3- Gene and genome duplication

Marc Robinson-Rechavi

SIB course Biodiversity bioinformatics 2023

@marcrr@ecoevo.social

www.bgee.org

Main types of duplication

Retroposition

- frequent

- most new paralogs dead on arrival

Tandem duplication

- frequent

- can lead to arrays of tandem duplicates

Genome duplication

- rare

- impacts whole genome

Discussion

What's a higher animal, how does it relate to gene or
genome duplication?

Hox cluster history
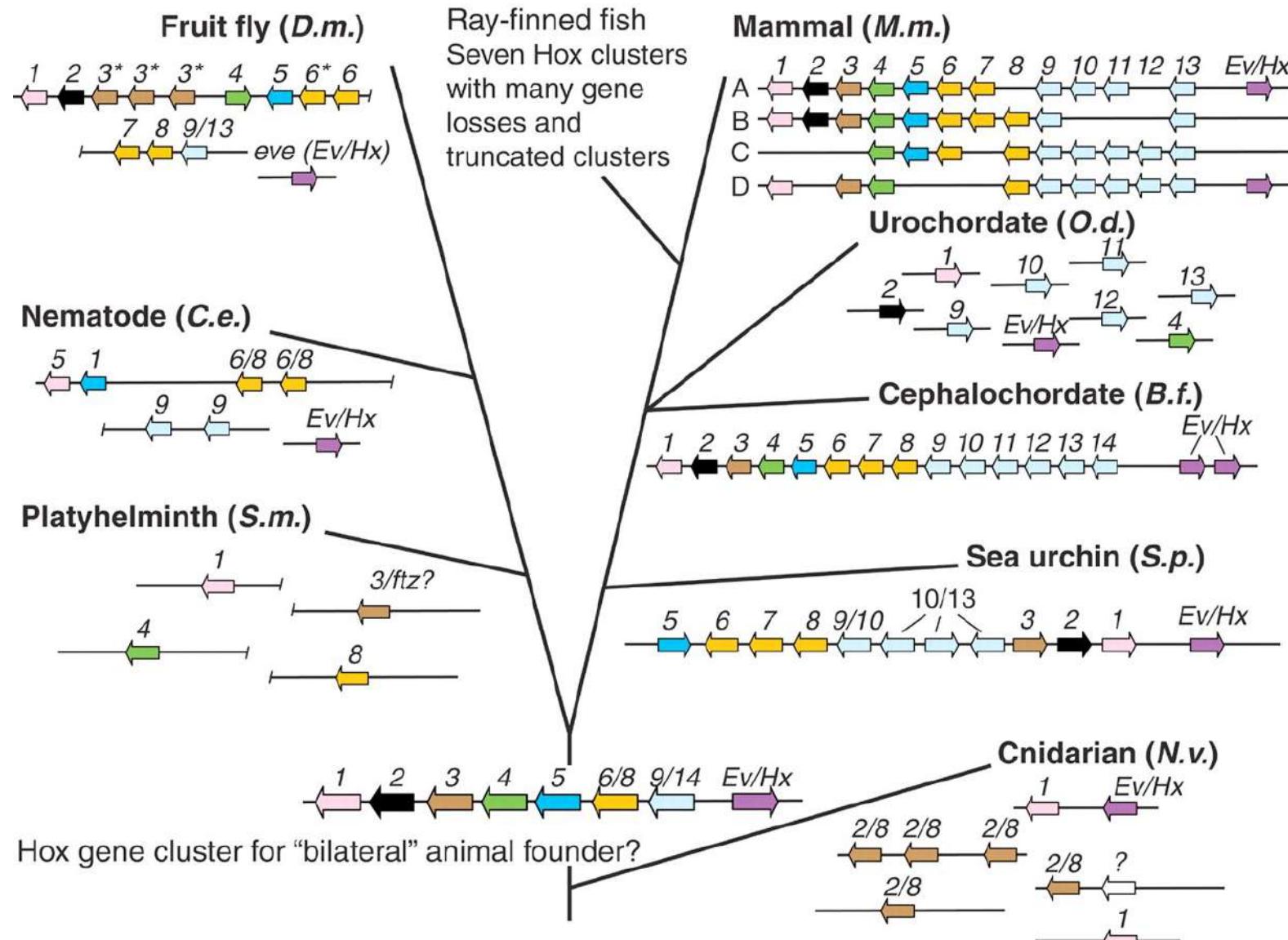
Mammal olfactory receptor genes
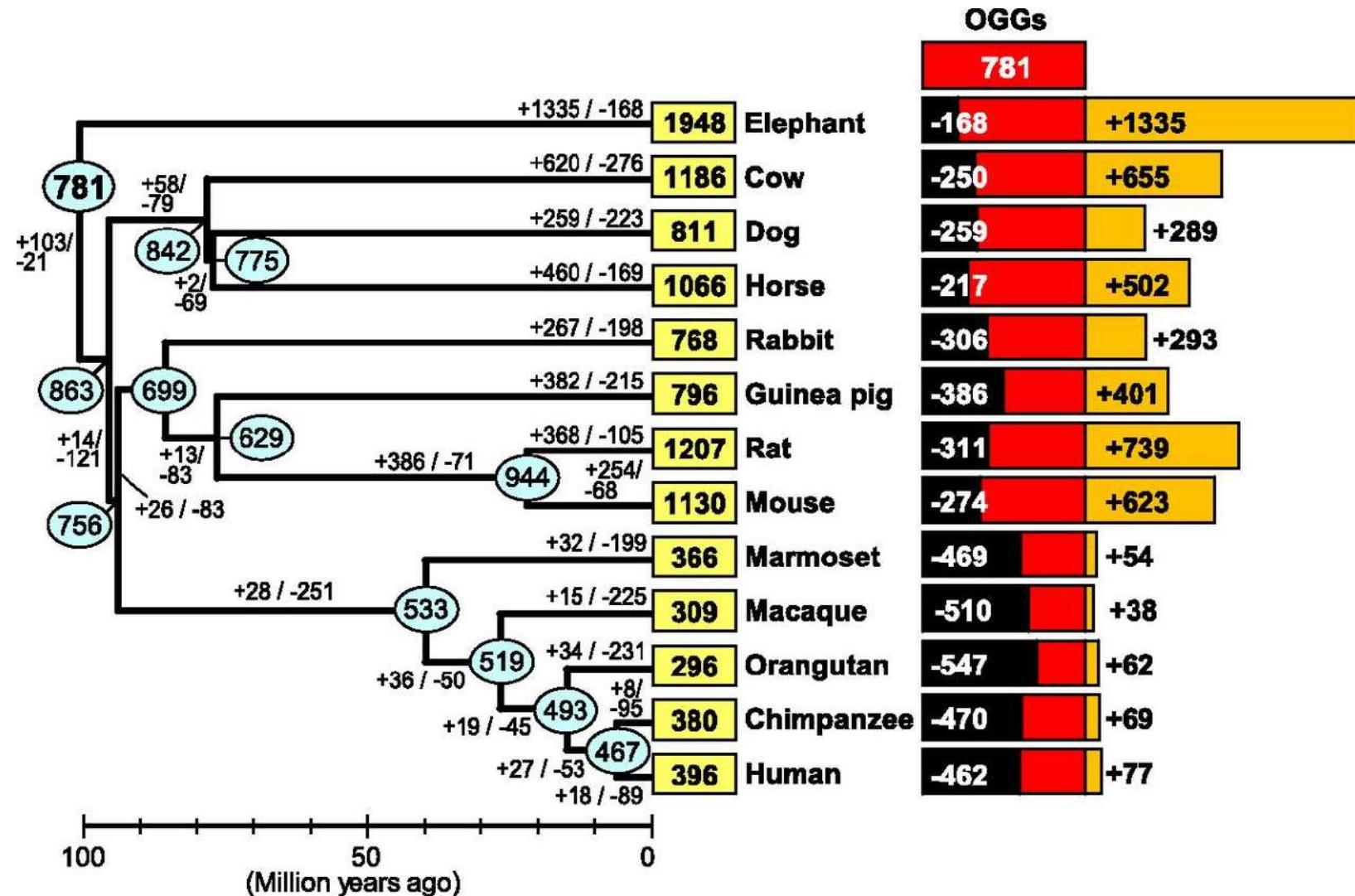
OGGs = ortholog gene groups

