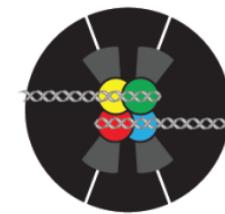


# Phylogenomics

**Maria Eleonora Rossi, Ana Riesgo**



Contact details:  
[mariaeleonora.rossi@universityofgalway.ie](mailto:mariaeleonora.rossi@universityofgalway.ie)



Gahan Lab



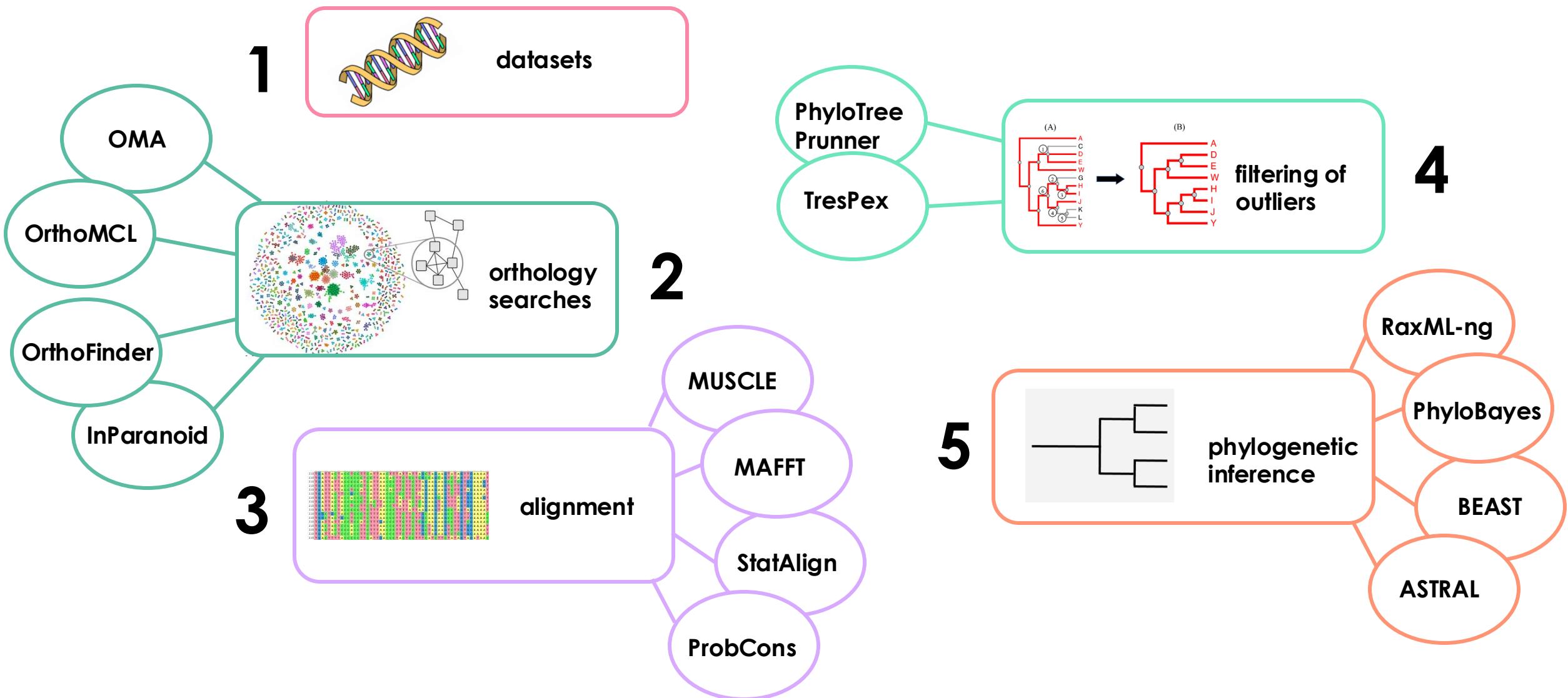
Centre for  
Chromosome  
Biology



OLSCOIL NA  
GAILLIMHE  
UNIVERSITY  
OF GALWAY

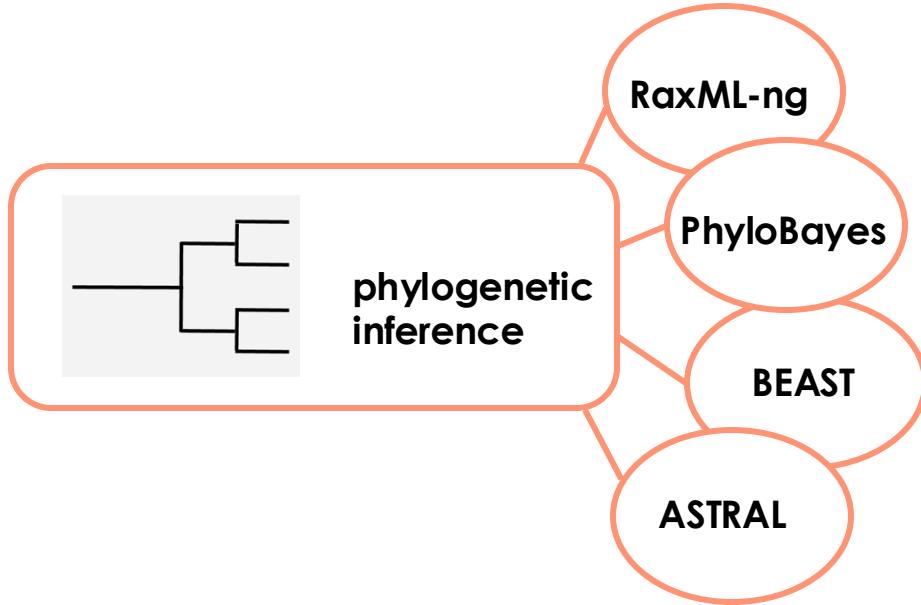


# workflow for a phylogenomic study



# How to do phylogenomics

5

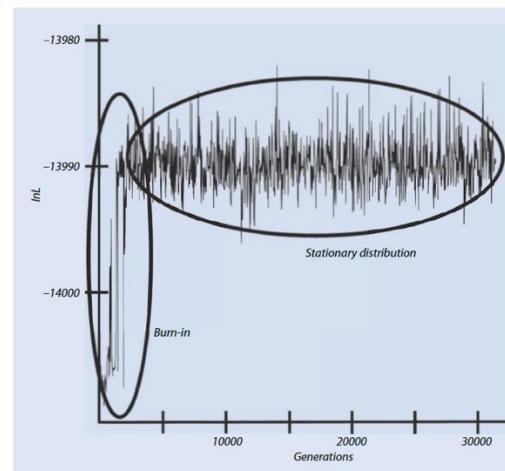
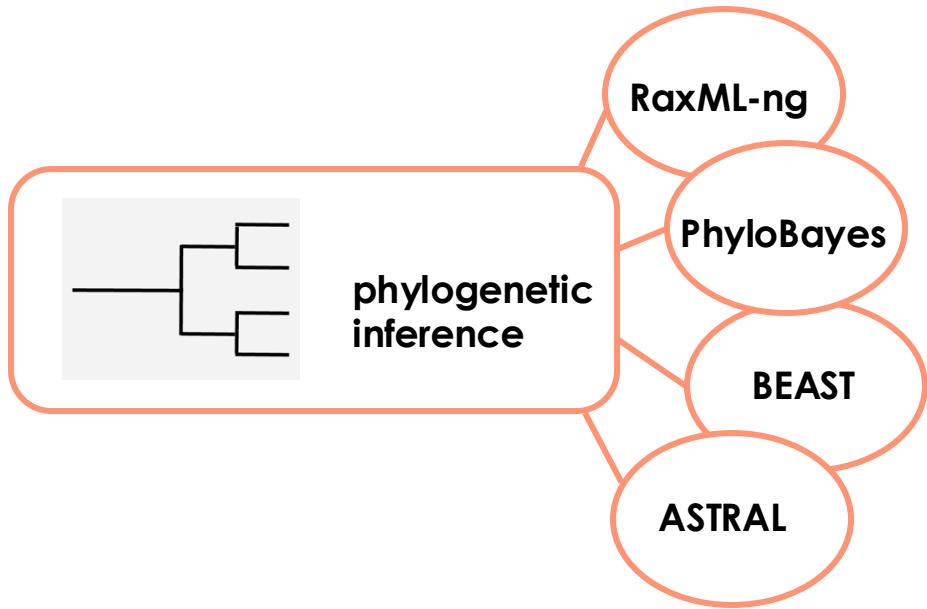


## Tree inference

**Maximum likelihood  
Bayesian inference  
Coalescent methods**

# How to do phylogenomics

5



## Maximum likelihood

- find a tree that maximizes the probability of the genetic data given the tree
- Model selection

Tree evaluation: Bootstrapping

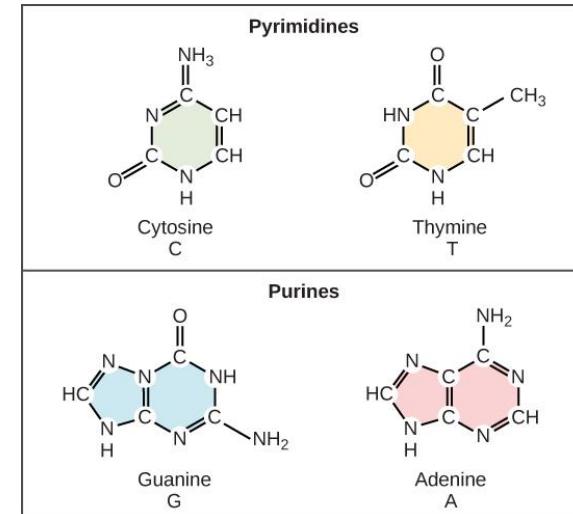
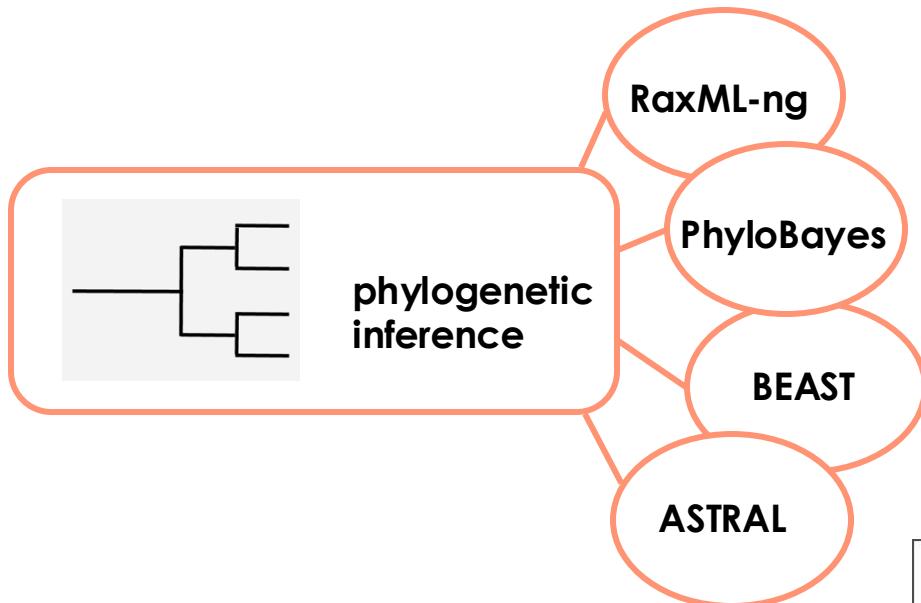
## Bayesian inference

