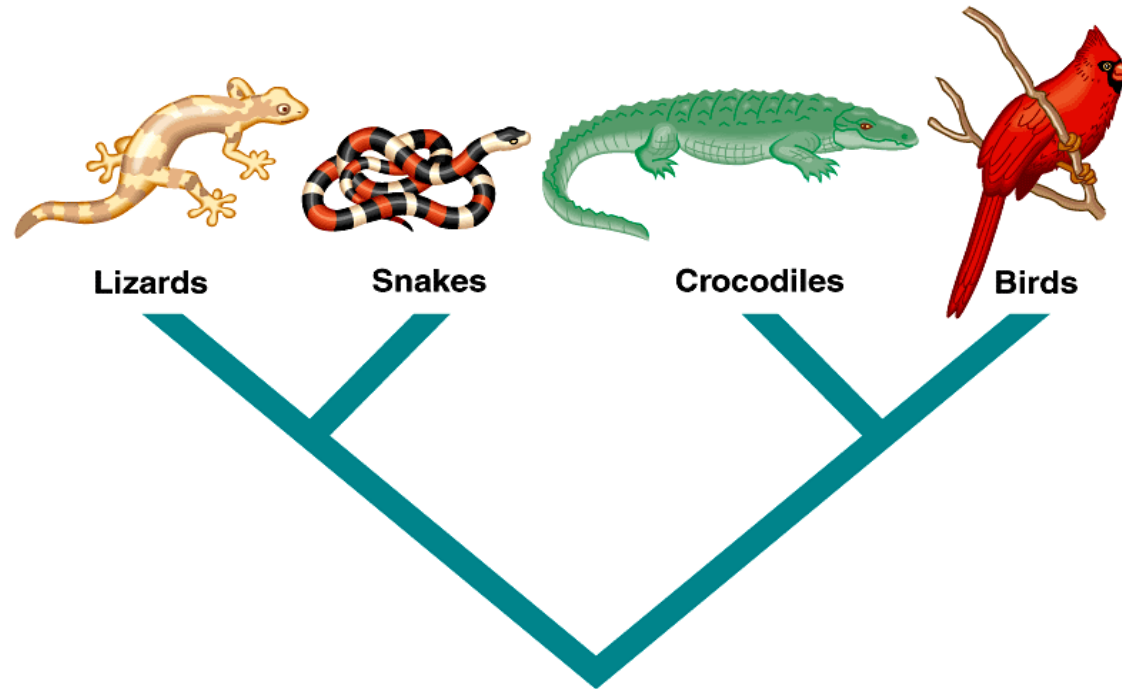# Today

- This morning
  - What is a phylogenetic tree?
  - How to "read" simple phylogenetic trees
  - How can you make a phylogeny?
  - How can you root a phylogeny?
- Gene Duplications & orthology

So far: one gene, one individual, one species
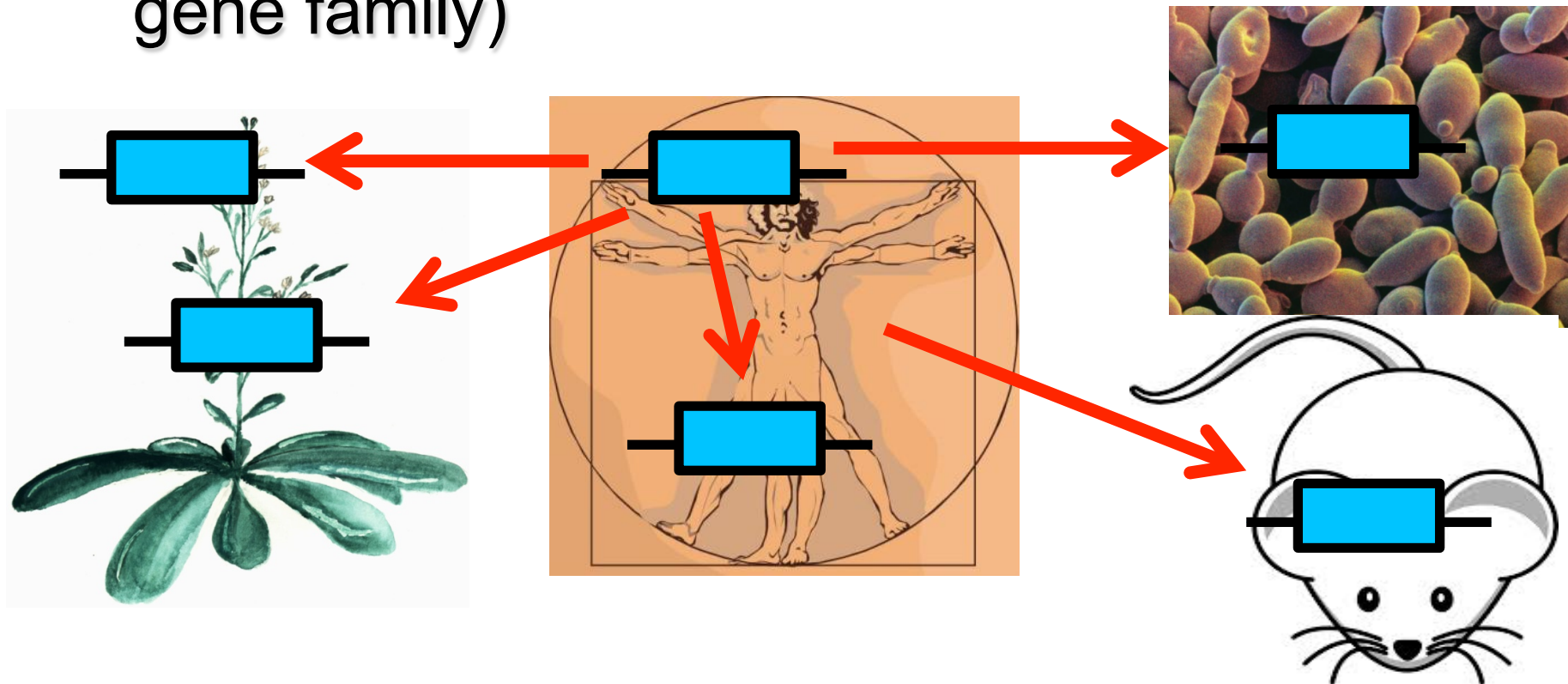

Lizards   Snakes   Crocodiles   Birds

Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.

- Biology is more complex
- Not all internal nodes are divergences between species

General idea of today

- Having all the genomes …

- What is the **relation** of my gene to *homologous* genes in the same and other organisms (what are the relations within the gene family)
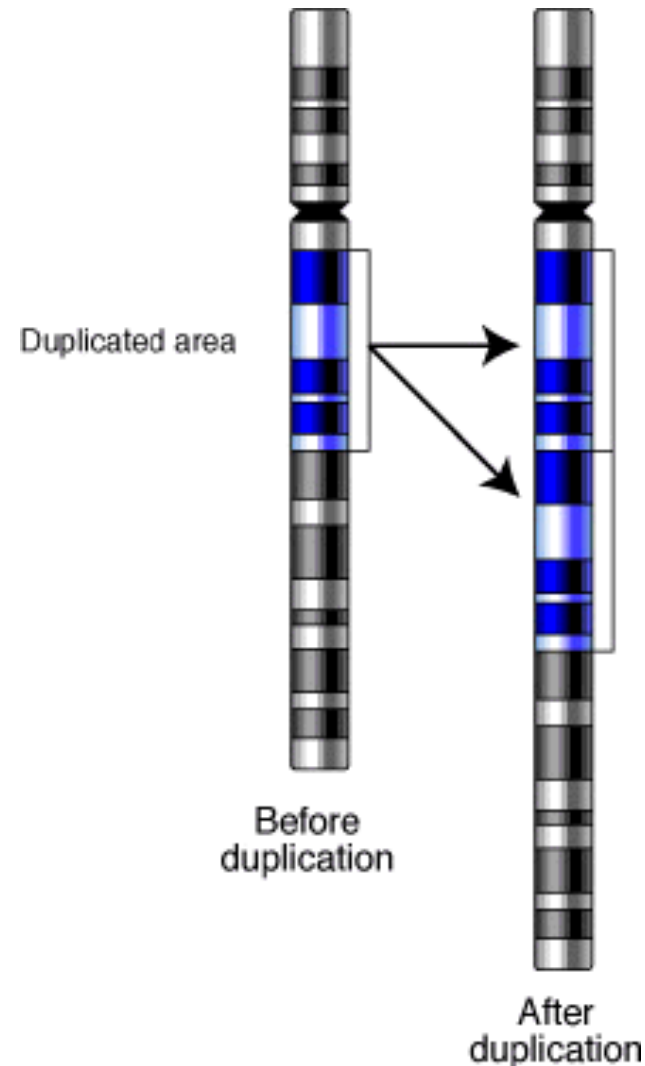
Not just divergence between species; what is up here?
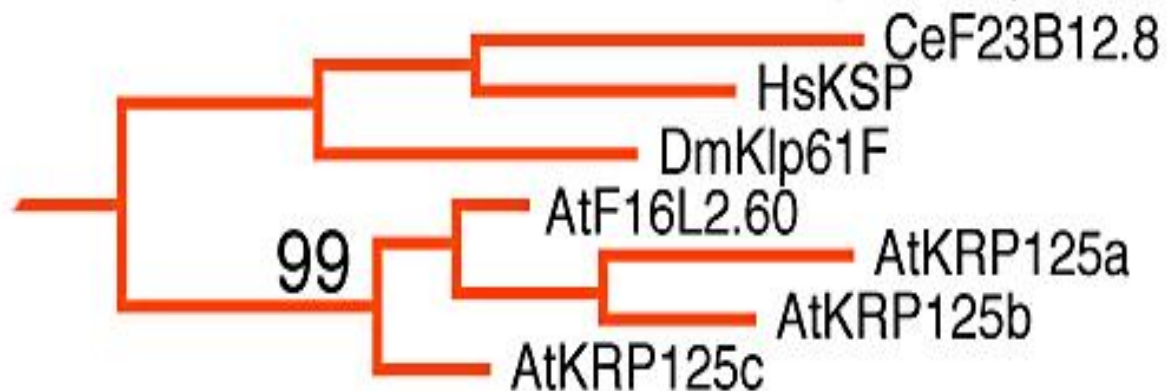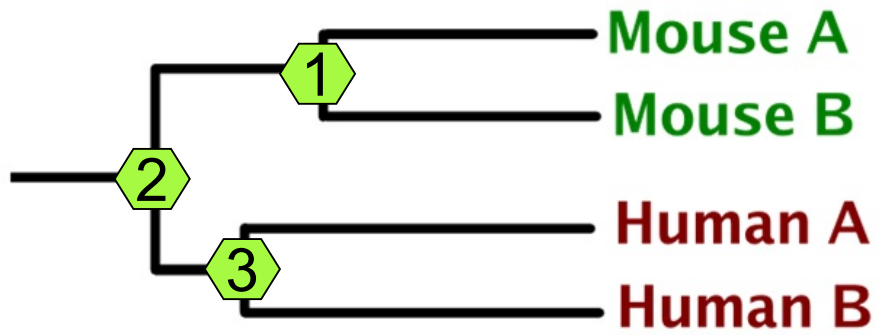

Kinase-5: essential for centrosome separation in mitosis

?b

Gene duplication

is the duplication of a region of DNA that contains a gene (important force in evolution of gene families / genome evolution)

Duplicated area

Before duplication

After duplication

Two genes per species: how to differentiate between one ancient or two recent duplications?

