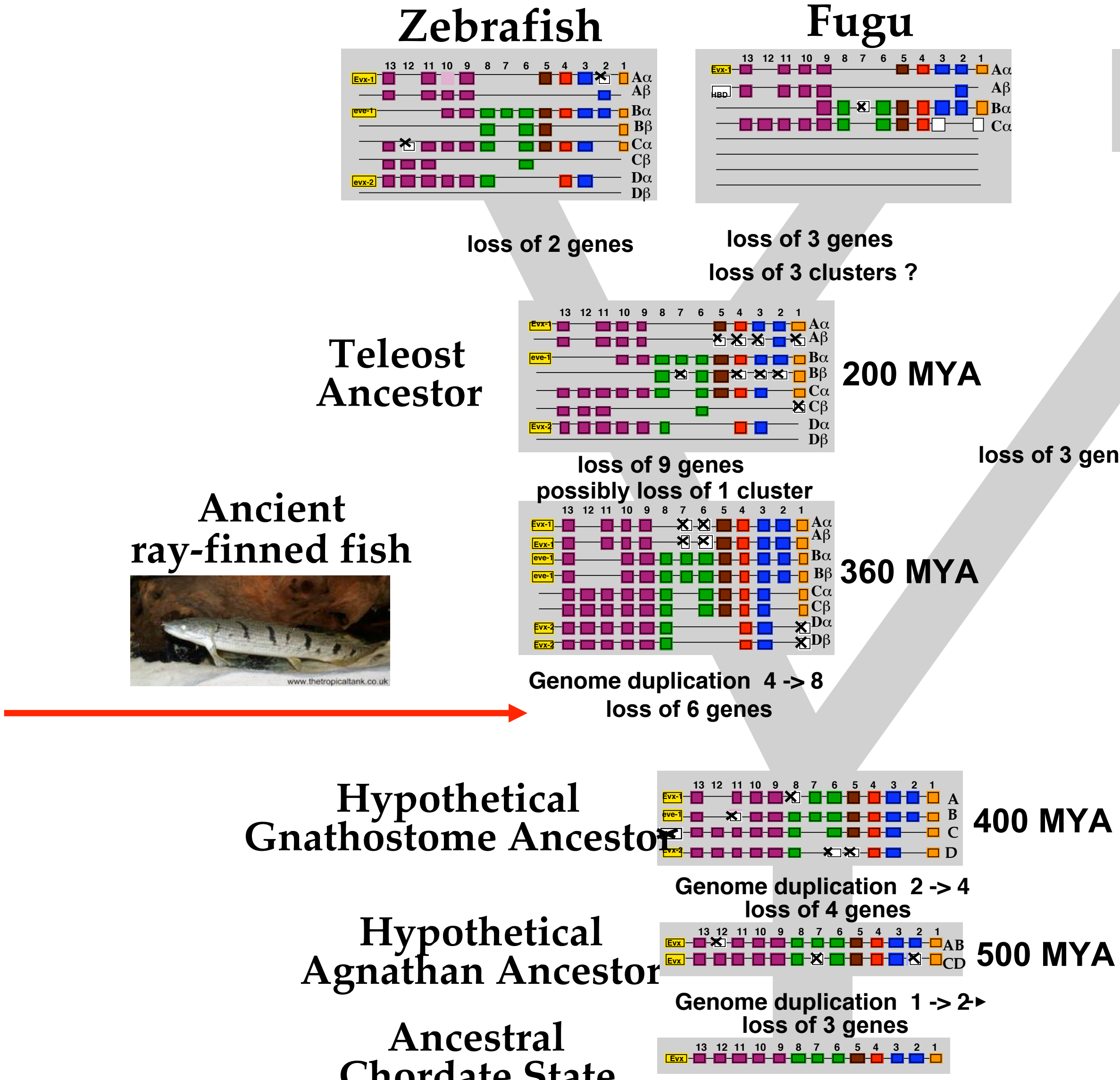




Meyer & Zardoya 2003, *Ann. Rev. Ecol. Syst.*

Evolution of vertebrate Hox gene clusters

Zebrafish

Fugu

Mouse

loss of 2 genes

loss of 3 genes
loss of 3 clusters ?

Teleost Ancestor

200 MYA

loss of 3 genes

loss of 9 genes
possibly loss of 1 cluster

Ancient ray-finned fish

360 MYA

Genome duplication 4 -> 8
loss of 6 genes

Hypothetical Gnathostome Ancestor

400 MYA

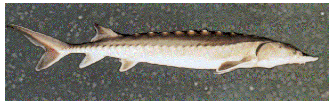
Genome duplication 2 -> 4
loss of 4 genes

Hypothetical Agnathan Ancestor

500 MYA

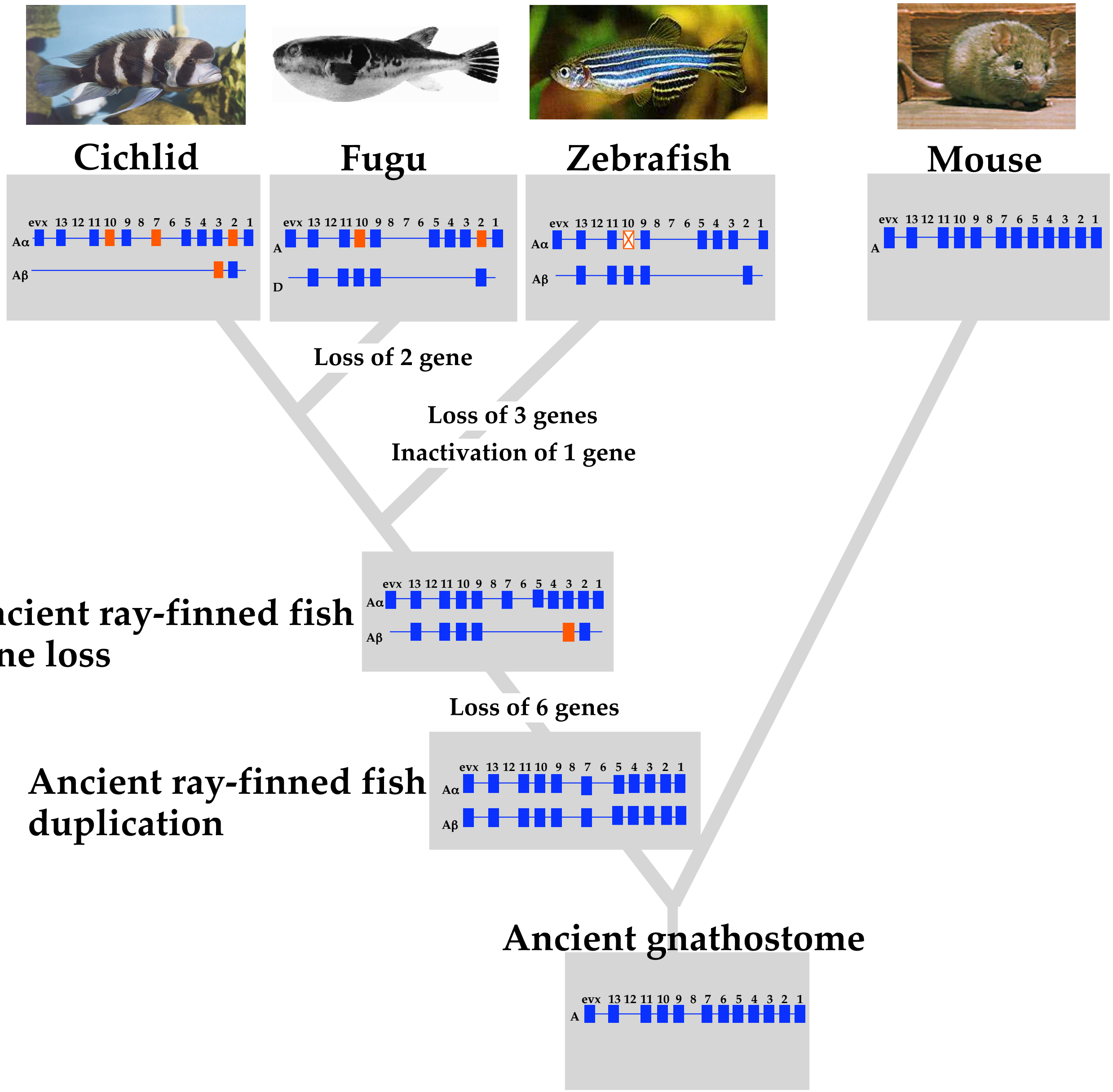
Genome duplication 1 -> 2
loss of 3 genes

Ancestral Chordate State



Modified from:
Meyer (1998) *Nature*
Widtbrodt, Meyer & Schartl (1998) *BioEssays*
Meyer & Malaga (1999) *Current Biology*
Malaga & Meyer (2001) *Am. Zool.*

Evolution of vertebrate Hox A_{alpha} clusters



Cichlid

Fugu

Zebrafish

Mouse

Loss of 2 gene

Loss of 3 genes

Inactivation of 1 gene

Ancient ray-finned fish gene loss

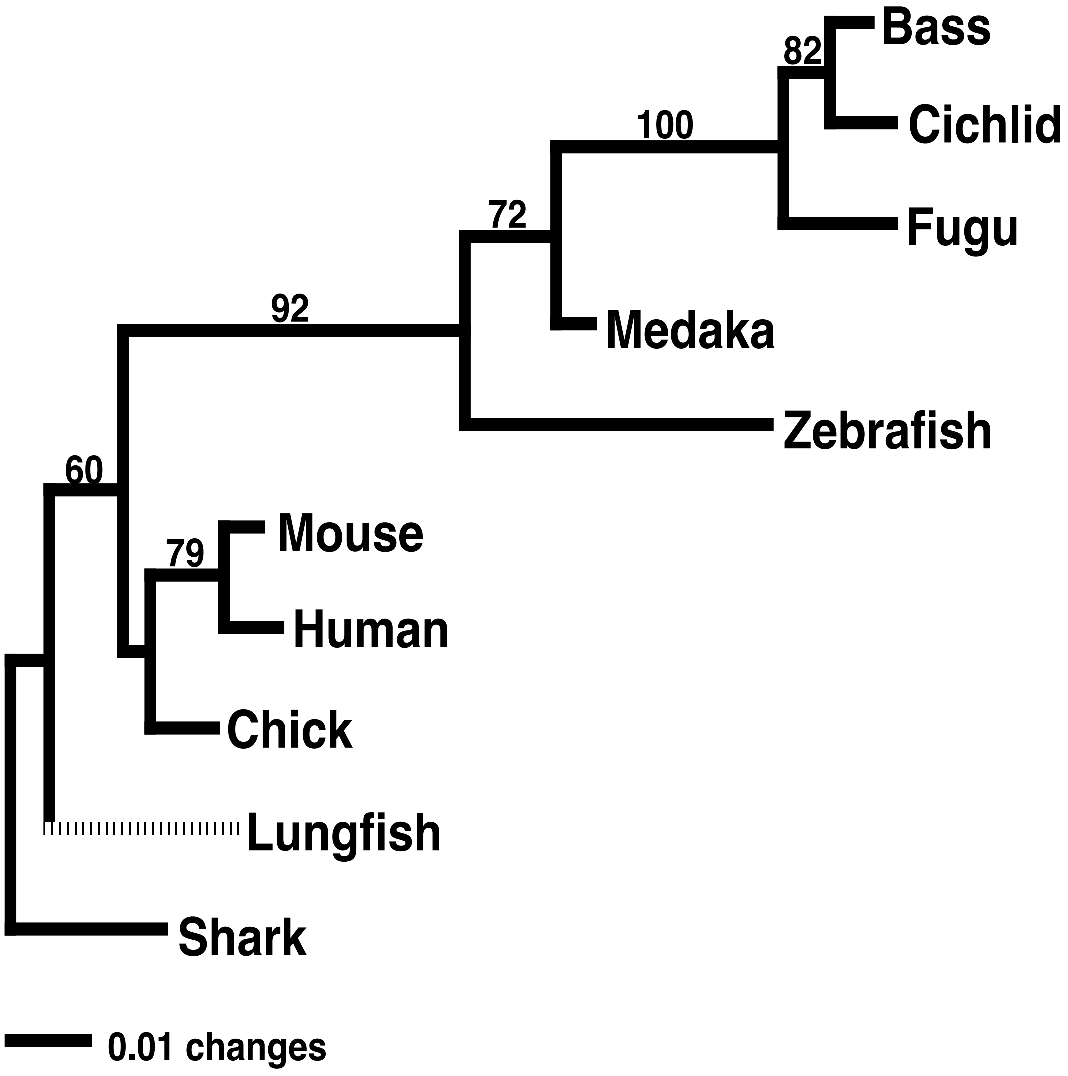
Loss of 6 genes

Ancient ray-finned fish duplication

Ancient gnathostome

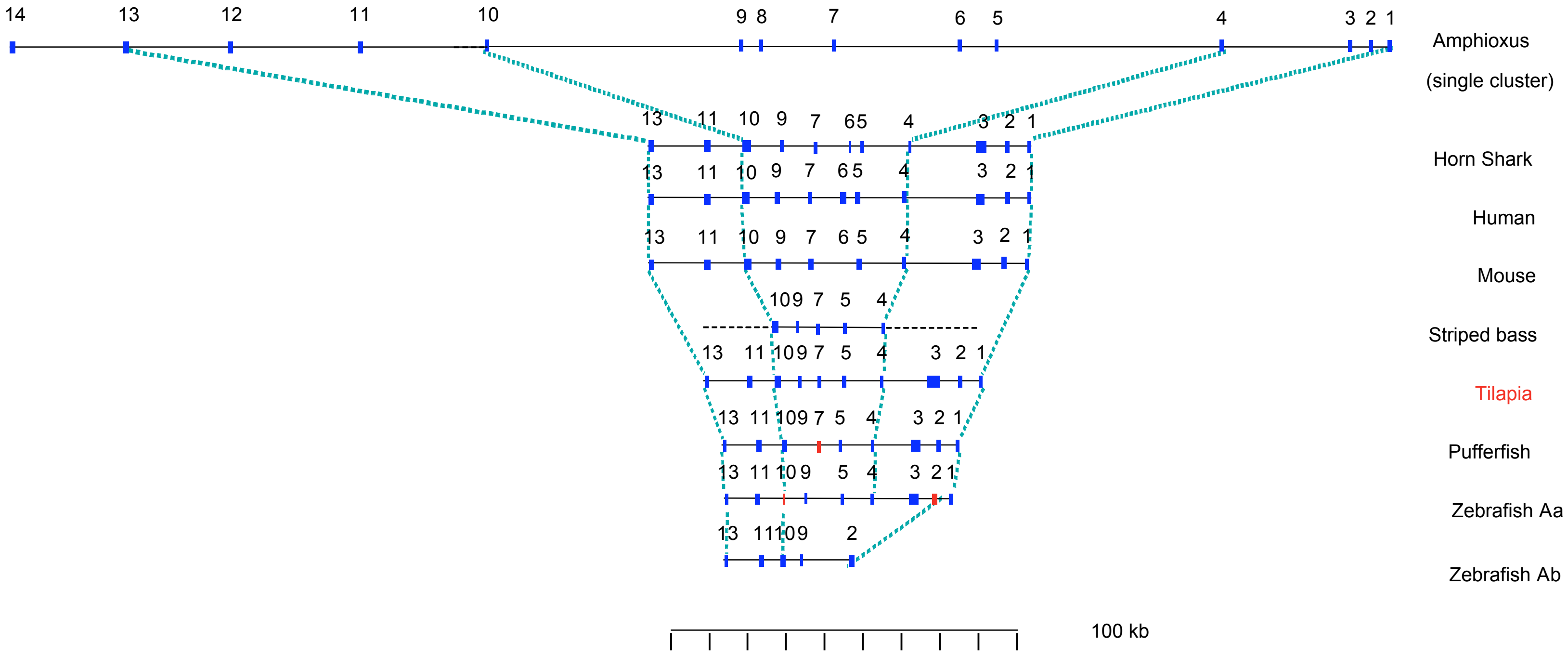
Malaga & Meyer (2001) *Am. Zool.*

Phylogenetic analysis of hoxa4/a9 sequences



Malaga & Meyer (2001) *Am. Zool.*

Length comparisons among HoxA clusters



Santini, Boore & Meyer (2003) *Genome Research*

Pipmaker (Schwartz *et al.*, 2000)

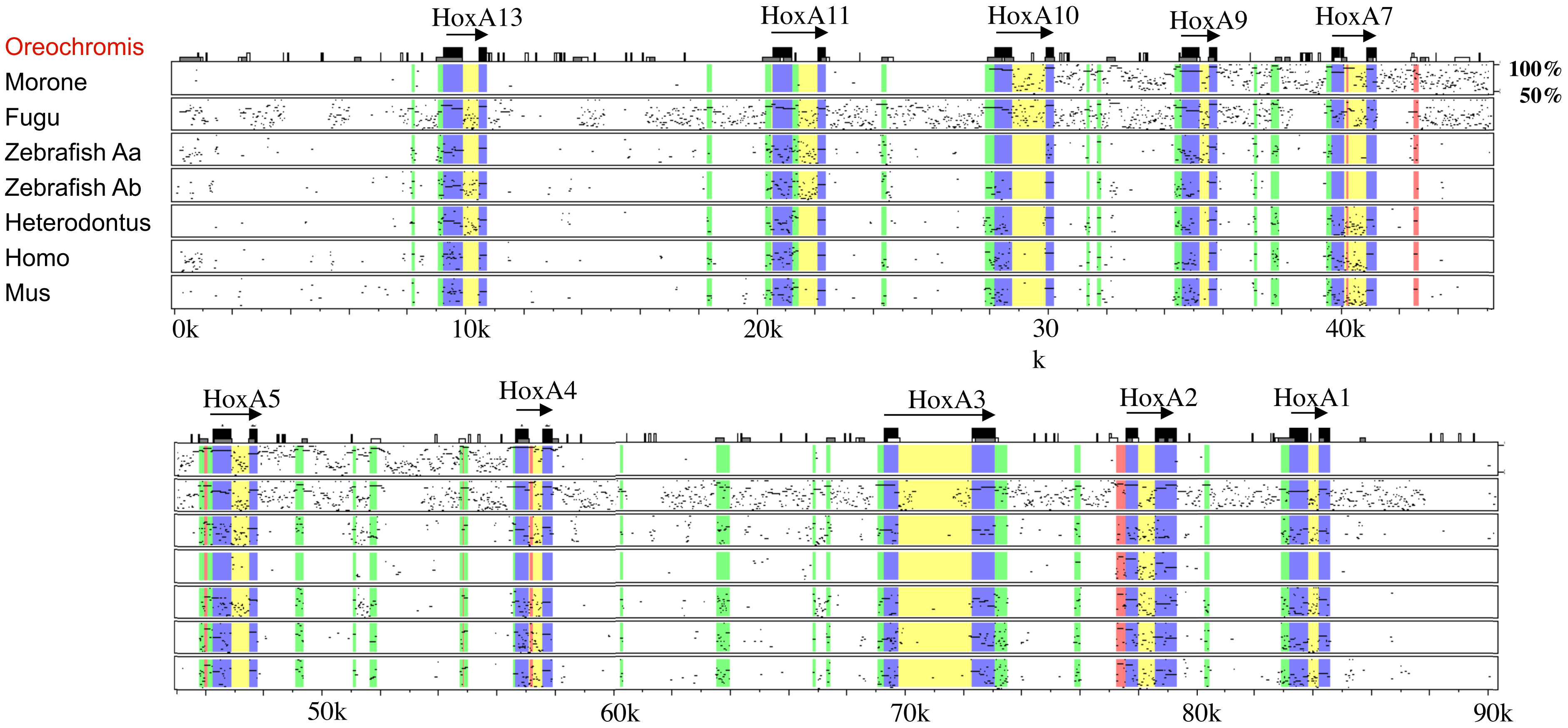
- computes alignments of similar regions in two or more DNA sequences - the alignment is summarized with a “percent identity plot” (pip) from 50-100%

- the alignment engine is BlastZ, an experimental variant of the Gapped Blast program (Altschul *et al*. 1997; Zhang *et al*. 1998), designed for aligning long sequences

- accommodates small overlaps between adjacent alignments

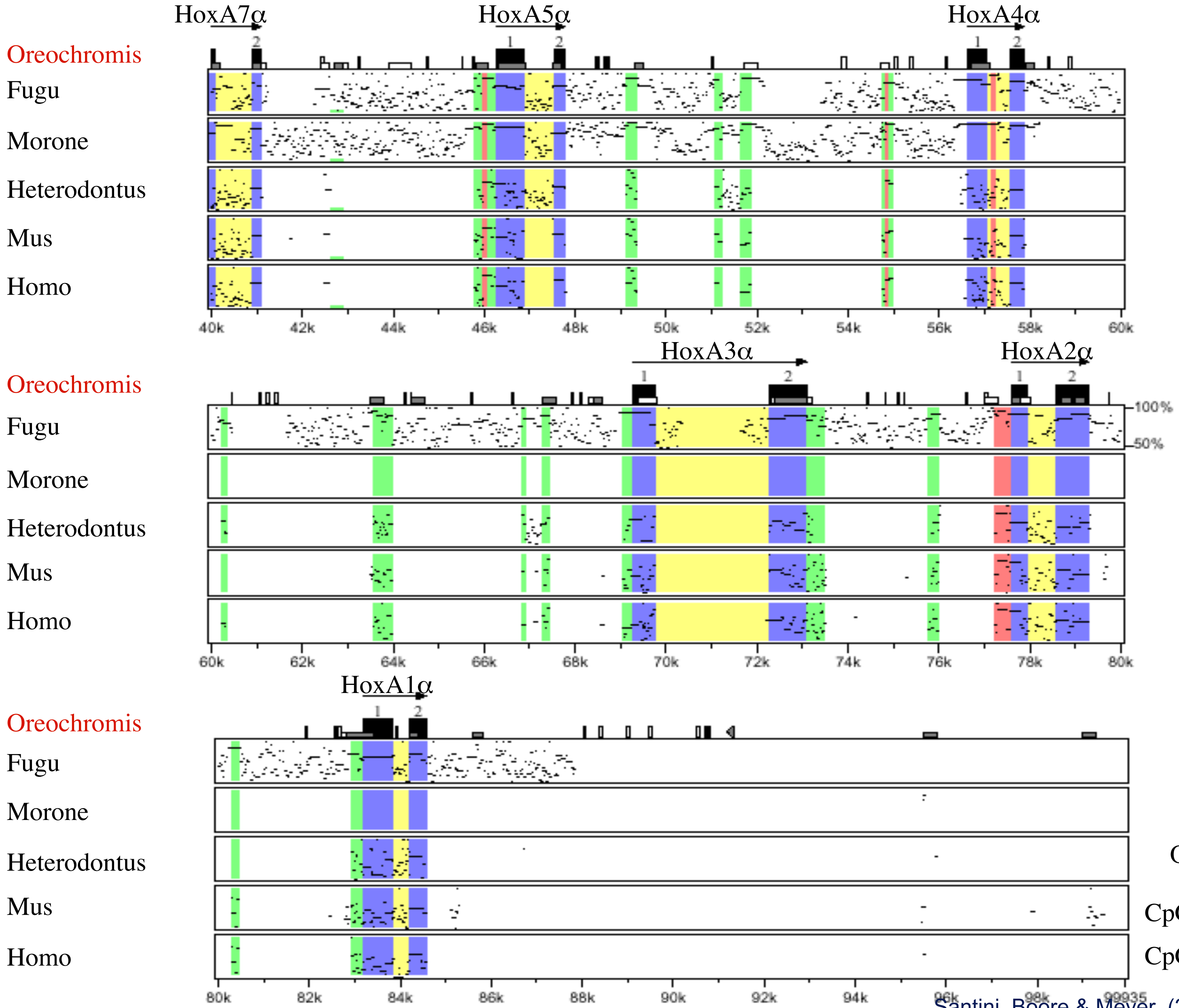
- permits a variety of “gap penalties” to be charged for offsets between two adjacent alignments in an optimal chain (particularly reducing the negative score due to gap opening and extension)