MCMC

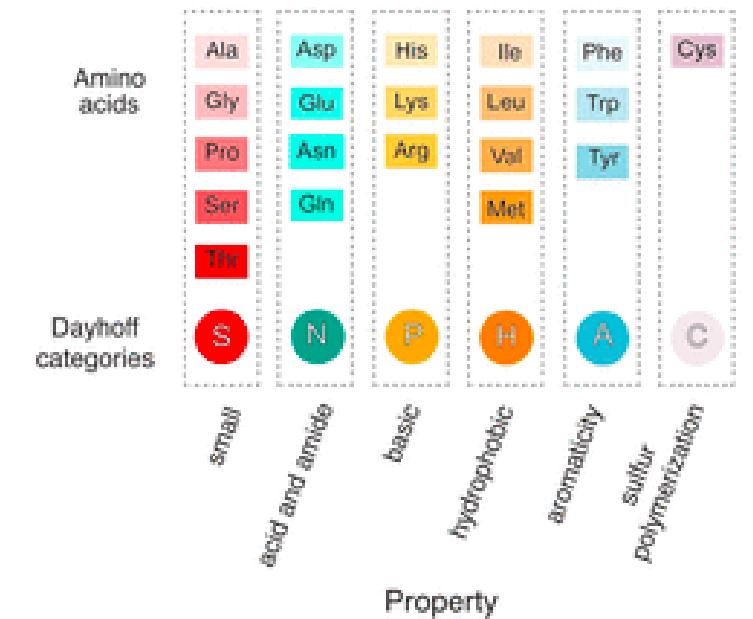
Tree evaluation: Posterior probability and convergence

# How to do phylogenomics

5



... but before this... MODEL  
SELECTION and recoding ...  
why?

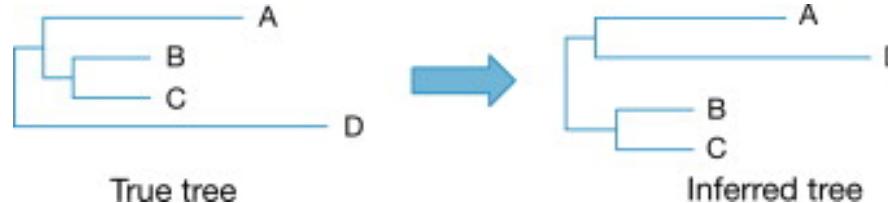


# How to do phylogenomics

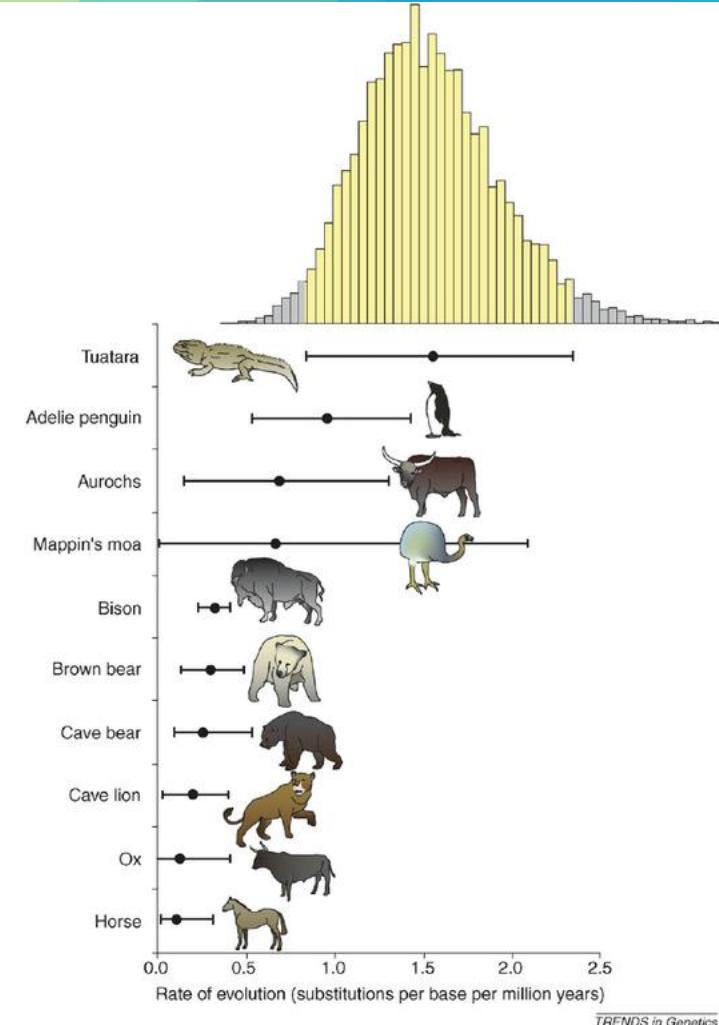
Main phylogenetic errors caused by unequal rates of evolution

**(site)Heterogeneity:** the tendency of genes or organisms to have unequal proportions of amino acids because of selective pressures and different translation efficacy:  
variation of the substitution process across sites

**Sequence saturation:** when there have been multiple amino acid substitutions at the same site washing out the evolutionary signal



Felsenstein 1978; Lawrence et al. 2019



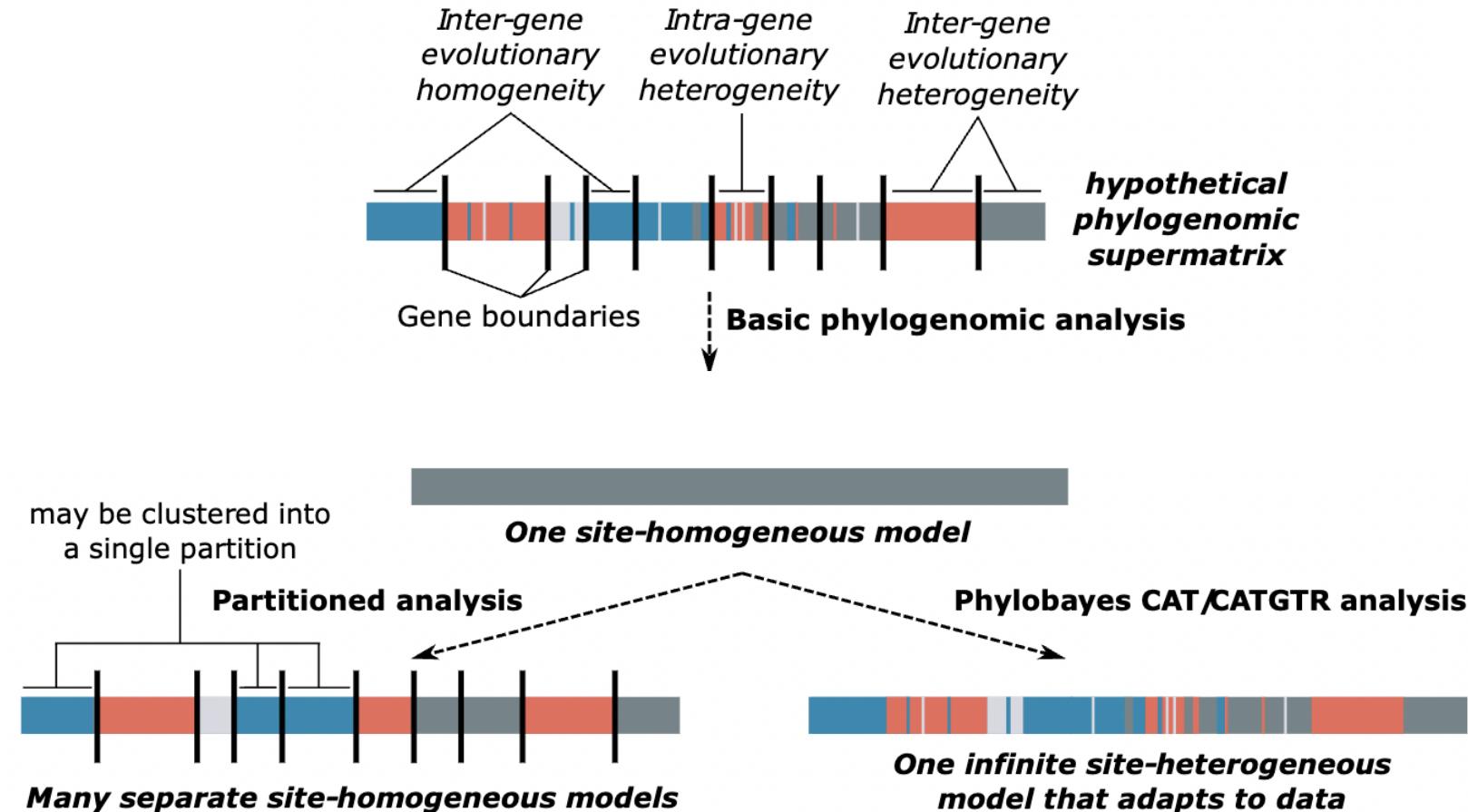
Hay et al. (2008). Trends Genet

# How to do phylogenomics

How do we deal with this?

