

Biodiversity bioinformatics

From large-scale phylogenomics to gene families and functions

30 Aug 2023: Orthology with OMA



Learning objectives

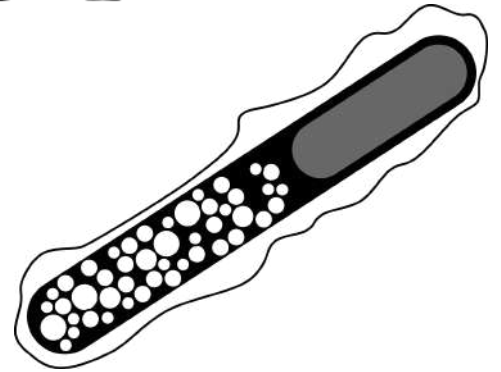
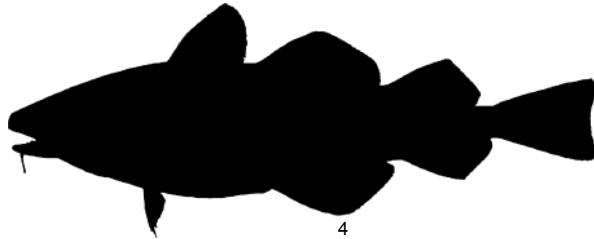
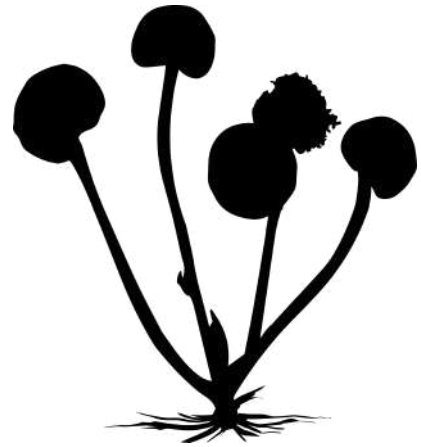
This course is centered around comparative genomics. After the course, you will be able to:

- Define orthology, paralogy and their subtypes
- Retrieve orthology information from the OMA database
- Map sequences quickly to their Hierarchical Orthologous Groups
- Infer orthologs on custom genomes using the FastOMA pipeline
- Construct and interpret phylogenetic species trees using OMA data