Comparison of HoxAa sequences



🟦 exon 🟨 intron 🟩 undescribed conserved region 🟥 described conserved region → gene ■ cds

Pipmaker analyses of fish and tetrapod Hox Aa clusters

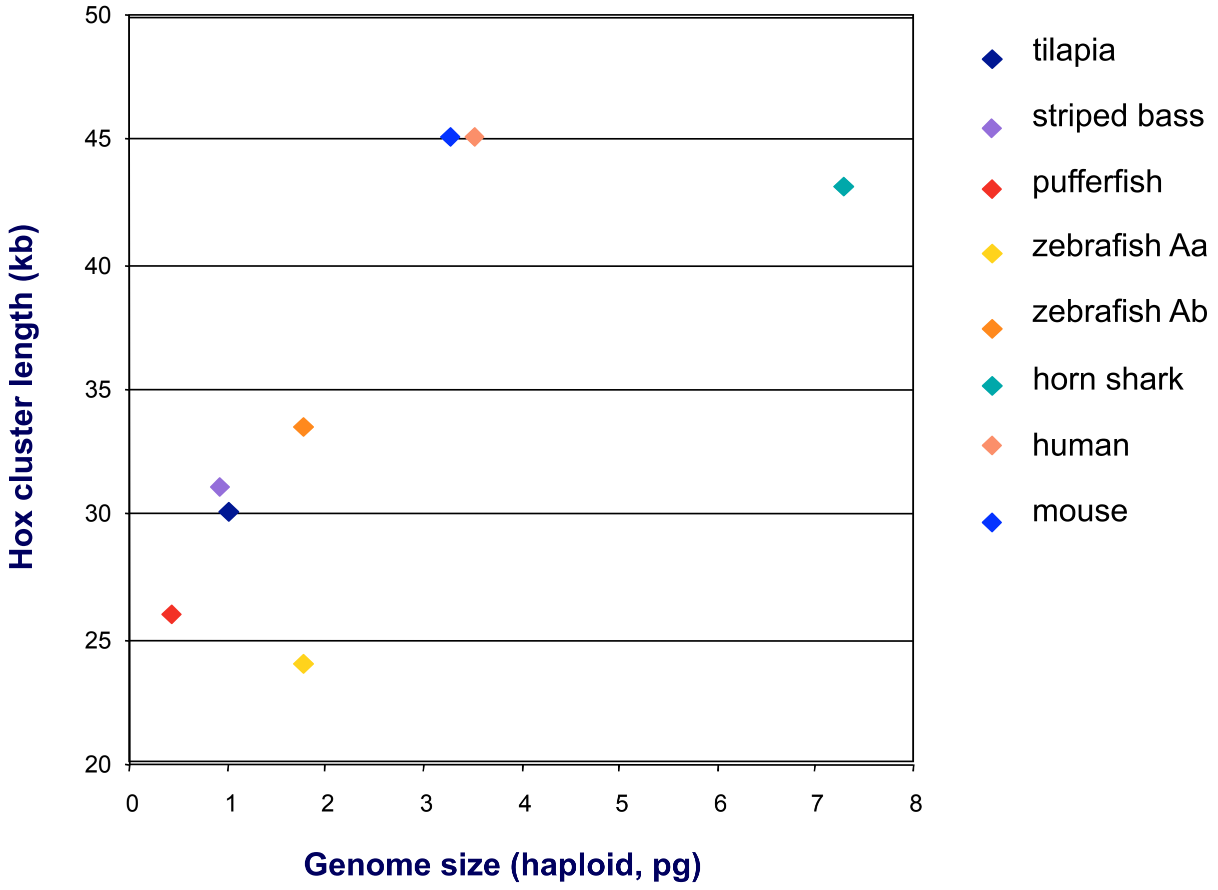


Blue:exon
Yellow :intron
Green: conserved
Red: known RE

Gene ➞
Exon ■
UTR ▨
RNA ⬠
Simple □
MIR ▶
Other SINE ▷
LINE 1 ▭
LINE 2 ▬
LTR ▬
Other repeats ▶
CpG/GpC≥0.60 ▭
CpG/GpC≥0.75 ▬

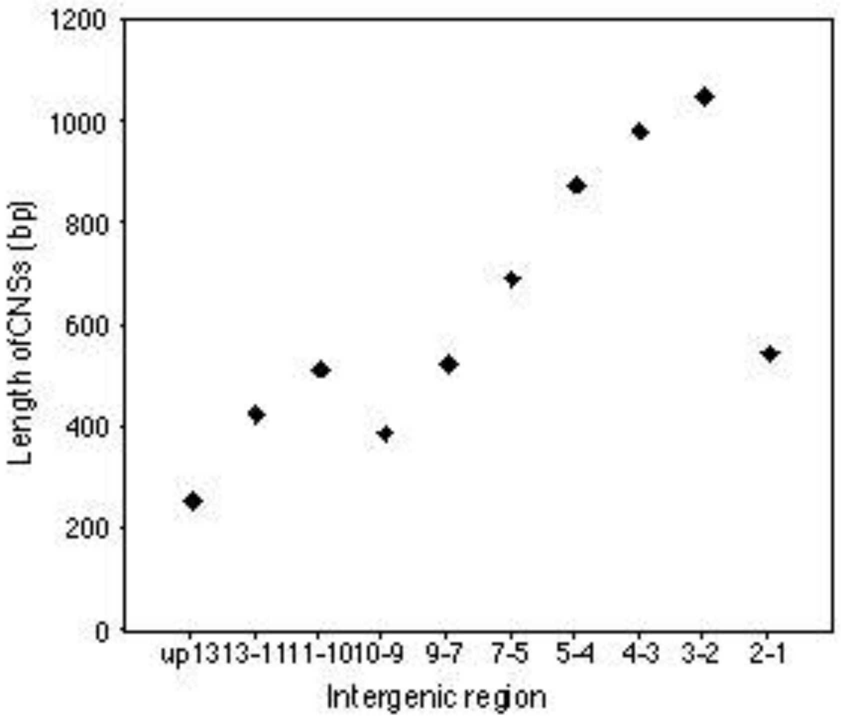
Santini, Boore & Meyer (2003) *Genome Research*

Variation in Hox A cluster size is correlated with genome size



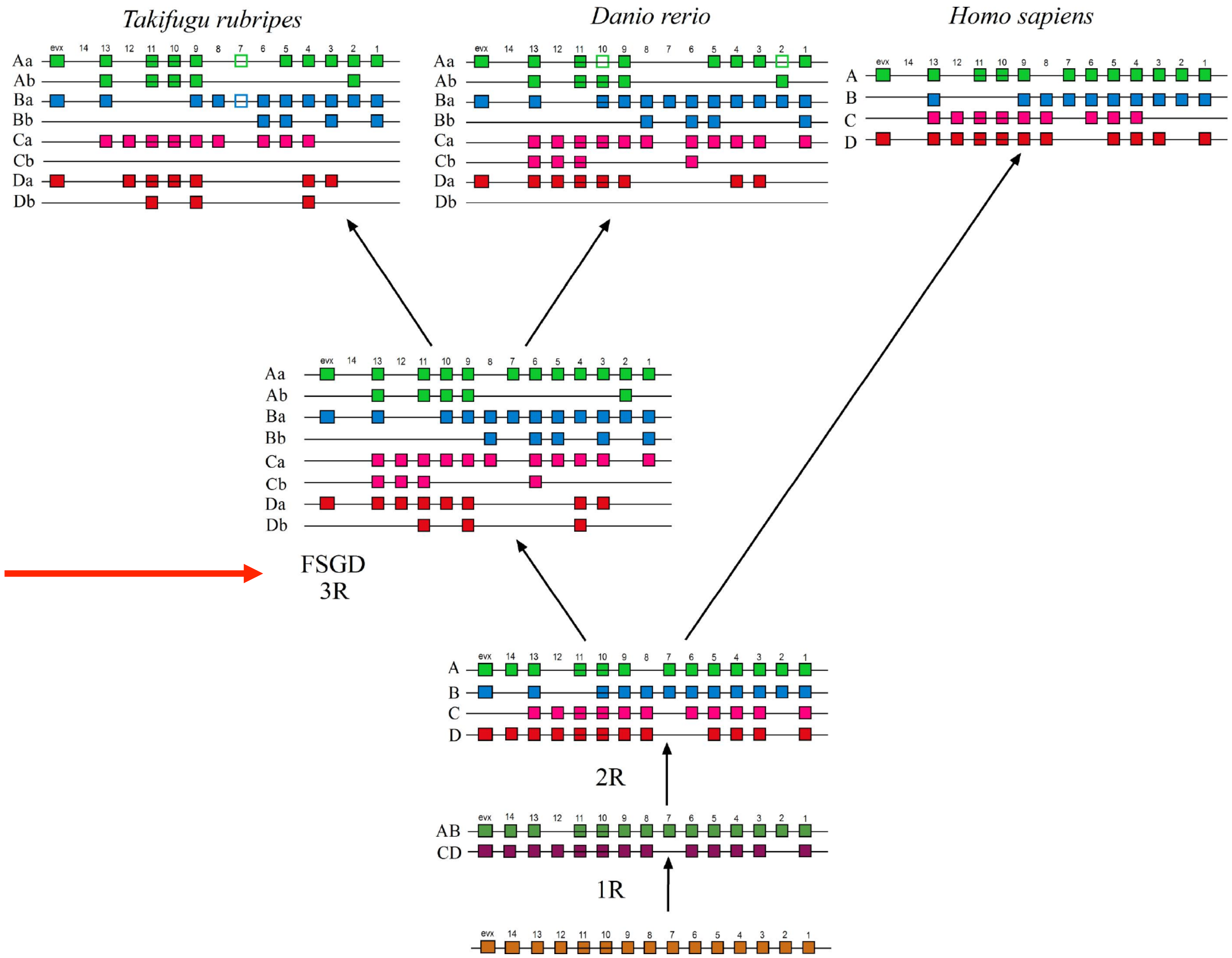
Santini, Boore & Meyer (2003) *Genome Research*

The lengths of intergenic conserved sequence blocks decreases from 3’ to 5’ in the in Hox A clusters



Santini, Boore & Meyer (2003) *Genome Research*

Evolution of vertebrate Hox gene clusters

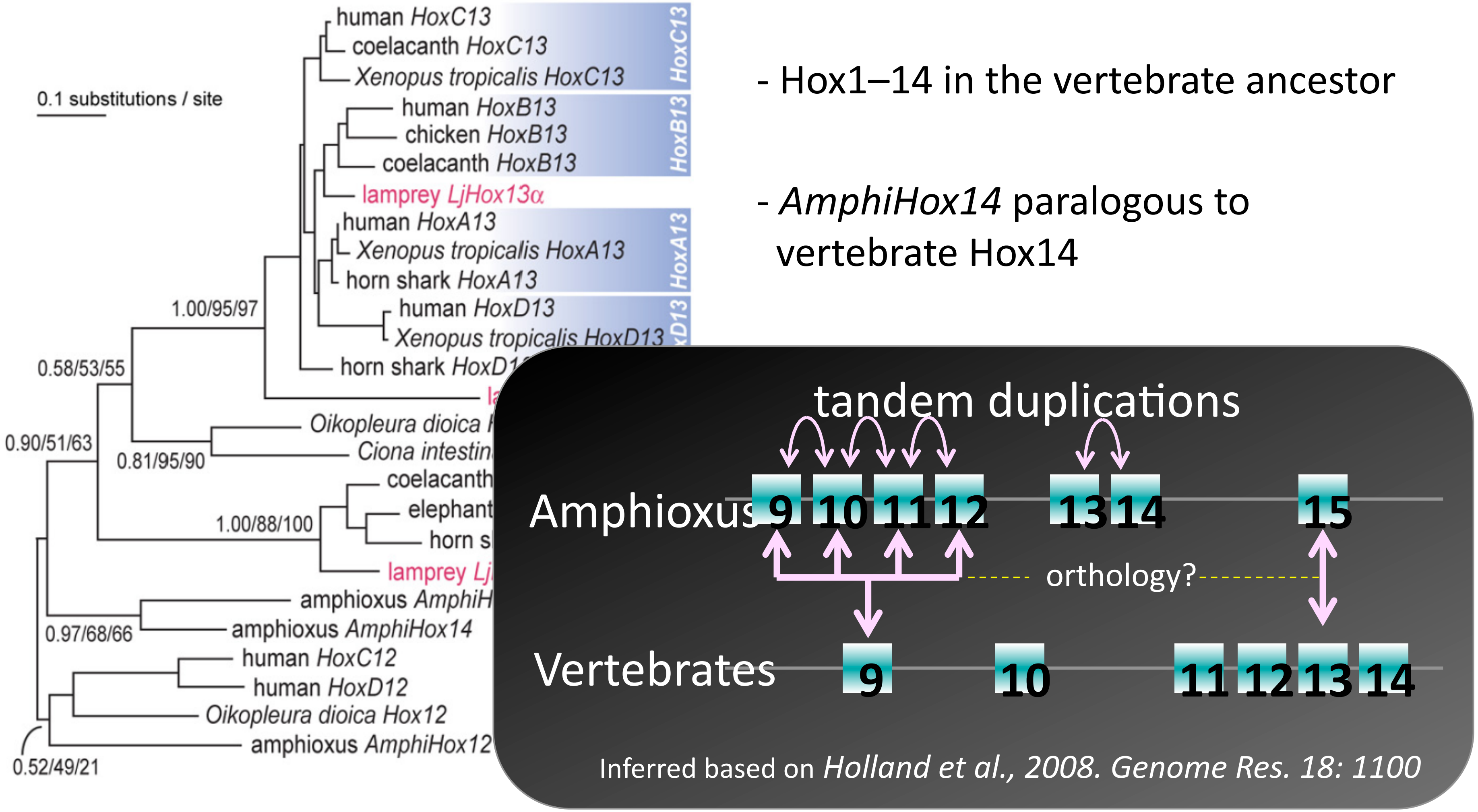


Meyer & Van de Peer *BioEssays* (2005)

Many gene losses in each lineage – clarification of "identity"

Example: Hox14

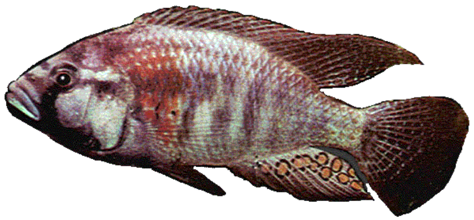
- Hox1–14 in the vertebrate ancestor
- *AmphiHox14* paralogous to vertebrate Hox14

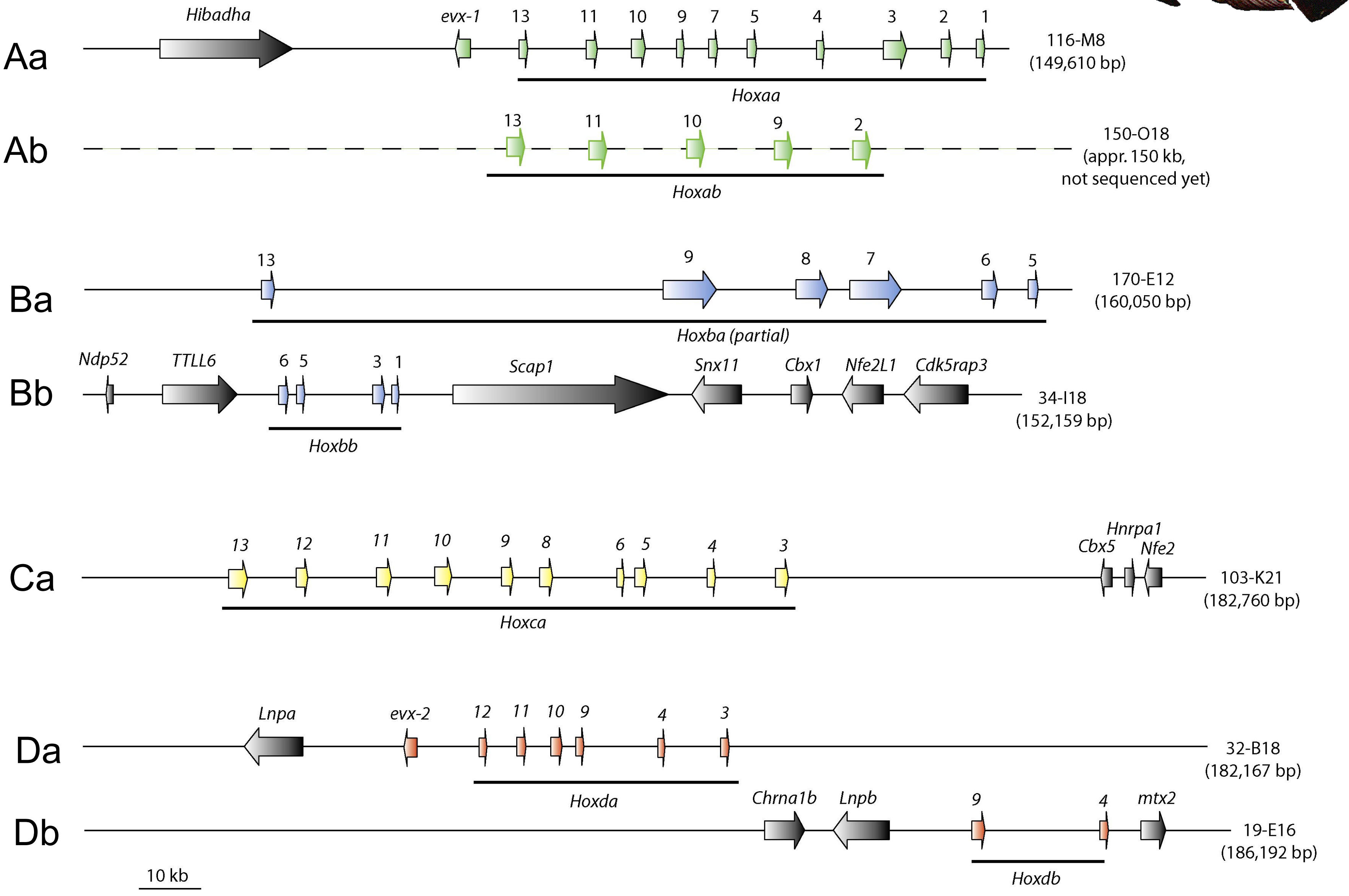


Inferred based on *Holland et al., 2008. Genome Res. 18: 1100*

Kuraku et al. PNAS 105: 6679 (2008)

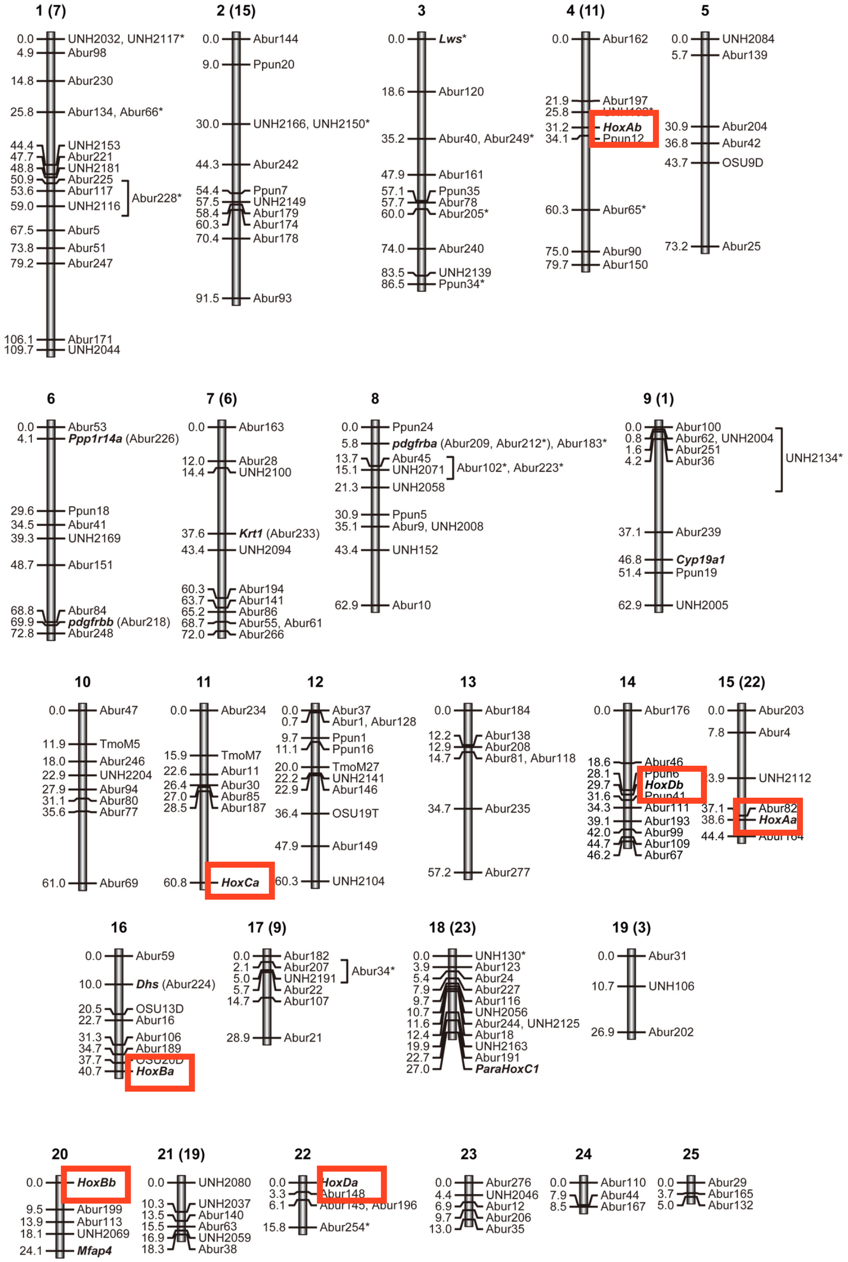
Sequenced Hox BAC clones *Astatotilapia burtoni*