- Two genes in Human chromosomes (Human A & Human B) & two genes in mouse chromosomes (Mouse A & Mouse B)
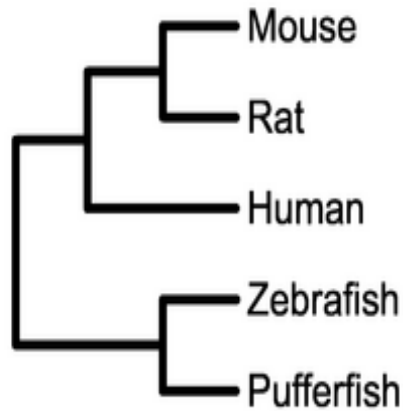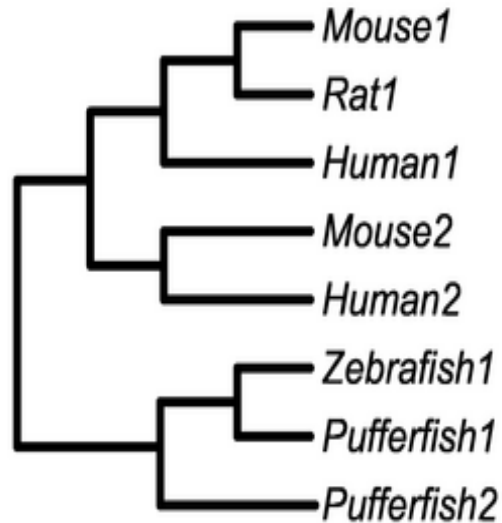
Duplications, Speciations

Speciation (red circle) **Gene Duplication** (blue square)

?

One more genome evolution process / gene family evolution process: what is up here? (besides duplication & speciation)
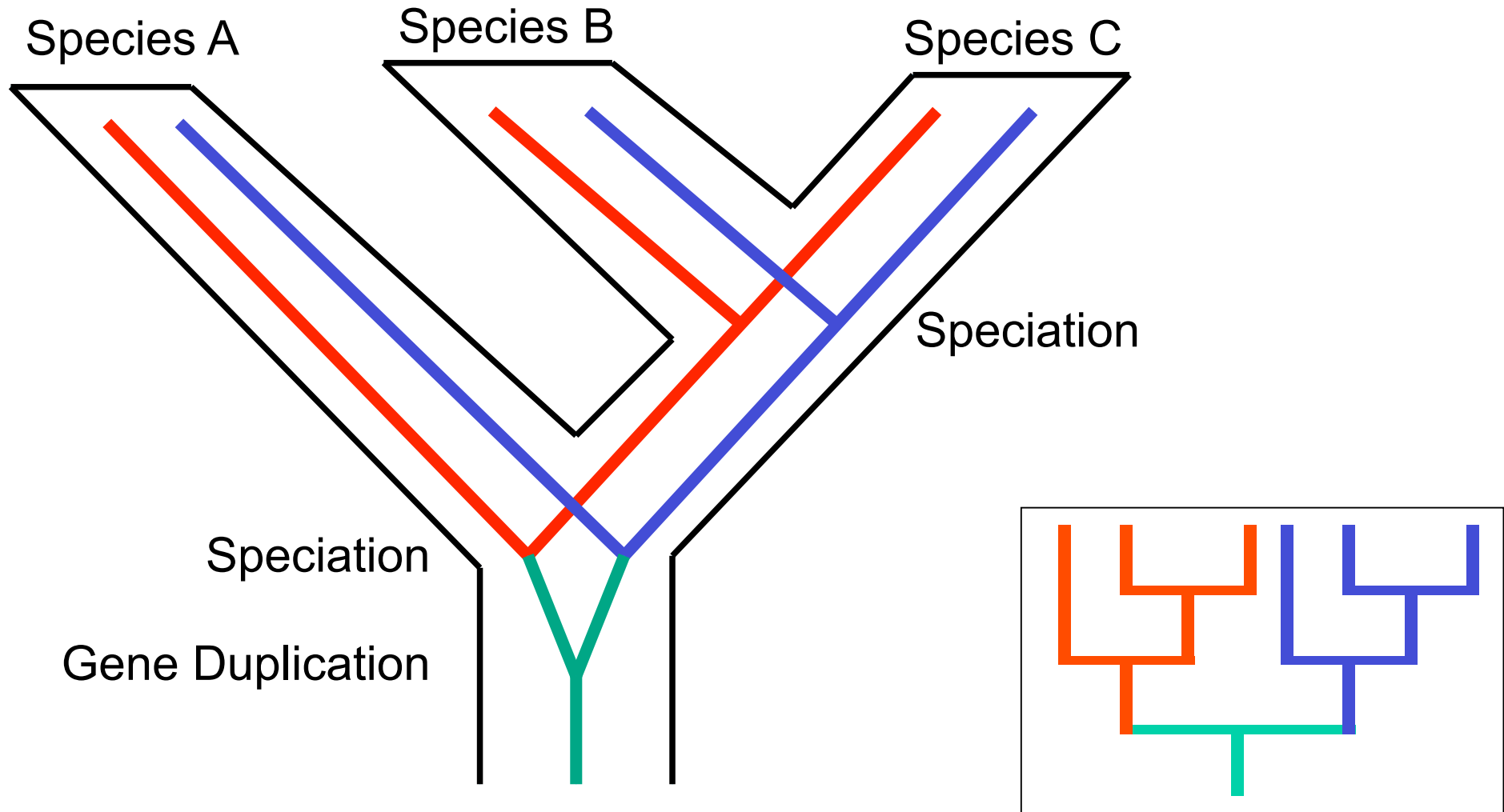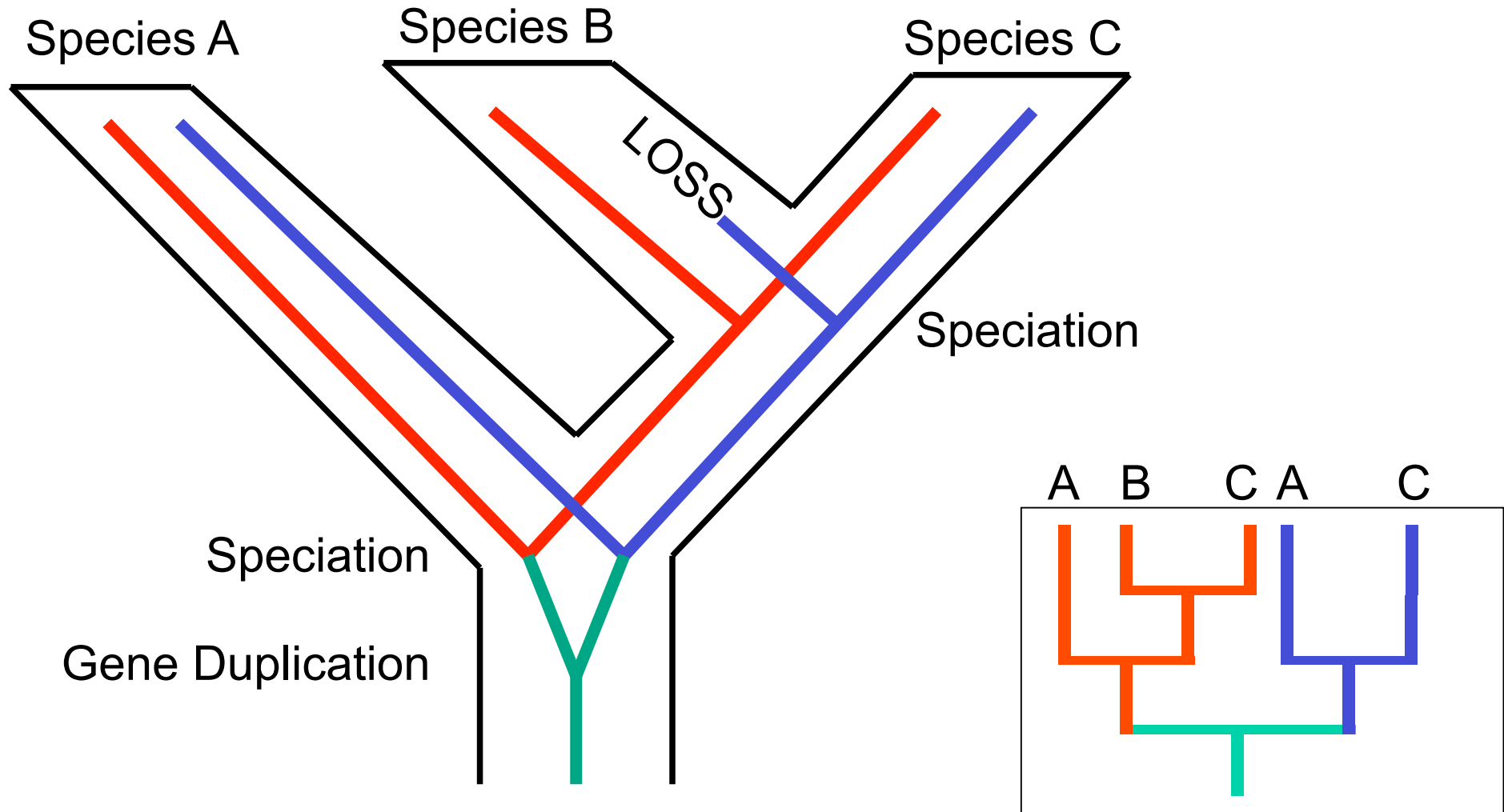

(a) Species Tree

(b) Gene Tree

Gene loss

b

Interpreting the tree: duplications vs speciations, gene tree inside species tree

Interpreting the tree: duplications vs speciations, gene tree inside species tree loss
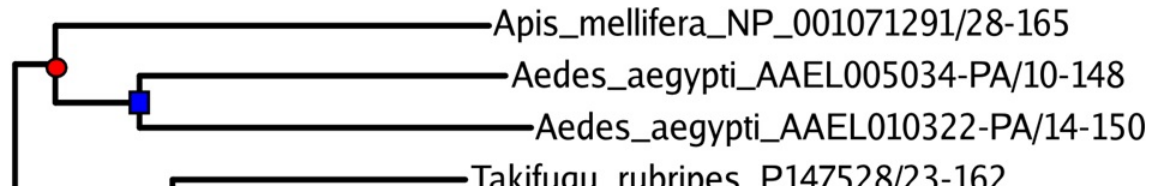
Species A   Species B   Species C

LOSS

Speciation

Speciation

Gene Duplication

A   B   C   A   C

Blast with a newly sequenced globin from frog

What kind of globin is it?