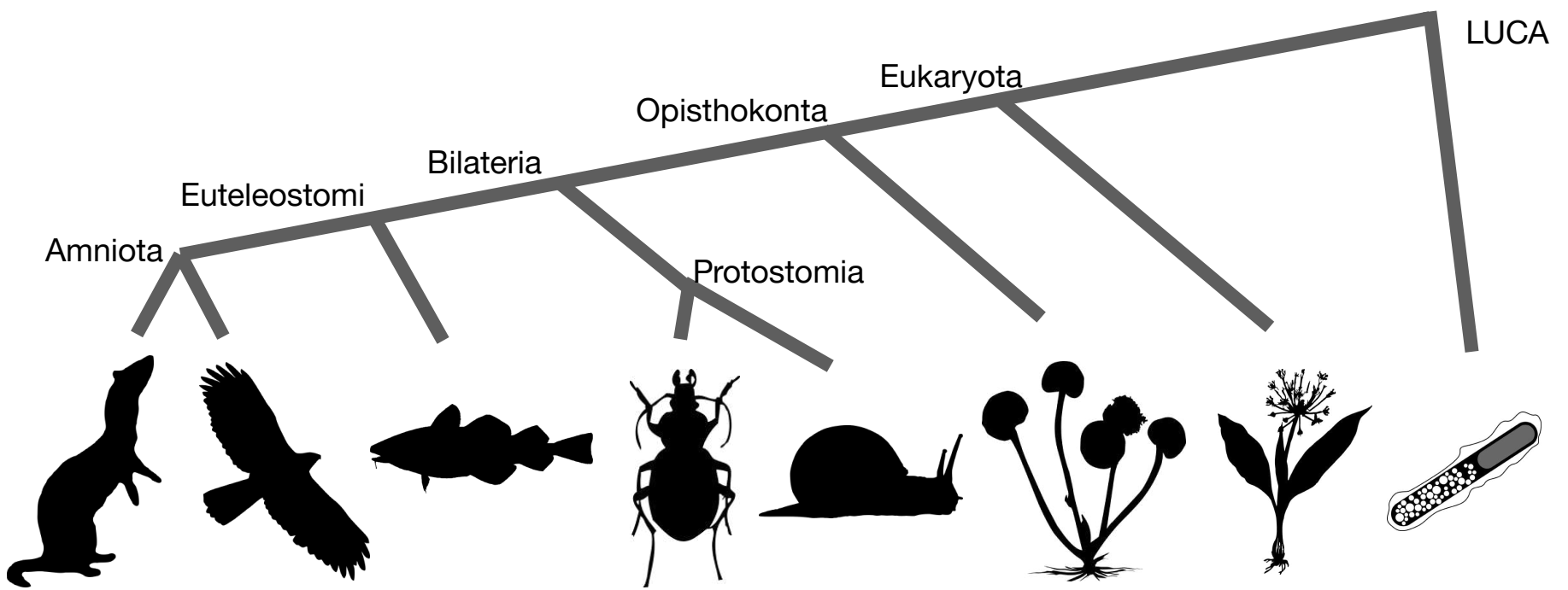
Schedule

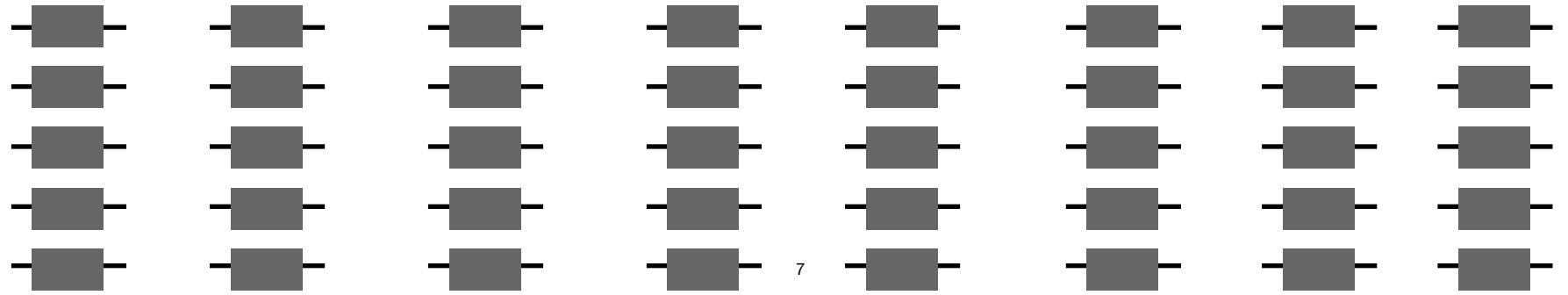
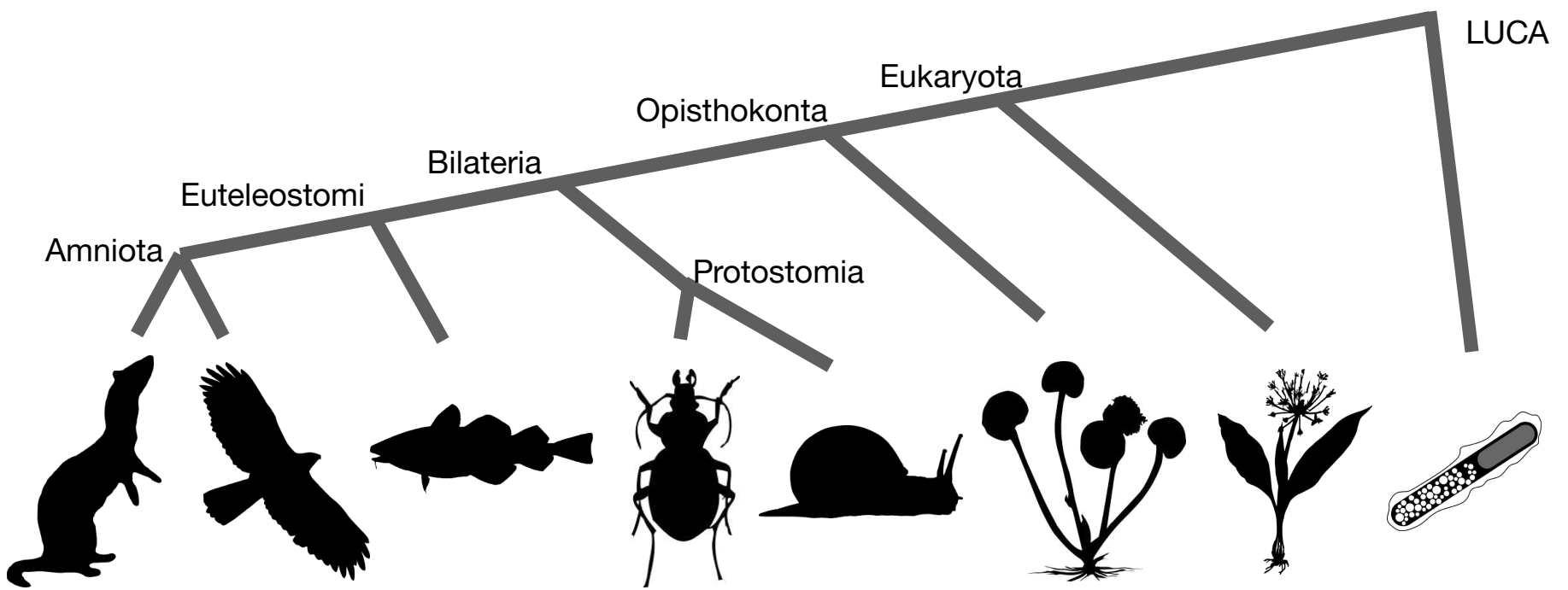
Schedule Aug 30 2023

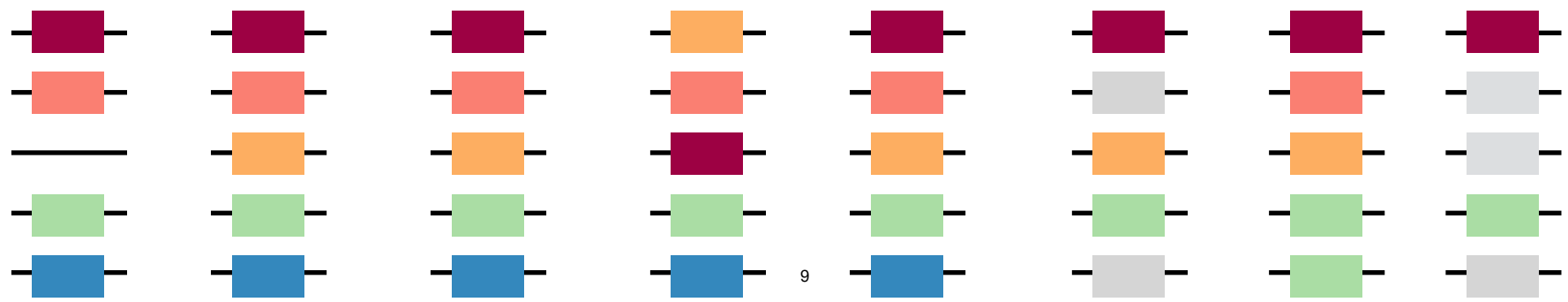
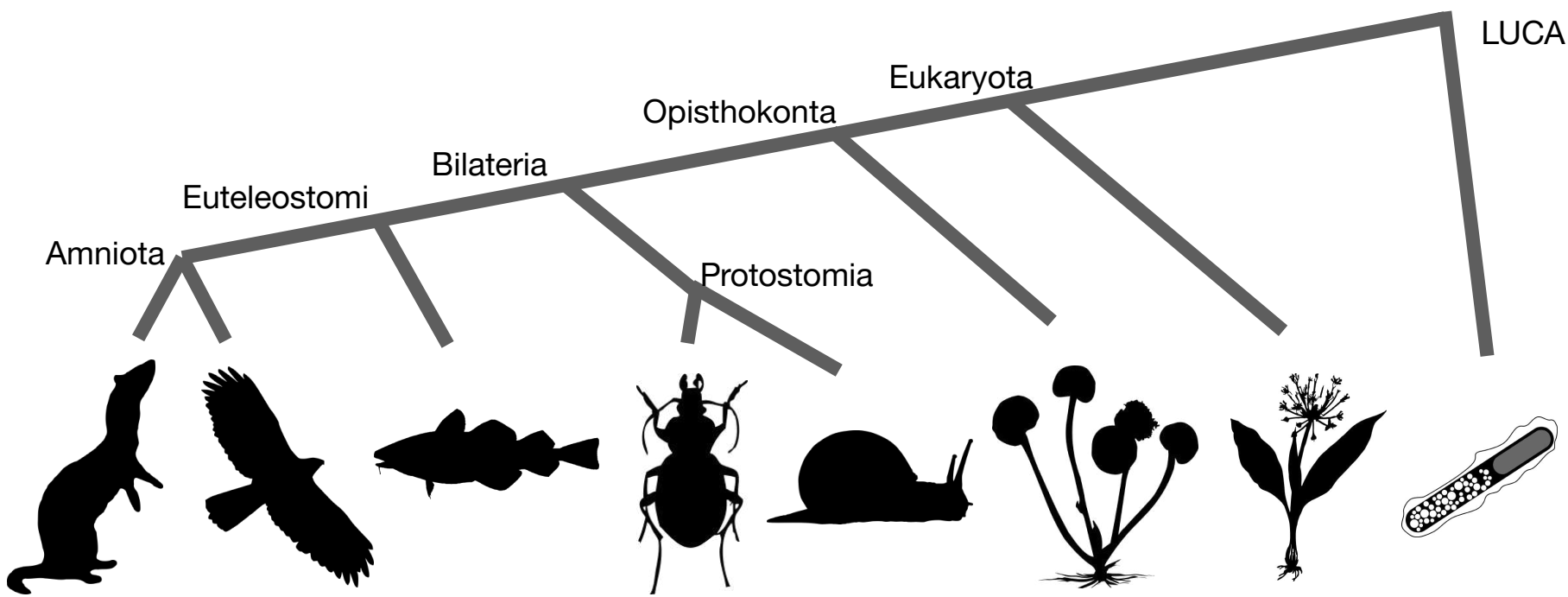
Time	Activity	In charge
9:00-9:30	Welcome, Introductions	Natasha Glover
9:30-10:00	Lecture: Overview, objectives, motivation, concept of orthologs, HOGs	Natasha Glover
10:00-10:45	Module 1: genes, groups, and genomes in the OMA Browser	Natasha Glover
10:45 - 11:15	Coffee break	
11:15-11:25	Go over results	Natasha Glover
11:25 - 11:30	Lecture: introduce OMAMer	Yannis Nevers
11:30 - 11:55	Module 2: Fast placement with OMAMer	Yannis Nevers
11:55-12:00	Go over results	Yannis Nevers
12:00-12:10	Lecture: FastOMA	Sina Majidian
12:10-12:30	Module 3 part 1: FastOMA (launch it before lunch)	Sina Majidian
12:30-13:30	Lunch	
13:30-14:00	Module 3 part 2: FastOMA	Sina Majidian
14:00-14:15	Go over results	Sina Majidian
14:15-14:30	Lecture: Gene trees and species trees	Christophe Dessimoz
14:30-15:00	Coffee Break	
15:00-16:00	Module 4: Building Species Trees	Christophe Dessimoz
16:00-16:15	Go over results	Christophe Dessimoz
16:15-16:45	OMA Clinic	Natasha, Yannis, Christophe
16:45-17:00	Wrap up	Natasha, Yannis, Christophe