Niimura et al https://doi.org/10.1101/gr.169532.113

# Elephant olfactory receptor genes
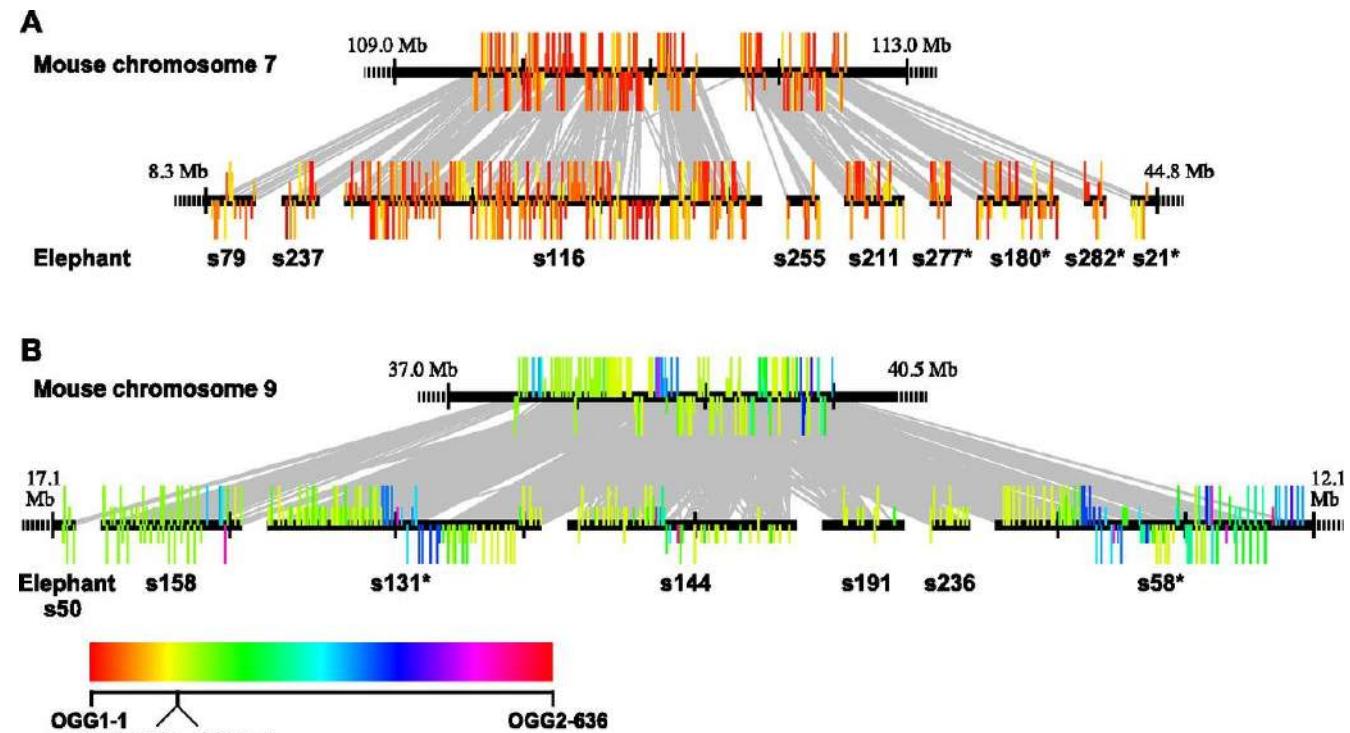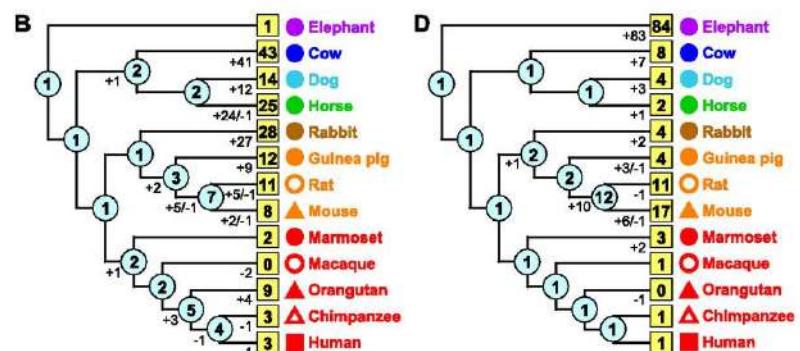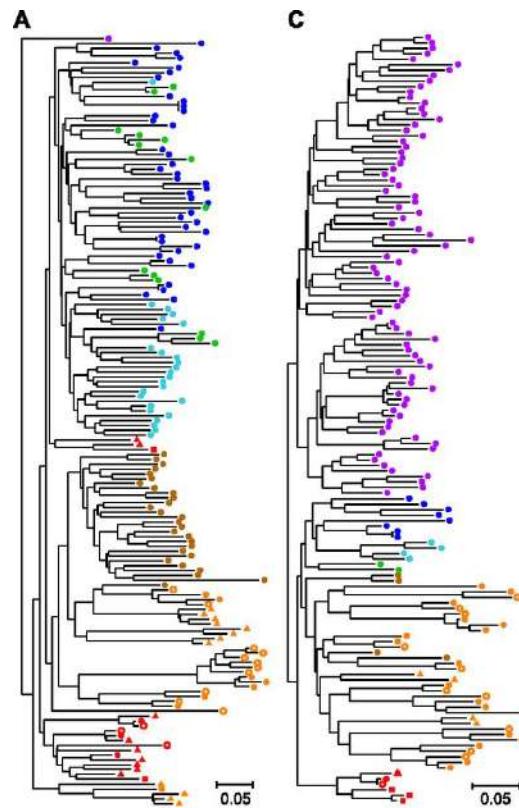
Visual opsin families in vertebrates
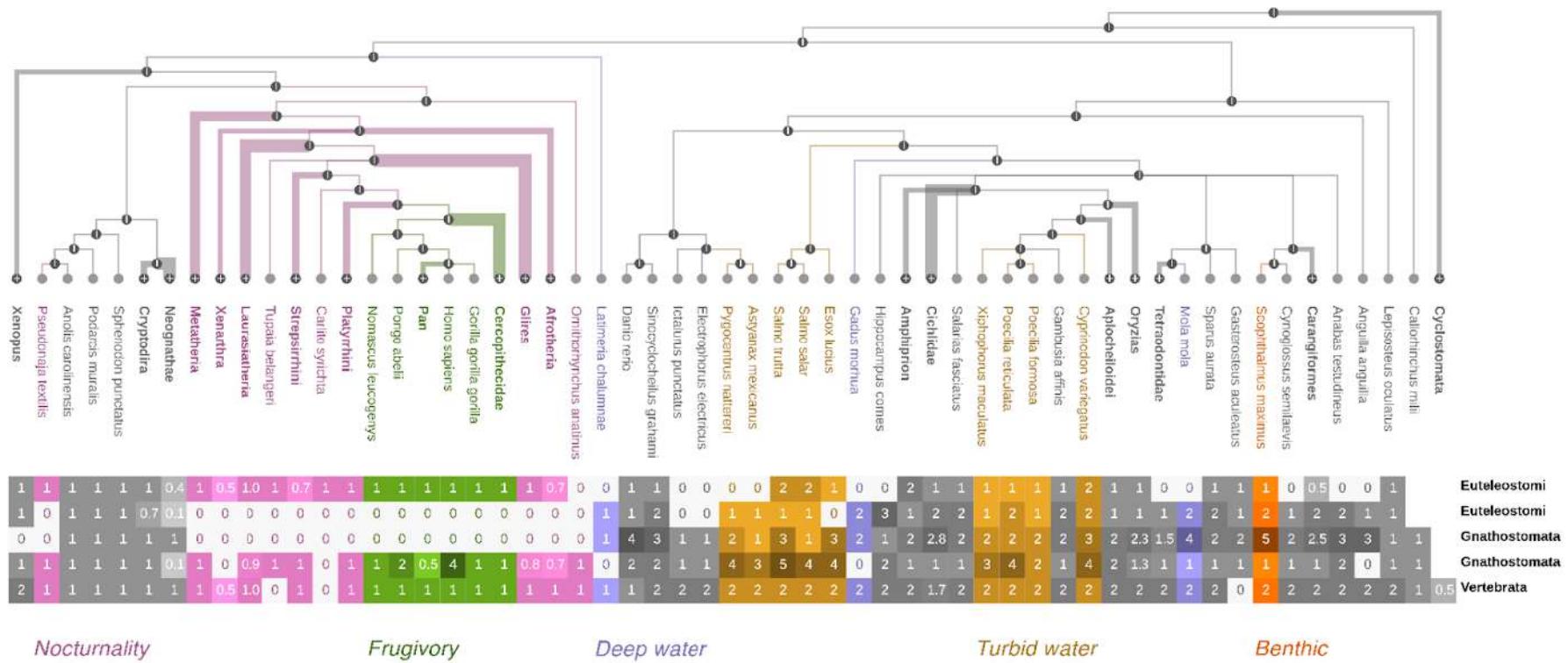
Genome duplication in teleost fishes
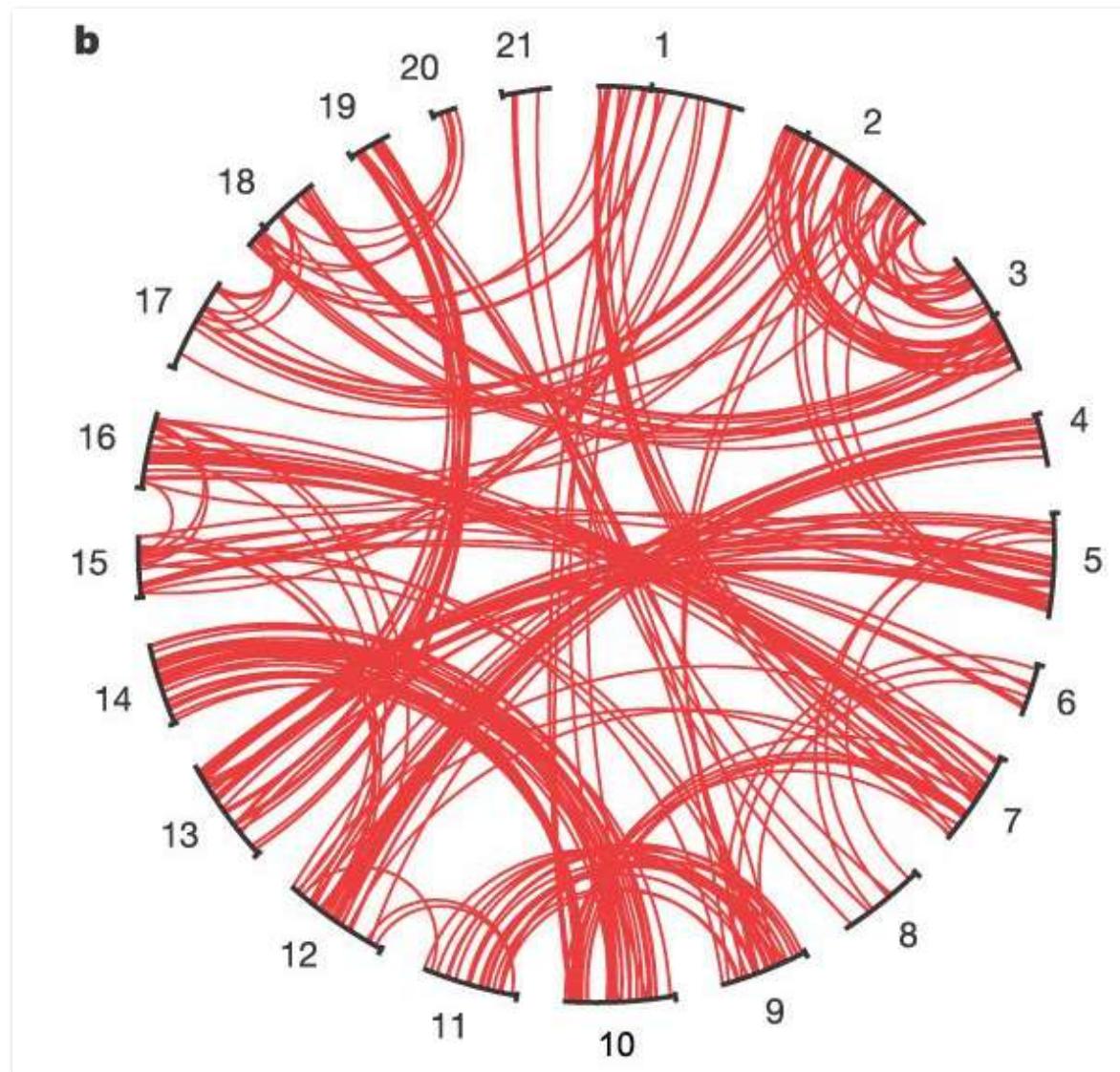
Global distribution of ancient duplicated genes (Ks > 0.35) in the Tetraodon genome