Globins

Apis_mellifera_NP_001071291/28-165
Aedes_aegypti_AAEL005034-PA/10-148
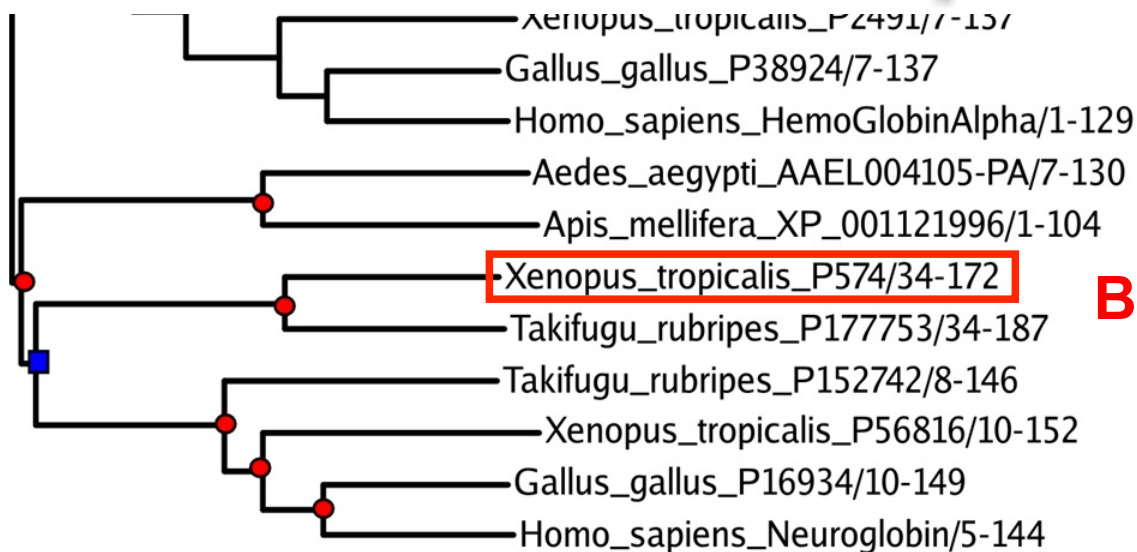Aedes_aegypti_AAEL010322-PA/14-150
Takifugu_rubripes_P147528/23-162

A Globin Gene of Ancient Evolutionary Origin in Lower Vertebrates: Evidence for Two Distinct Globin Families in Animals

Anja Roesner,* Christine Fuchs,† Thomas Hankeln,† and Thorsten Burmester*

*Institute of Zoology, Johannes Gutenberg University, Mainz, Germany; and †Institute of Molecular Genetics, Johannes-Gutenberg-University, Mainz, Germany

Hemoglobin, myoglobin, neuroglobin, and cytoglobin are four types of vertebrate globins with distinct tissue distributions and functions. Here, we report the identification of a fifth and novel globin gene from fish and amphibians, which has apparently been lost in the evolution of higher vertebrates (Amniota). Because its function is presently unknown, we tentatively call it globin X (GbX). Globin X sequences were obtained from three fish species, the zebrafish

unknown, we tentatively call it globin X (GbX).

Xenopus_tropicalis_P2491/7-137
Gallus_gallus_P38924/7-137
Homo_sapiens_HemoGlobinAlpha/1-129
Aedes_aegypti_AAEL004105-PA/7-130
Apis_mellifera_XP_001121996/1-104
Xenopus_tropicalis_P574/34-172
Takifugu_rubripes_P177753/34-187
Takifugu_rubripes_P152742/8-146
Xenopus_tropicalis_P56816/10-152
Gallus_gallus_P16934/10-149
Homo_sapiens_Neuroglobin/5-144

α chain
Heme

(b) Hemoglobin

Blast query

NOT orthologous relative to
the human-mouse speciation
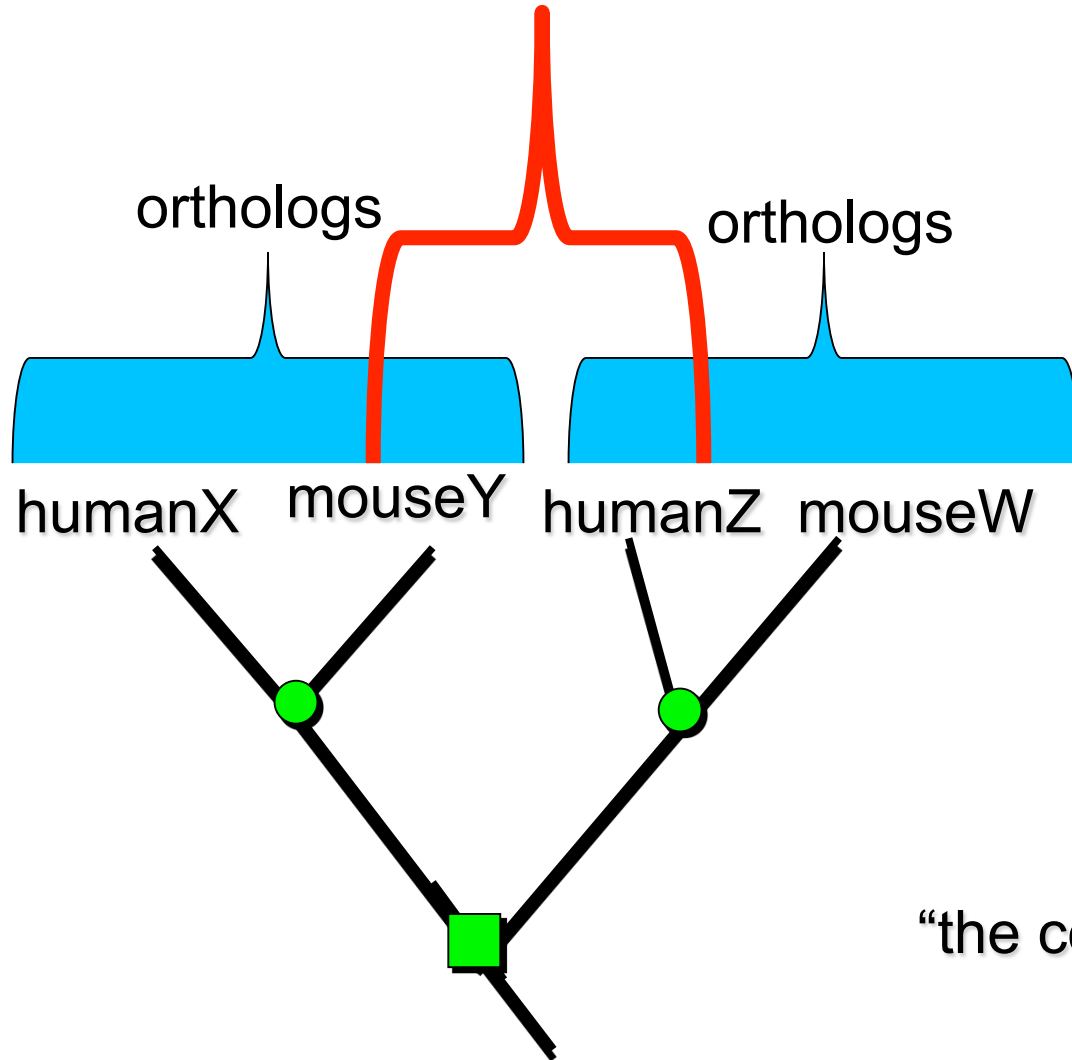
Fitch 1970
Two genes in two
species are
orthologous if they
derive from one gene
in their *last* common
ancestor; **This means
that the "node"' in
the tree where they
diverge is a
speciation node.**

orthologs   orthologs

humanX   mouseY   humanZ   mouseW

"the corresponding gene"

often have the same function

Genes can diverge by

● Speciation, or

■ Duplication

Retraction in MBE because of duplication / speciation, orthology / paralogy

- http://retractionwatch.wordpress.com

- Quite simply, our study targetted specifically a comparative genomics analysis of salmonid species. However, because the whole family went through a relatively recent whole genome duplication event, sorting out paralogs from true orthologs remains a serious challenge. We realised that some targets retained in our analyses were likely to be paralogs, which could cause some biases in our results. We now need to revisit that very carefully using alternative analytical approaches /

- They wanted to compare the sequences of genes across different species, and thought that's what they had done. However, in these species, apparently a lot of genes are duplicated, such that there are two relatively similar copies (call them gene 1a and gene 1b). If you're looking at a copy of a gene in a species, it's difficult to tell whether it's gene 1a or gene 1b. So presumably what happened is that they thought they were looking at gene 1a in both species, but realized they were looking at gene 1a in one species, but gene 1b in another species. This is an easy mistake to make, and could definitely lead to major problems. Though again, without the text of the original article, it's tough to be more precise than that.