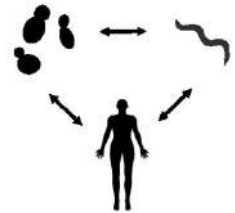
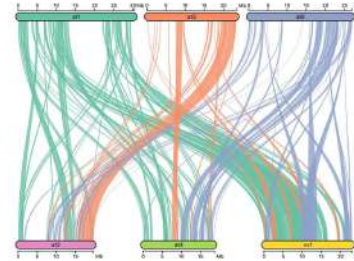
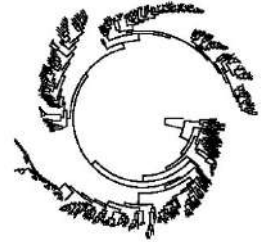
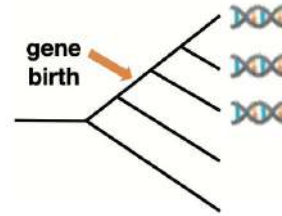




Orthology is fundamental to comparative genomics

Applications:

- Building phylogenetic trees
- Transferring gene function
- Finding lineage-specific genes
- Studying ancestral genomes
- Finding co-evolving genes
- Exploring synteny

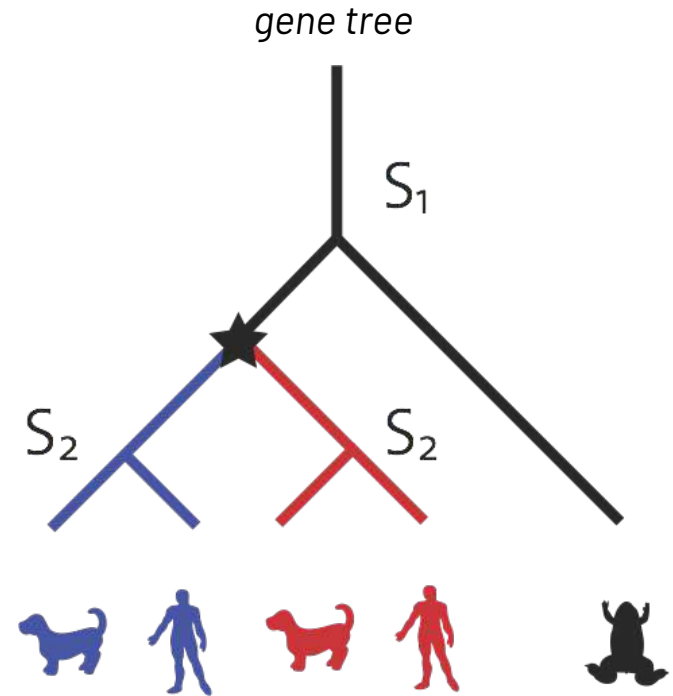


Orthology: What is it?

Homology

- The study of genetic material almost always starts with identifying, within or across species, **homologous regions**—regions of common ancestry.
- Homologs = **gene families**
- It is useful to distinguish between two classes of homologous genes.

Orthologs vs. paralogous



Orthologs vs. paralogs

- **Orthologs:** pairs of genes that started diverging via evolutionary **speciation**
- **Paralogs:** pairs of genes that started diverging via gene **duplication**