**Site-Heterogeneous models**

CAT/CATGTR  
infinite mixture model

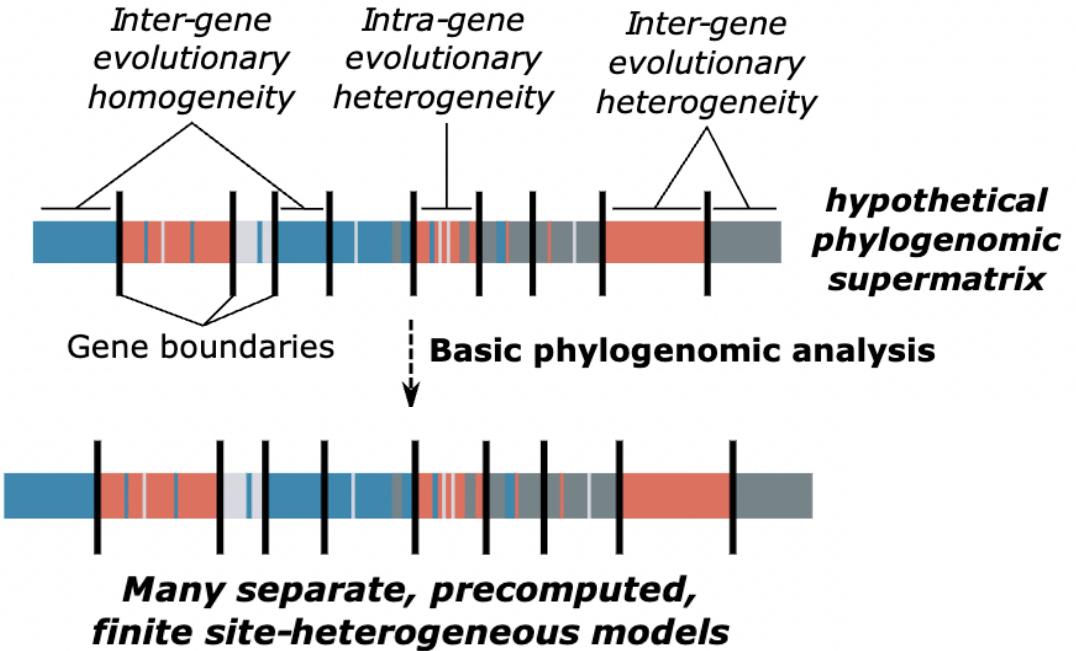


# How to do phylogenomics

How do we deal with this?

**Site-Heterogeneous models**

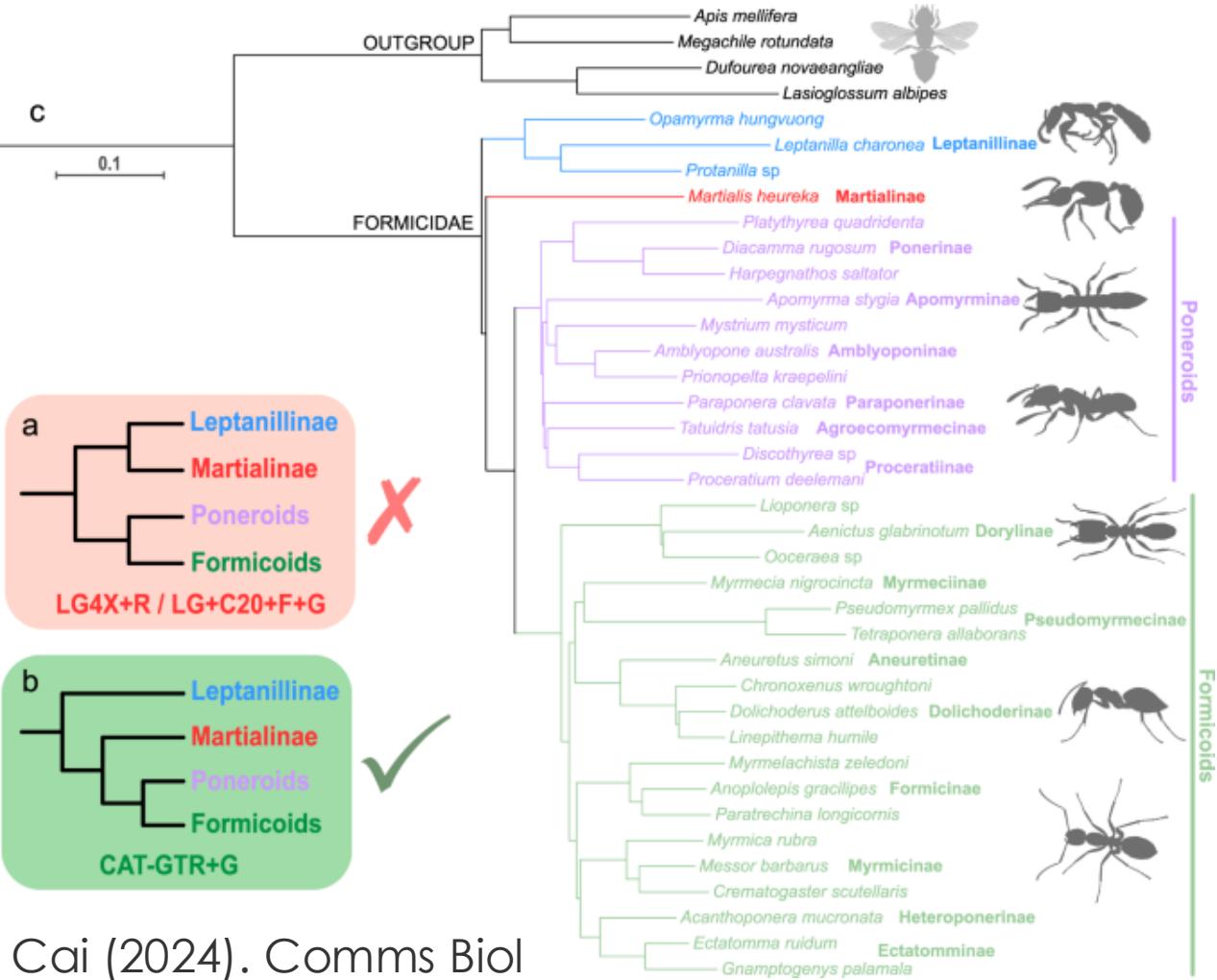
CAT/CATGTR  
infinite mixture model



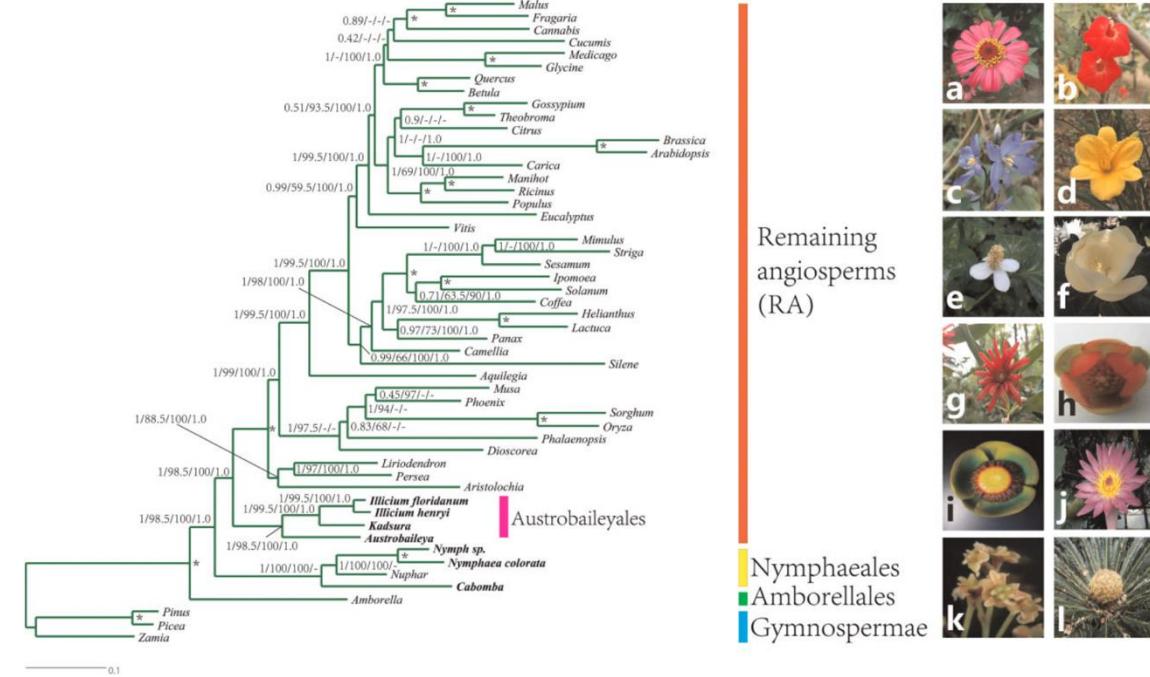
Has this been useful?

# How to do phylogenomics

## Evolution of ants

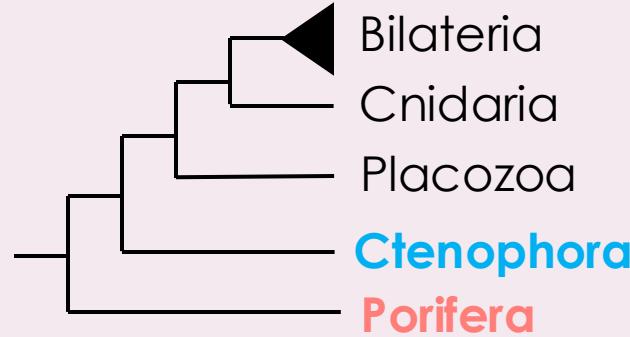


## Evolutionary Root of Angiosperms



# Phylogenomics and comparative genomics

## The phylogenomic wars: Porosis and Ctenosis



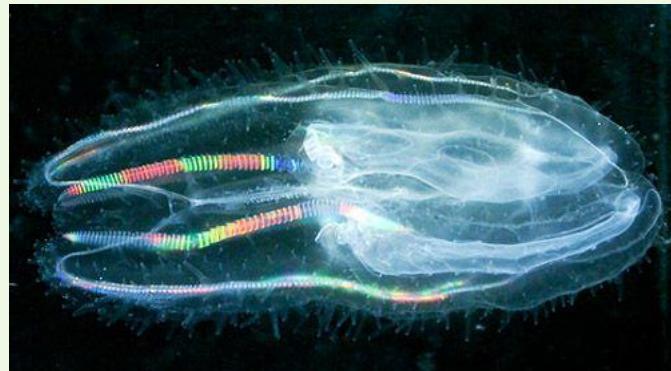
“Porifera sister”

Most parsimonious explanation

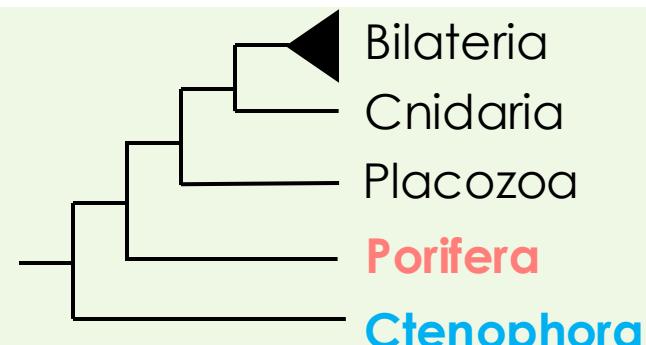


Nosenko et al. 2013, Philippe et al. 2009,  
Pick et al. 2010, Pisani et al. 2015,  
Feuda et al. 2017, Redmond & McLysaght 2021