InParalogs (/ Co-orthologs)
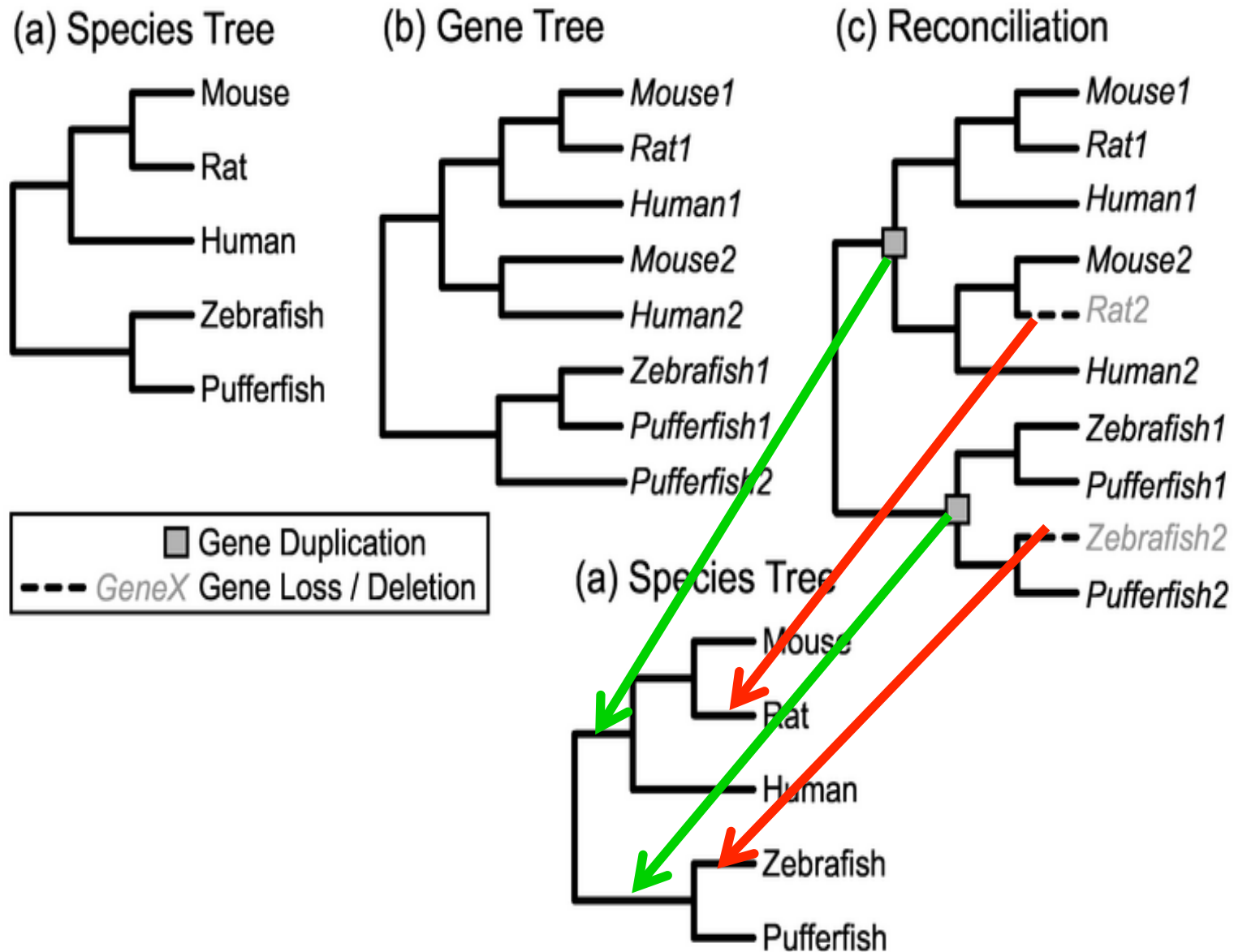## (Sonnhammer & Koonin TiG)
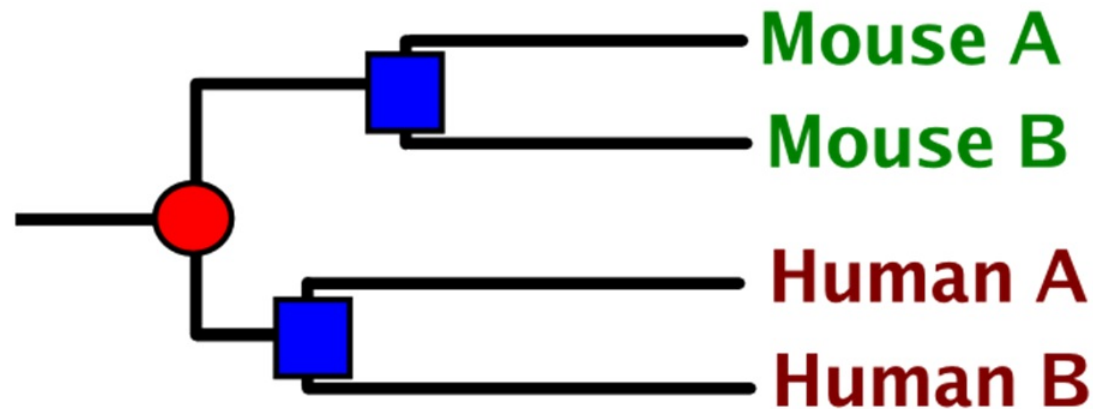
If the node where two genes diverged was a speciation than two genes are orthologs; if it was a duplication they are paralogs

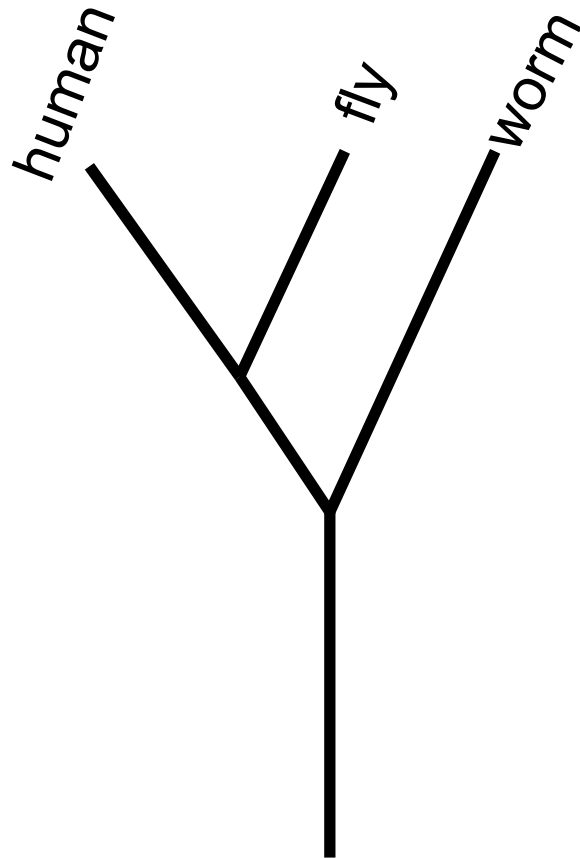2 genes in mouse are both orthologs to two genes in human!!
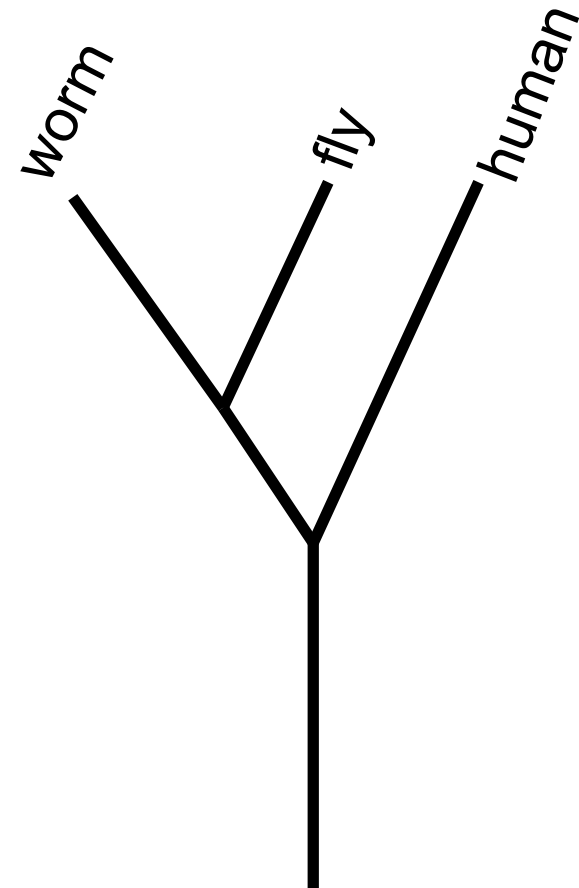
"Phylogenetic timing of duplicates"

How to time duplications and detect losses given that single gene trees are often based on too little data to reliably infer their evolutionary history?
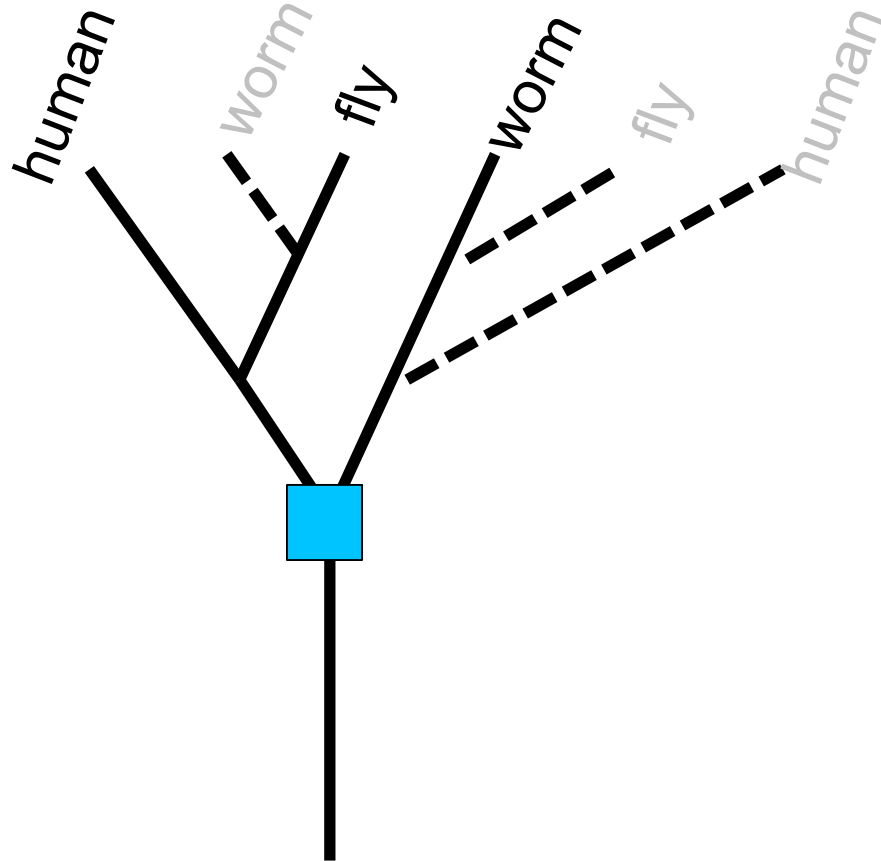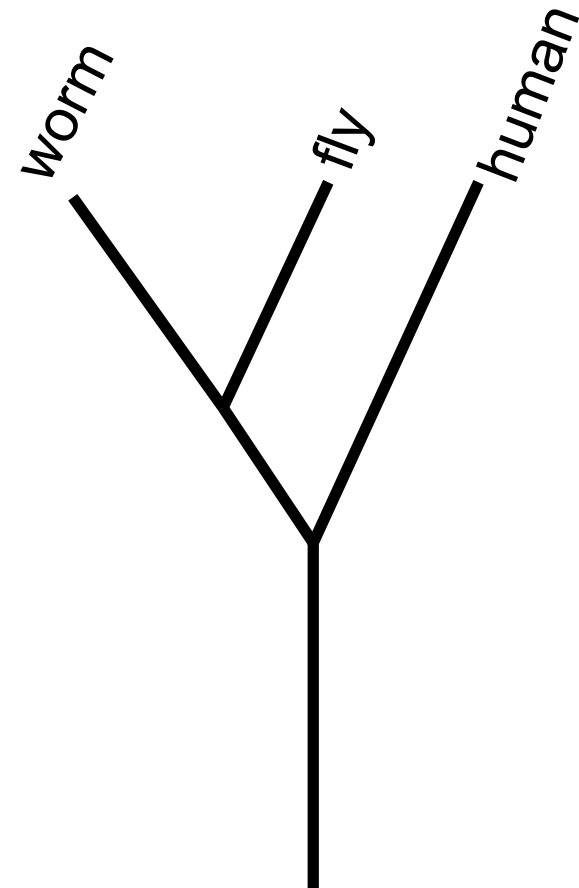
Gene tree