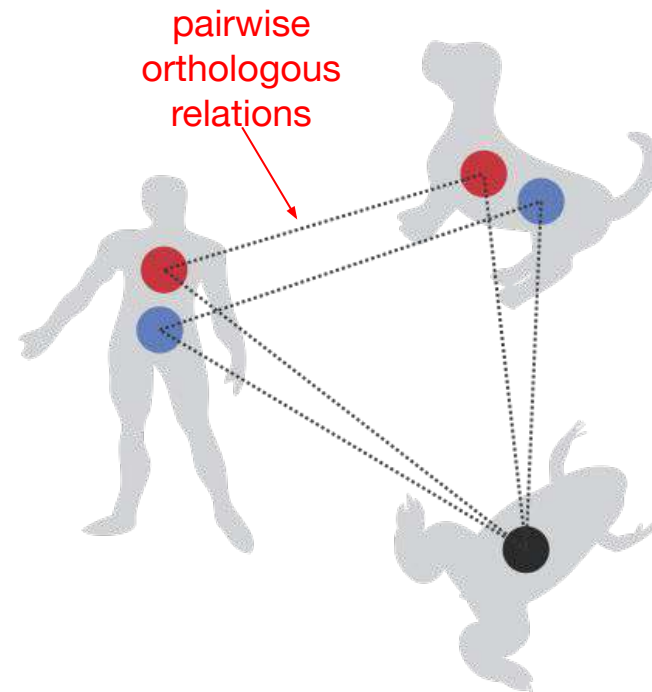
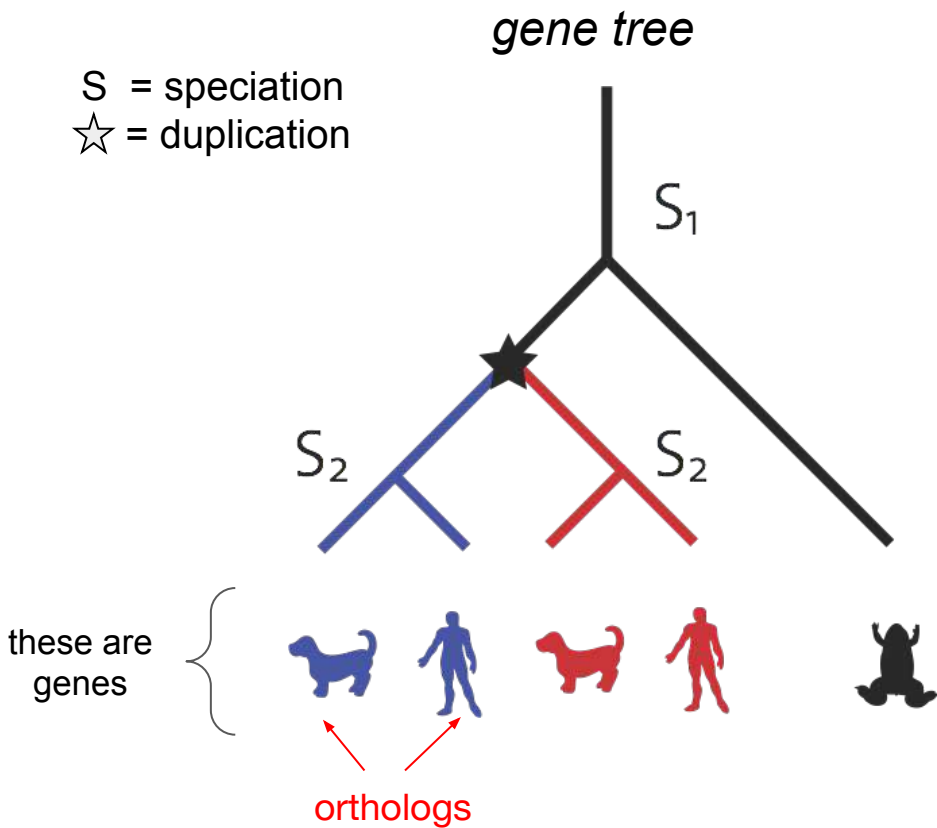
- **Orthology:** A relation between pairs of genes that started diverging via evolutionary **speciation**
- **Paralogy:** A relation between pairs of genes that started diverging via gene **duplication**

Ortho = exact

Para = beside/next to

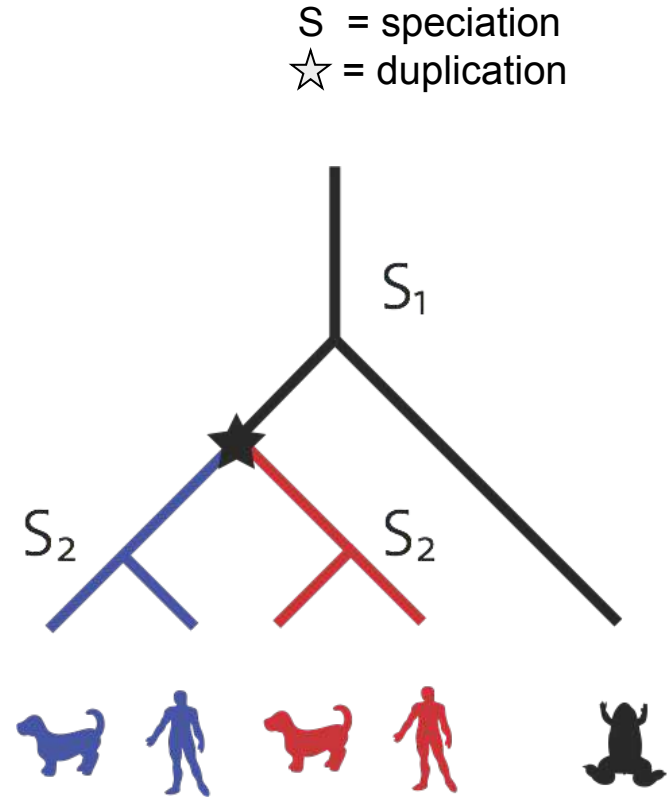
Orthologs

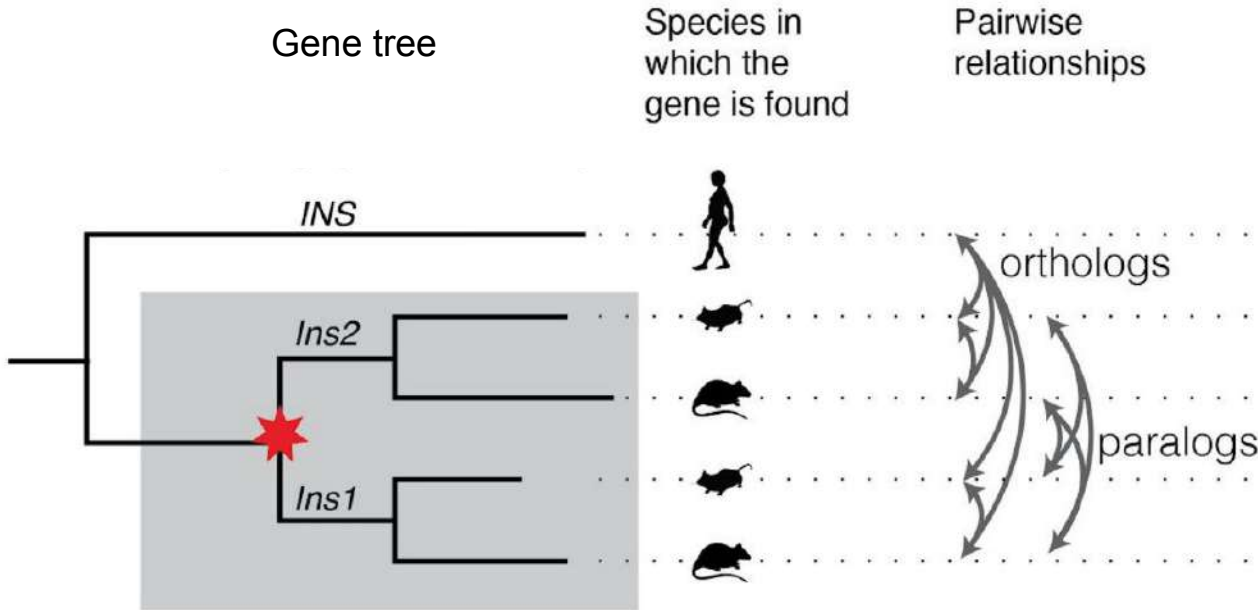
- Two genes in two species are **orthologous** if they derive from one gene in their last common ancestor
- Can be thought of as “corresponding genes” between species



Common misconceptions

- Orthologs don't have to be one-to-one
- Orthology can also be a one-to-many, many-to-one, or many-to-many relationship
- Paralogs don't have to be in the same species





Difficult to interpret pairwise relationships when referring to groups of genes in several species...