Site-heterogeneous models



Site-homogeneous models

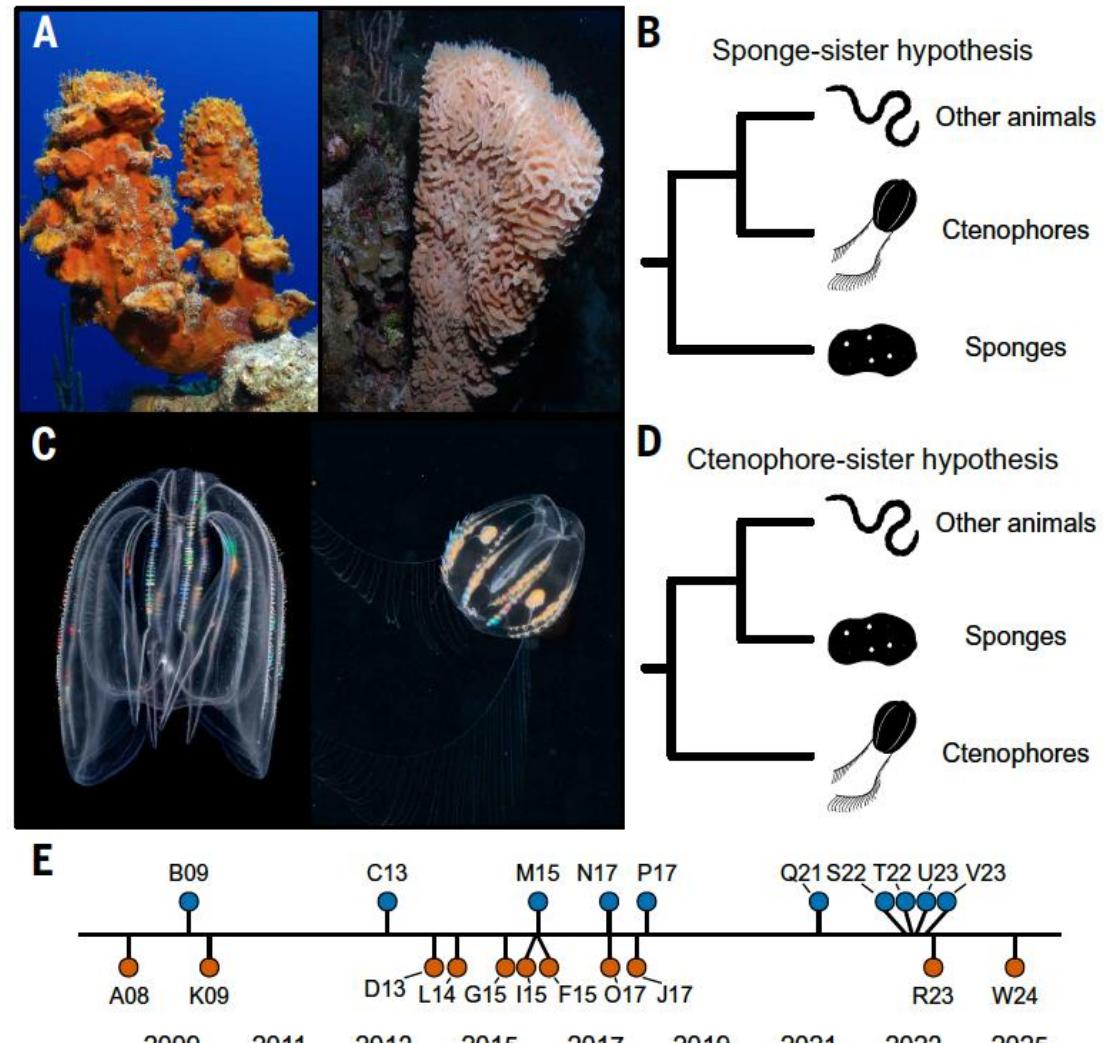
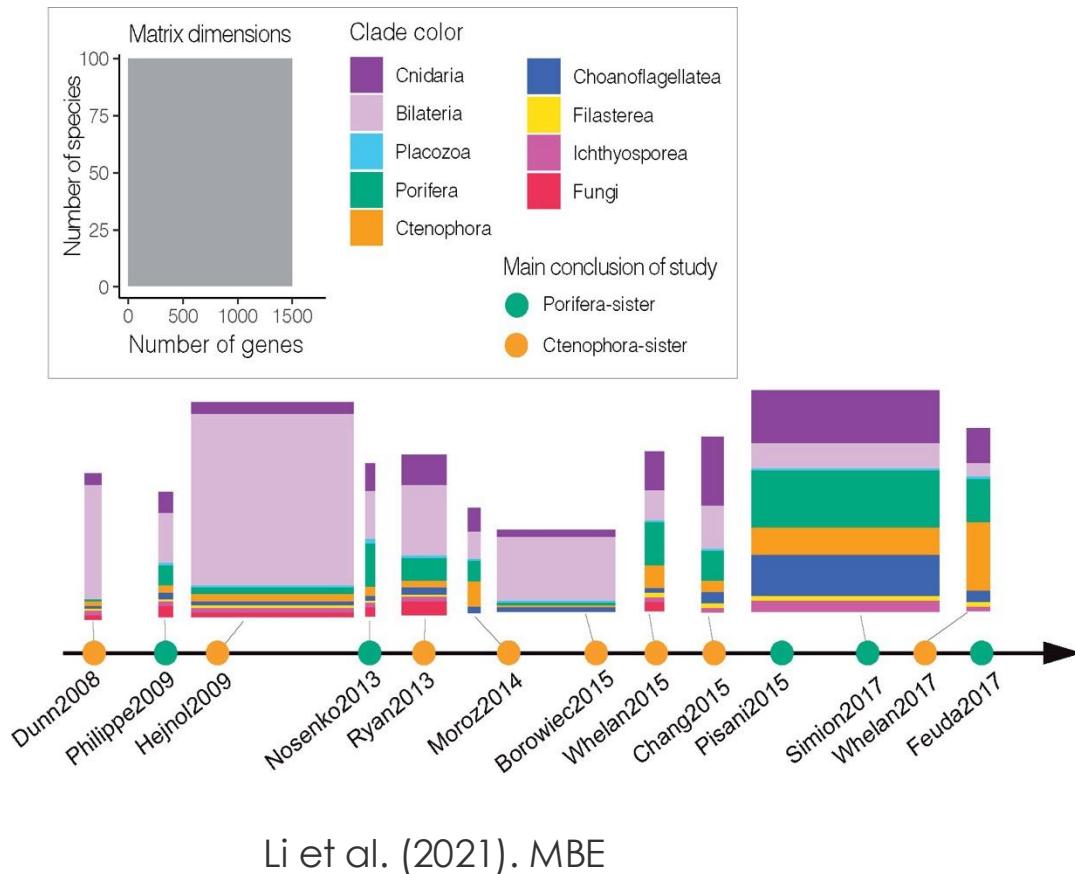


“Ctenophora sister”

1. Gene loss in **Porifera**
2. Parallel evolution in **Ctenophora**

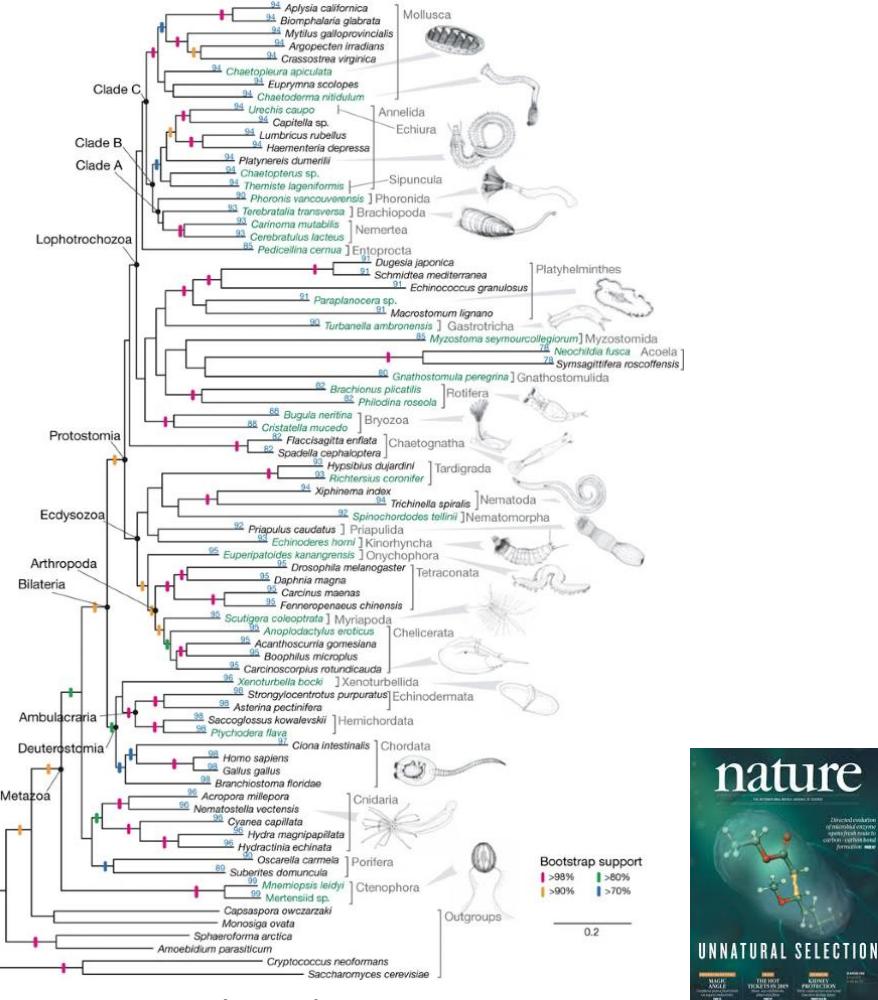
Dunn et al. 2008, Hejnol et al. 2009,  
Ryan et al. 2013, Borowiec et al. 2015,  
Whelan et al. 2015, Schultz et al. 2021