Species tree

human

fly

worm

worm

fly

human

How to time duplications and detect losses given that single gene trees are often based on too little data to reliably infer their evolutionary history?

Strict reconciliation?
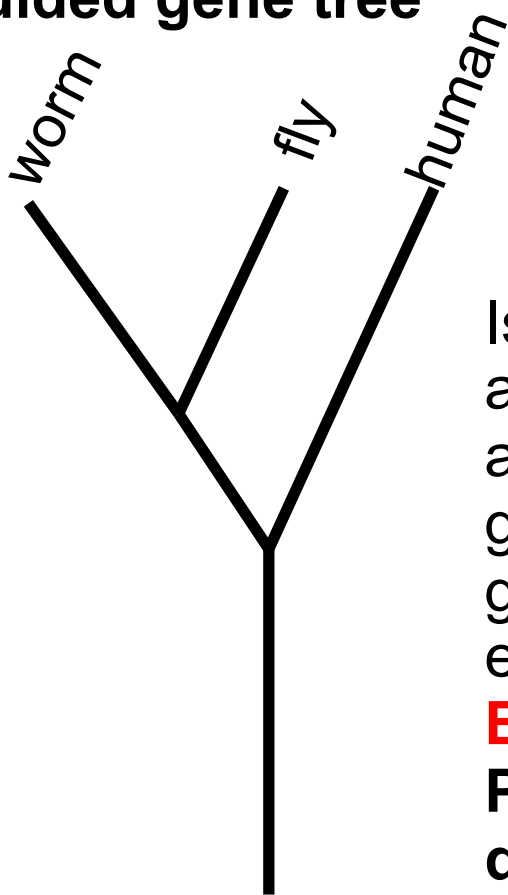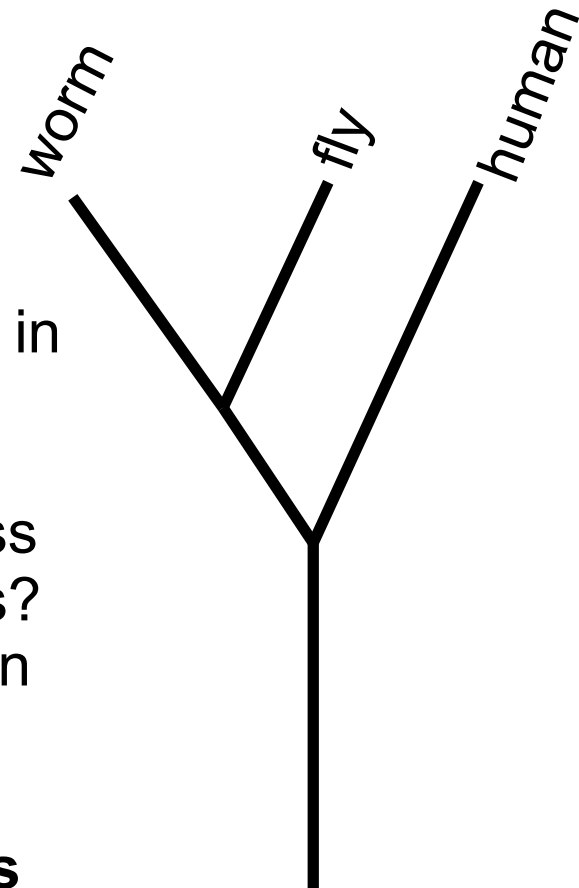

Reconciled Gene tree

Species tree

How to time duplications and detect losses given that single gene trees are often based on too little data to reliably infer their evolutionary history?

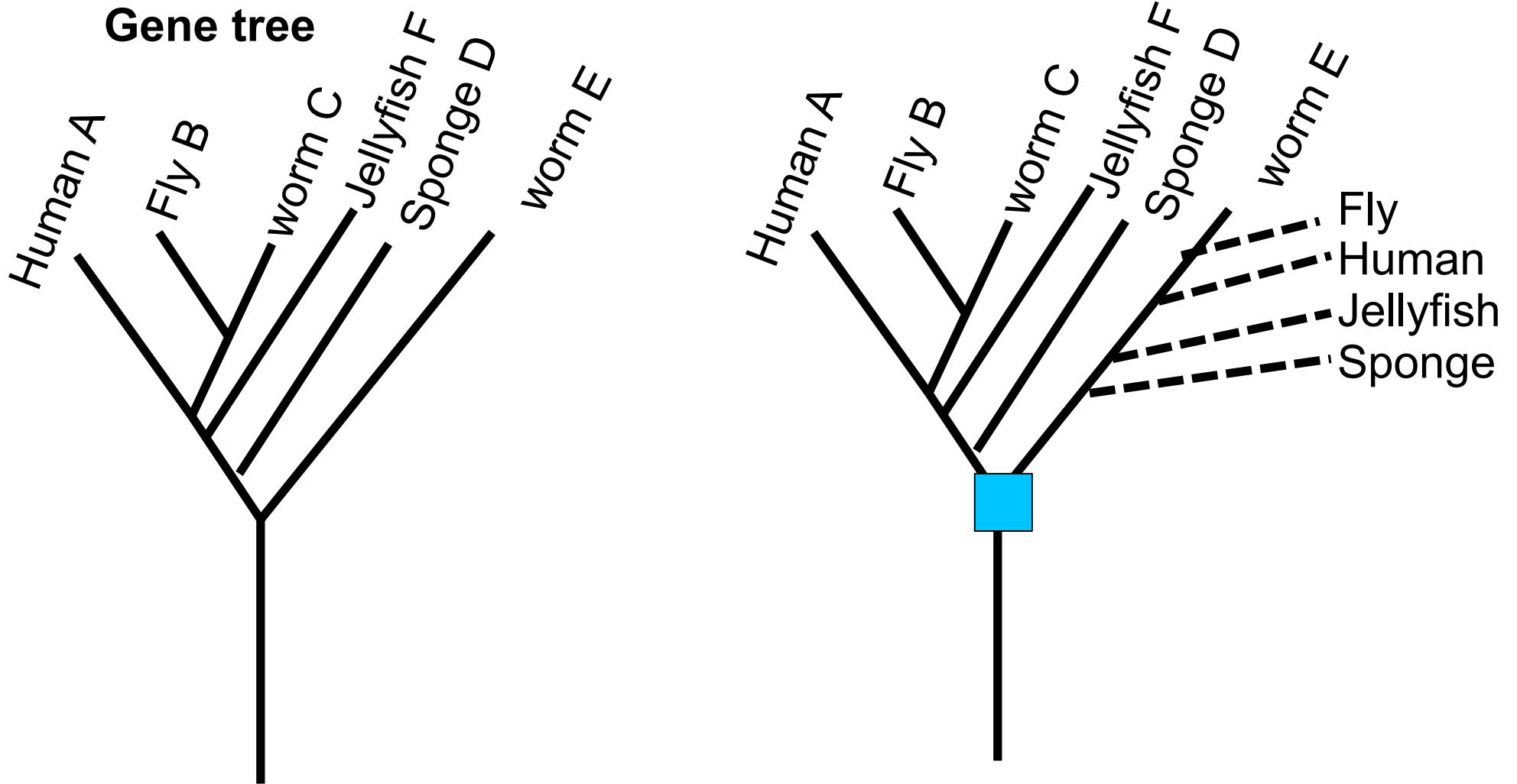## …. Species tree guided tree building

Species tree guided gene tree

worm

fly

human

Species tree

worm

fly

human

Is there significant signal in aspects of the data (the alignment) to support a gene tree that implies less genome evolution events? e.g. **TreeBEST** as used in **ENSEMBL COMPARA** Prevents spurious duplications and losses