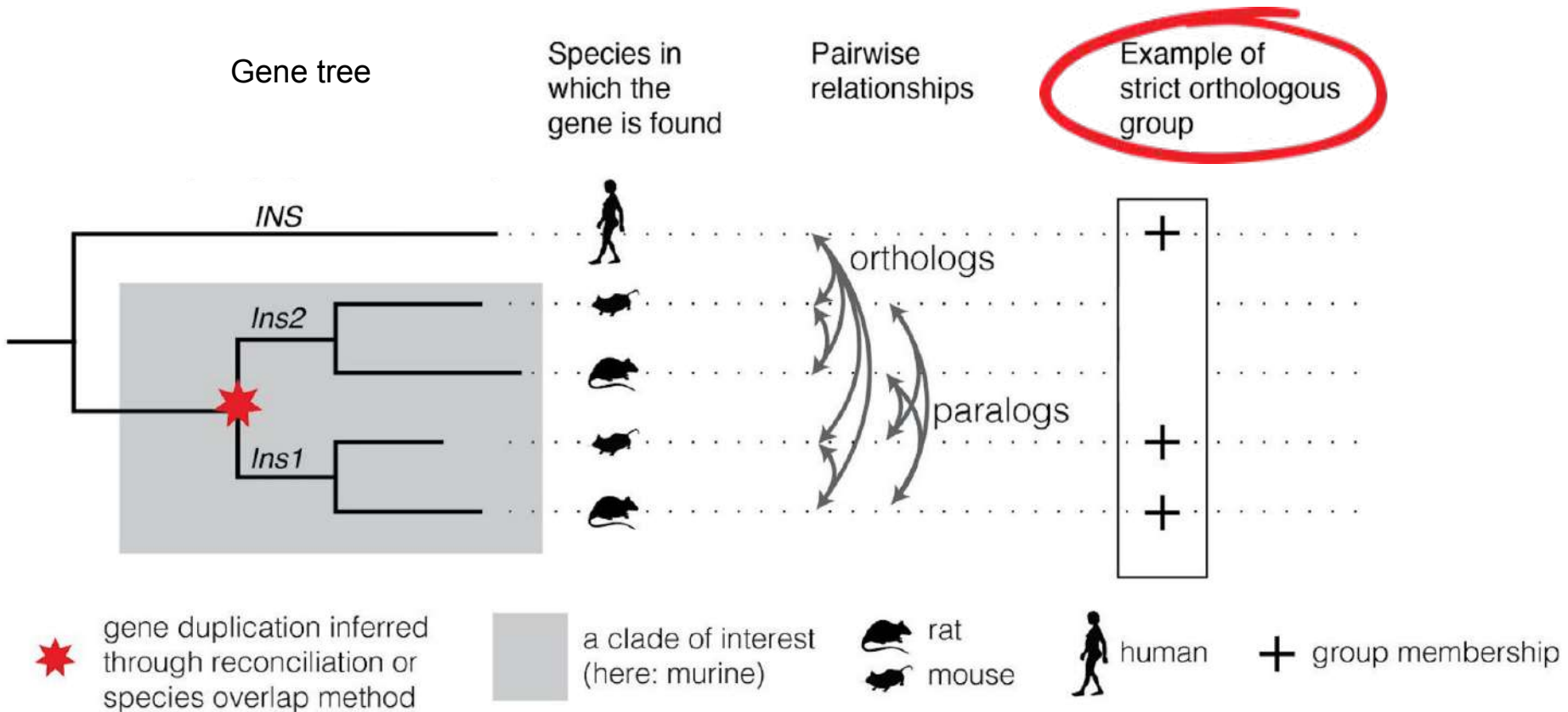
★ gene duplication inferred through reconciliation or species overlap method

■ a clade of interest (here: murine)

🐭 rat
 🐁 mouse

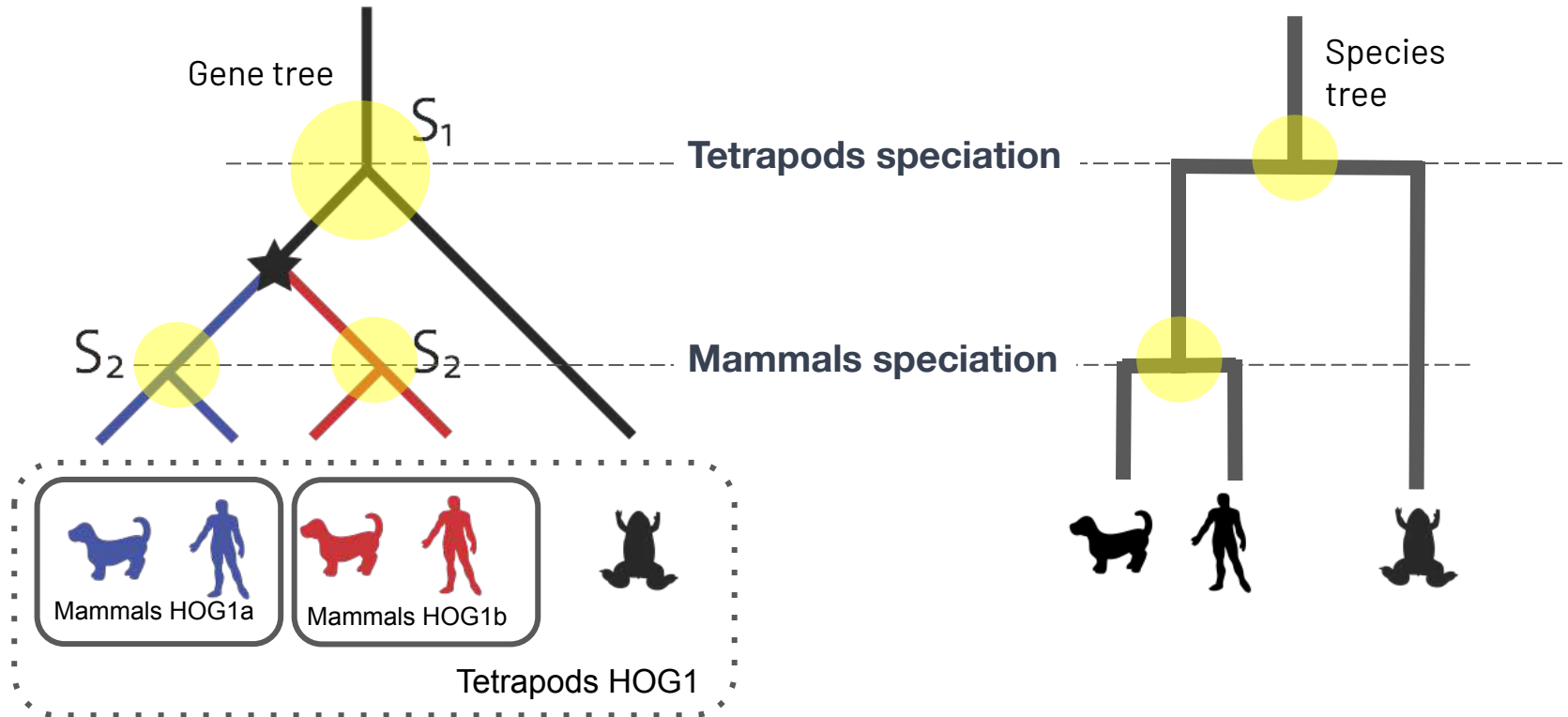
Orthologous Groups (i.e. OMA Groups/OGs)