# Whole genome duplication in yeast followed by gene loss
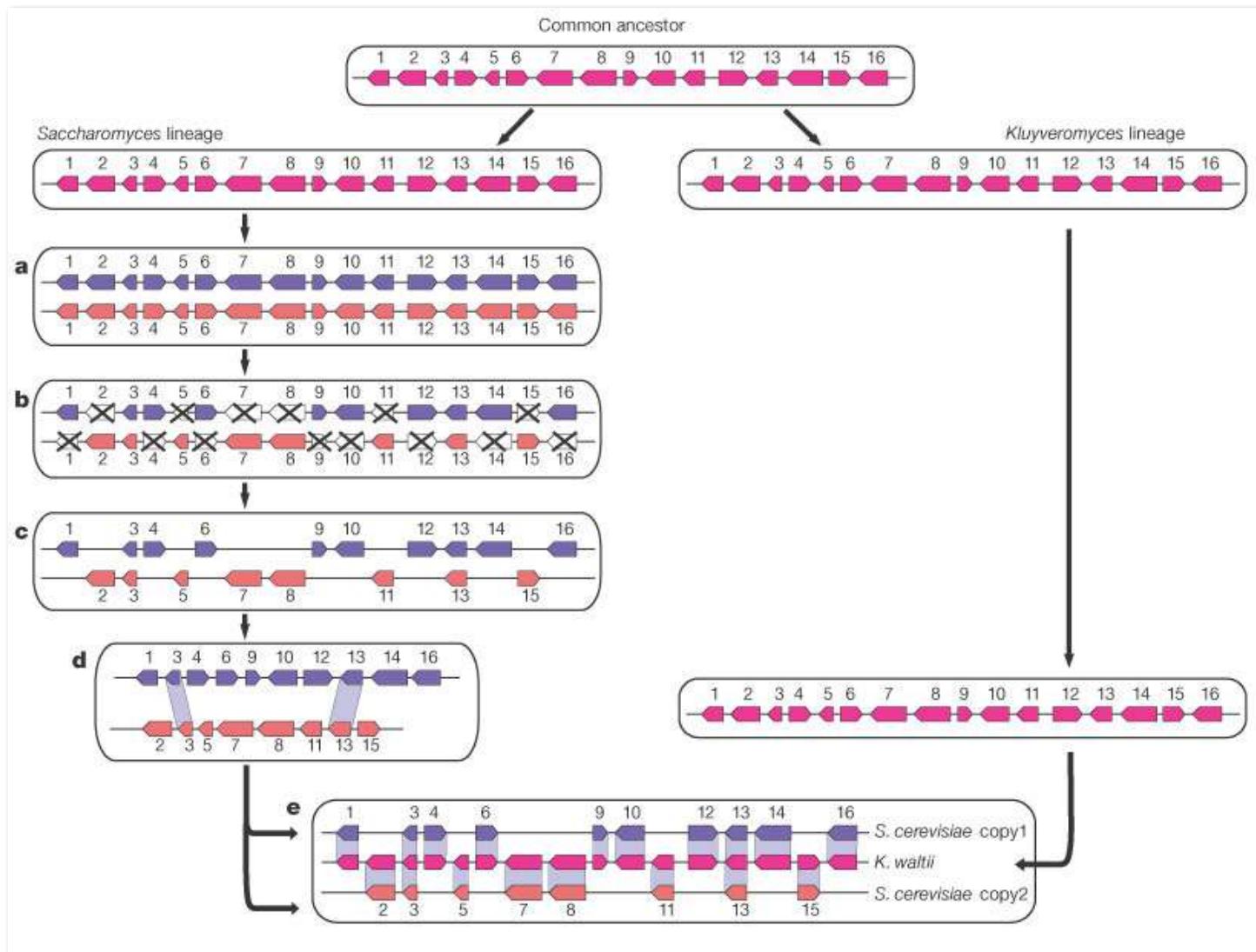
Genome duplications across eukaryotes