Example of difference between species tree guided reconstruction vs strict tree reconciliation for duplication
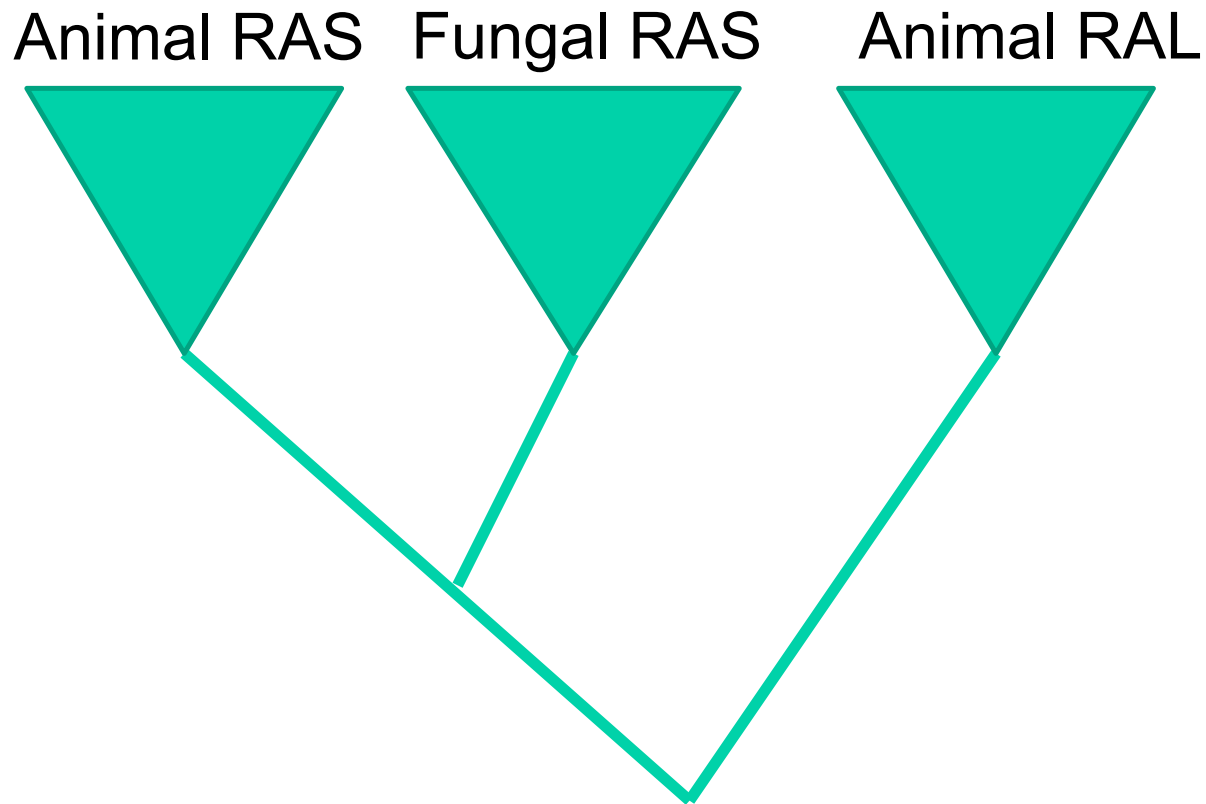

Gene tree

Human A, Fly B, worm C, Jellyfish F, Sponge D, worm E

Strict tree reconciliation

Human A, Fly B, worm C, Jellyfish F, Sponge D, worm E

Fly
Human
Jellyfish
Sponge