- Groups of genes which are all orthologous to each other
- Strict orthologous group
- Not necessarily 1:1 orthology, but each group contains at most one gene per species

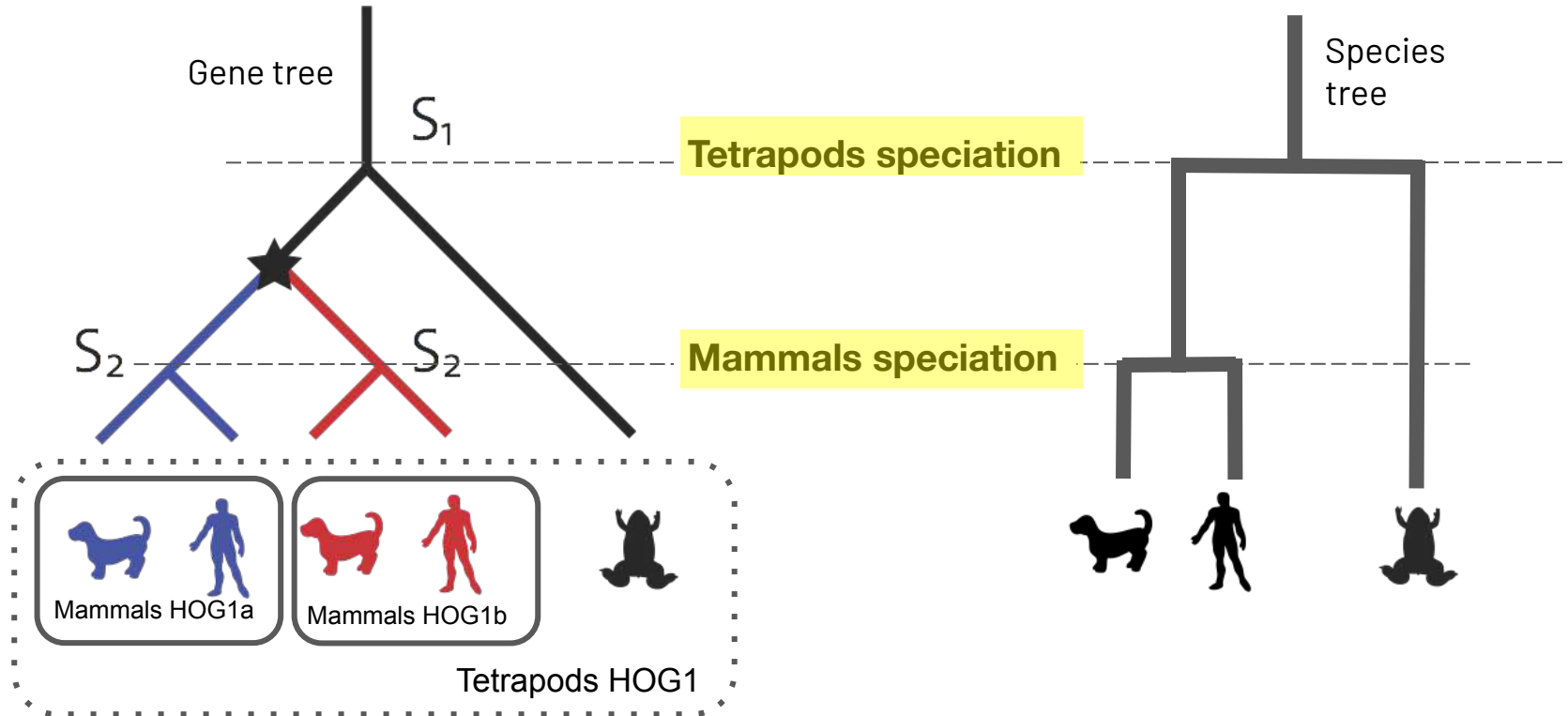


Hierarchical Orthologous Groups (HOGs)

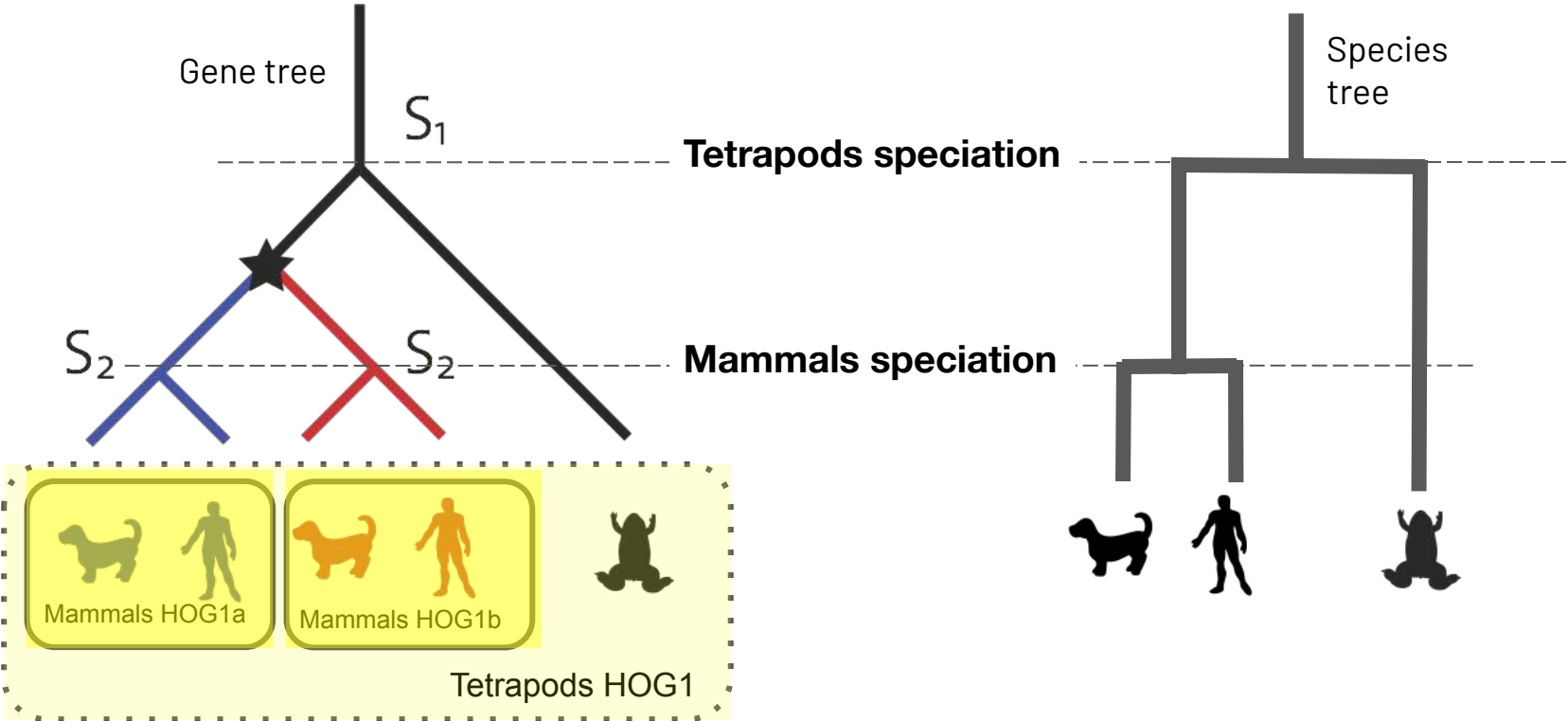
HOGs = Sets of genes that descended from a common ancestral gene in a given ancestral species