Van de Peer et al. https://doi.org/10.1038/nrg2600
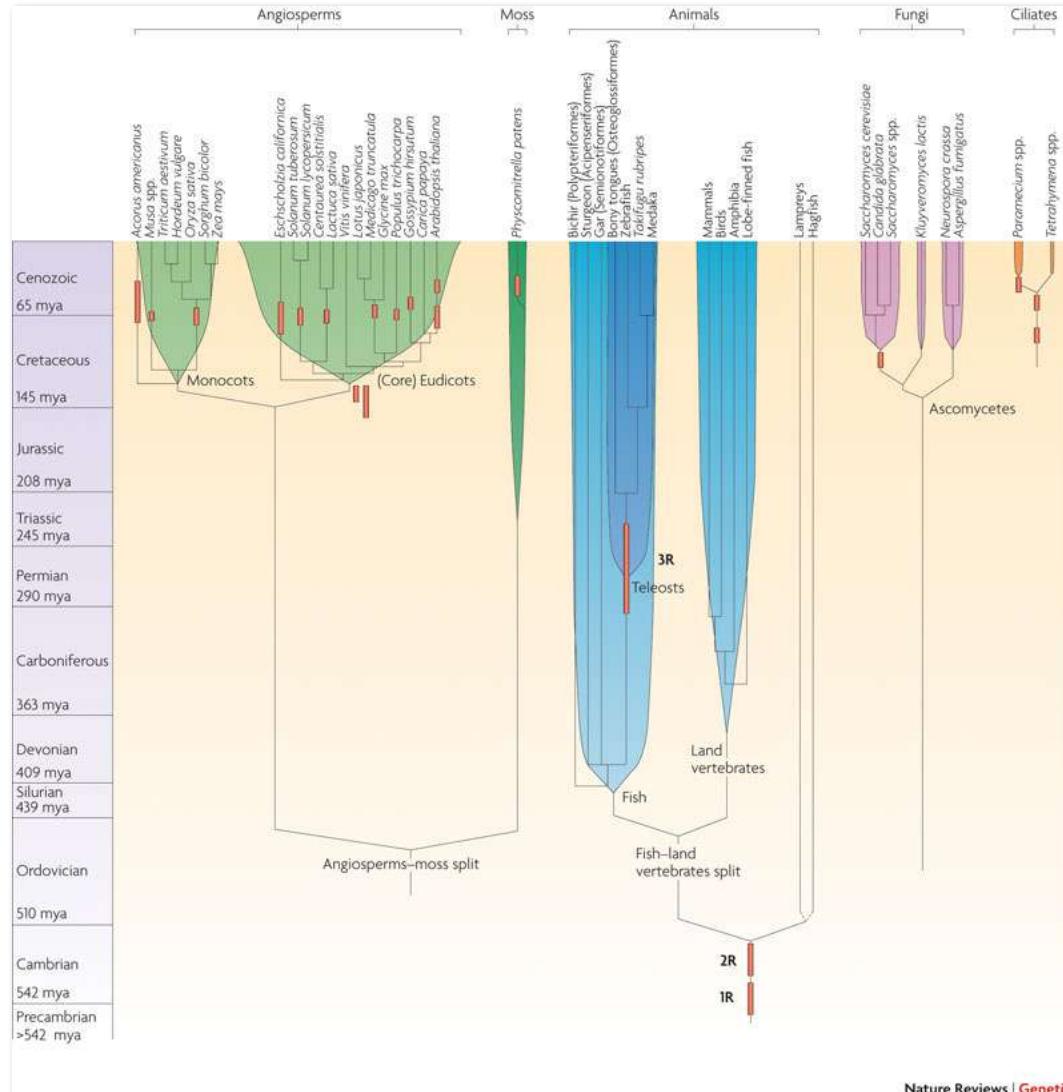
# Genome duplications in plants
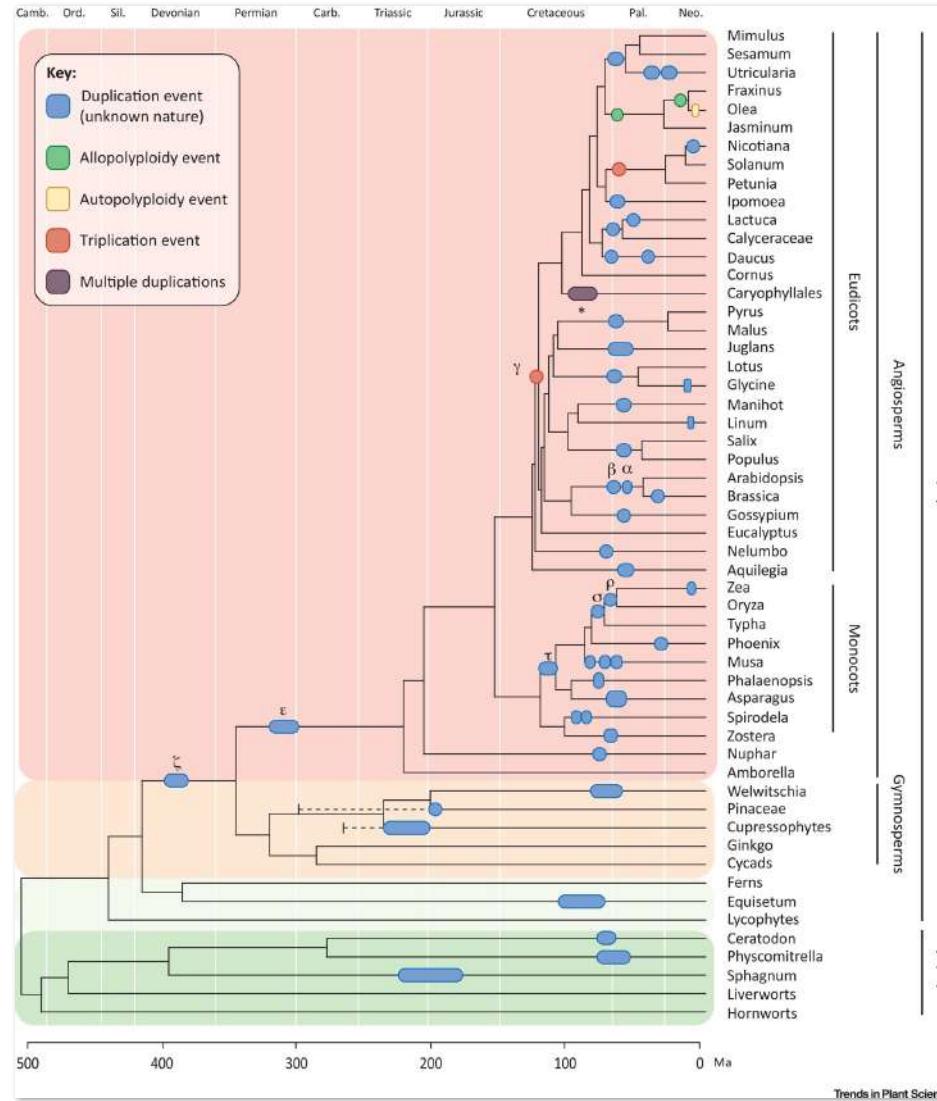
# Genome duplications in vertebrates
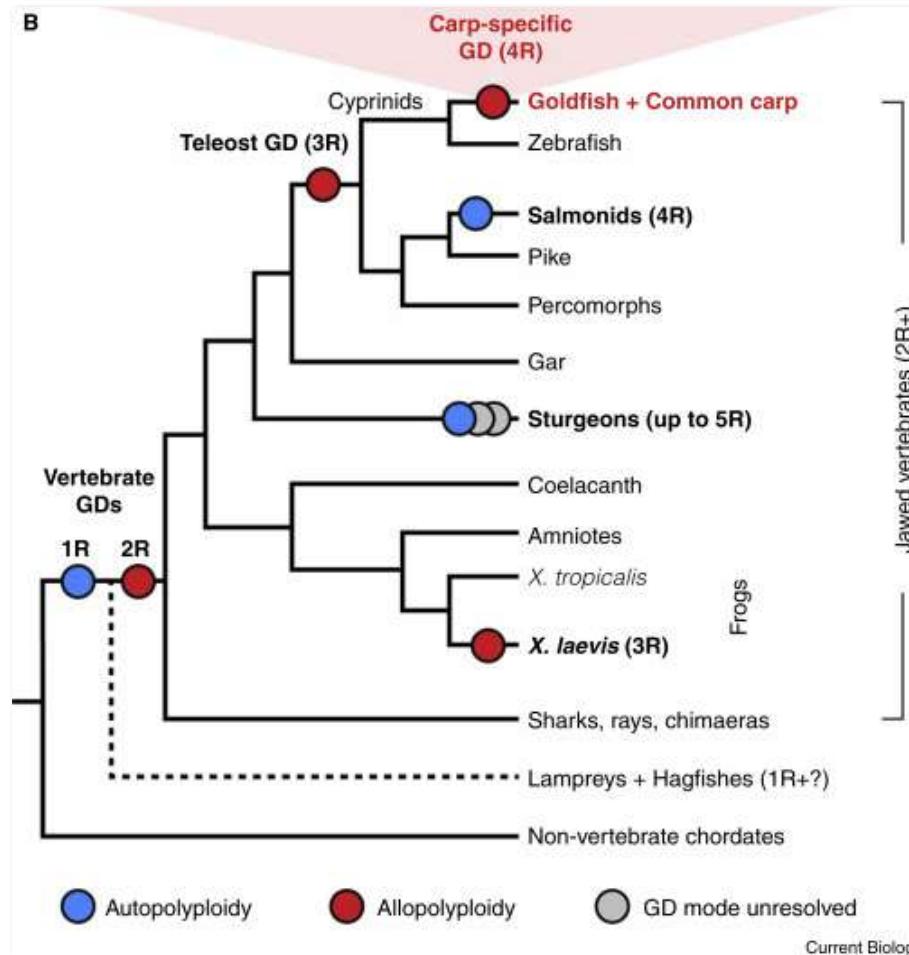
Evolution after duplication