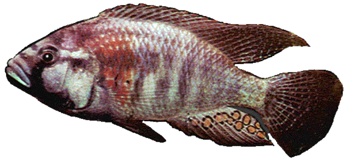




Hoegg, Boore, Kuehl and Meyer (2007) *BMC Genomics*

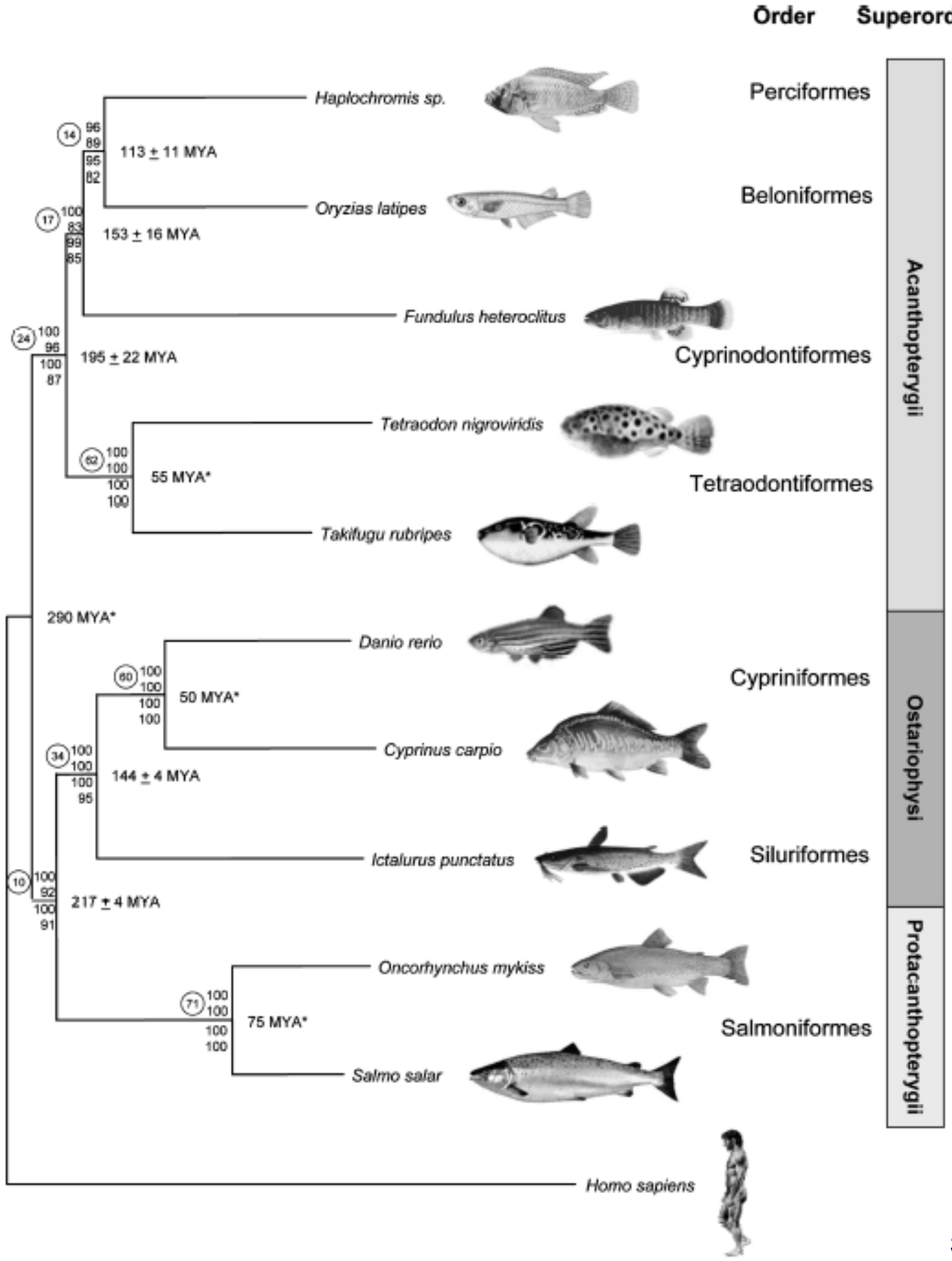
Seven Hox clusters mapped to seven chromosomes in *Astatotilapia burtoni*

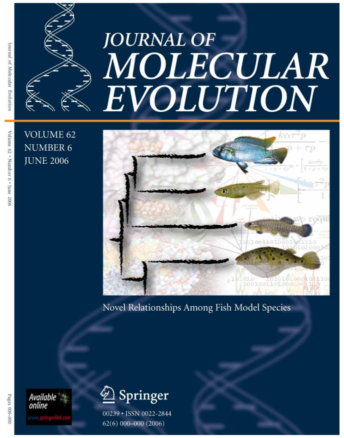




Sanetra, Henning, Fukamachi and Meyer (2009) *Genetics*

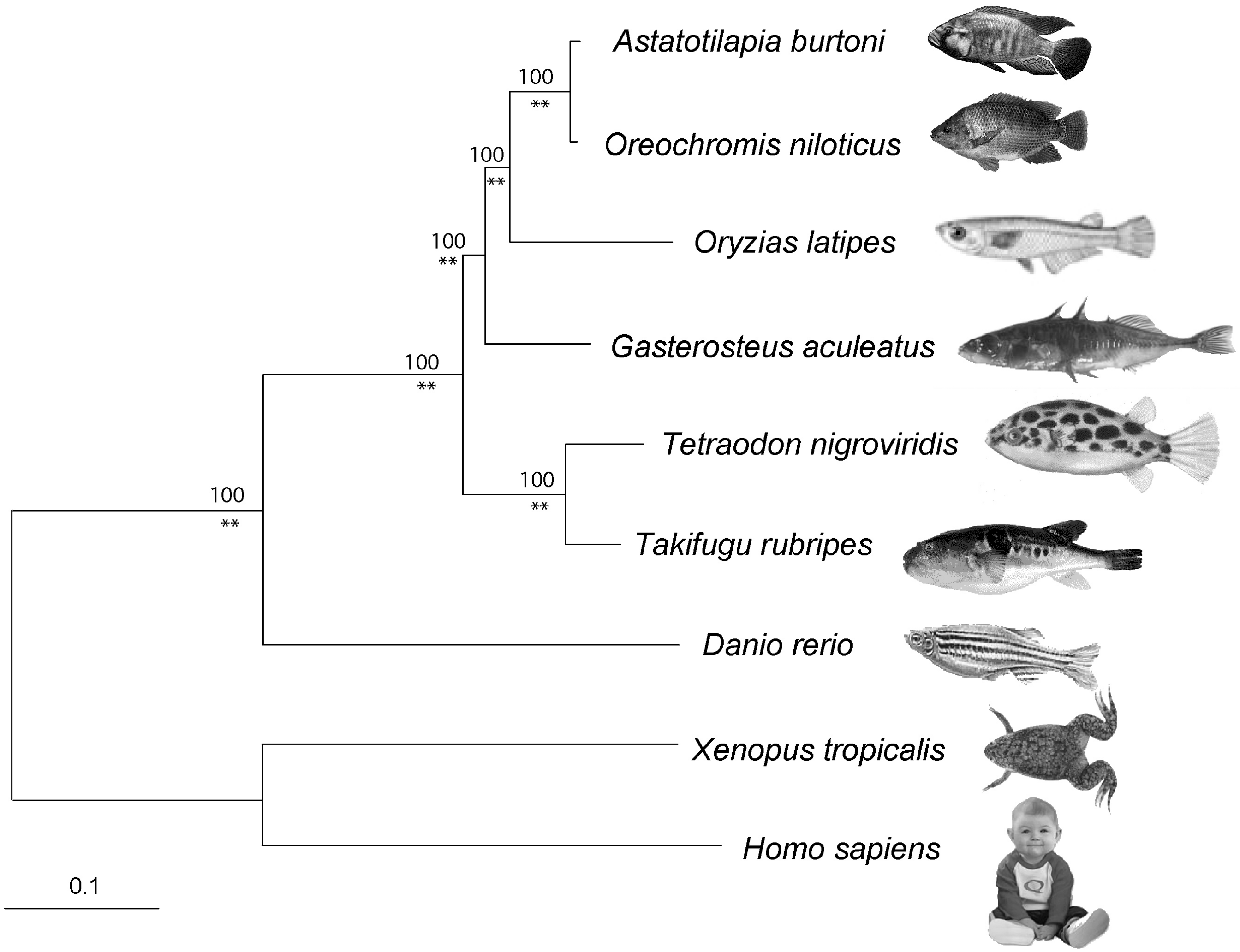
(Novel) Phylogenetic relationships among 10 fish model systems - based on ESTs





Steinke, Salzburger & Meyer (2006) *Journal of Molecular Evolution*

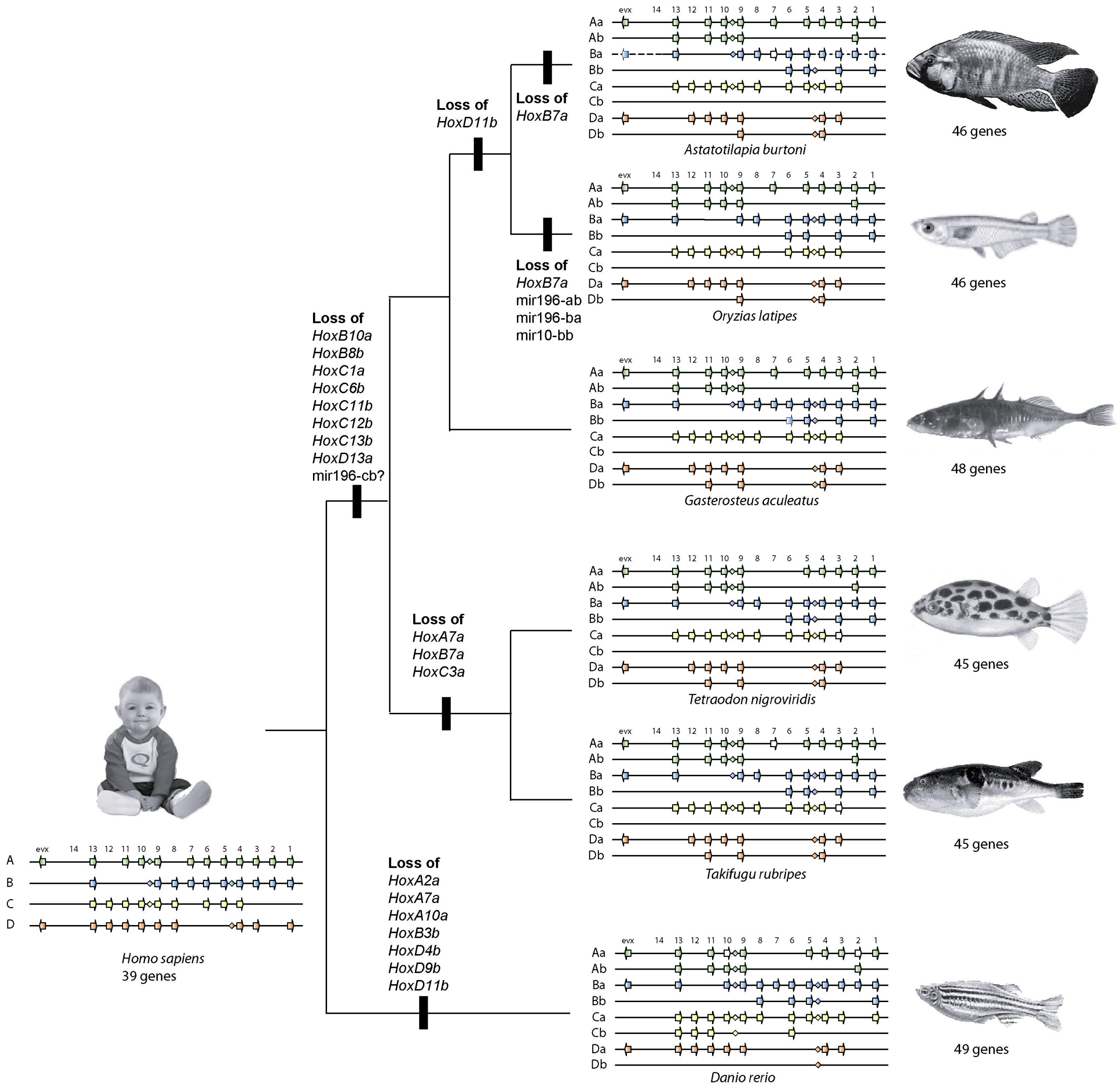
Phylogenetic relationships among teleost species



ML tree based on 24 Hox genes (20009 bp)

Hoegg, Boore, Kuehl and Meyer (2007) *BMC Genomics*

Gain and loss of Hox genes during teleost evolution


Loss of
HoxD11b

Loss of
HoxB7a

Loss of
HoxB7a
mir196-ab
mir196-ba
mir10-bb

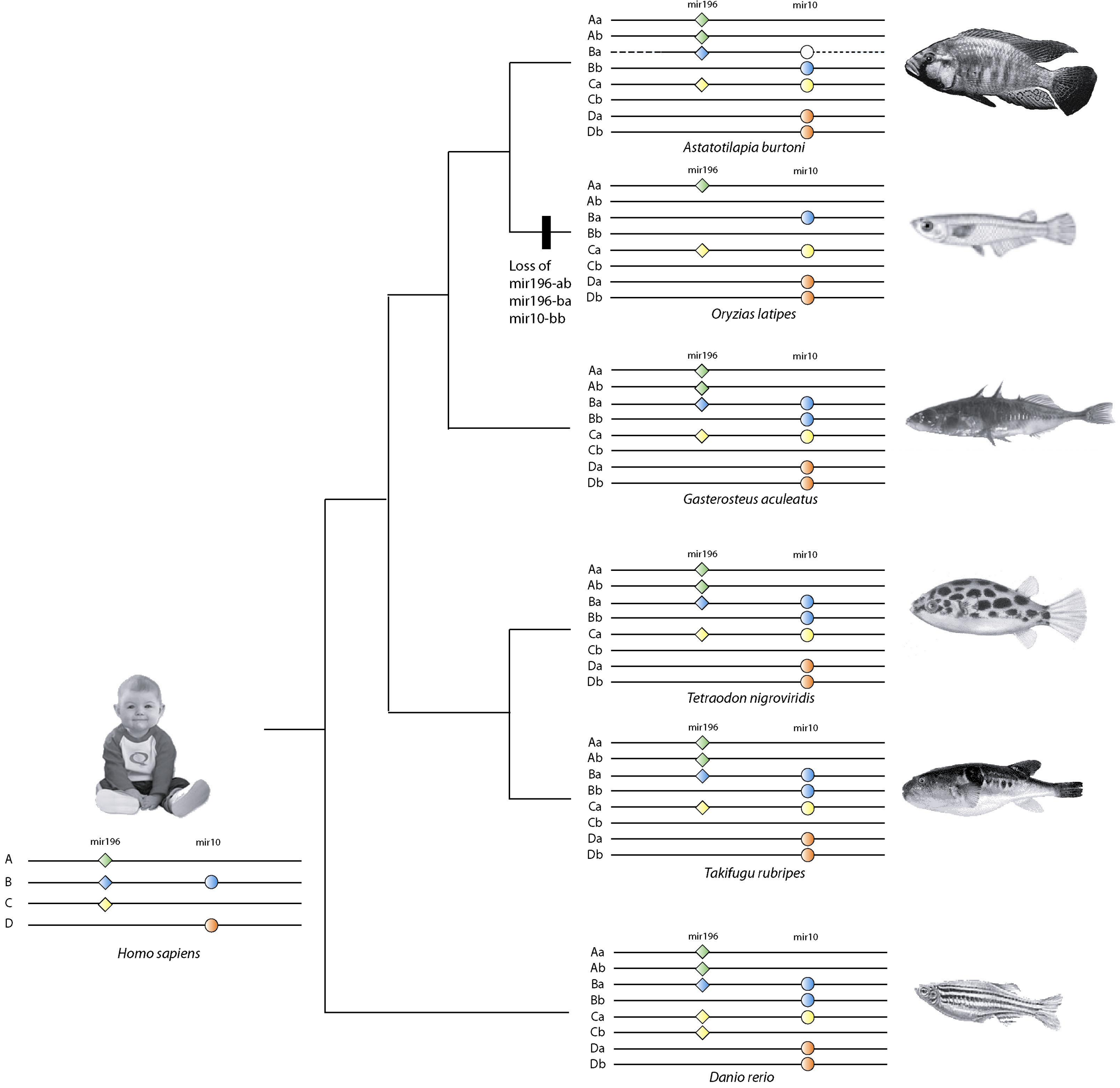
Loss of
HoxB10a
HoxB8b
HoxC1a
HoxC6b
HoxC11b
HoxC12b
HoxC13b
HoxD13a
mir196-cb?

Loss of
HoxA7a
HoxB7a
HoxC3a

Loss of
HoxA2a
HoxA7a
HoxA10a
HoxB3b
HoxD4b
HoxD9b
HoxD11b

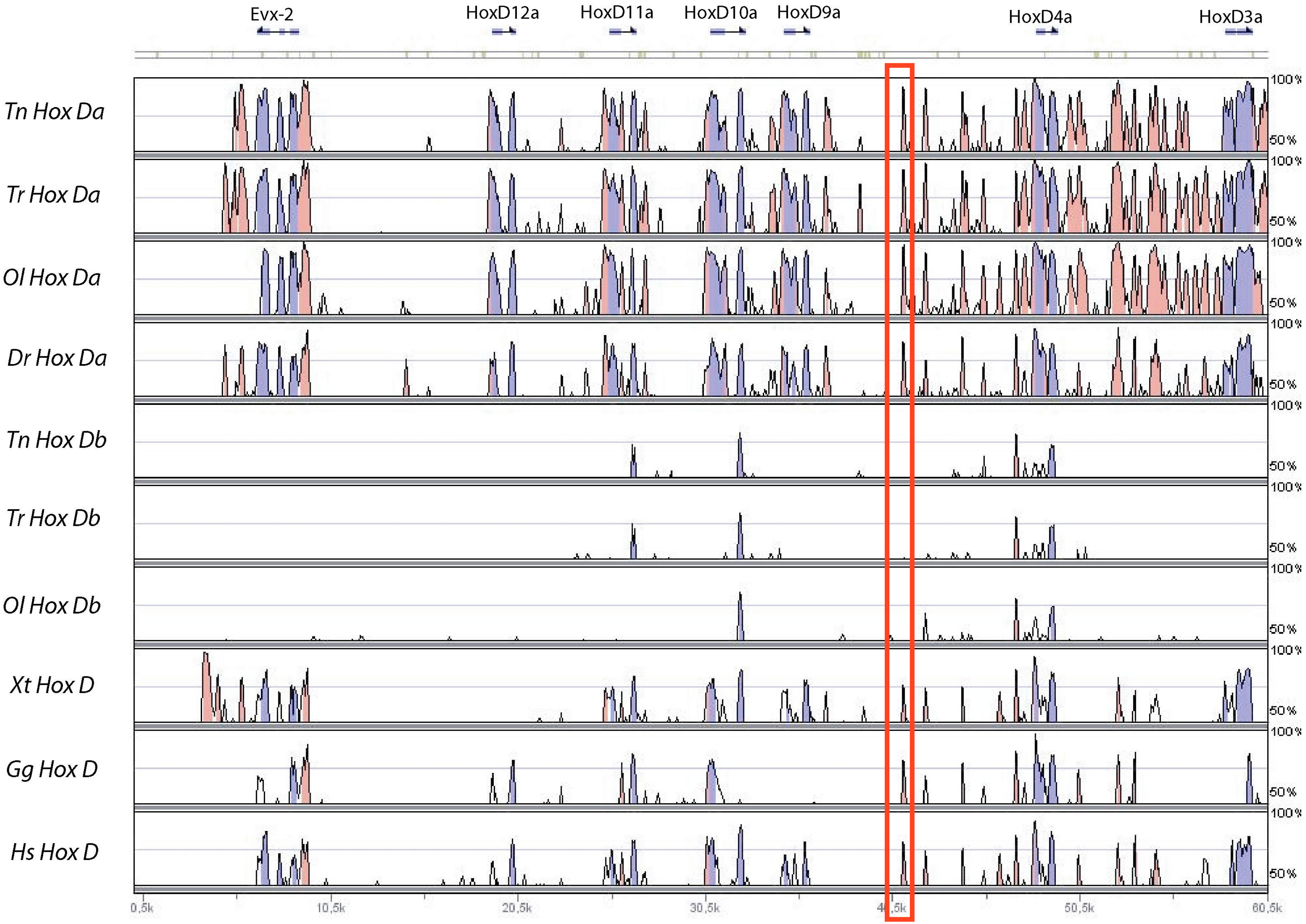
Hoegg, Boore, Kuehl and Meyer (2007) *BMC Genomics*