# Phylogenomics and comparative genomics



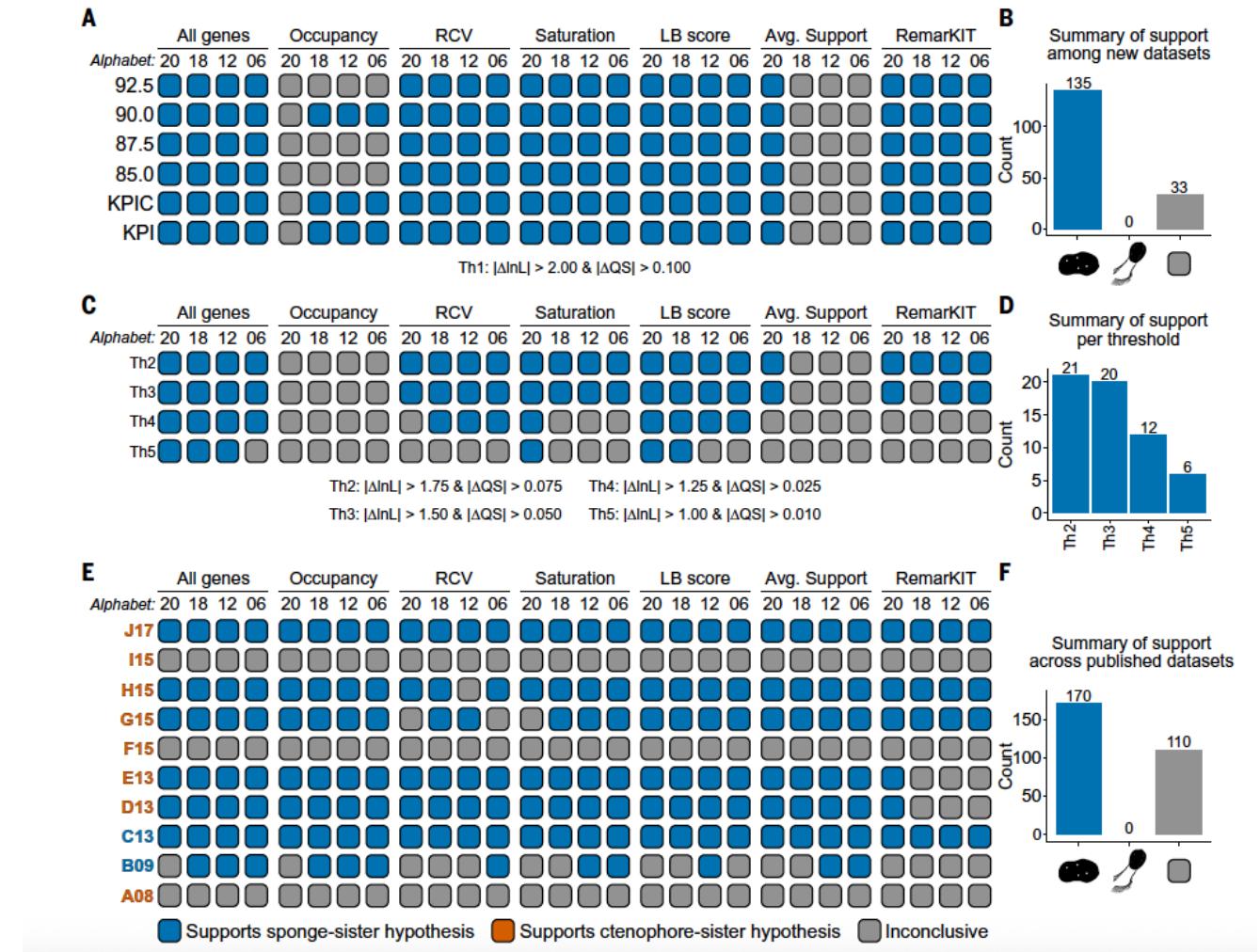
# Phylogenomics and comparative genomics

How we started...



Dunn et al. (2008). Nature

How we ended...



Stynweek & King, 2025 (Science)

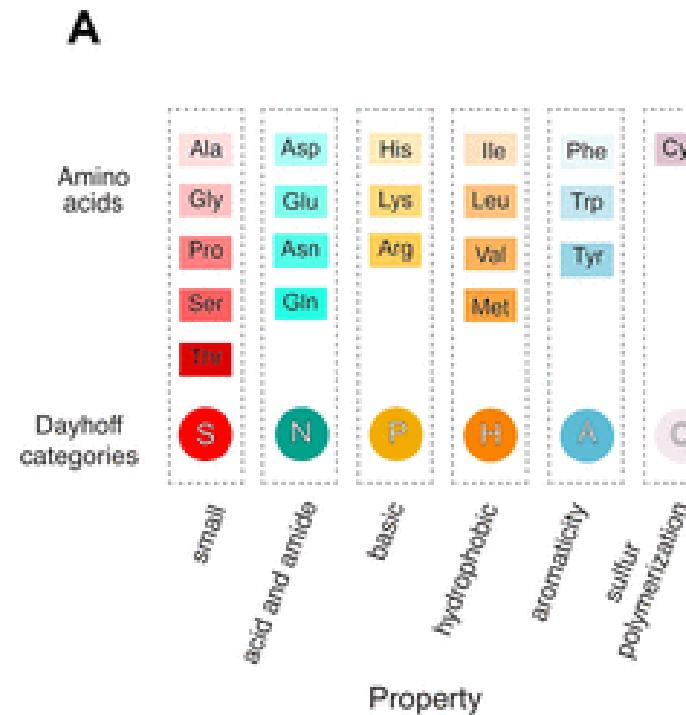
# How to do phylogenomics

We need to account for heterogeneity: and sometimes complex evolutionary models are not enough

## Matrix recoding

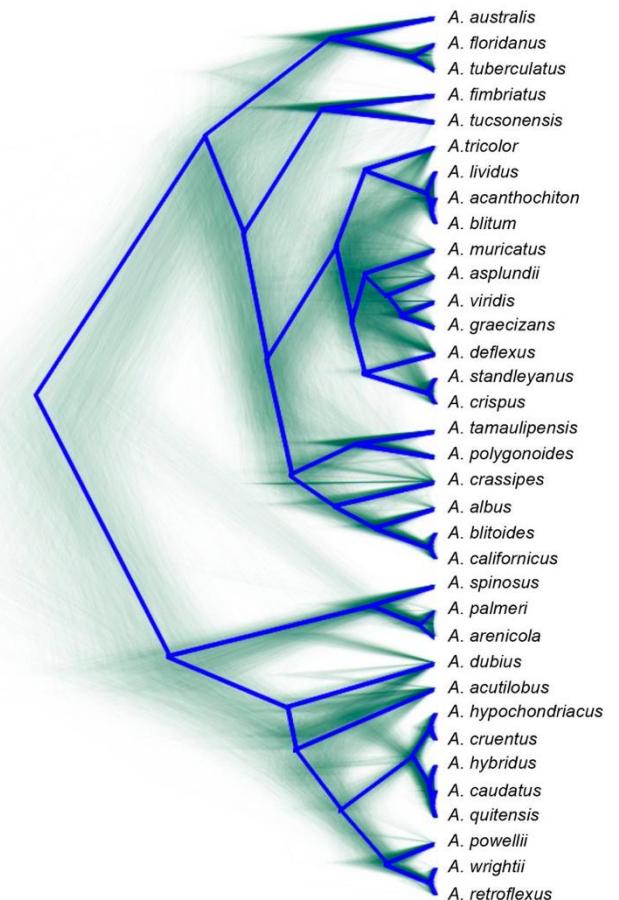
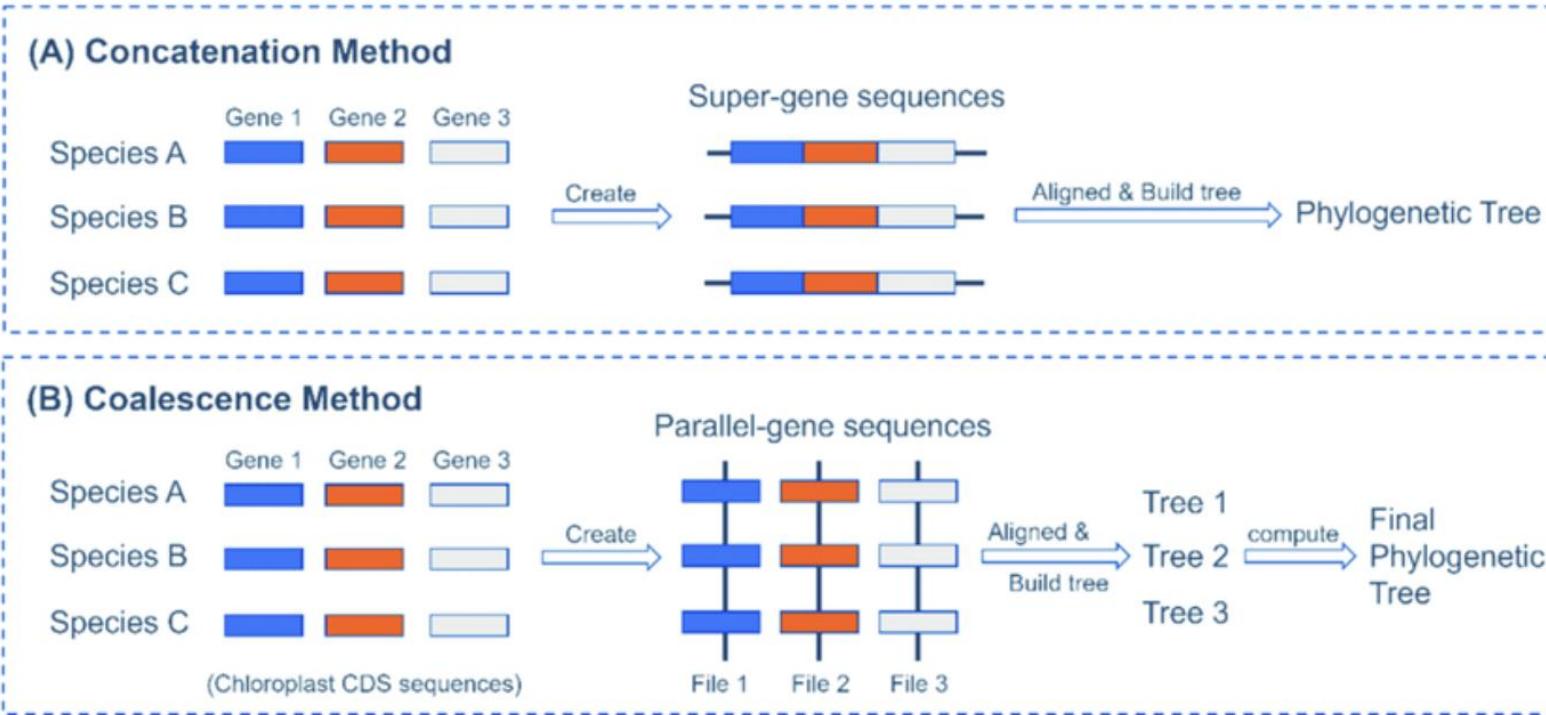
### Dayhoff-6

Or different numbers of categories for recoding:  
9, 12, 15, 18...