Biased retention of duplicate genes

duplicate 1   duplicate 2
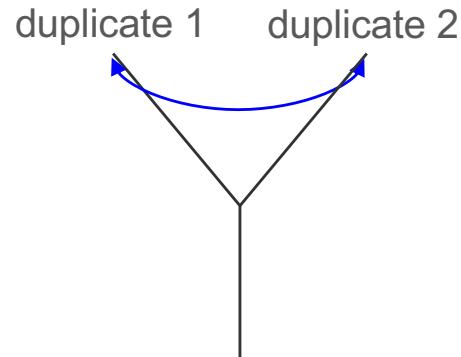
dN = original rate of gene
+ possible change due to duplication

Study Species - where duplicate and singleton genes are identified. (C. elegans, S cerevisiae)

D. melanogaster   *A. gambiae*

Outgroup species in which representative pairs of orthologs are identified.
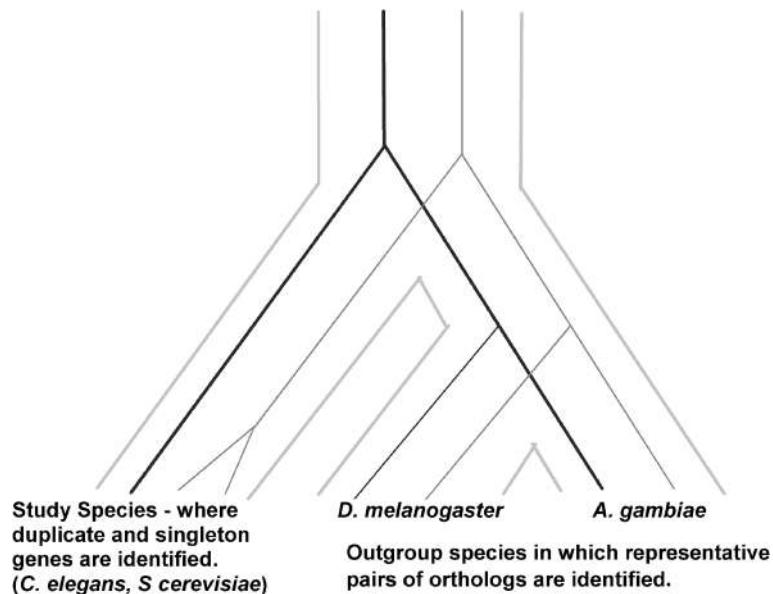
dN Drosophila / Anopheles = independent of
changes after duplication in yeast or nematode

A

dN (Representative Pairs)

0.3
0.2
0.1
0

Singletons   Duplicates

S. cerevisiae

B

dN (Representative Pairs)