MicroRNA evolution in teleost fish



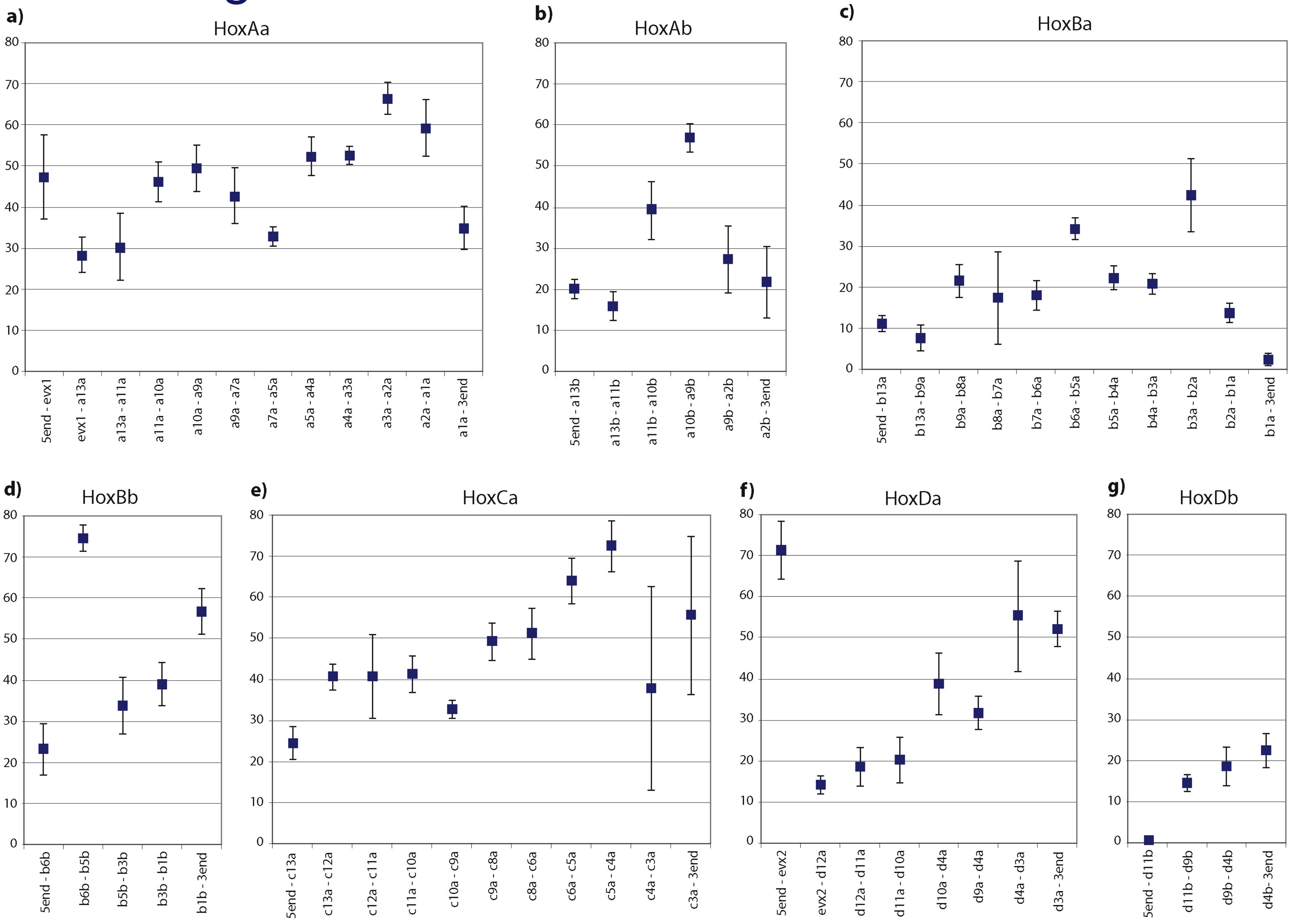
Hoegg, Boore, Kuehl and Meyer (2007) *BMC Genomics*

Hox clusters are rich in conserved non-coding sequences



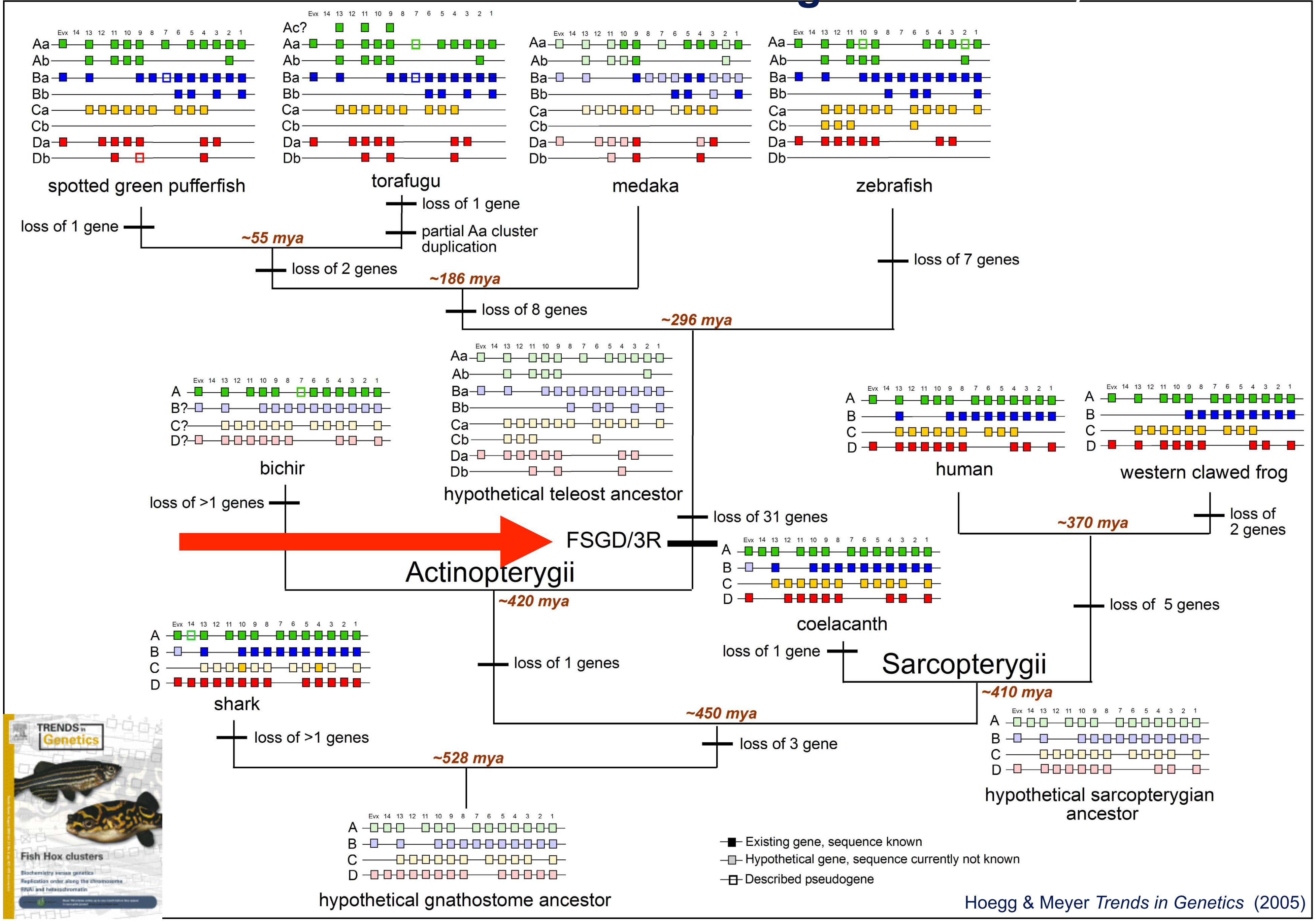
Hoegg, Boore, Kuehl and Meyer (2007) *BMC Genomics*

CNS regions in Hox clusters of several fish models



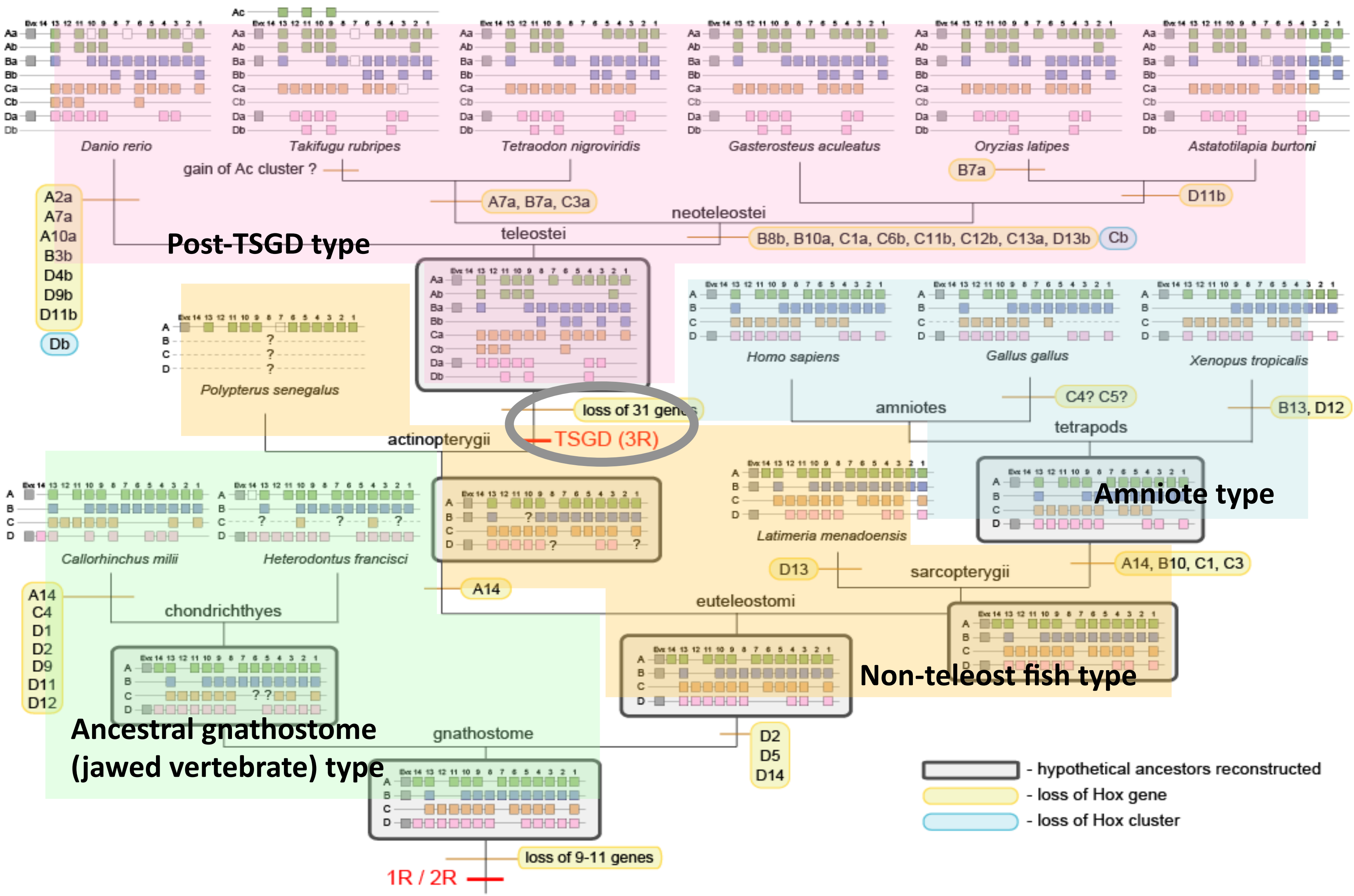
Hoegg, Boore, Kuehl and Meyer (2007) *BMC Genomics*

Evolution of vertebrate Hox gene clusters



Hoegg & Meyer *Trends in Genetics* (2005)

Evolution of *Hox* clusters in vertebrates



Post-TSGD type

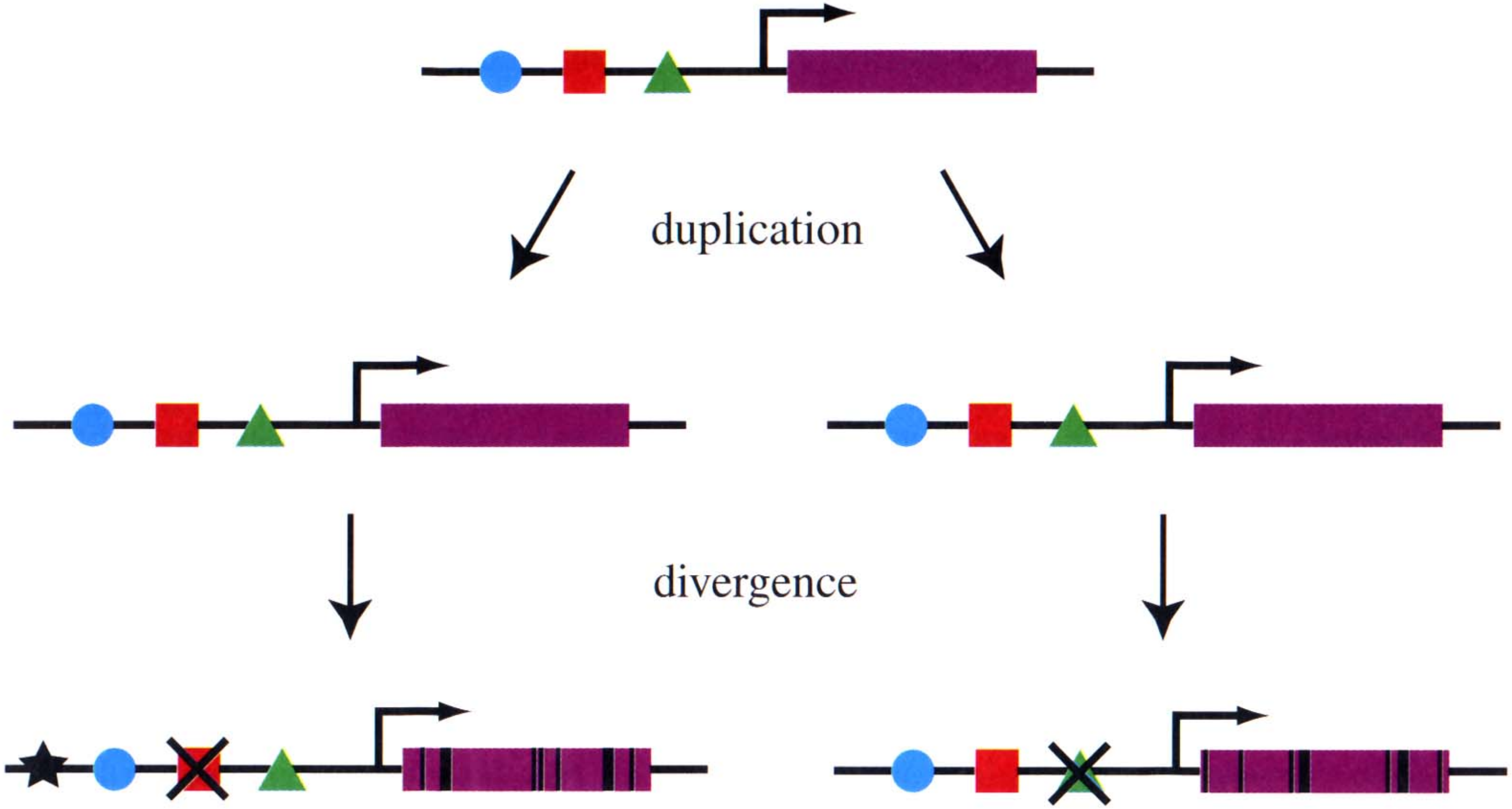
Amniote type

Non-teleost fish type

Ancestral gnathostome (jawed vertebrate) type

Kuraku & Meyer 2009

One mechanism of divergence in gene function - combining gene duplication and regulatory evolution

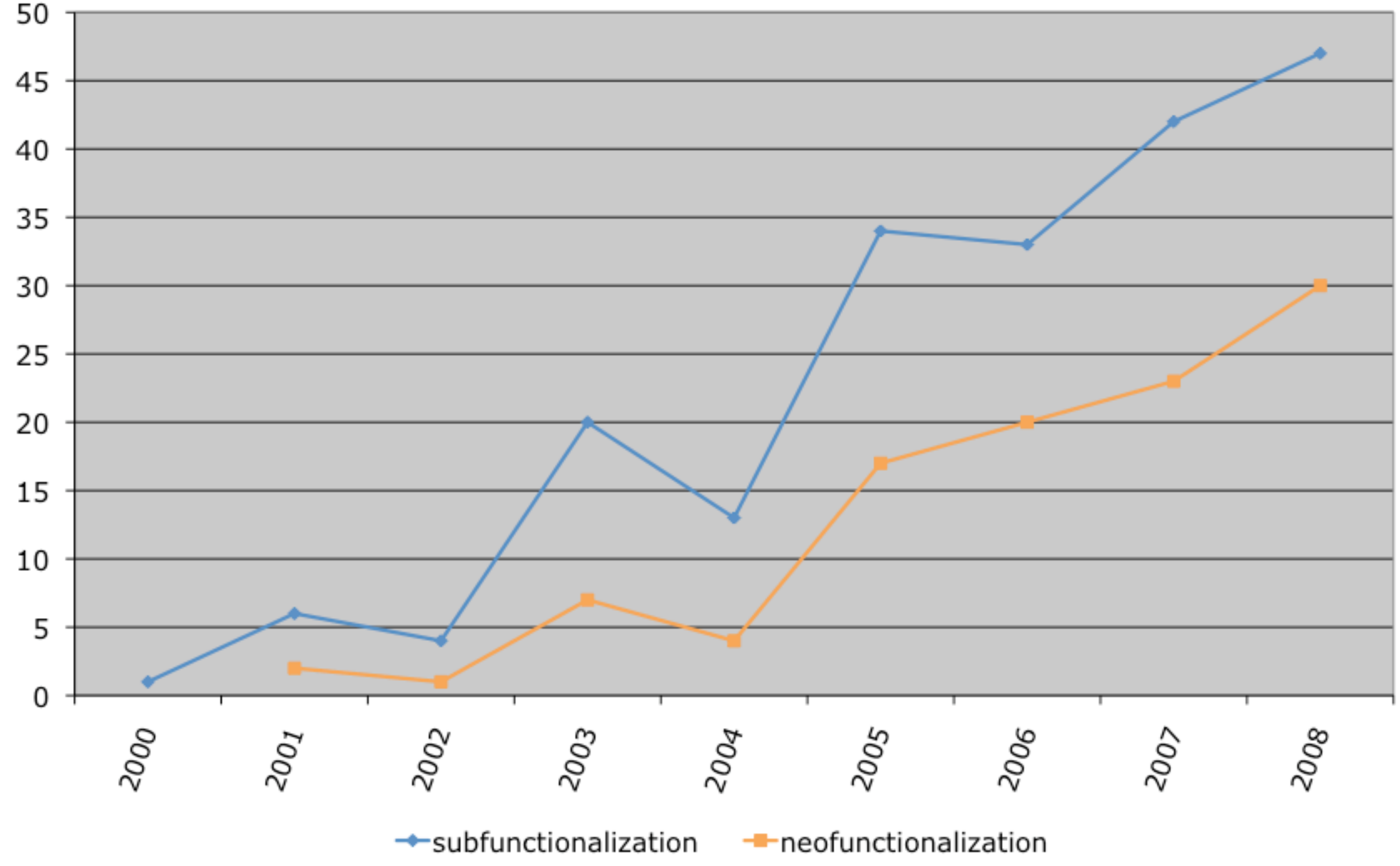


Force et al. (1999) *Genetics*

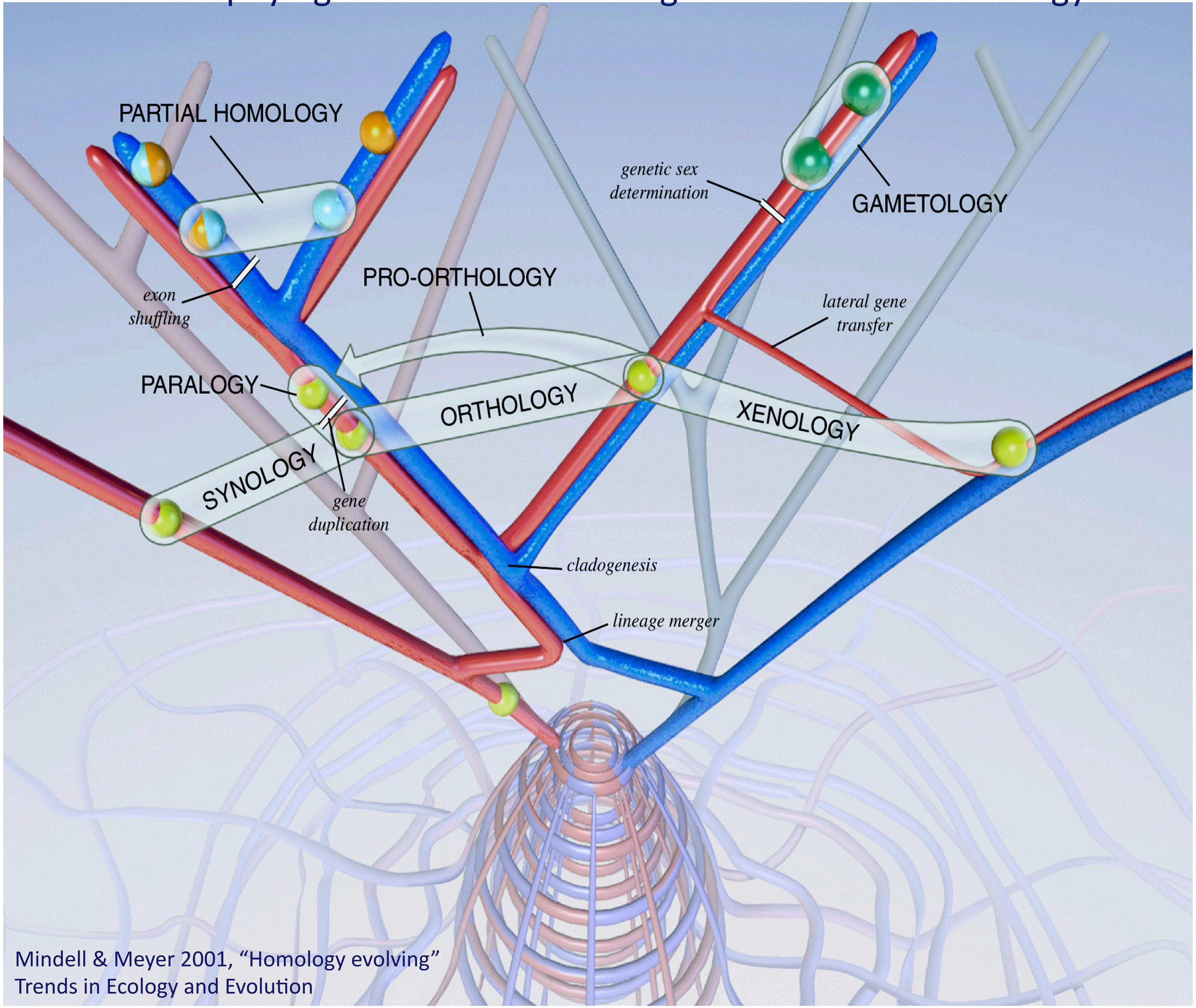
Carroll et al. (2001) “From DNA to Diversity”