Example of difference between species tree guided reconstruction vs strict tree reconciliation for duplication
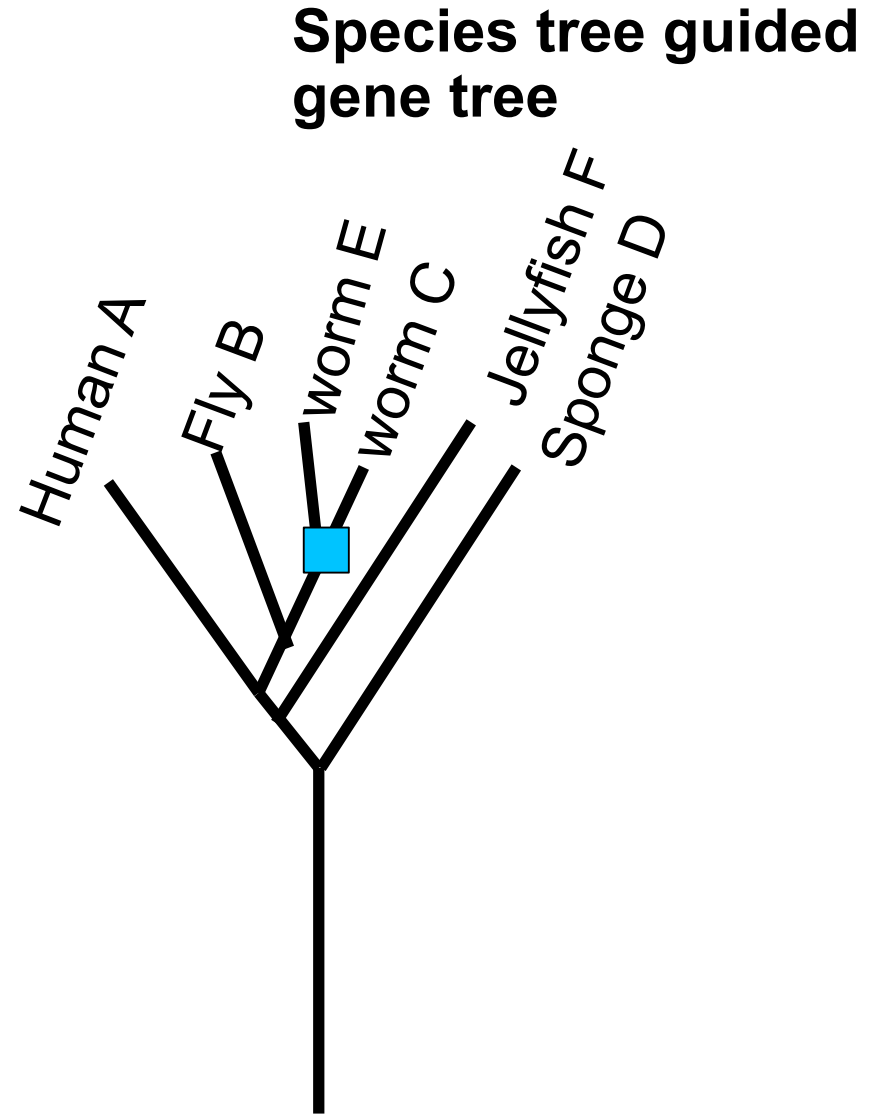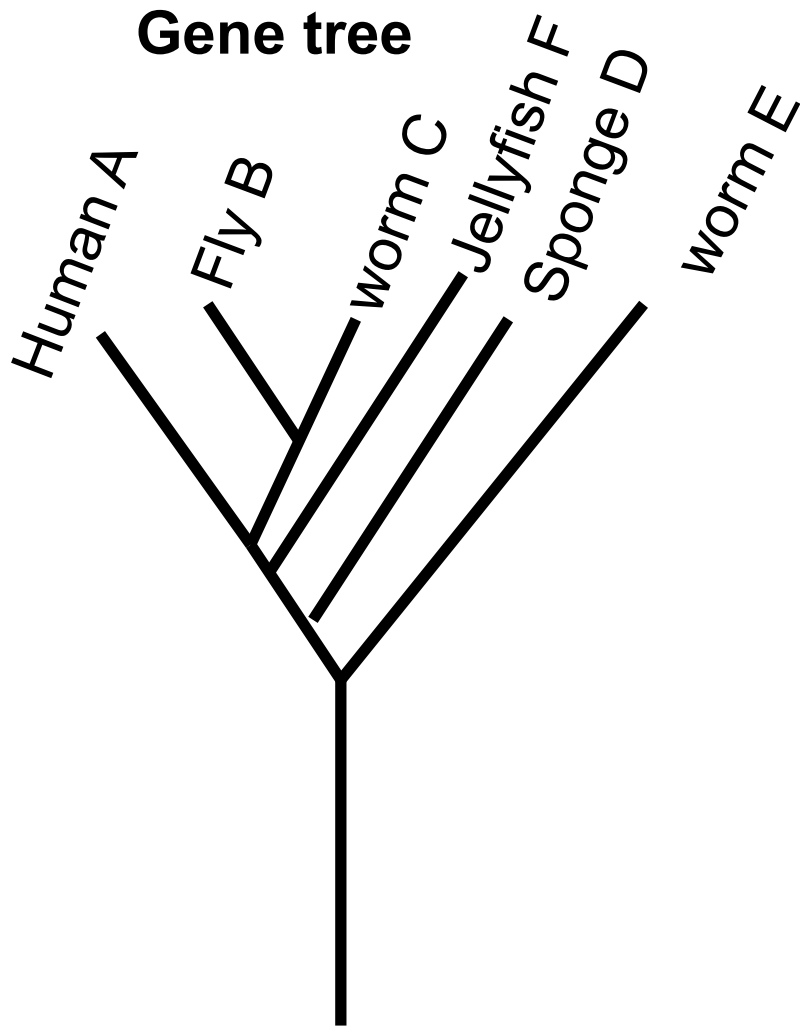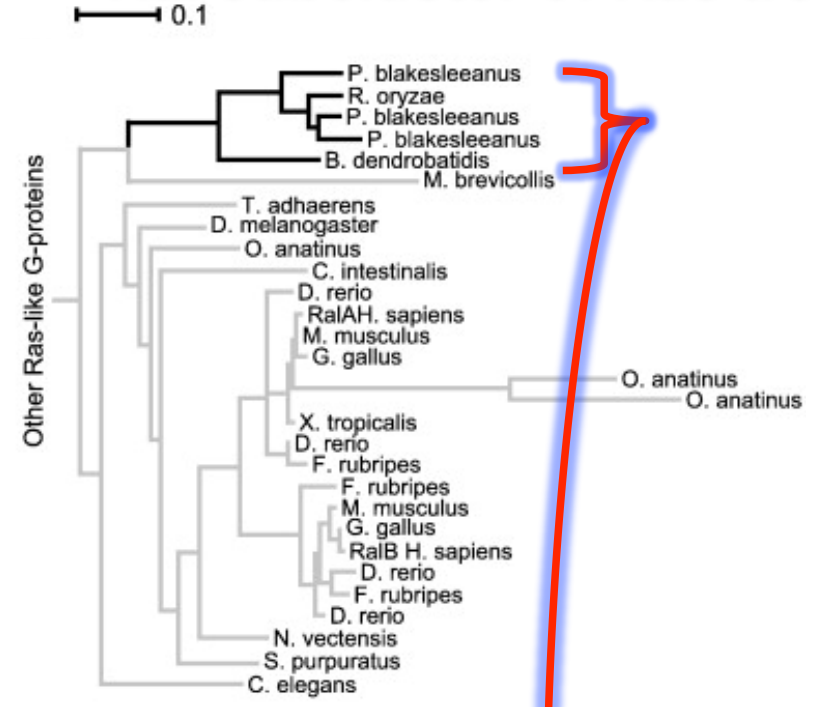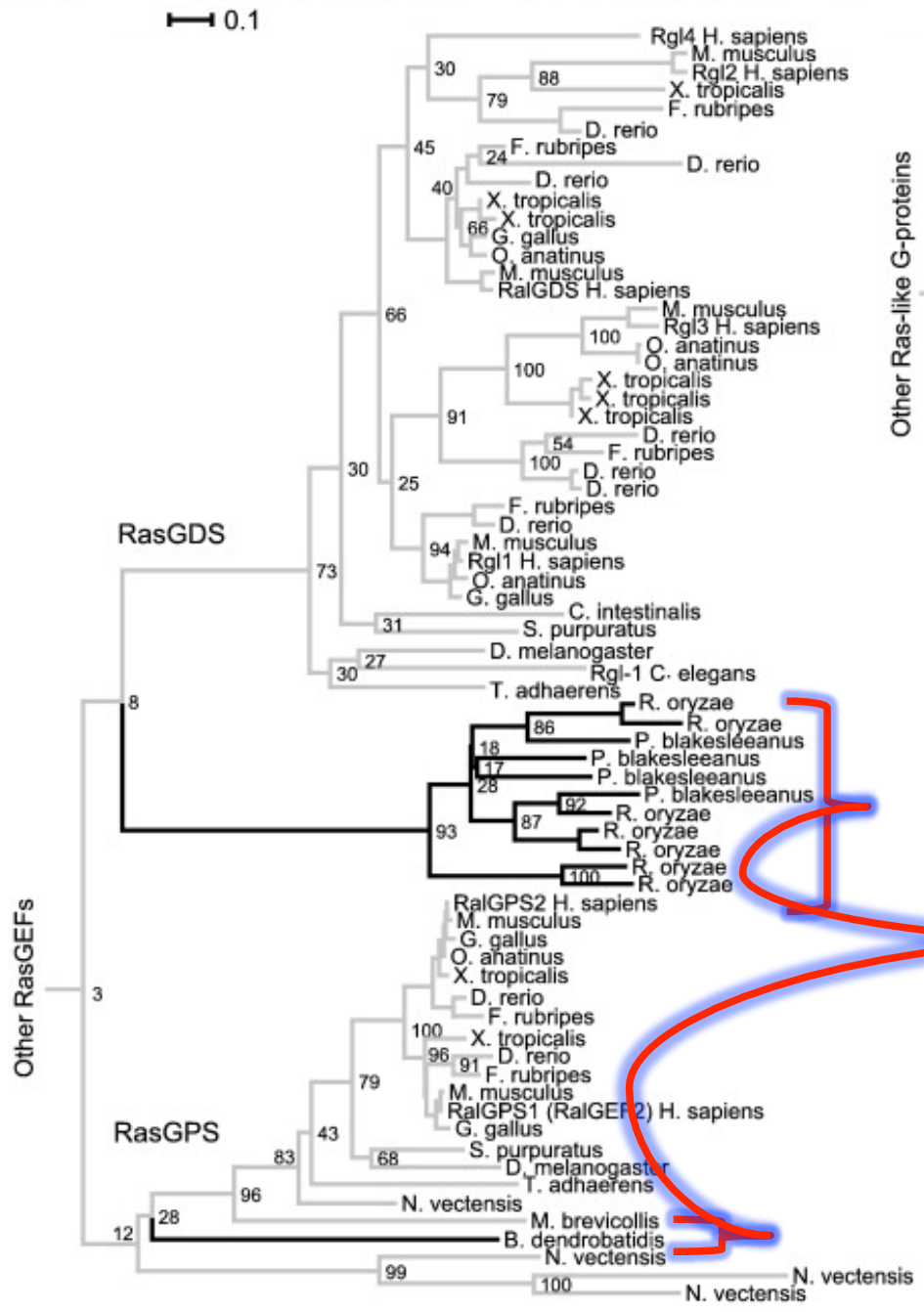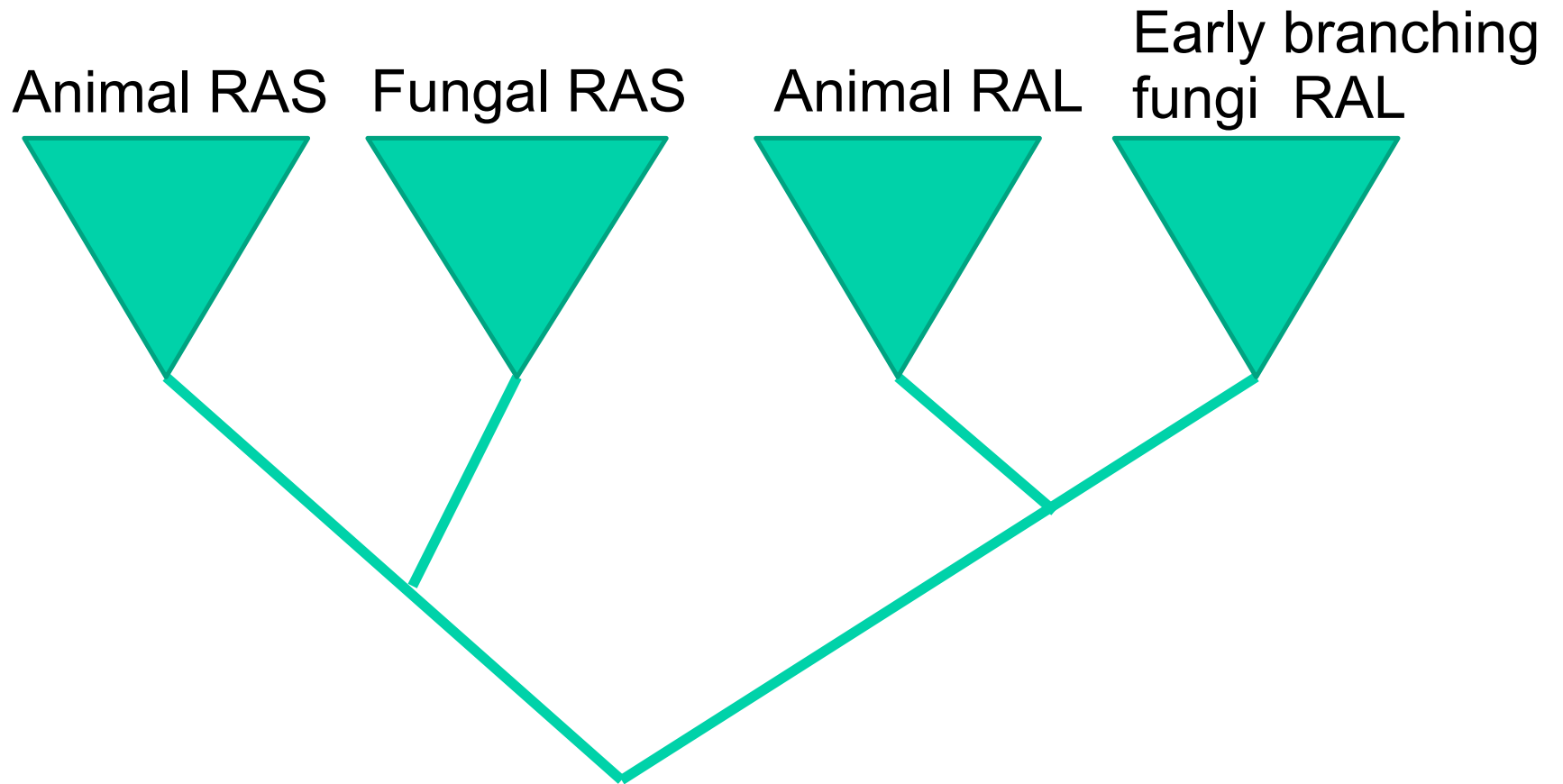
RAL evolution?