0.3
0.2
0.1
0

Singletons   Duplicates

C. elegans
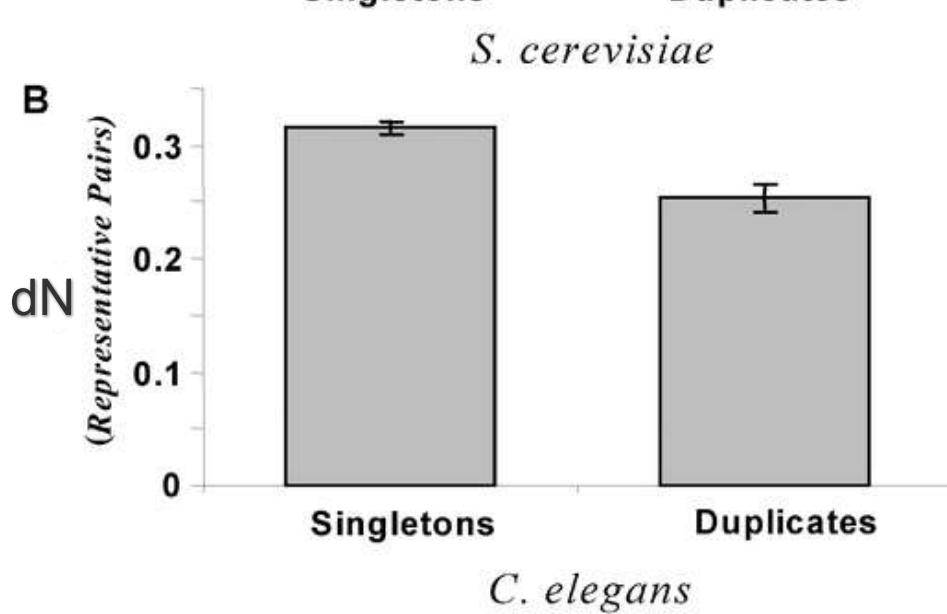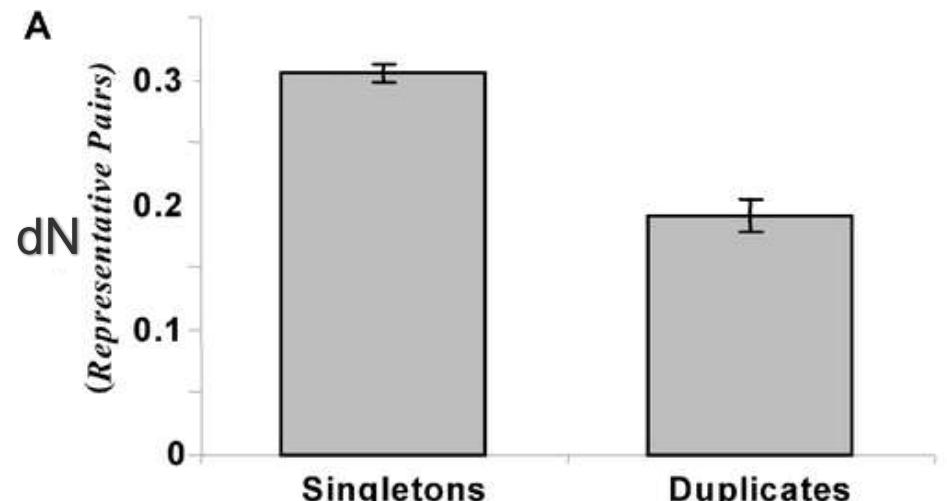
Impact of small- and large-scale duplications
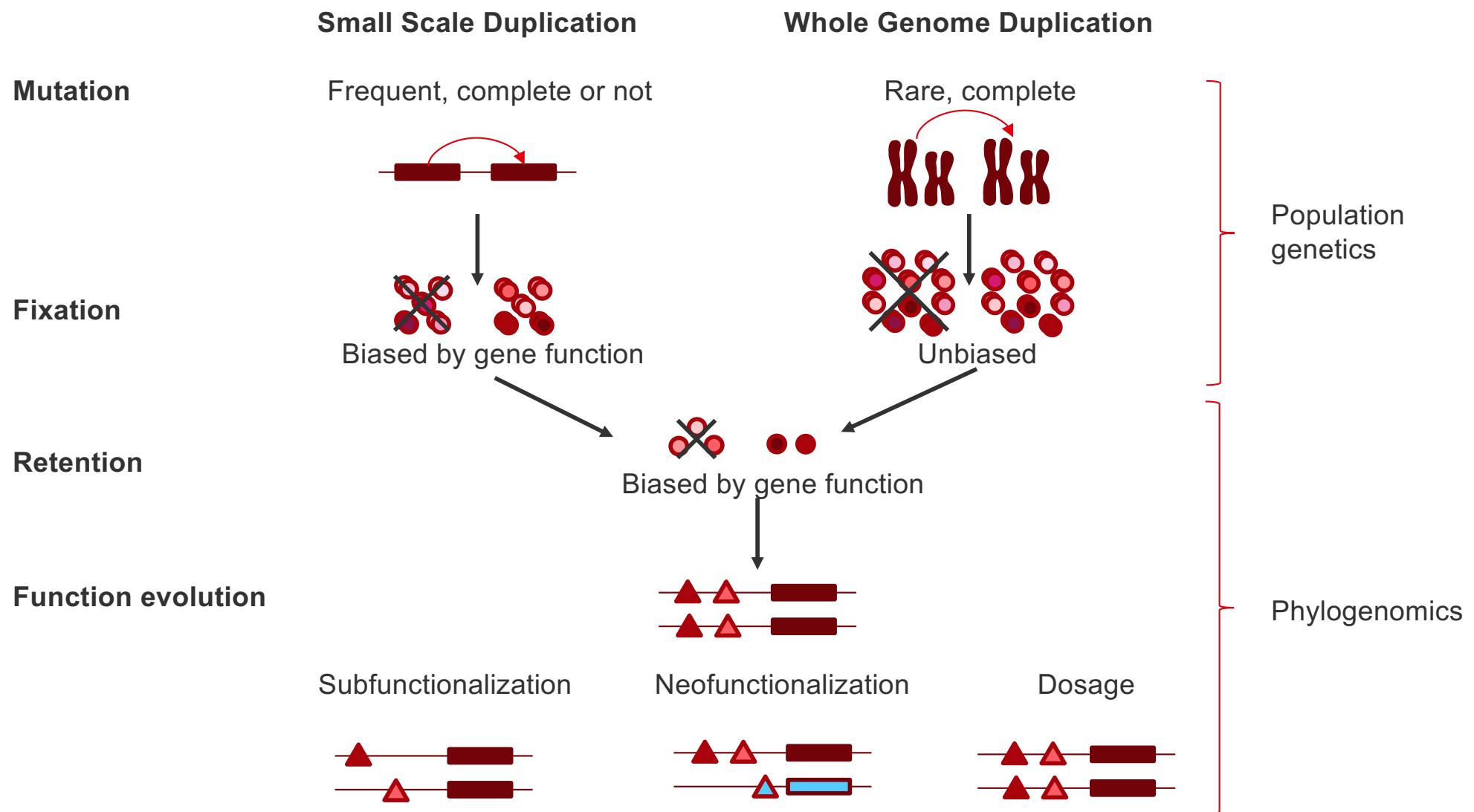
Selection and duplication
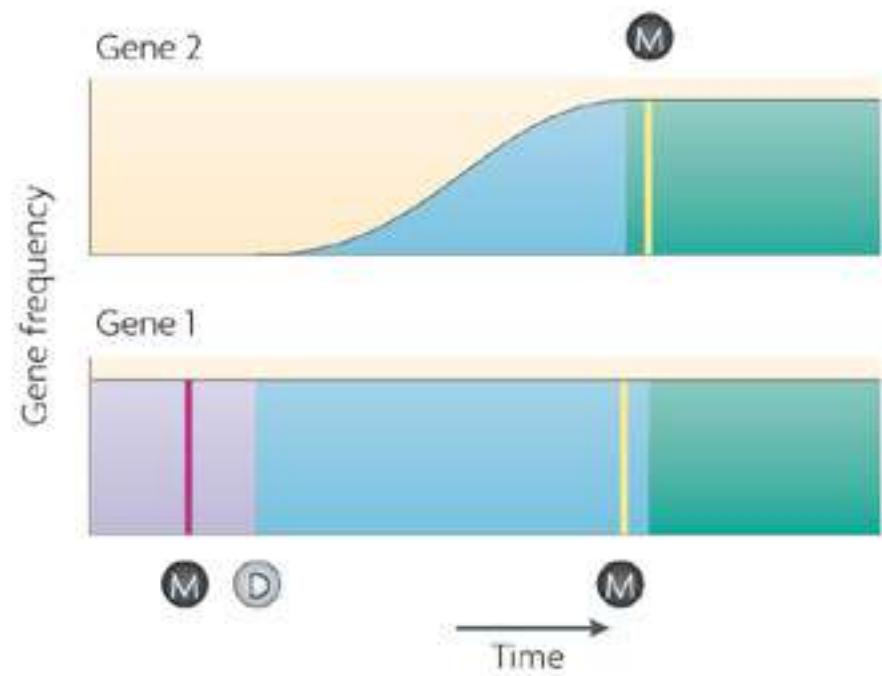
Conant & Wolfe https://doi.org/10.1038/nrg2482