Since Ohno's 1970 book
"Evolution by gene duplication"

numbers of papers on sub- and neofunctionalization

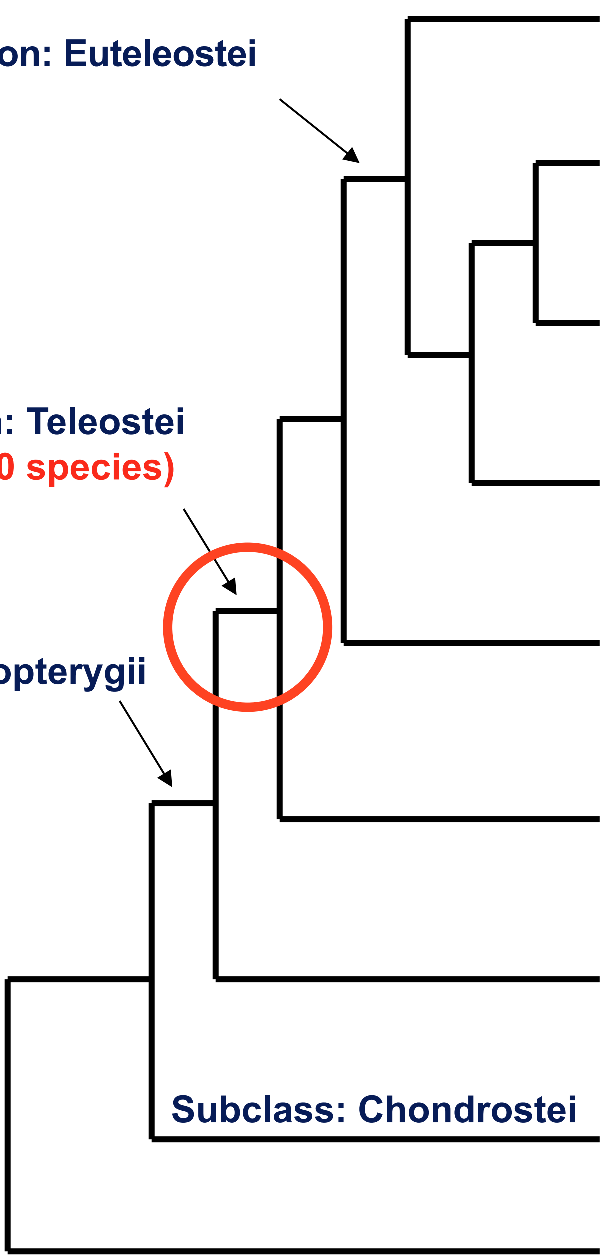


Idealized phylogenetic tree illustrating seven forms of homology



Mindell & Meyer 2001, “Homology evolving”
Trends in Ecology and Evolution

Evolution of the fish genome - phylogenetic timing of the FSGD



Subdivision: Euteleostei

Division: Teleostei
(25,000 species)

Subclass: Neopterygii

Subclass: Chondrostei

Zebrafish

Pufferfish

Cichlid

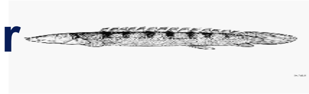
Medaka

Eel

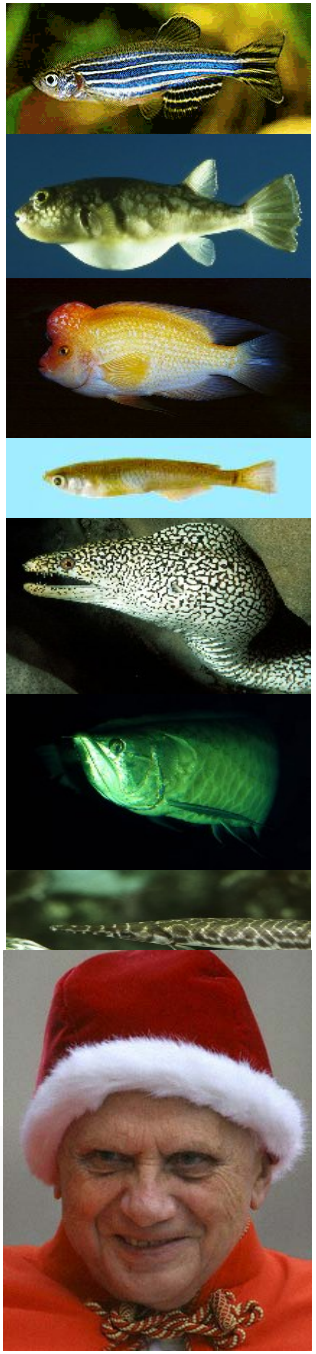
Arapaima

Gar

Bichir



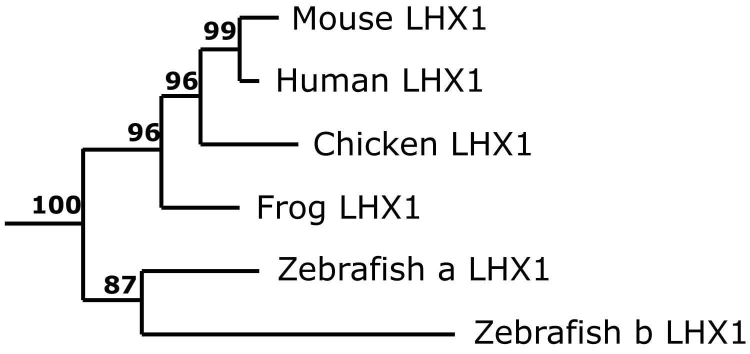
Land vertebrates

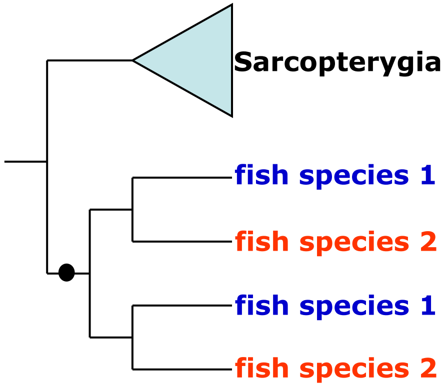


Hoegg, Brinkmann & Meyer Journal of Molecular Evolution (2004)

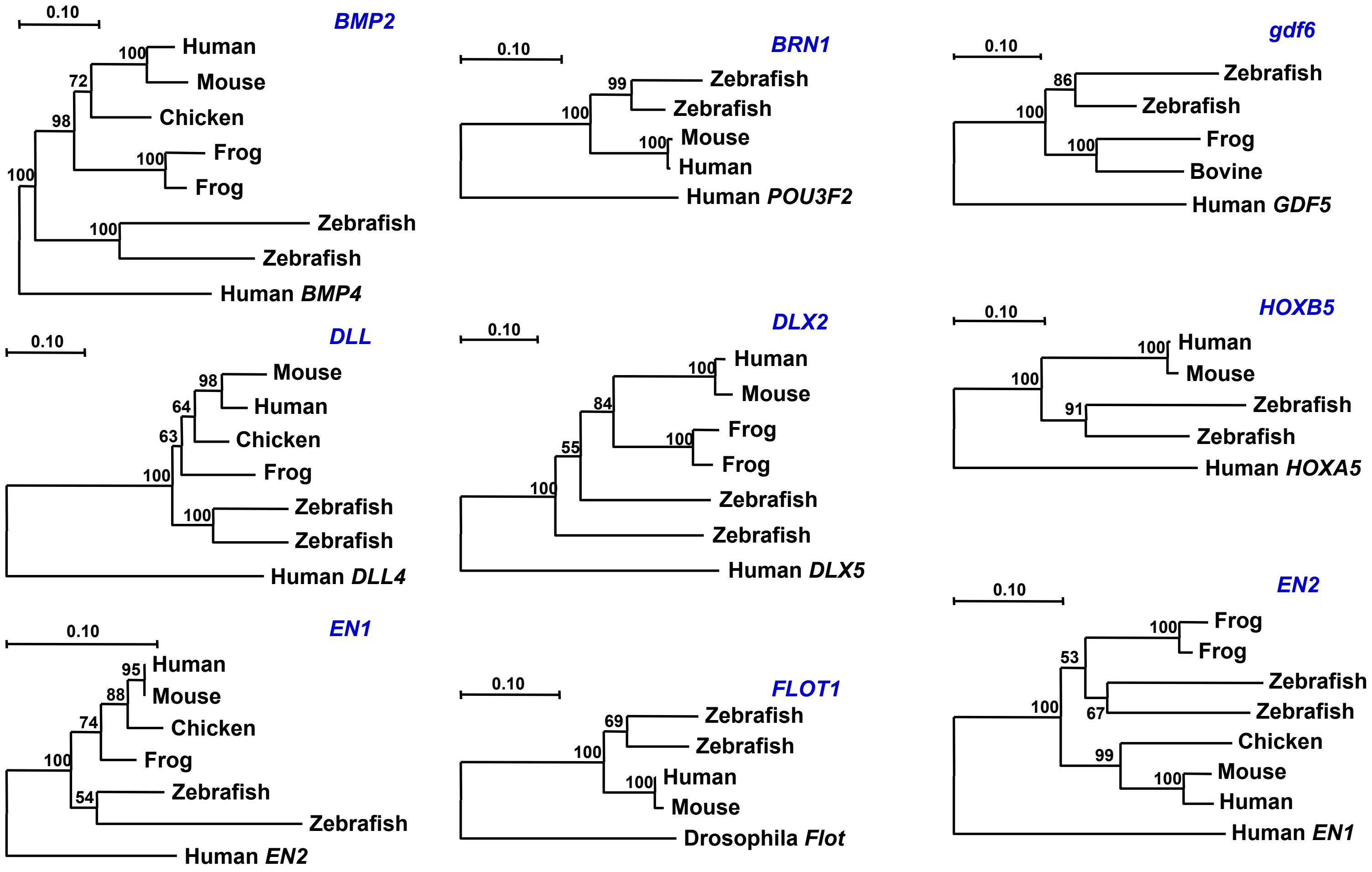
Does comparative genomics provide evidence for an ancient genome duplication event in fish ?

(not only Hox cluster data?)



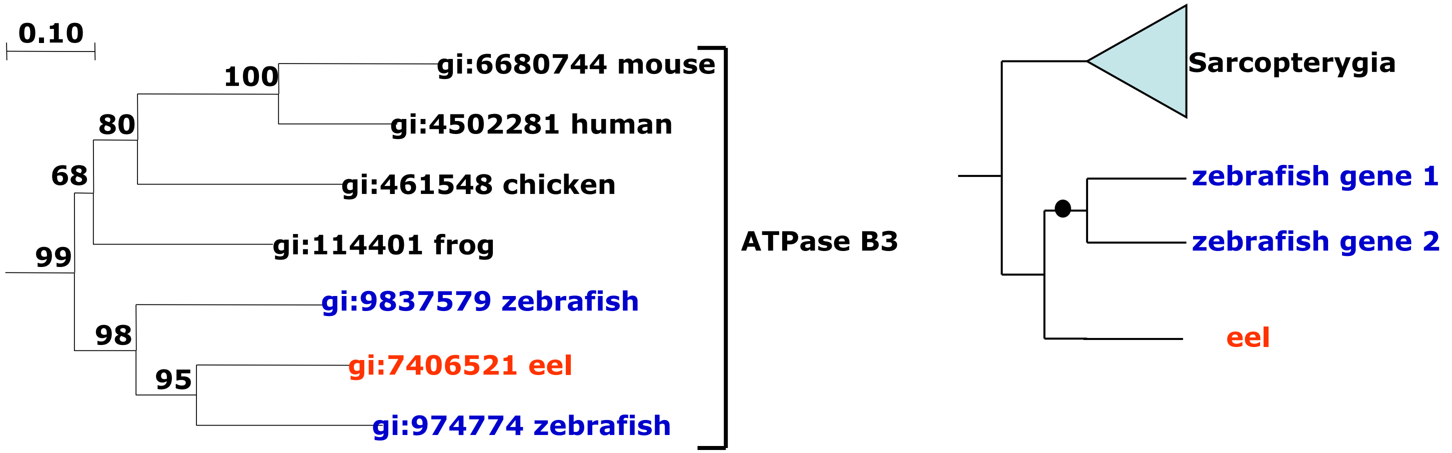


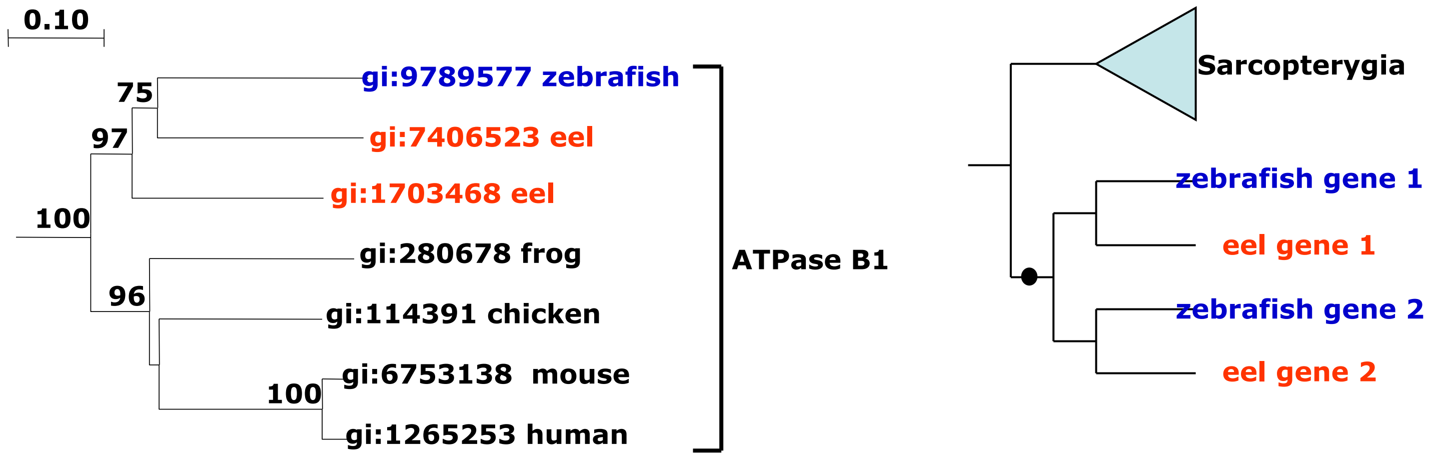
Genes with two copies in zebrafish



Taylor, Van de Peer, Braasch & Meyer (2001) *Phil. Trans. Roy. Soc.*

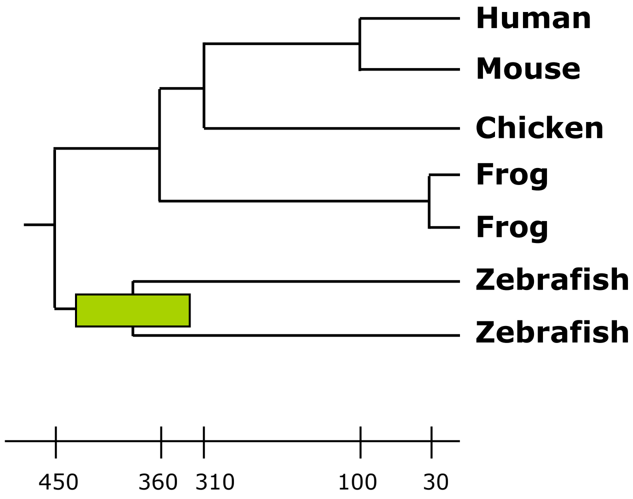
Genome duplication shared by all fish ?

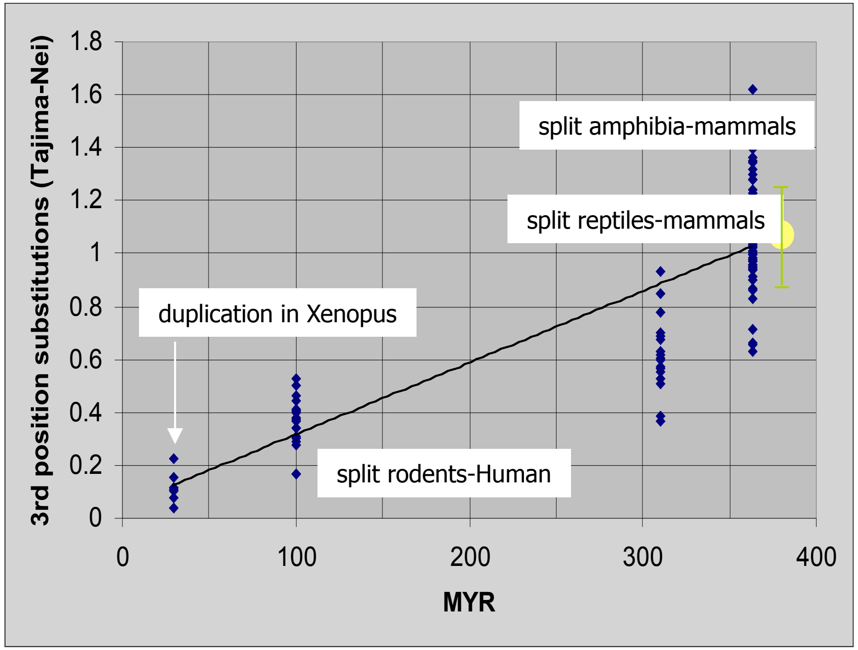




Van de Peer, Frickey, Taylor & Meyer (2002) *Gene*

Dating the fish-specific genome duplication

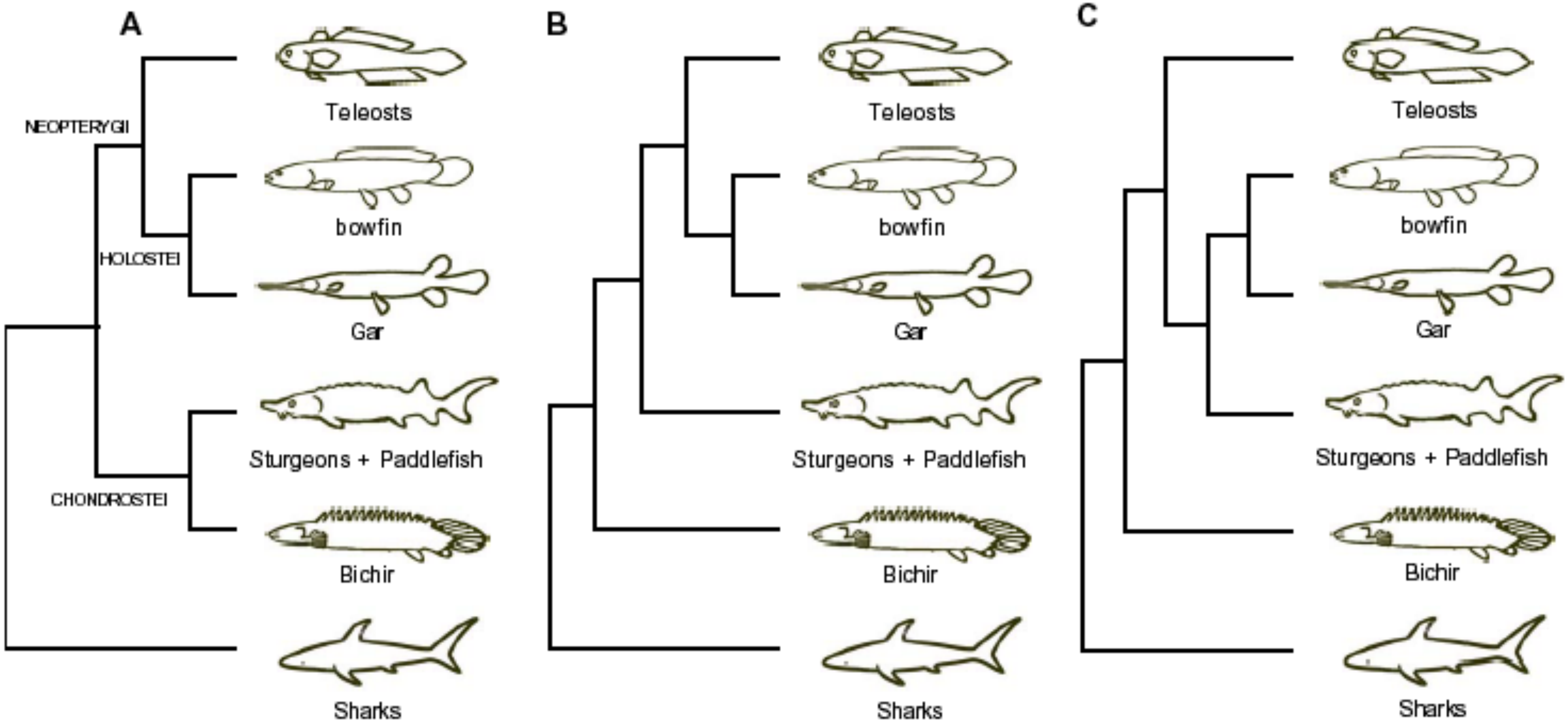




Van de Peer, Taylor, Braatsch & Meyer (2001) *Journal of Molecular Evolution*

Major phylogenetic relationships among basal Actinopterygii:

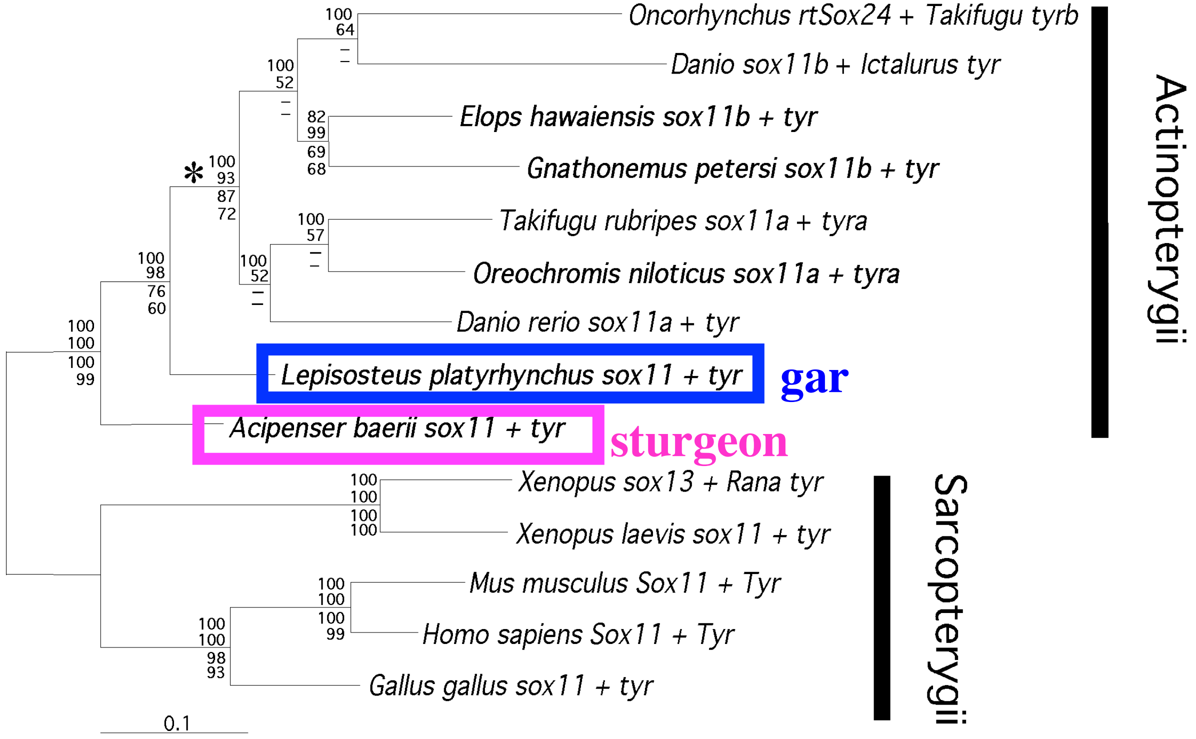
Dating the fish-specific genome duplication



Meyer & Zardoya 2003, *Ann. Rev. Ecol. Syst.*

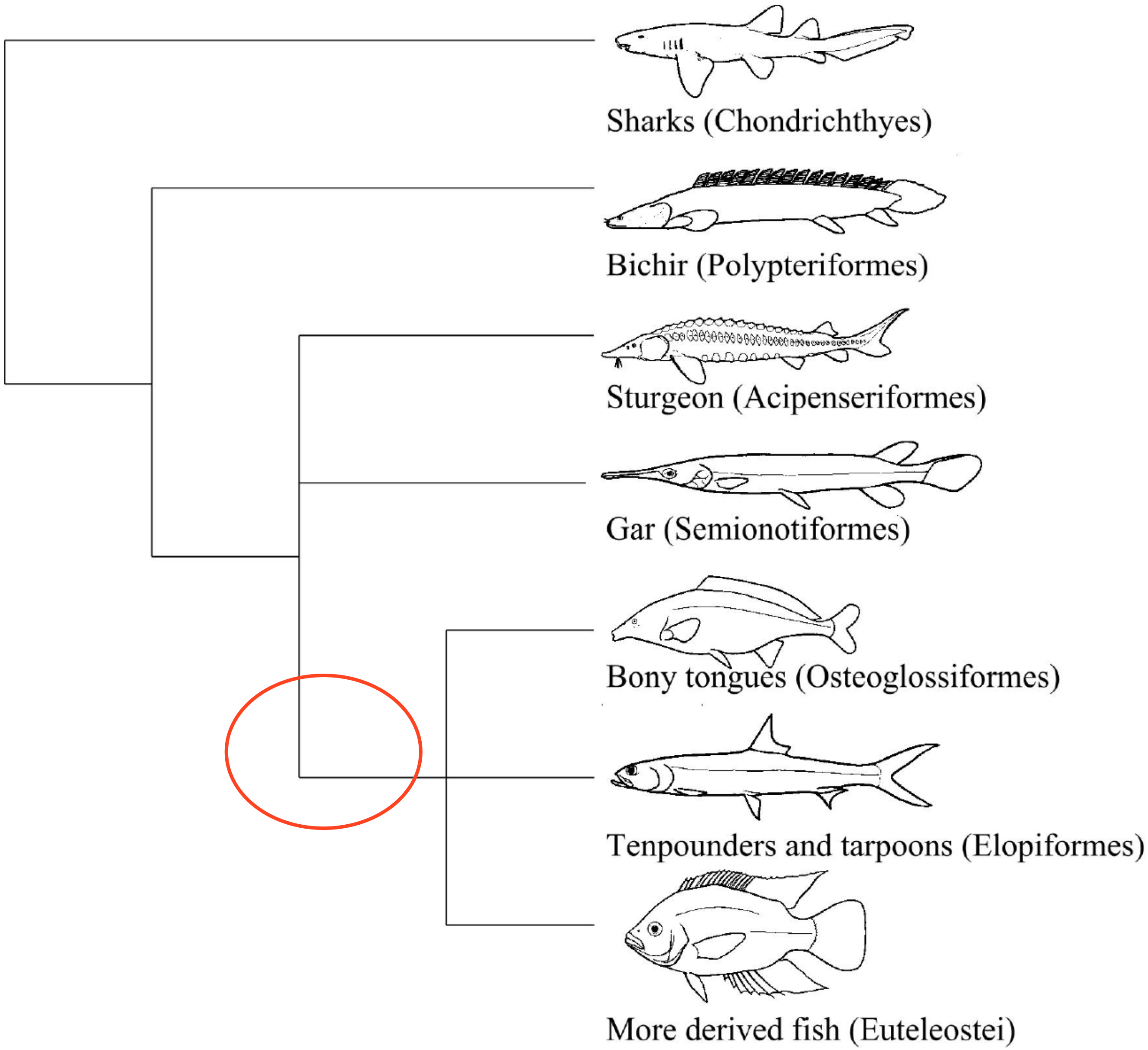
Fish-specific genome duplication (FSGD) precedes major diversification of ray-finned fishes

sox and *tyr* data



Hoegg, Brinkmann, Taylor & Meyer *Journal of Molecular Evolution* (2004)

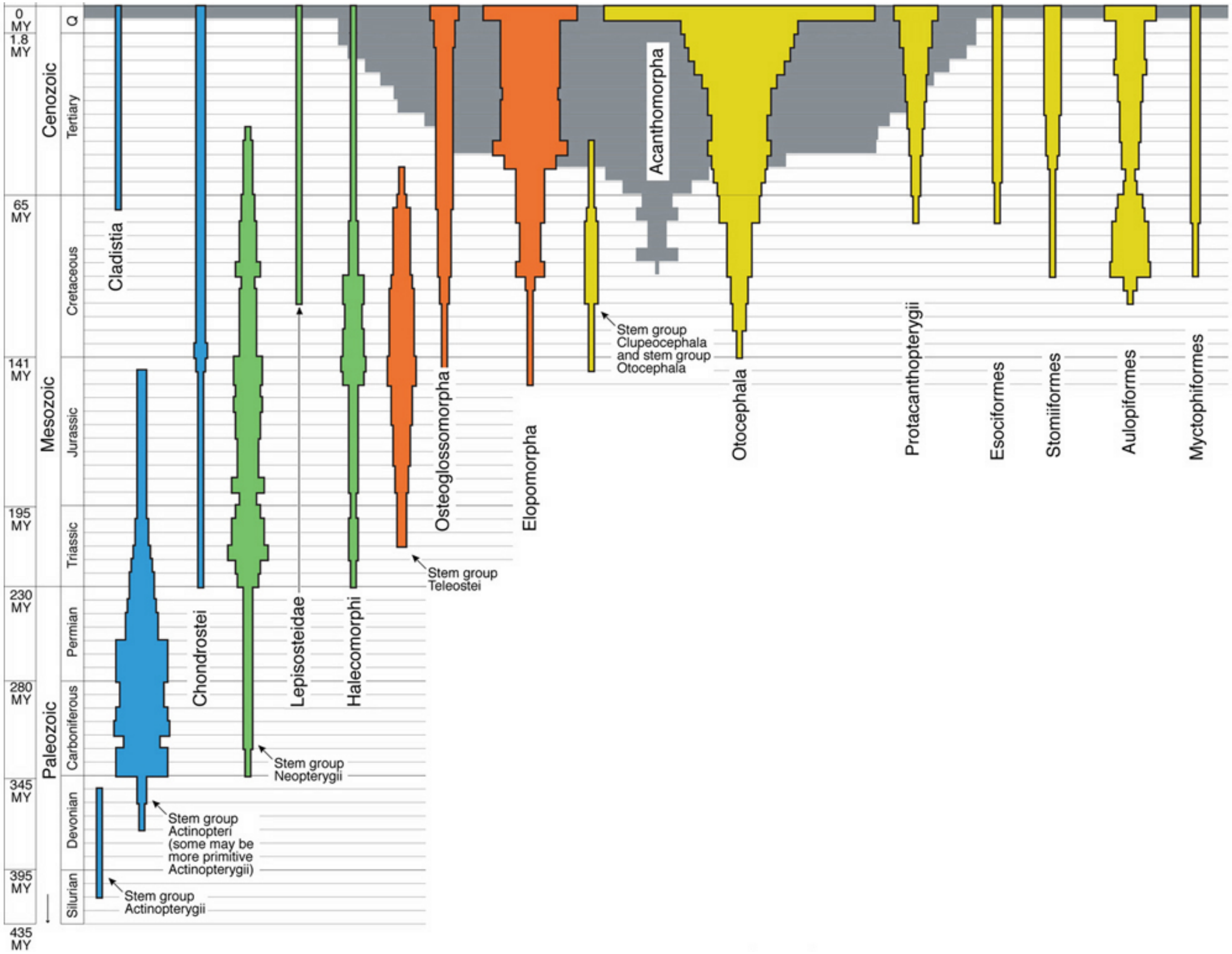
Dating the fish-specific genome duplication



Hoegg, Brinkmann & Meyer *Journal of Molecular Evolution* (2004)

Evolution of the fish genome

Fish: origin, age and diversity through time

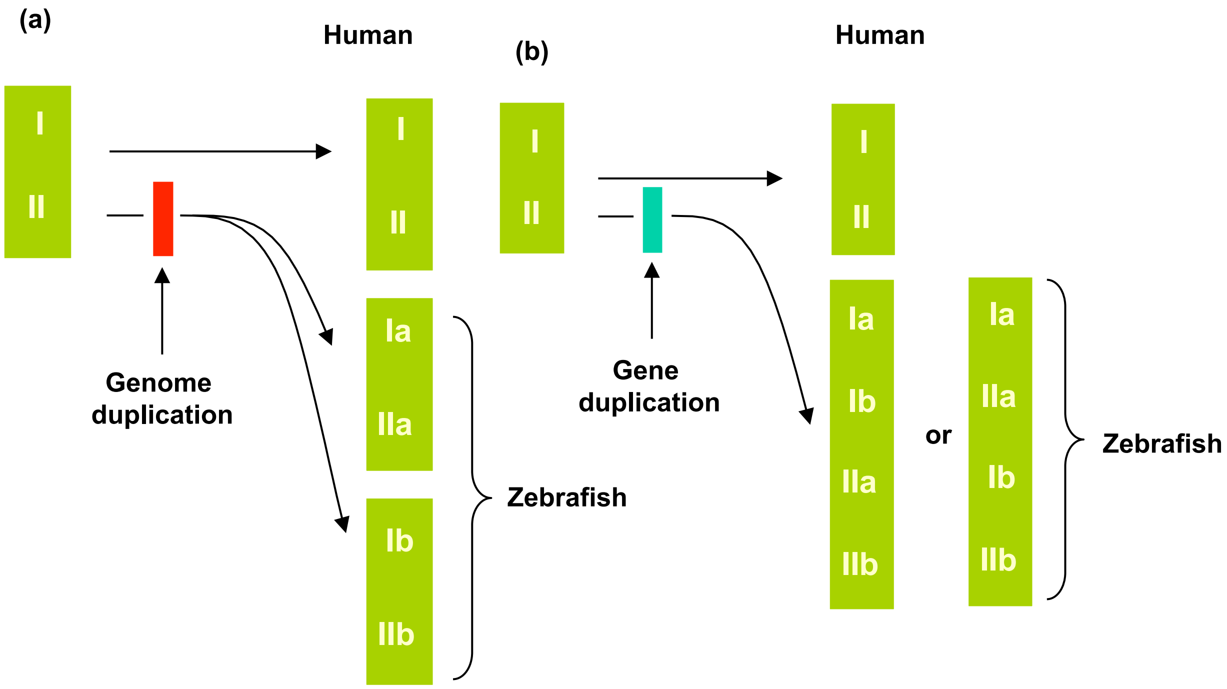


Meyer & Van de Peer (2005) *BioEssays*

Gene or genome duplication?

Synteny:

paralogous chromosomes



(a)

I
II

Genome duplication

Human

I
II

Ia
IIa

Ib
IIb

Zebrafish

(b)

I
II

Gene duplication

Human

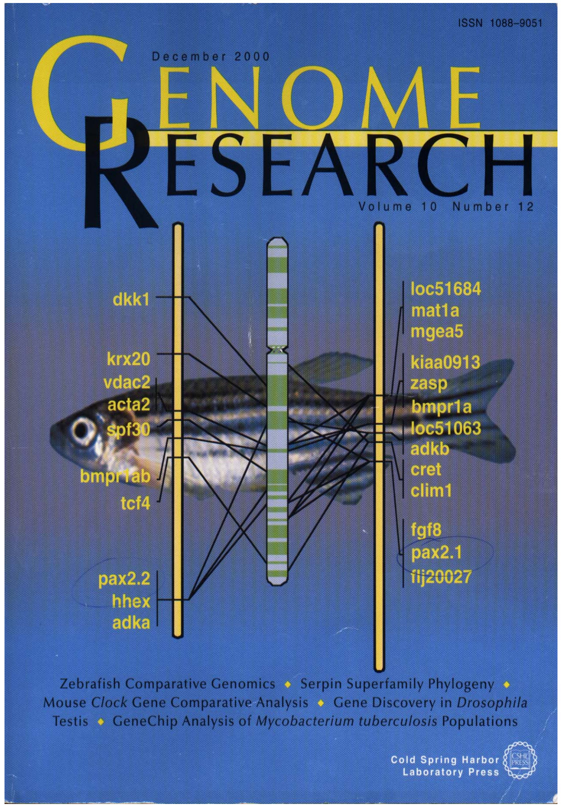
I
II

Ia
Ib
IIa
IIb

or

Ia
IIa
Ib
IIb

Zebrafish



ISSN 1088-9051

December 2000

GENOME RESEARCH

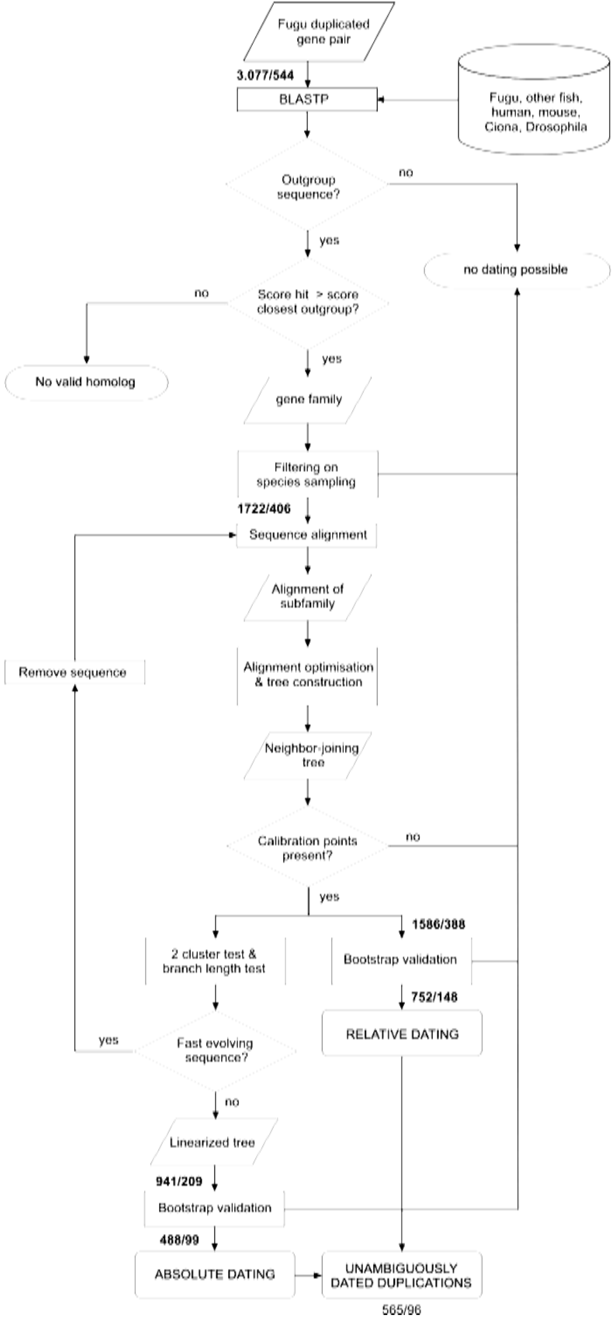
Volume 10 Number 12

Zebrafish Comparative Genomics ◆ Serpin Superfamily Phylogeny ◆ Mouse *Clock* Gene Comparative Analysis ◆ Gene Discovery in *Drosophila* Testis ◆ GeneChip Analysis of *Mycobacterium tuberculosis* Populations

Cold Spring Harbor Laboratory Press

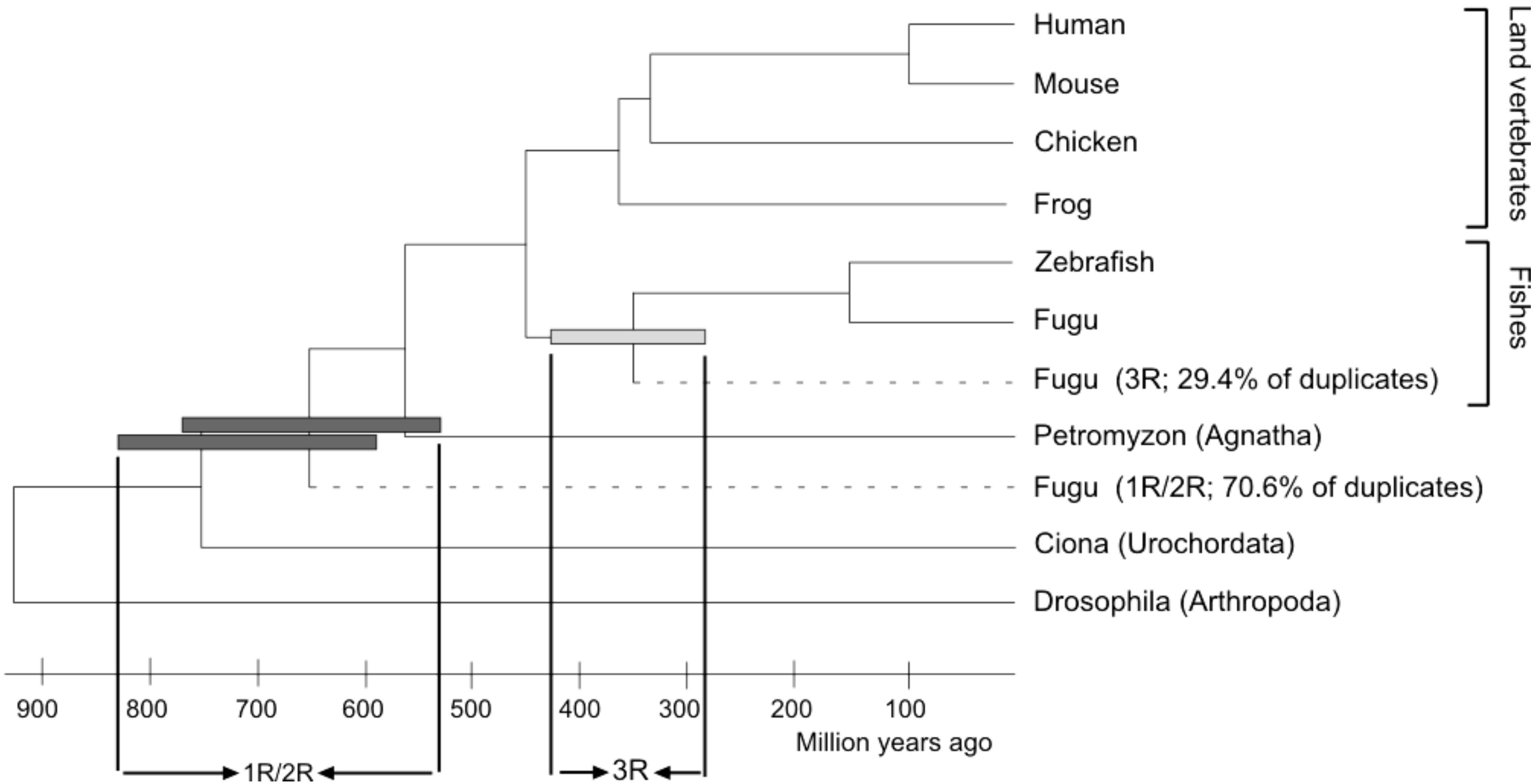
Meyer & Schartl (1999) *Curr. Op. Cell. Biol.*