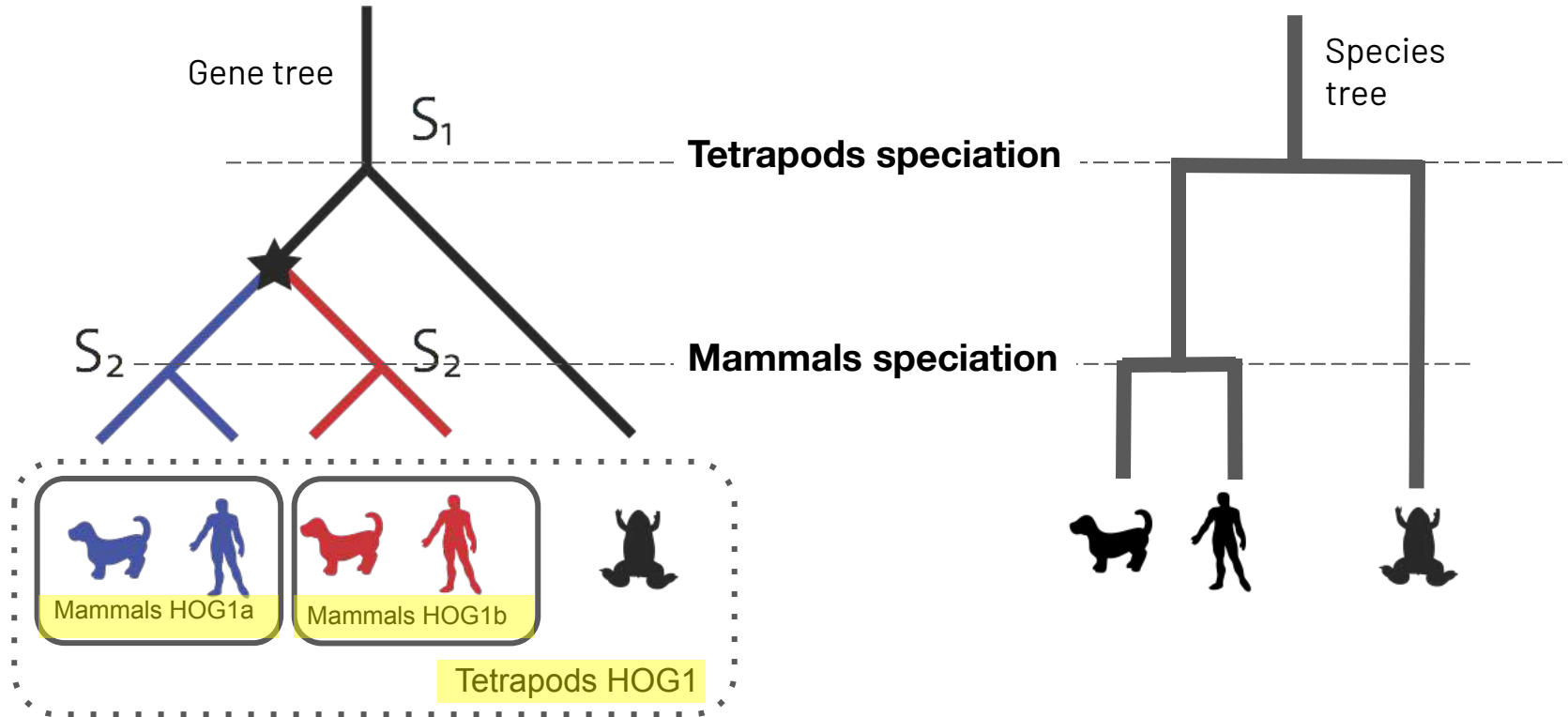
HOGs are defined with respect to specific clades



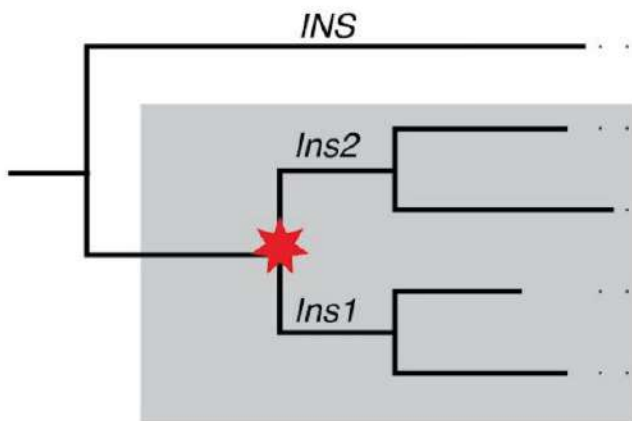
HOGs are hierarchical because groups defined with respect to deeper clades subsume multiple groups defined on their descendants



HOGs are gene families; SubHOGs are nested subfamilies



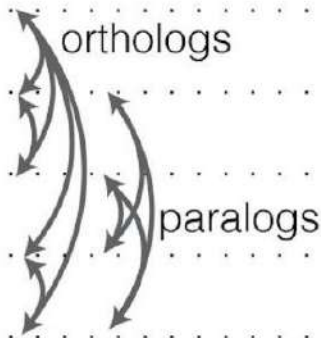
Labelled gene trees
(marked internal nodes are duplication nodes; others are implicitly speciation nodes)



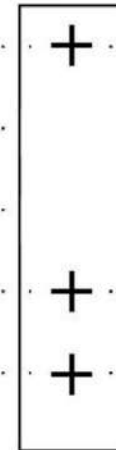
Species in which the gene is found



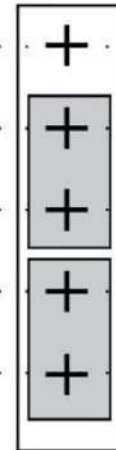
Pairwise relationships



Example of strict orthologous group



Hierarchical Orthologous Groups



★ gene duplication inferred through reconciliation or species overlap method

■ a clade of interest (here: murine)

🐭 rat
🐁 mouse

🧑 human

+ group membership

How to infer orthology?