Evolution of expression after fish WGD

genome duplications:
Teleost
Salmonid

Ts3R

Ss4R

Salmonids

Spotted gar (*Lepisosteus oculatus*)

Zebrafish (*Danio rerio*)

Stickleback (*Gasterosteus aculeatus*)

Pike (*Esox lucius*)

Coregonus + Prosopium + Stenodus

Thymallus

Brachymystax + Hucho

Oncorhynchus + Salvelinus + Salvethymus

Atlantic salmon (*Salmo salar*)

400   300   200   100   0
Million years ago
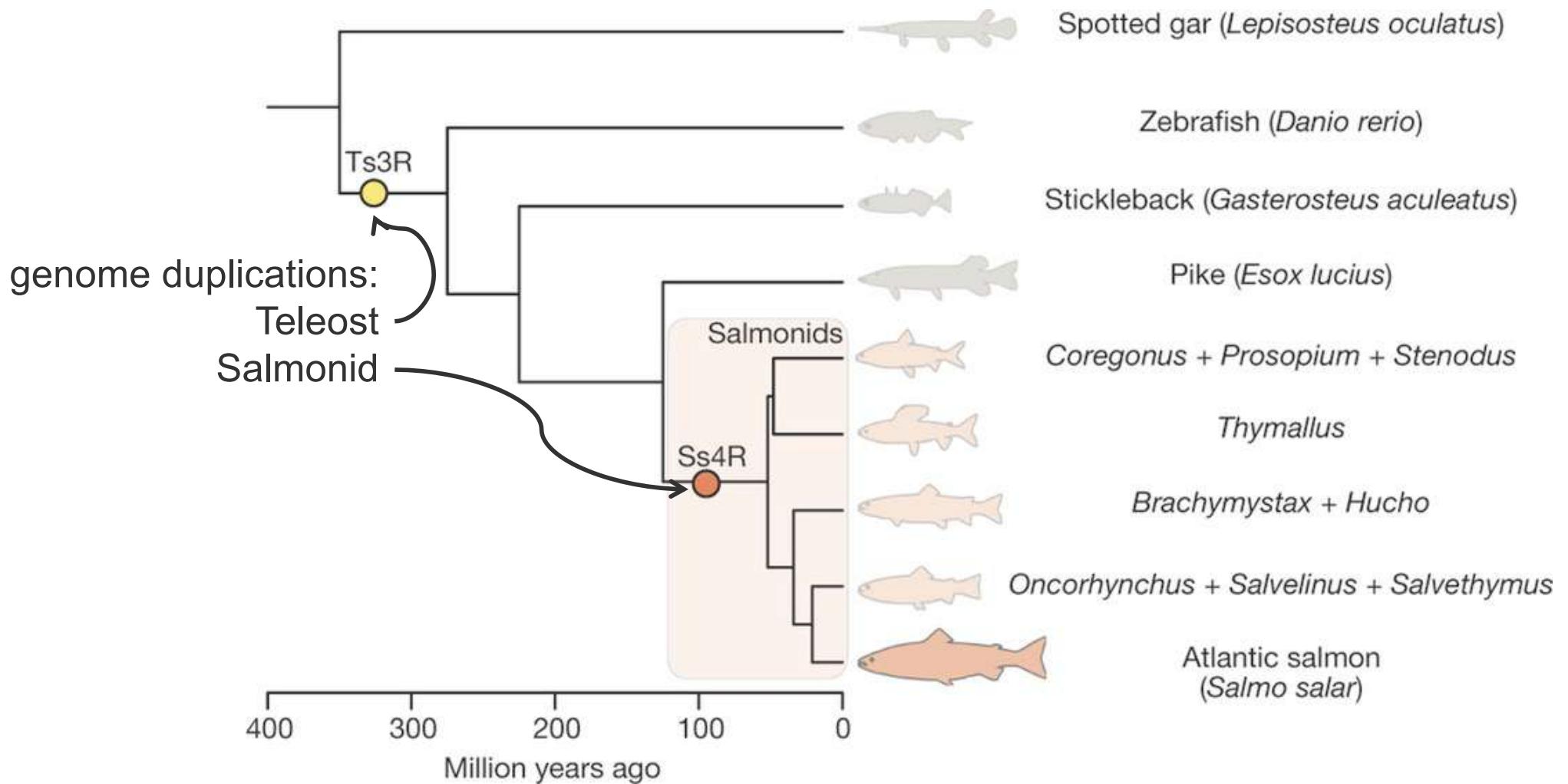
Support for sub-functionalization from teleosts
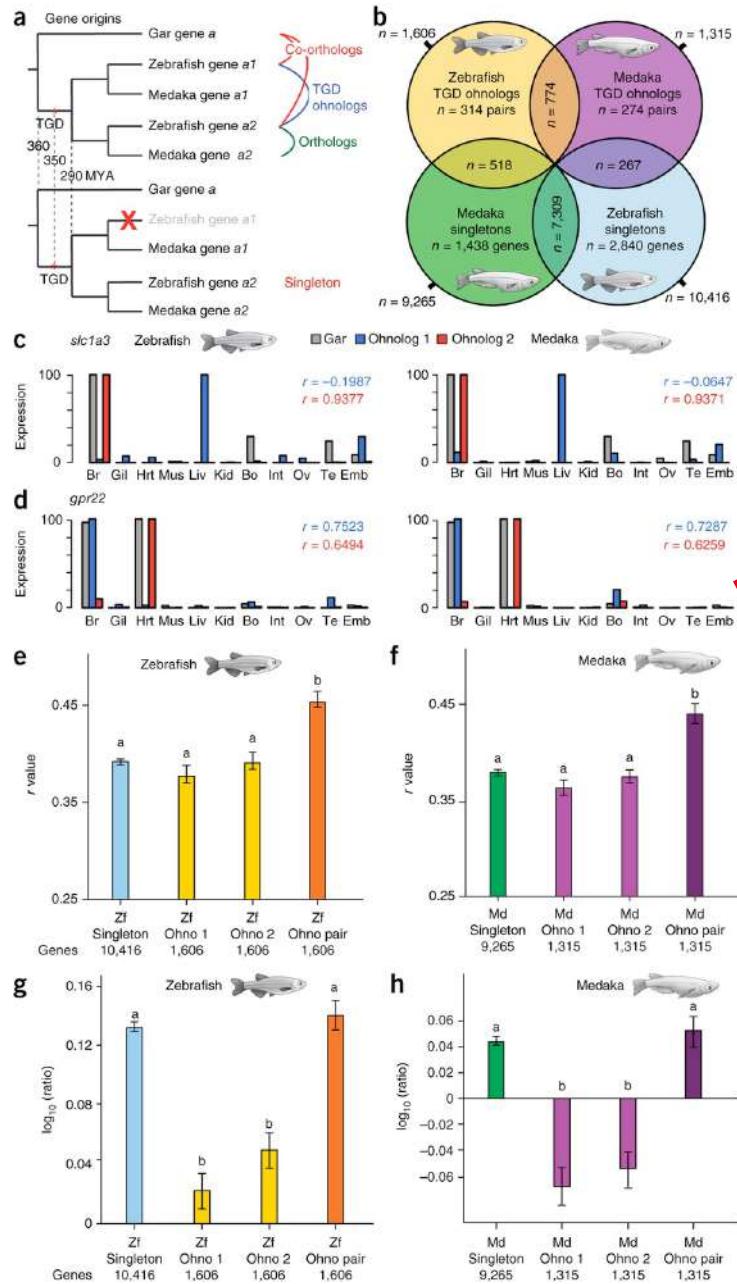
Comparison of expression:
RNA-seq in 10 organs or embryos of gar, medaka and zebrafish