Evidence for a fish-specific (3R) genome duplication - analyses of the paranome



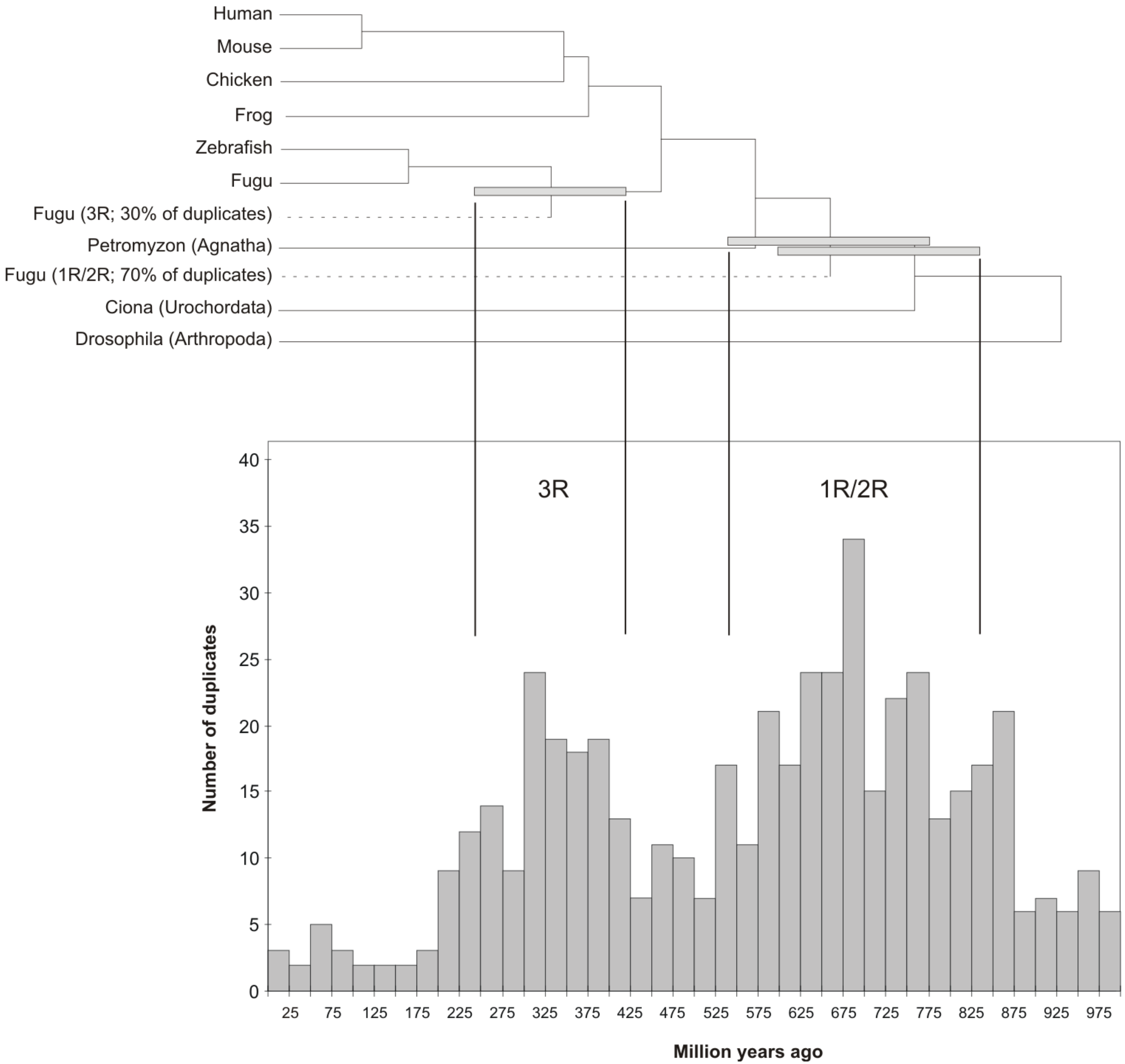
Vanderpoele, Wouter, Taylor, Meyer and Van de Peer *PNAS* (2004)

Evidence for a fish-specific (3R) genome duplication



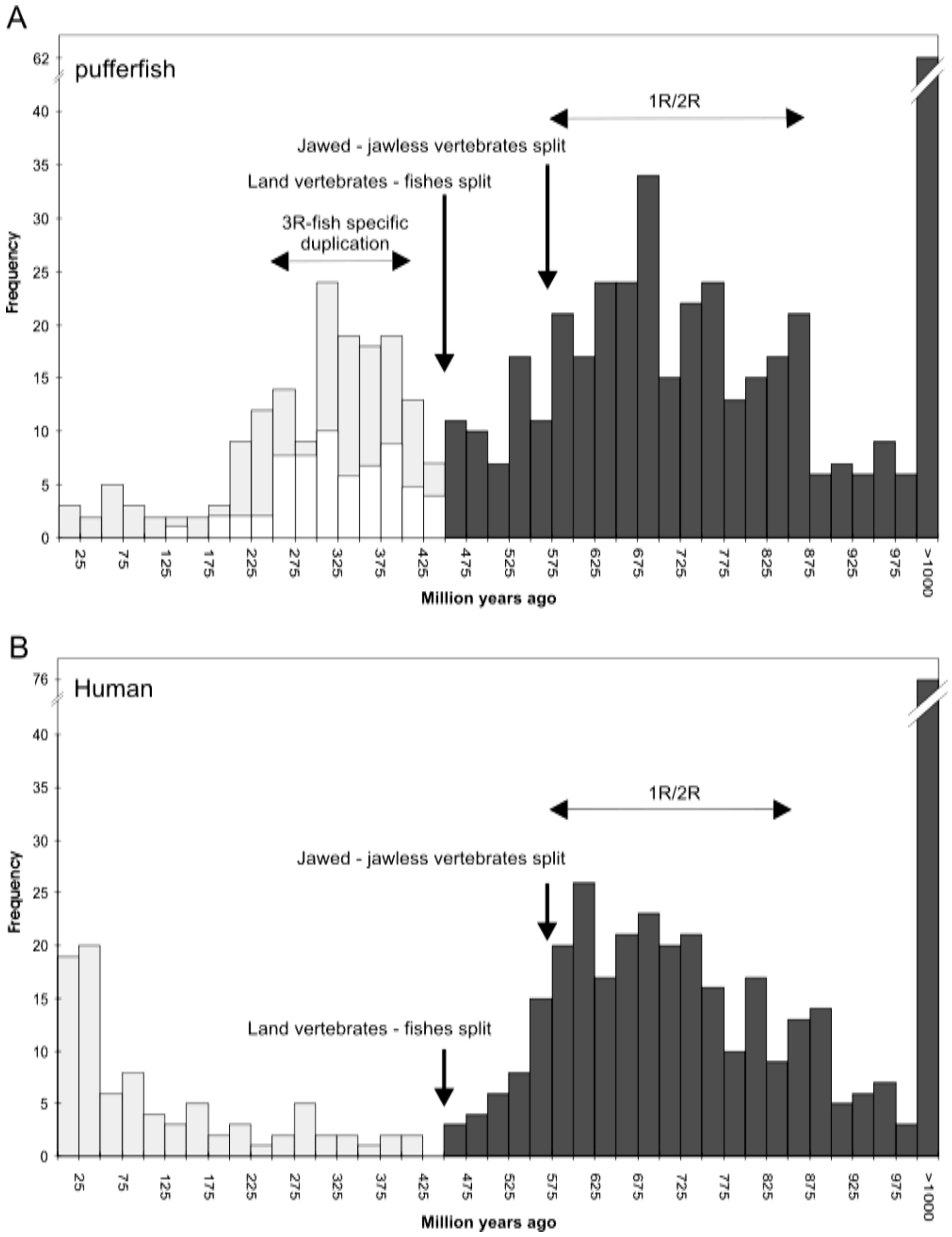
Vanderpoele, Wouter, Taylor, Meyer and Van de Peer *PNAS* (2004)

Evidence for a fish-specific (3R) genome duplication



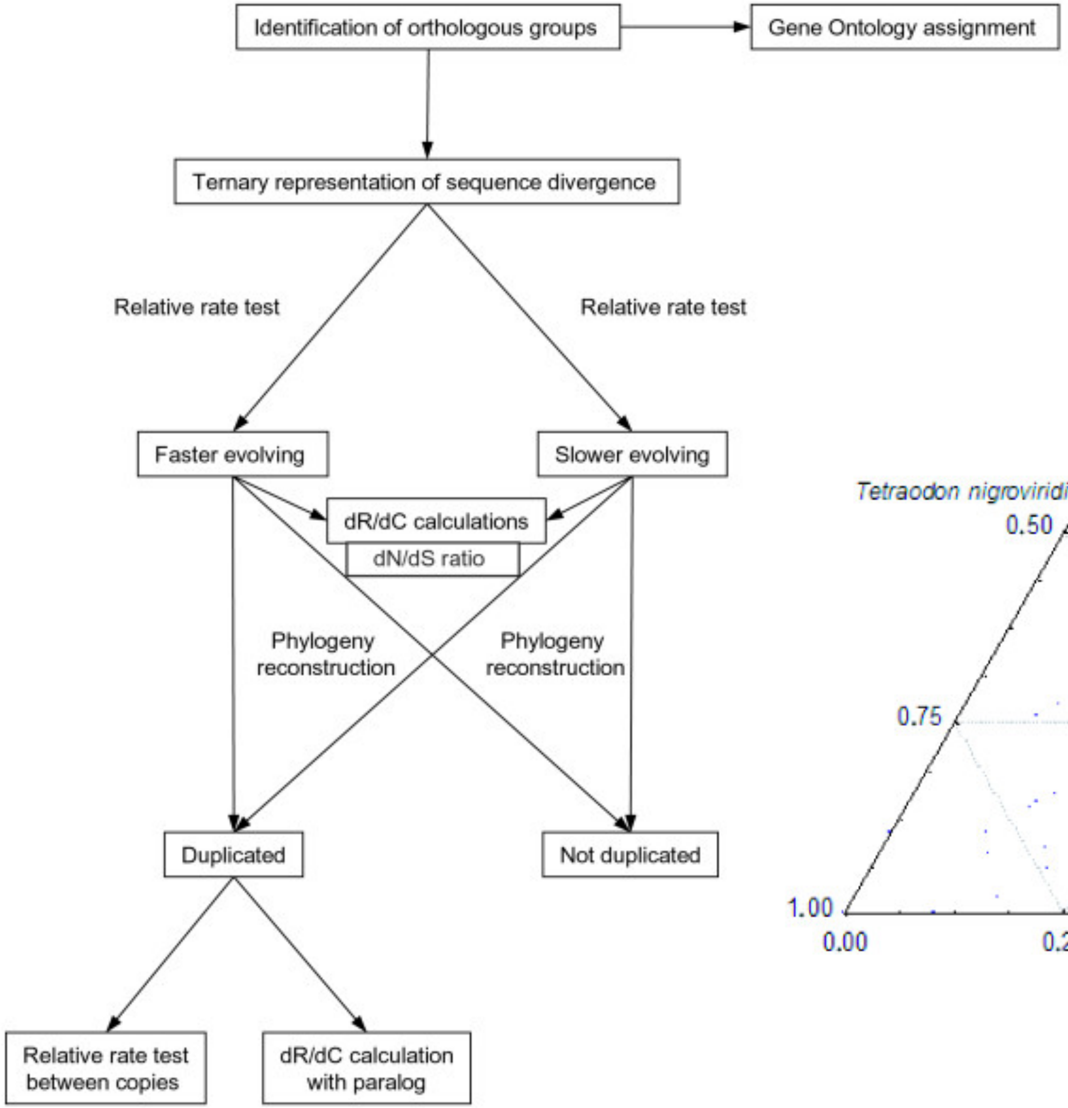
Meyer and Van de Peer *BioEssays* (2005)

Evidence for a fish-specific (3R) genome duplication
- estimated ages of duplicate genes (paralogs)



Vanderpoele, Wouter, Taylor, Meyer and Van de Peer *PNAS* (2004)

Asymmetric rates of duplicated genes in fish



Identification of orthologous groups

Gene Ontology assignment

Ternary representation of sequence divergence

Relative rate test

Relative rate test

Faster evolving

Slower evolving

dR/dC calculations

dN/dS ratio

Phylogeny reconstruction

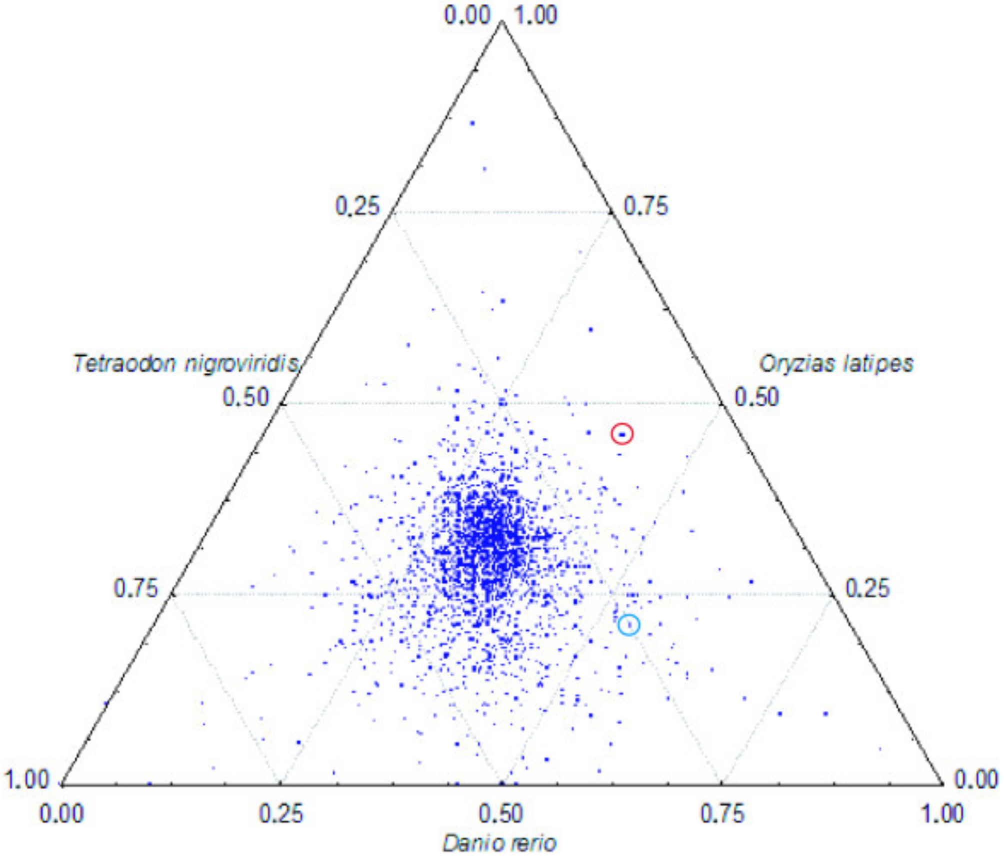
Phylogeny reconstruction

Duplicated

Not duplicated

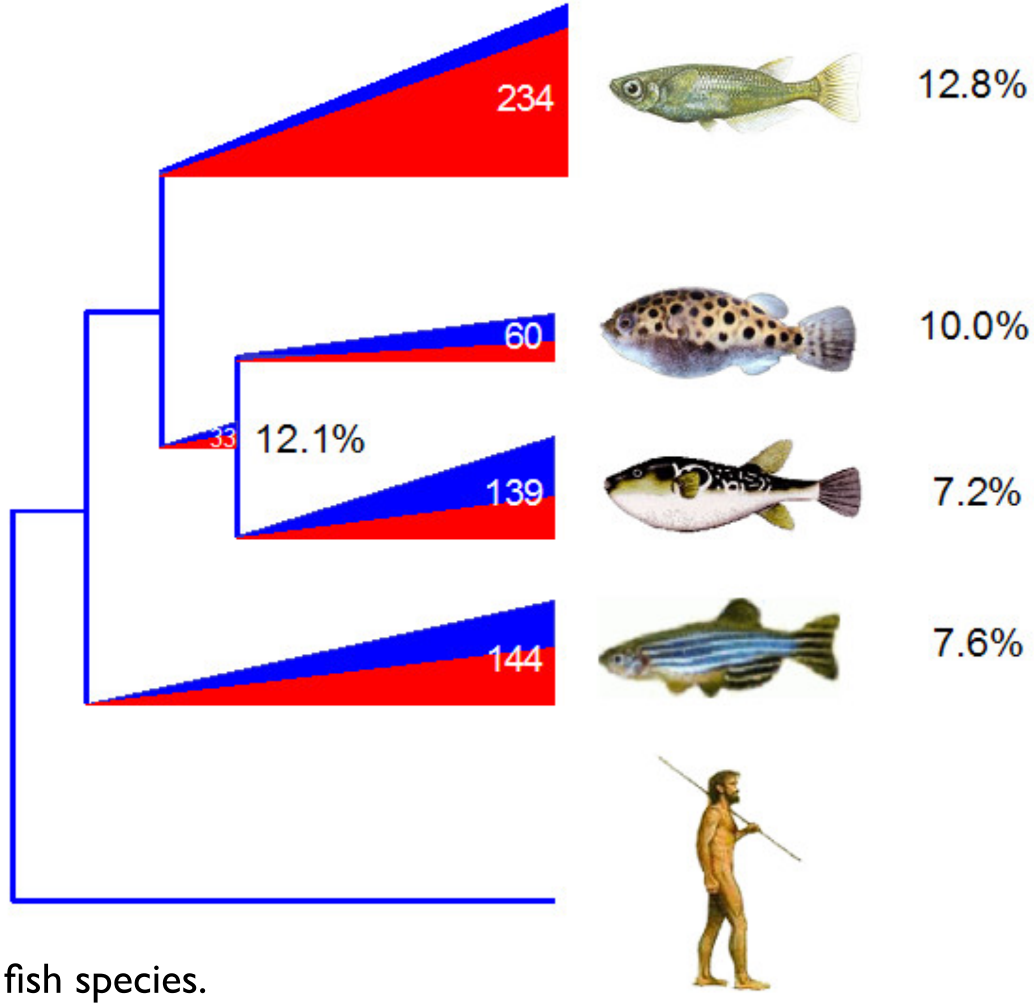
Relative rate test between copies

dR/dC calculation with paralog



Steinke, Salzburger Braasch & Meyer *BMC Genomics* (2006)

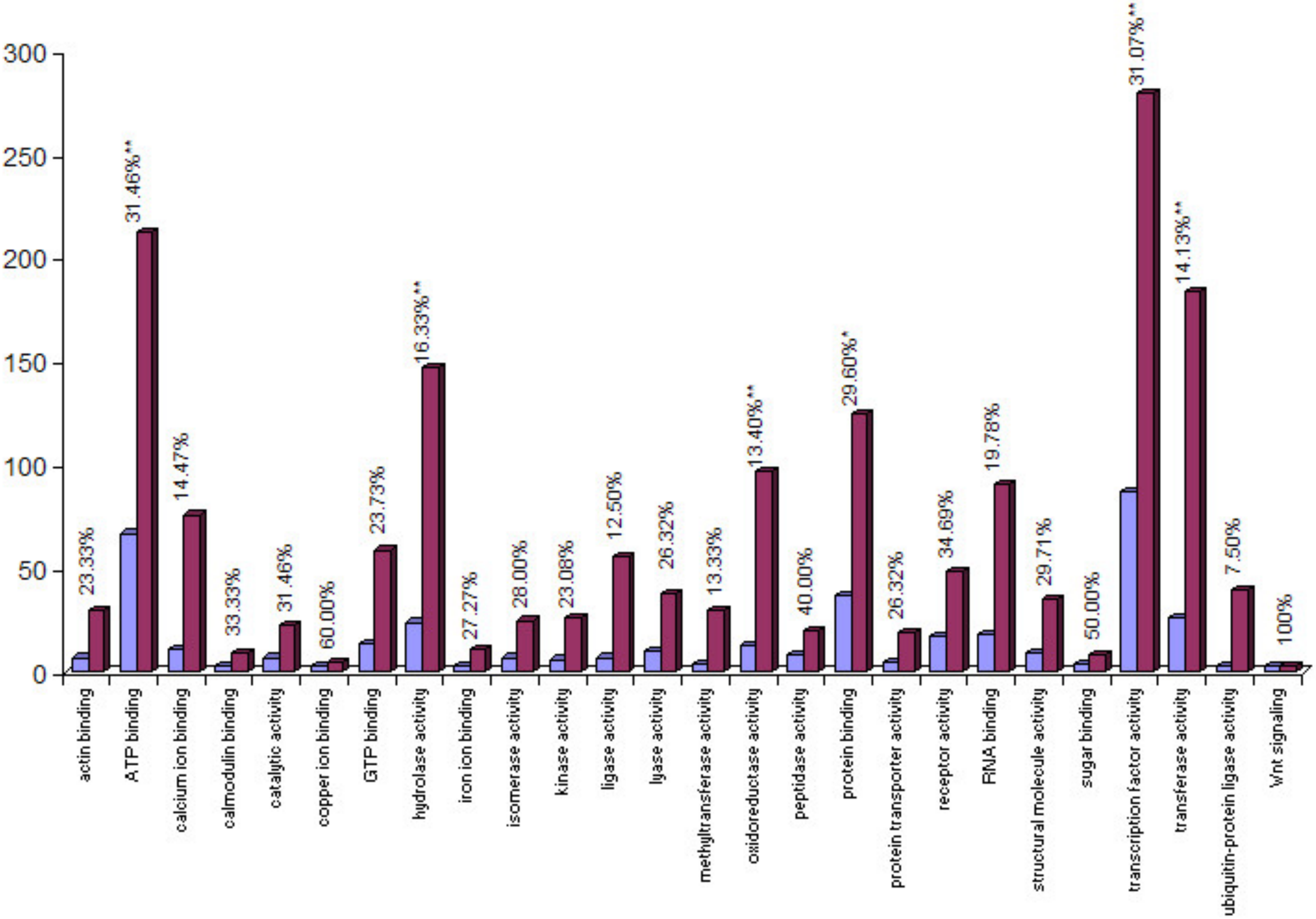
Asymmetric rates of duplicated genes in fish



Phylogeny of the studied fish species.
The proportion of genes with significantly lower (red) or higher (blue) distance to the human ortholog than to other fish species, are mapped onto the phylogeny as proportional triangles.
Numbers within the triangle represent the total abundance of those genes.
The percentages represent the corresponding proportion of transcription factors.