# How to do phylogenomics

## Phylogenetic inference



# How to do phylogenomics

## Concatenated approaches

Problems with Bayesian Inference

Painfully long computing times for large datasets

JOURNAL ARTICLE

## Practical Speedup of Bayesian Inference of Species Phylogenies by Restricting the Space of Gene Trees

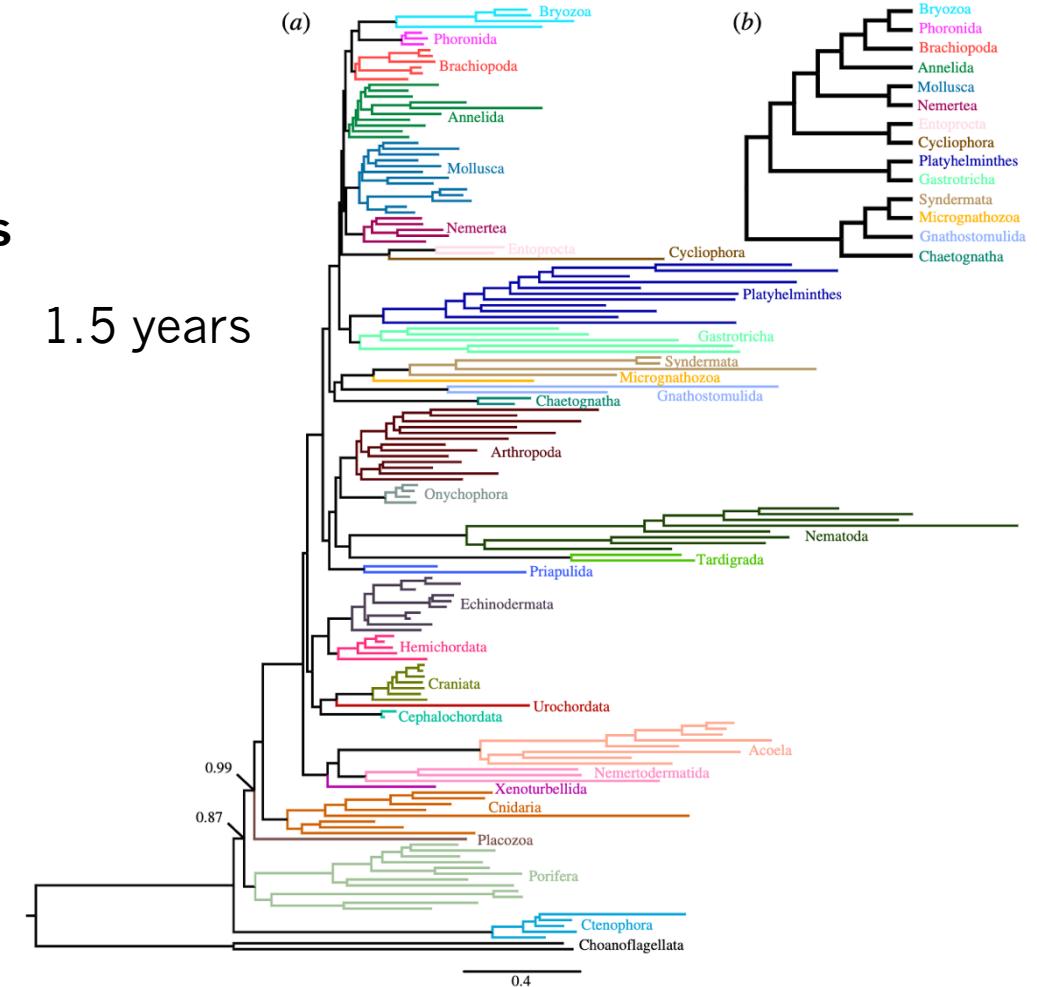


Yaxuan Wang, Huw A Ogilvie, Luay Nakhleh ✉

*Molecular Biology and Evolution*, Volume 37, Issue 6, June 2020, Pages 1809–1818,

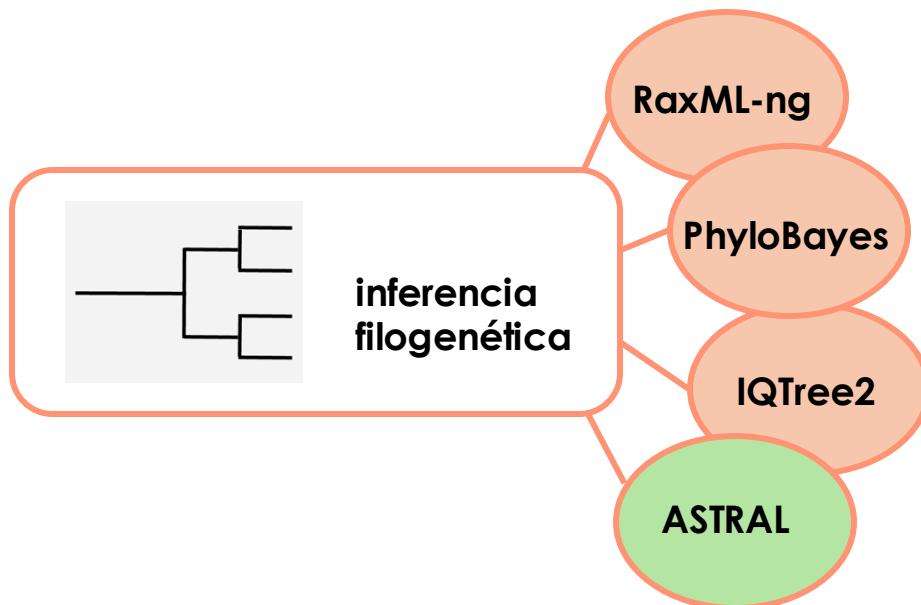
<https://doi.org/10.1093/molbev/msaa045>