Orthology Databases

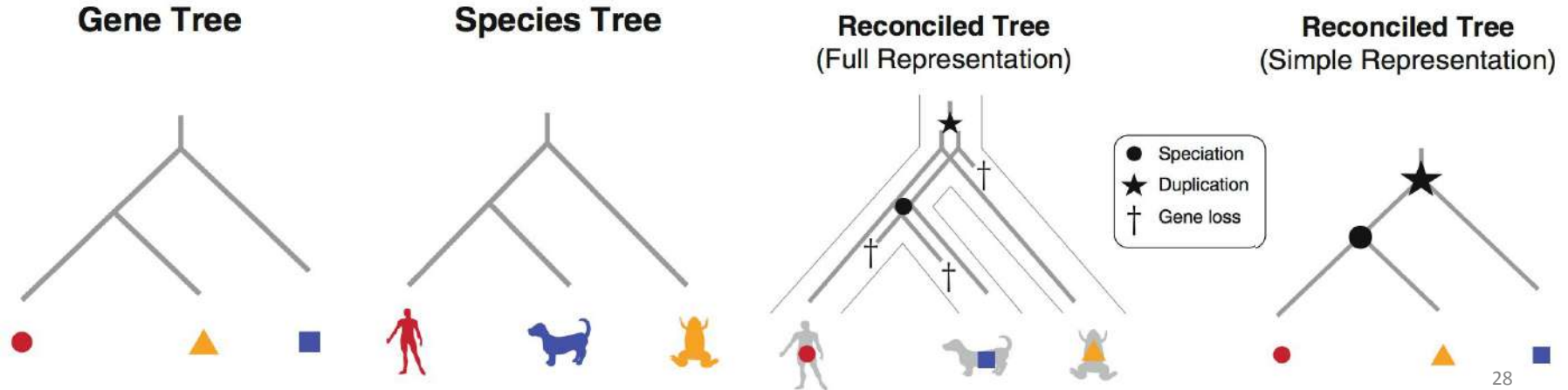
- OMA: <https://omabrowser.org/>
- OrthoDB: <https://www.orthodb.org/>
- EggNOG: <http://eggnogdb.embl.de>
- InParanoid/Hieranoid:
<http://hieranoidb.sbc.su.se/>
- COG: <https://www.ncbi.nlm.nih.gov/COG/>
- PANTHER: <http://www.pantherdb.org/genes/>
- PhylomeDB: <http://phylomedb.org/>
- And more!

Many different methods:
*Namely tree-based or
graph-based*

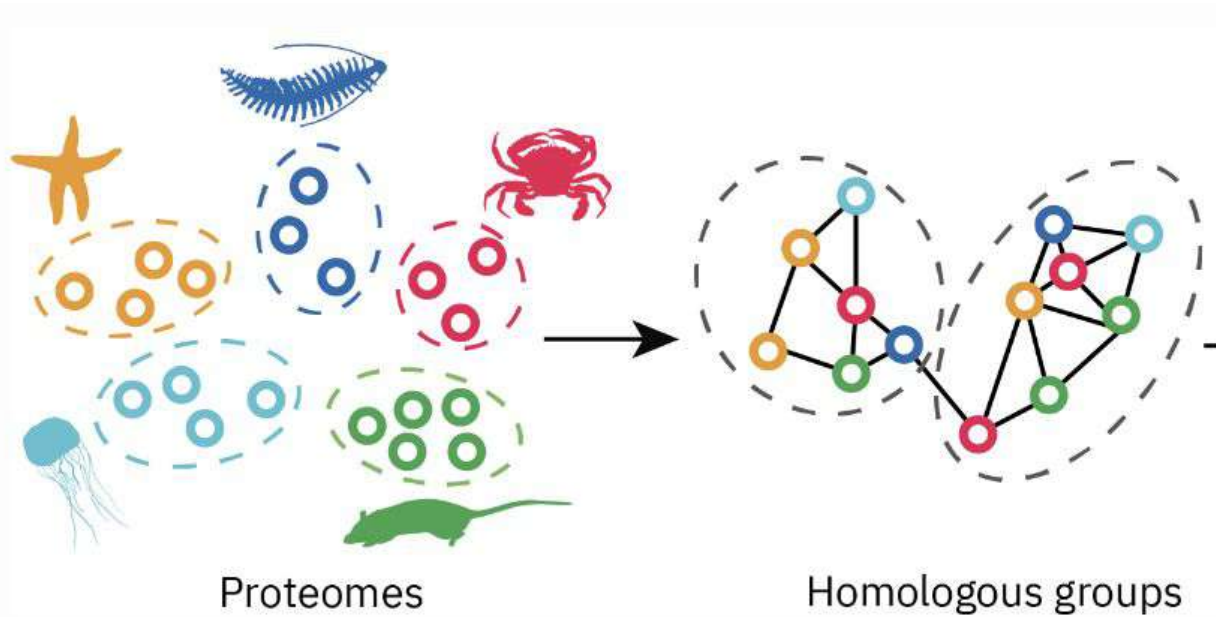
Tree-based methods

- Start with a group of homologous sequences
- Reconstruct a gene tree
- Infer the type of evolutionary event represented by each internal node of the tree

Gene tree/species tree reconciliation:



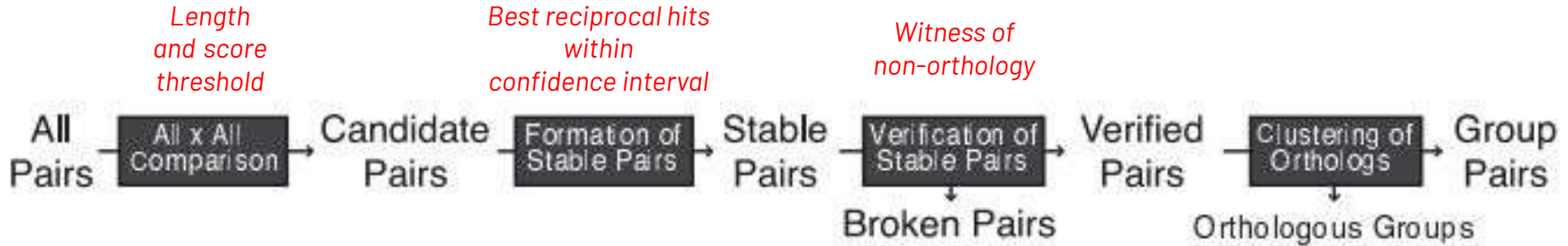
Graph-based methods



- Compare pairs of genes within and between species
- For pairs of genes between 2 species, orthologs tend to be the pairs of sequences that have diverged the least



Pipeline:



FastOMA, our new tool

Input Proteomes

```
>HA1  
MADTSHL..  
>HA2  
MHPYSTQ..  
#Human
```

OMAmer
Mapping sequences
on OMA gene families
based on k-mers



Root HOGs (Gene families)

```
>HA1 ●  
>HA2 ●  
>MA ●  
>CA1 ●  
.. #rootHOG1  
#rootHOG2  
#rootHOG3  
...
```

SubHOG/event
inference
(in parallel)

Hierarchical Orthologous
Groups (HOGs)