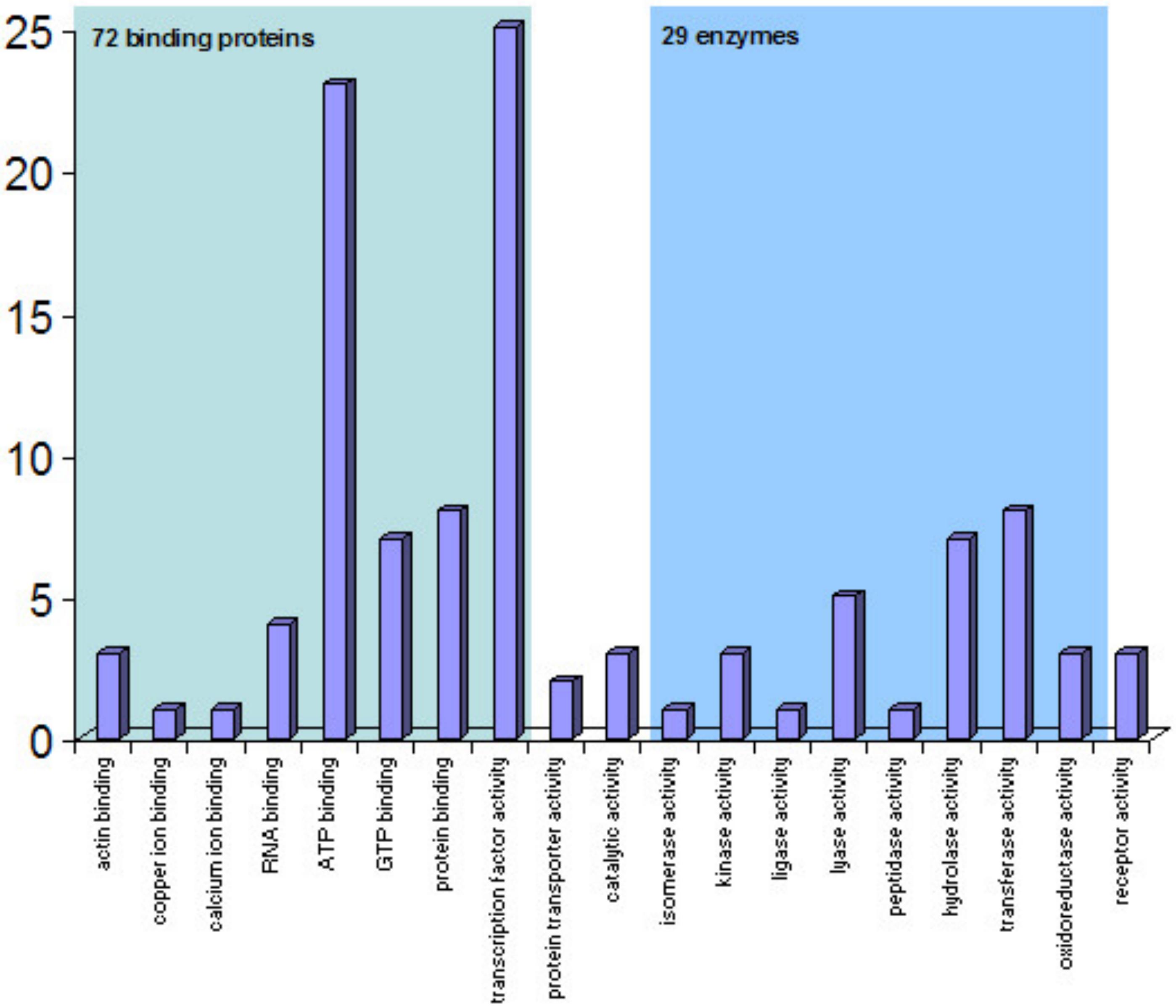
Steinke, Salzburger Braasch & Meyer *BMC Genomics* (2006)

Asymmetric rates of duplicated genes in fish



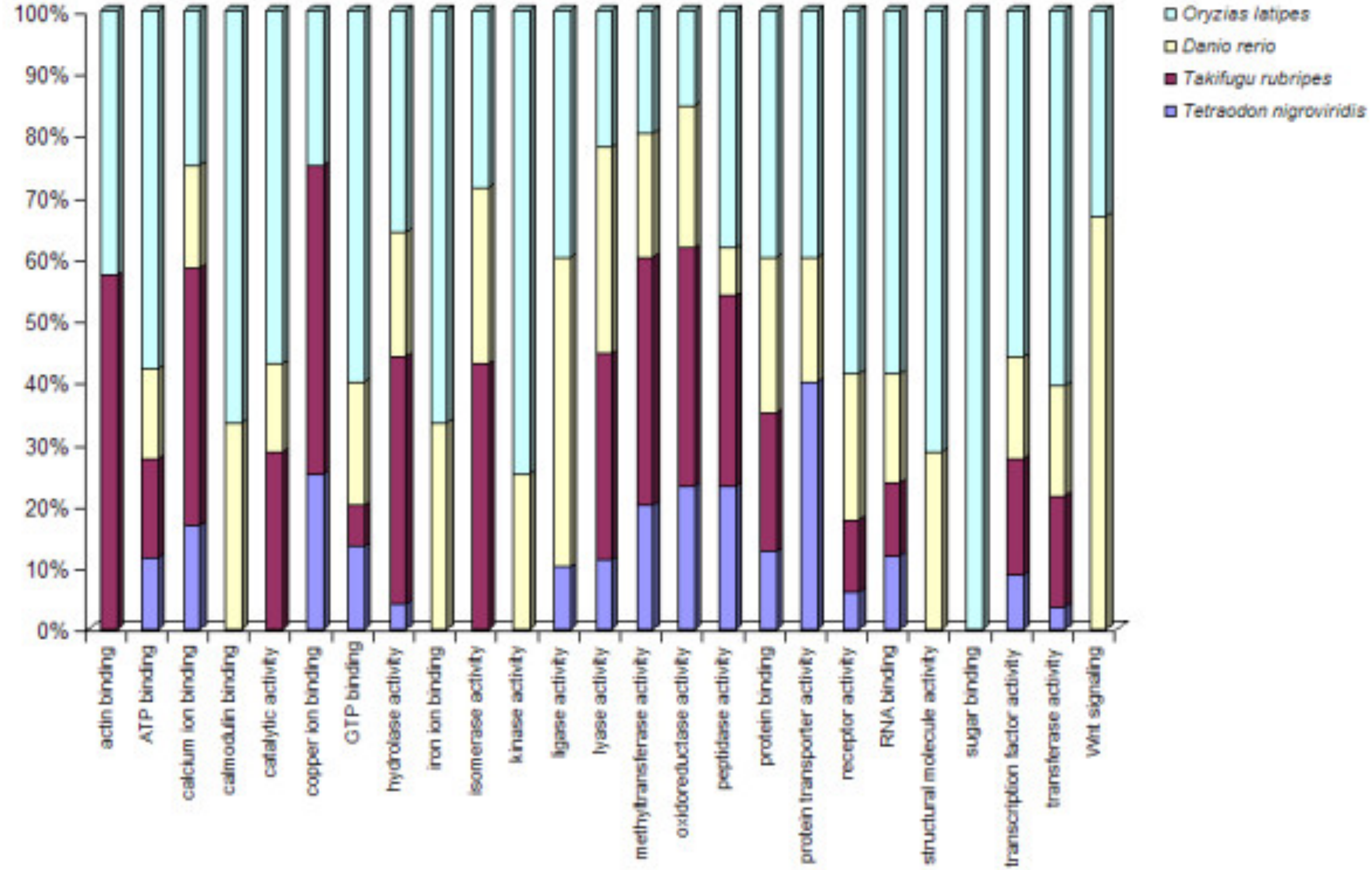
Steinke, Salzburger Braasch & Meyer *BMC Genomics* (2006)

Asymmetric rates of duplicated genes in fish



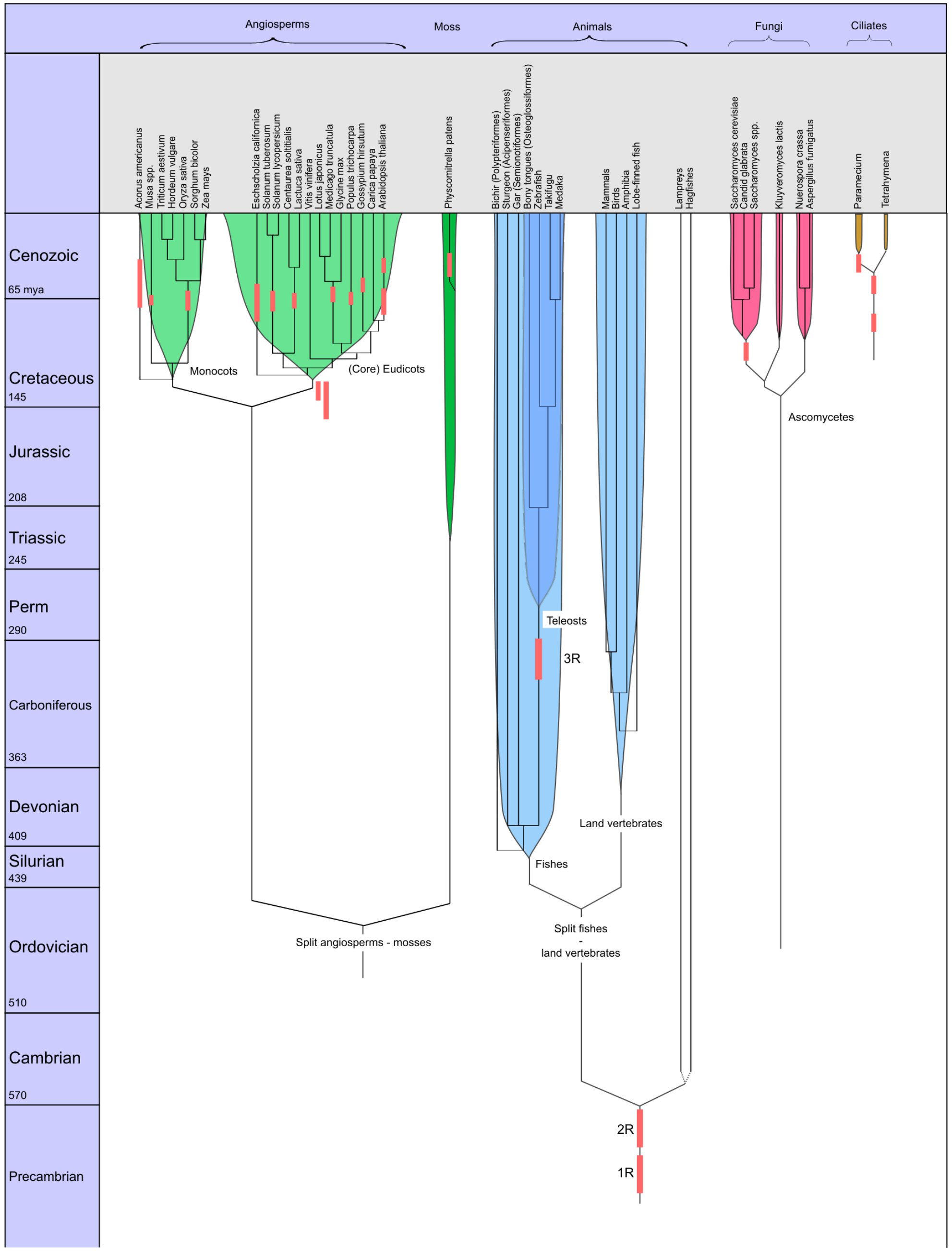
Steinke, Salzburger Braasch & Meyer *BMC Genomics* (2006)

Asymmetric rates of duplicated genes in fish



Steinke, Salzburger Braasch & Meyer *BMC Genomics* (2006)

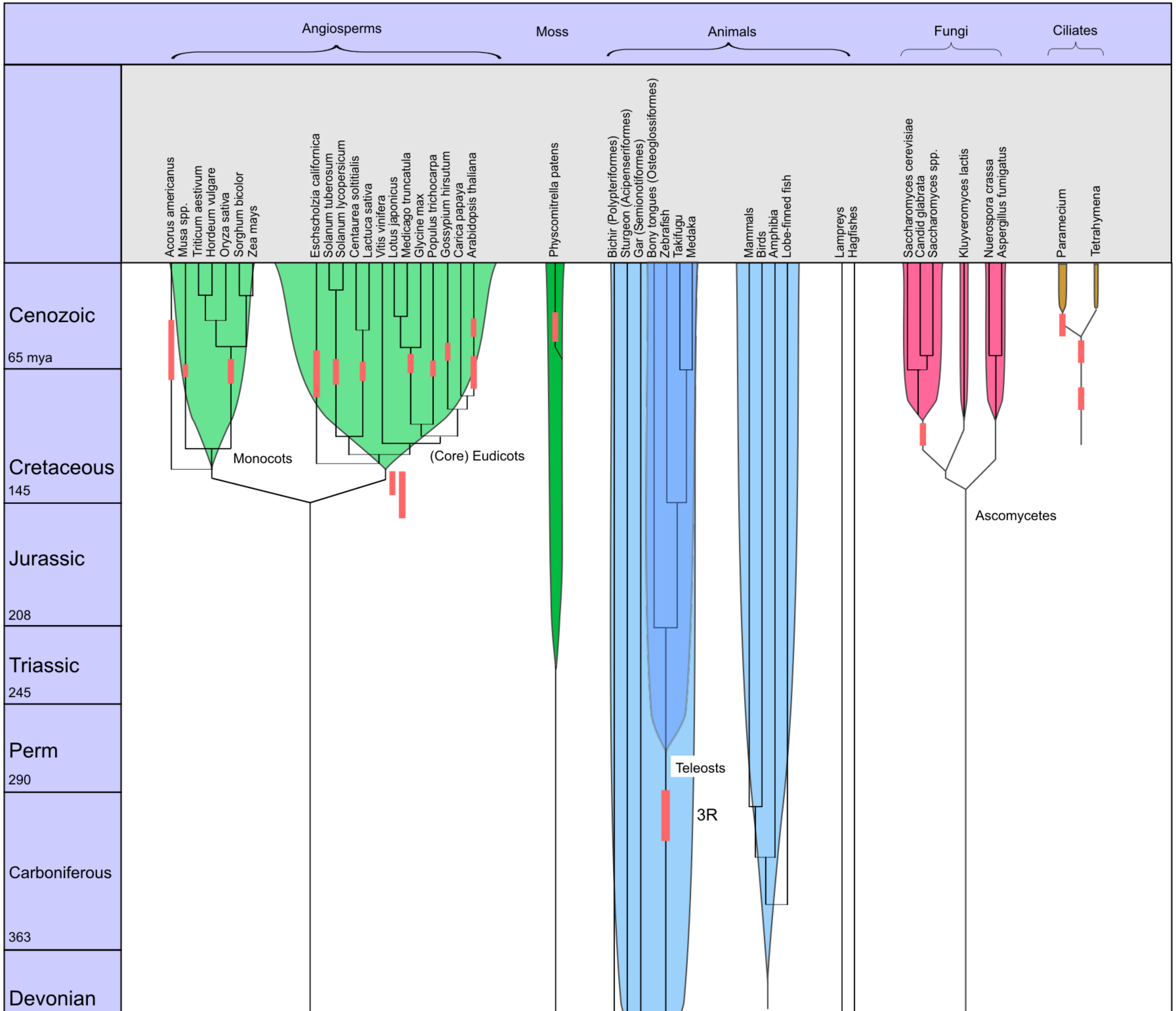
Known genome duplications in eukaryotes



Whole genome duplications across the phylogeny of eukaryotes. Schematic and heavily pruned phylogenetic tree of eukaryotes, showing taxa that have experienced whole genome duplications. Paleopolyploidy events are indicated as red bars, and are based on studies published previously for plants [6, 21, 28], fishes [11, 61, 62], vertebrates [120], fungi [22], and ciliates [14].

Van de Peer, Maere & Meyer *Nature Review Genetics* (2009)

Known genome duplications in eukaryotes



Van de Peer, Maere & Meyer *Nature Review Genetics* (2009)

Reciprocal gene loss or subfunctionalization and speciation

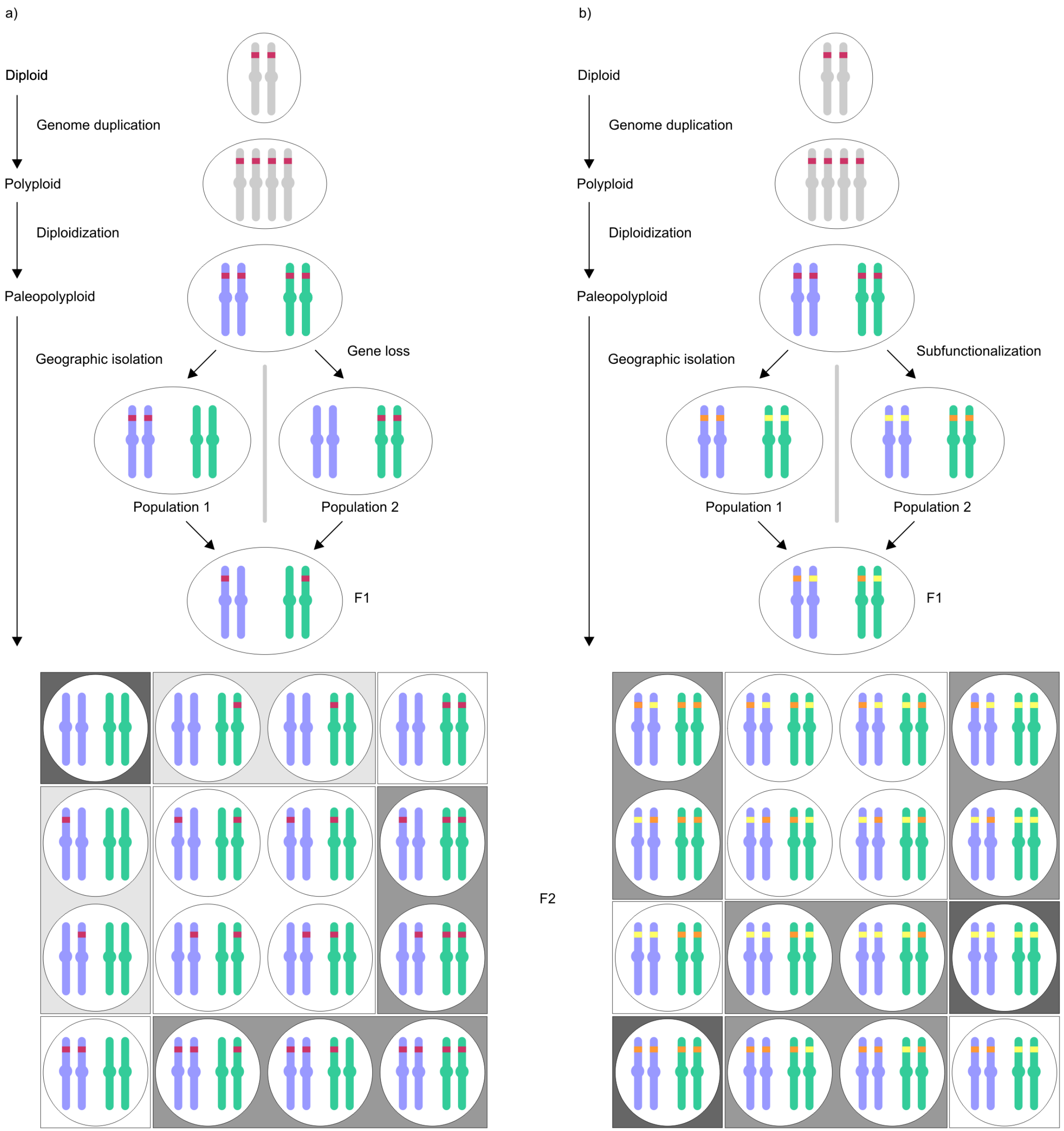


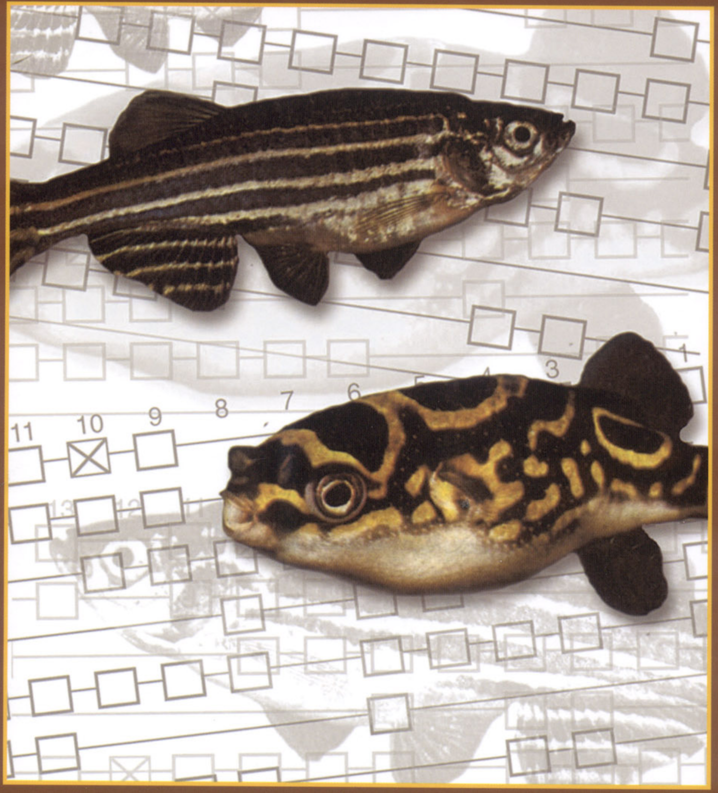
Figure 3: Reciprocal gene loss or subfunctionalization facilitates speciation. (a) Red bands on chromosomes represent a locus that is duplicated (along with all other loci) during a tetraploidization event. In this hypothetical example, following diploidization, the duplicated gene is present on two different chromosomes. After geographic isolation, both populations have lost the duplicate on different chromosomes. If individuals from isolated populations mate, their 'hybrid' progeny would be heterozygous, possessing a functional allele at each locus of the duplicated gene. However, crosses between the F1 individuals produce some (approximately 6%) 1/16th F2 individuals with only null alleles at both loci in question (black square), and therefore lacking viability and/or fertility. Others would receive from one allele (light grey squares), which might reduce functionality when a gene is haploinsufficient to three or four functional alleles (dark grey squares), which might have a negative dosage effect. All these might lead to post-mating reproductive isolation [52]. (b) Red bands again represent a locus that is duplicated during a tetraploidization event but this time, following diploidization and geographic isolation, the duplicated genes in the different populations have subfunctionalized (orange and yellow bands on chromosomes). Hybrids between the two populations should in general develop normally but in the F2, a portion of their offspring will be homozygous for alleles lacking one essential subfunction (1/16th of the F2) or the other (another 1/16th) (black squares), thus reducing the fitness of hybridizing individuals. Other F2 individuals might, as in (a) show reduced fitness due to dosage of haploinsufficiency effects.

Van de Peer, Maere & Meyer *Nature Review Genetics* (2009)

Genome Evolution

Gene and Genome Duplications and the Origin of Novel Gene Functions

Axel Meyer and Yves Van de Peer (eds.)



Kluwer Academic Publishers