Published: 20 February 2020

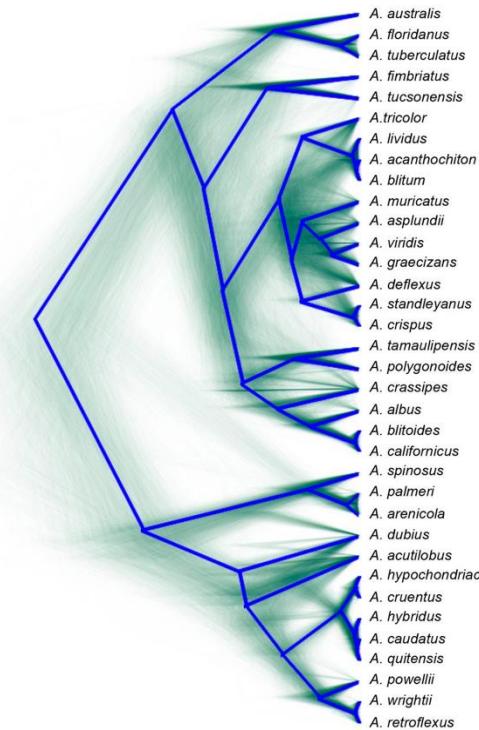


Laumer et al. (2019). Proc Roy Soc B

# How to do phylogenomics



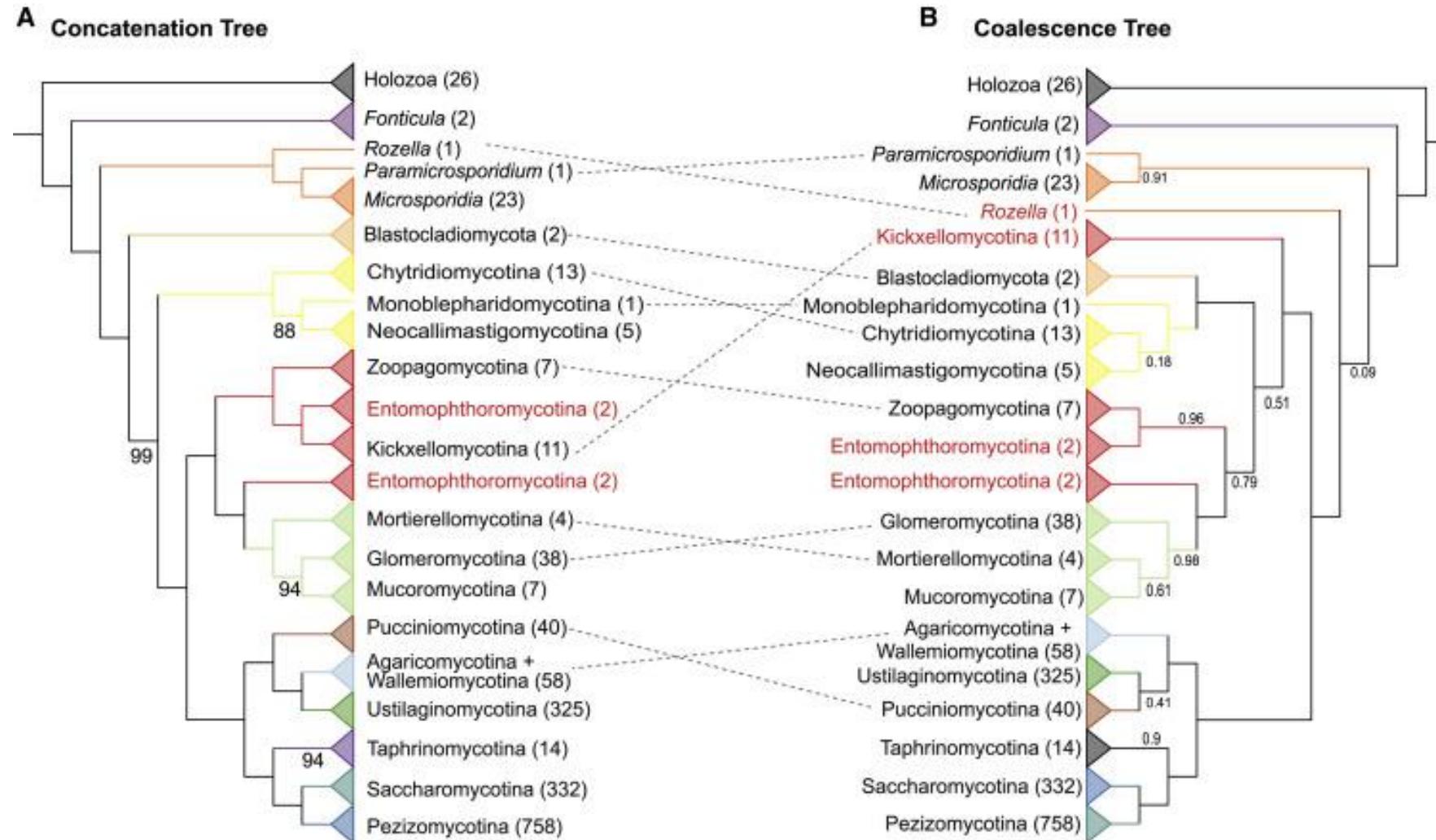
Concatenation-based analysis



Coalescence-based analysis

# How to do phylogenomics

Resolving the tree of Fungi with 1644 species and 290 genes



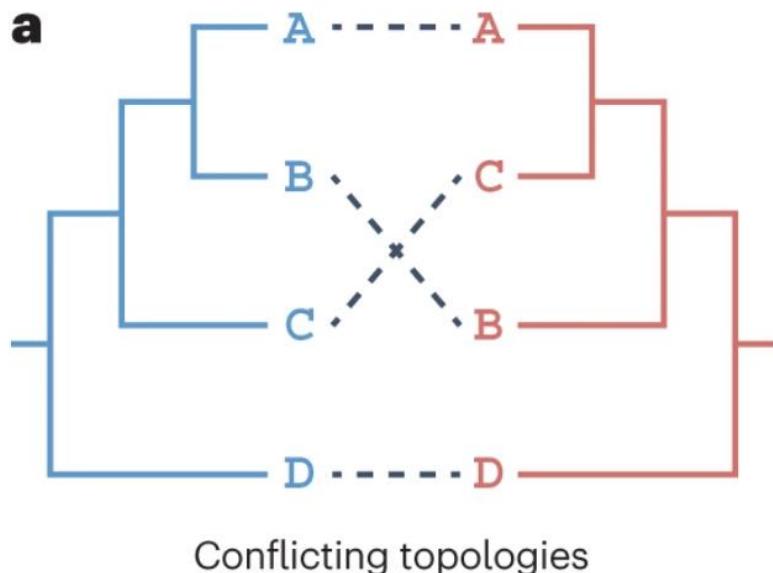
# How to do phylogenomics

## Phylogenomic inherent problems



This is probably your worst enemy

incongruence — the inference of conflicting evolutionary histories



### 1. Biological factors:

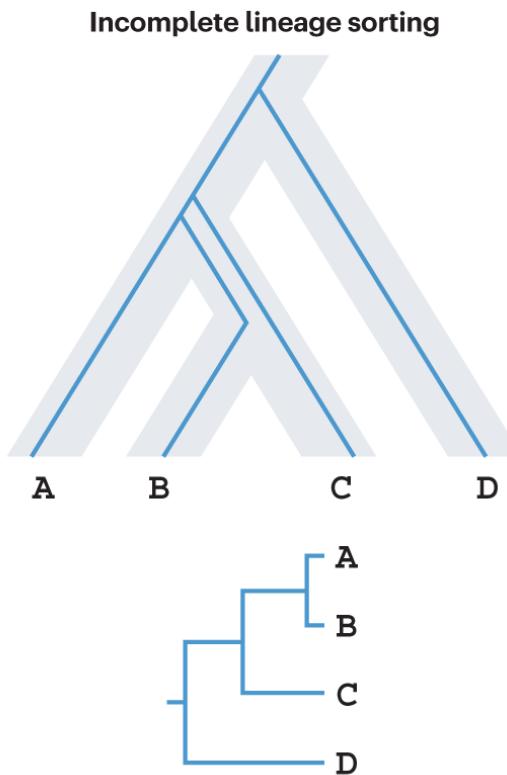
- incomplete lineage sorting
- horizontal gene transfer
- hybridization
- convergent molecular evolution

### 2. Analytical factors:

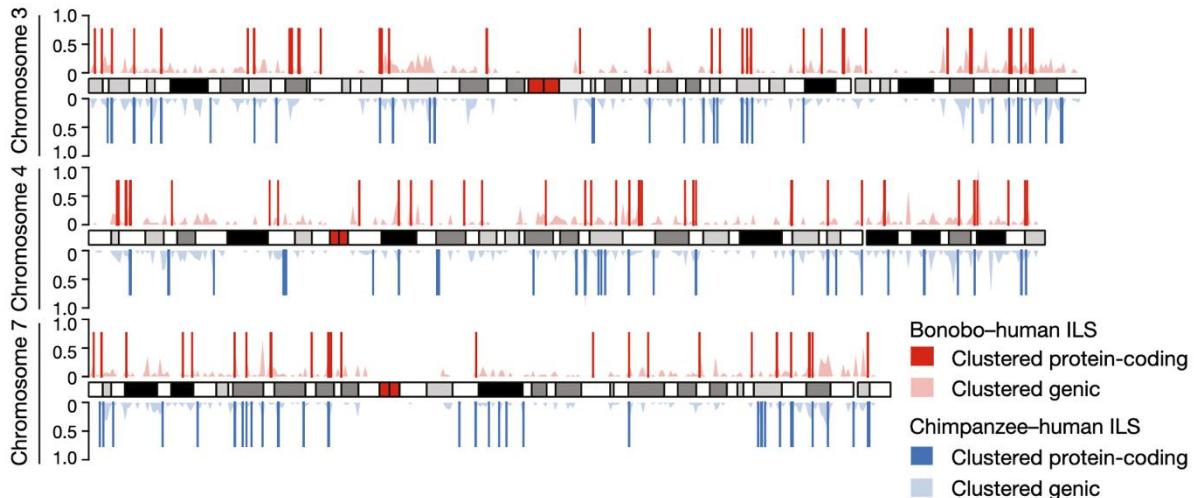
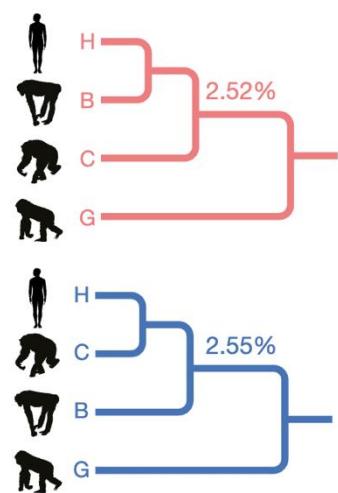
- stochastic errors
- hidden paralogy
- systematic and treatment errors

# How to do phylogenomics

## Incomplete lineage sorting (ILS) or retention of ancestral polymorphism

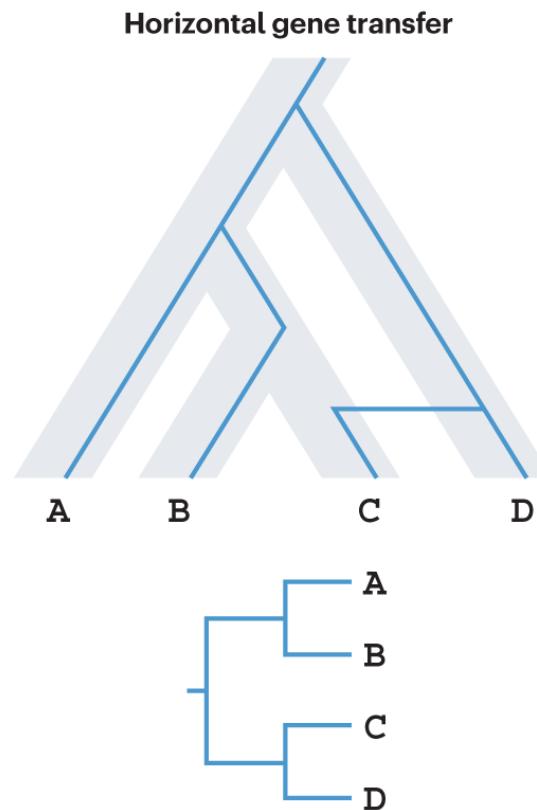


the average divergence time between genes may differ from the divergence time between species