Animal invention and wrong tree ("consensus" in the RAS field) OR old duplication and loss

RalGEF subcluster of RasGEF tree   Ral subcluster of Ras tree

RAL evolution?

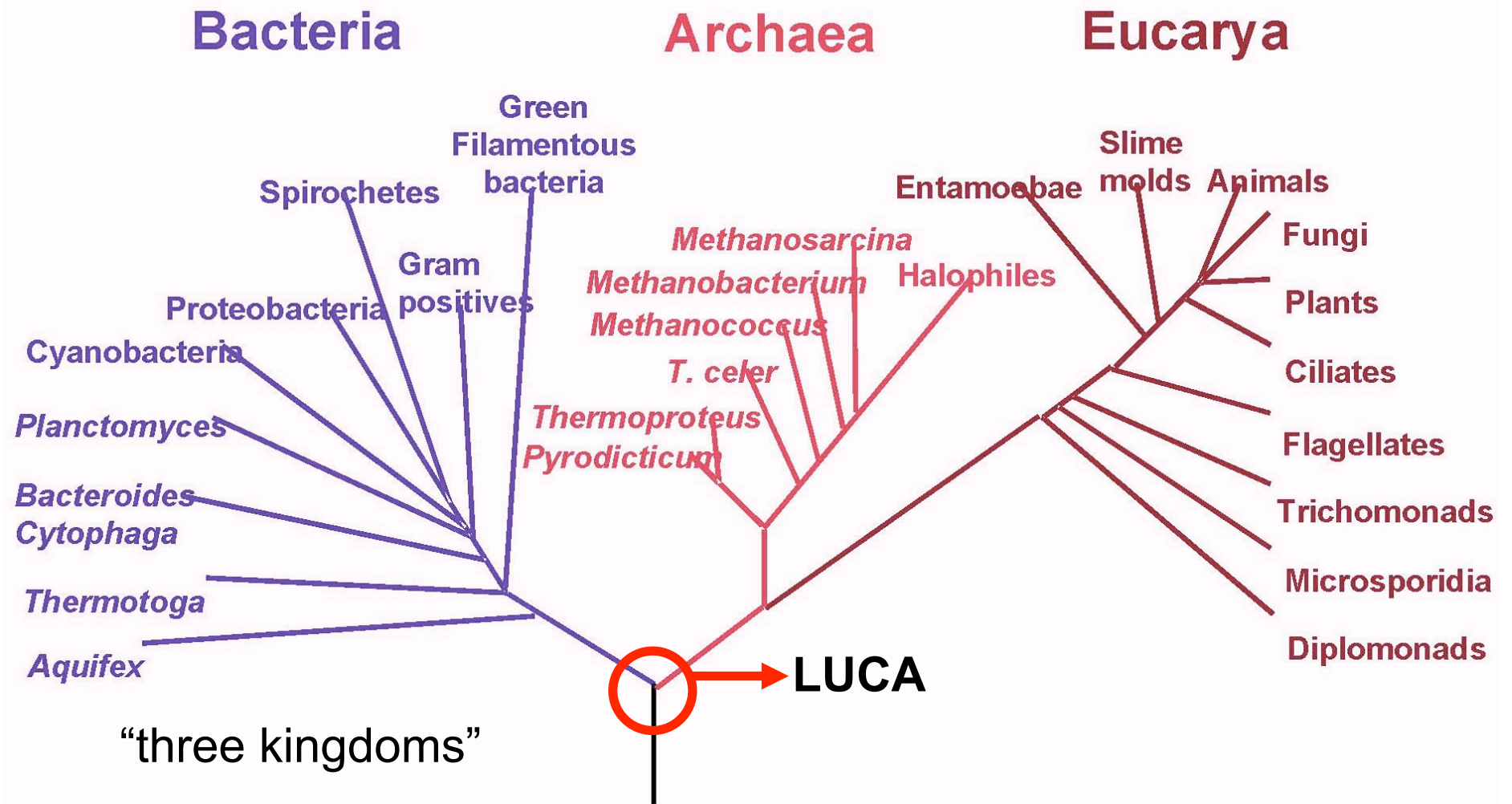Animal RAS   Fungal RAS   Animal RAL   Early branching fungi RAL
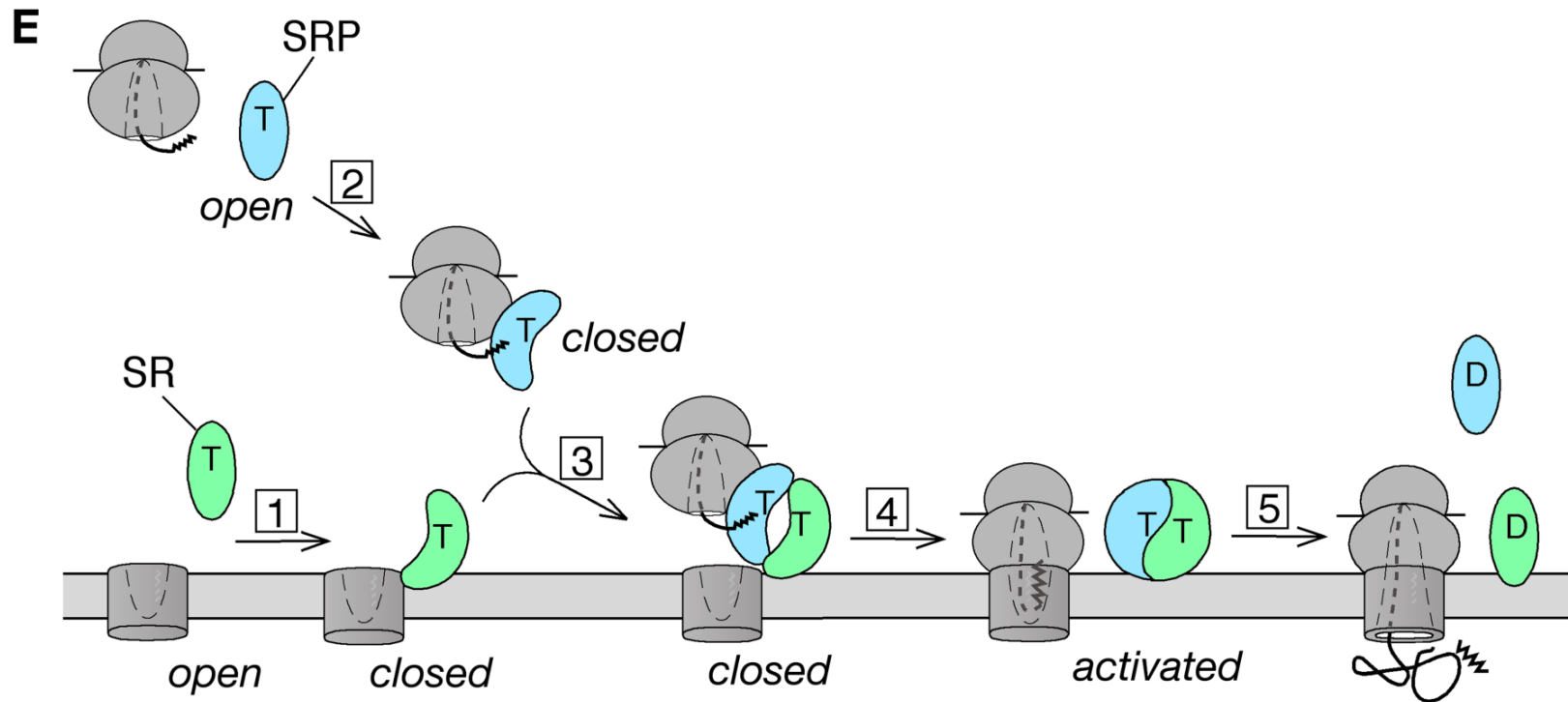
Old duplication and loss. No more OR.

Rooting 2: via gene duplications: a not unimportant example
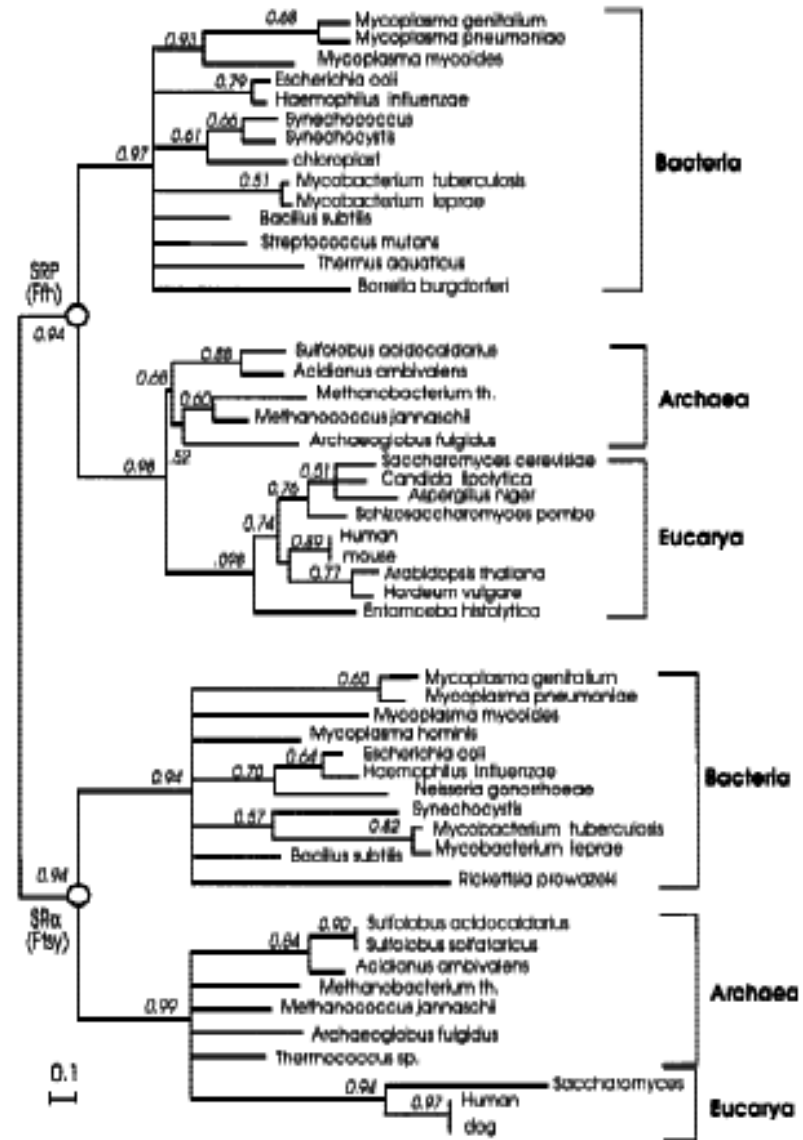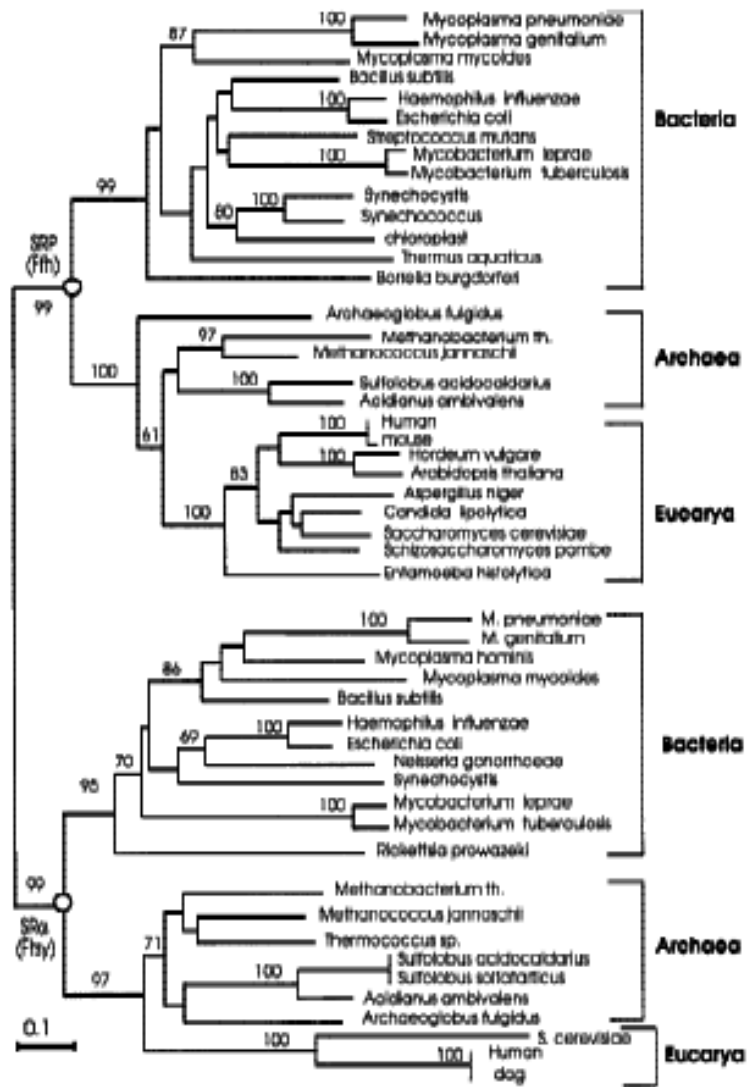

Phylogenetic Tree of Life

How to root the tree of life?
1: Find paralogs that duplicated before the LUCA

6 found so far

How to root the tree of life? 2: Make a tree of paralogs that duplicated before the LUCA

Griblado 1998 J Mol Evol

Zero losses

One gene duplication

The root of the tree of life

A poor alternative, tree reconciliation if e.g. euk vs bac & arch

Six losses

Three gene duplication