Braasch et al. https://doi.org/10.1038/ng.3526
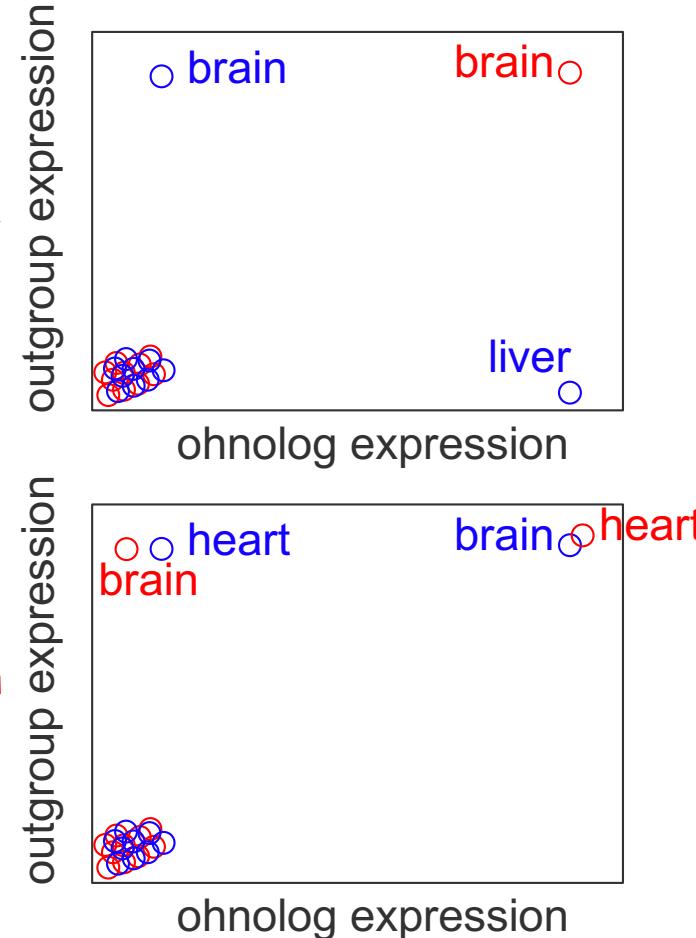
Support for neo-functionalization from salmonids
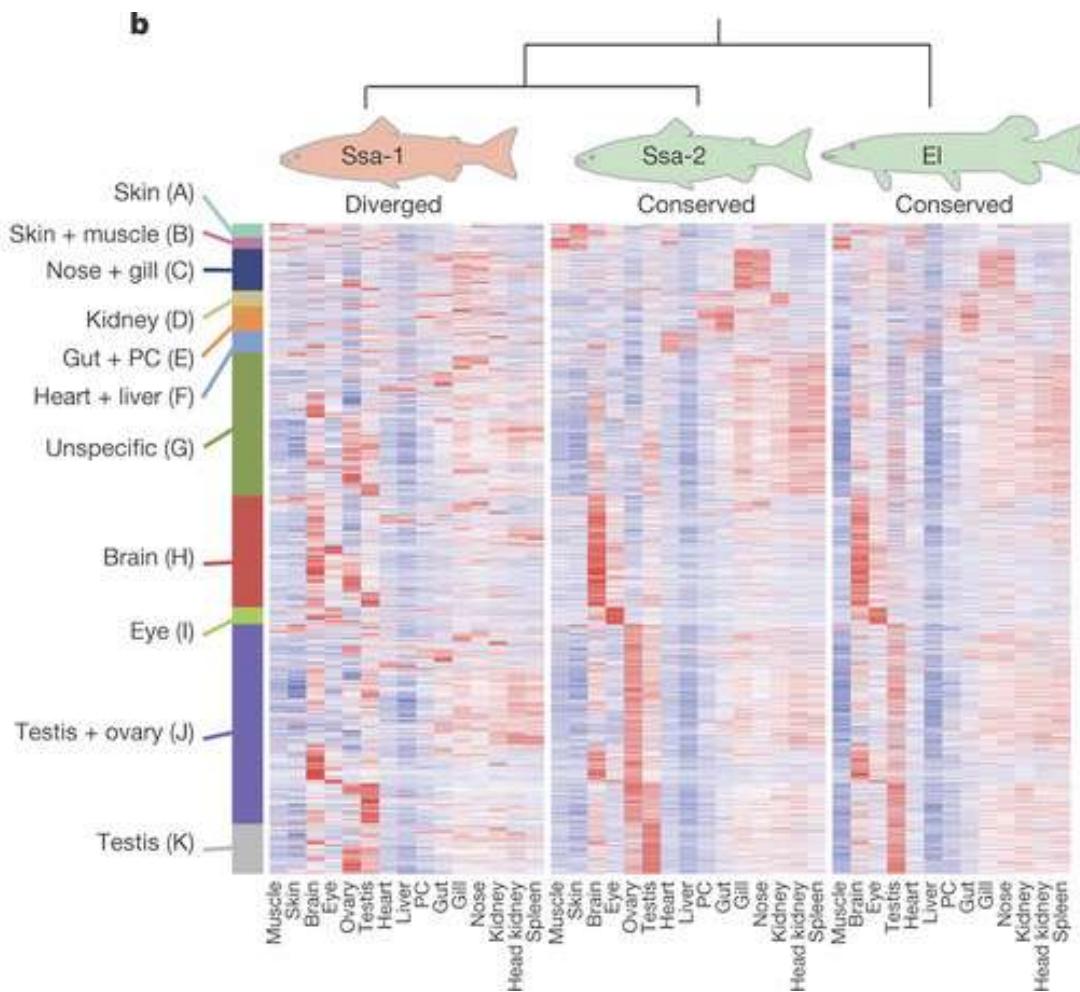
Comparison of expression: RNA-seq in 13 organs or embryos of salmon (Ssa = *Salmo salar*) and pike (El = *Esox lucius*)

Evolution after genome duplication

- Source of information: expression

- Both sub- and neo-functionalization

- Many genes do not diverge significantly

Take home

Discussion

What do you take home from today?

Diversity of genomes makes sense in the light of gene function