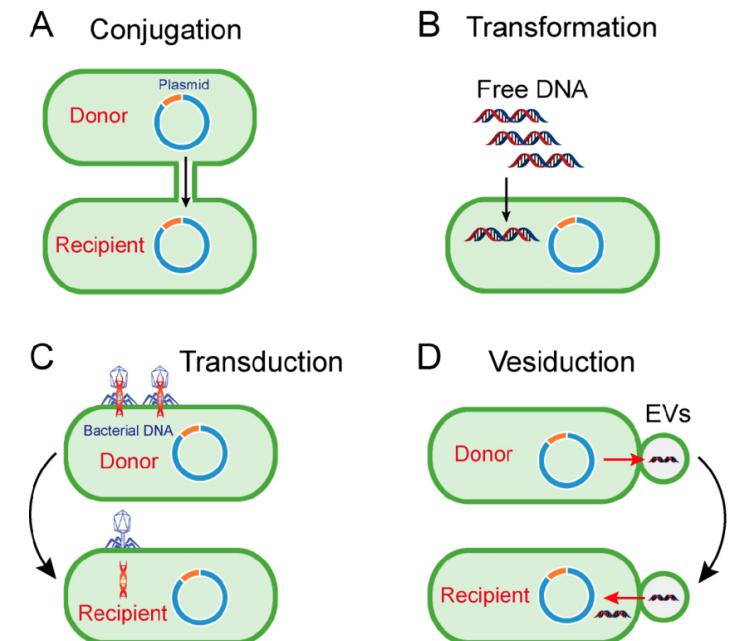
# How to do phylogenomics

## Horizontal gene transfer (HGT)



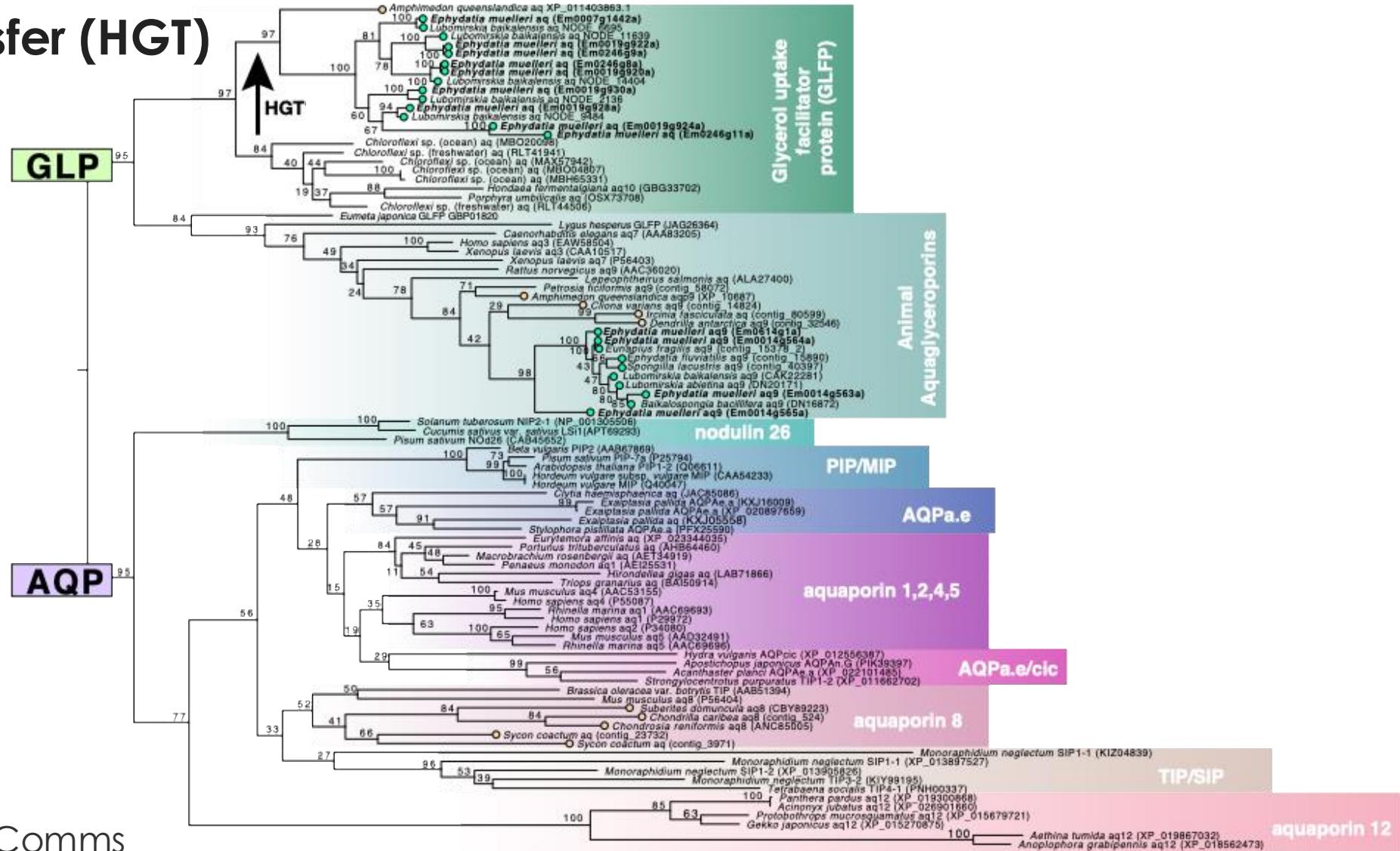
the non-sexual movement of genetic information between genomes

- Topology tests
- Sequence-based analysis
- GC-content
- Presence of introns
- SpeciesRax



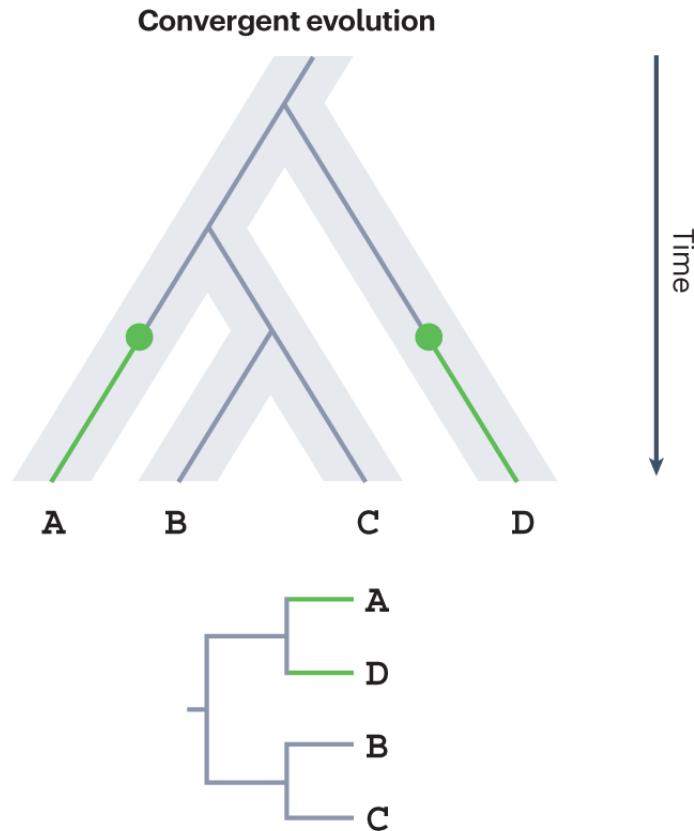
# How to do phylogenomics

## Horizontal gene transfer (HGT)

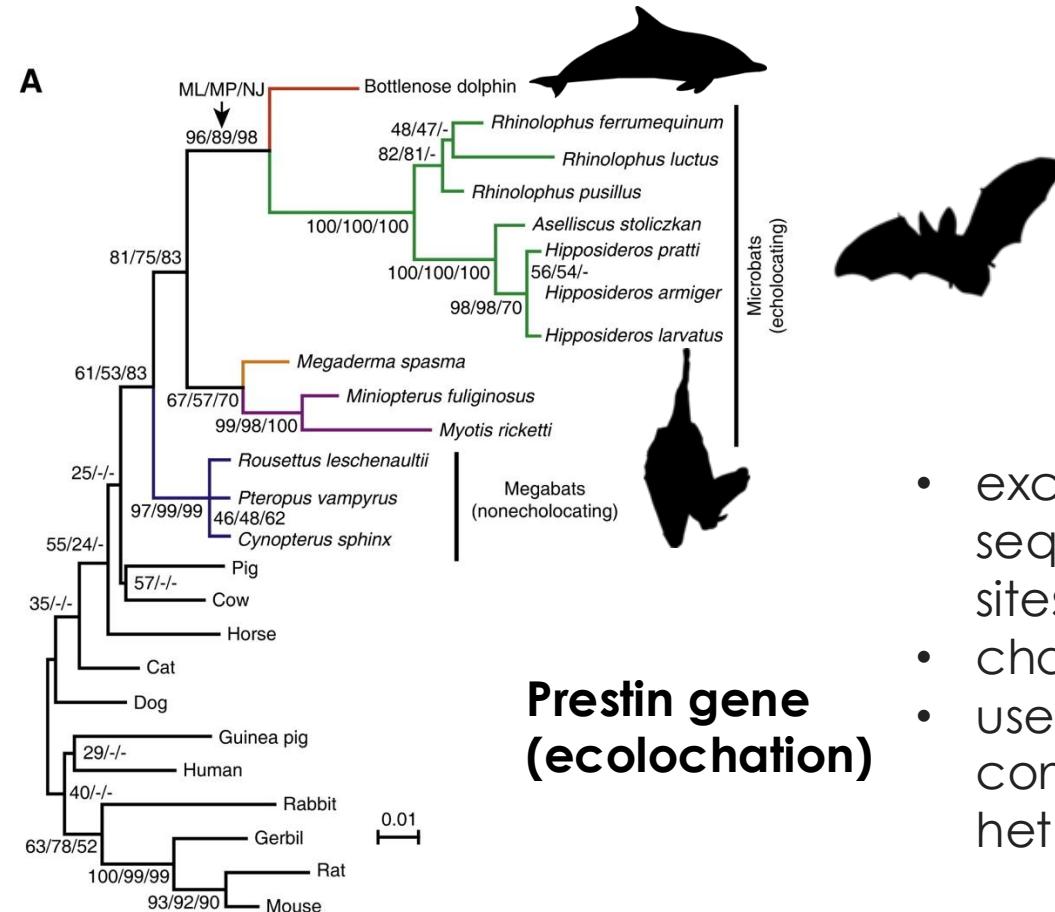


# How to do phylogenomics

## Convergent evolution



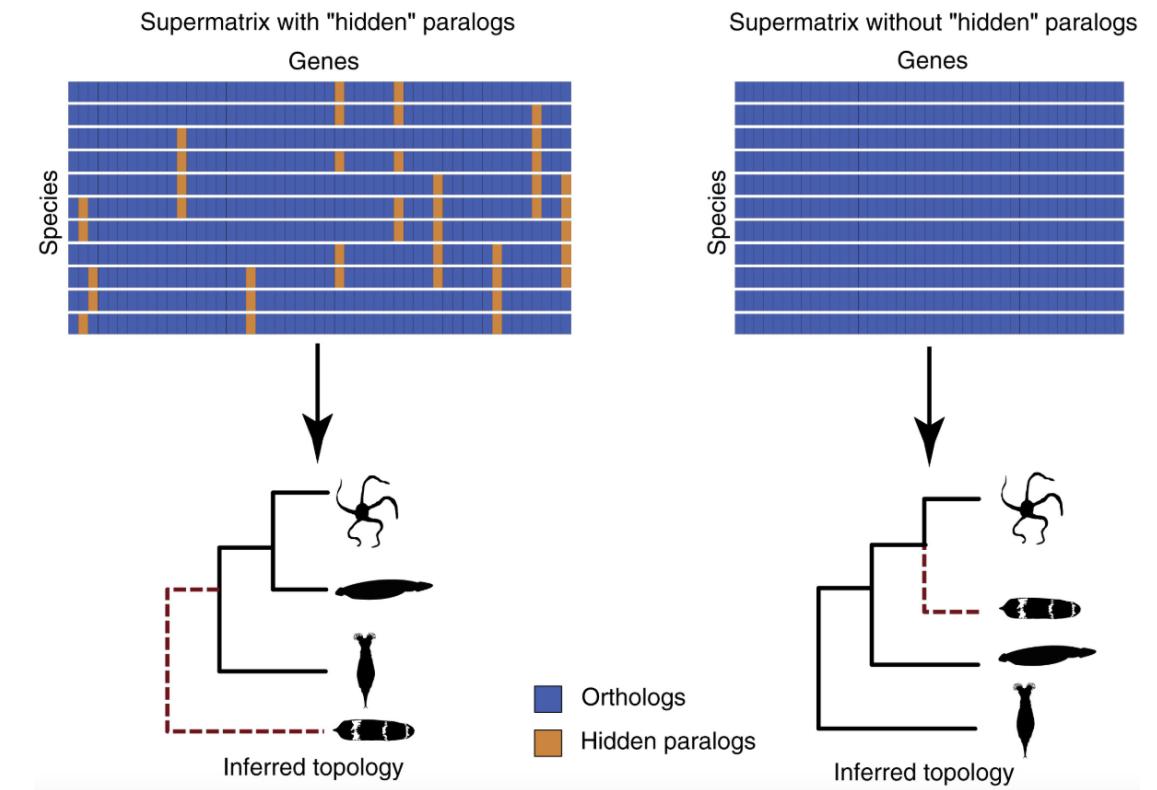
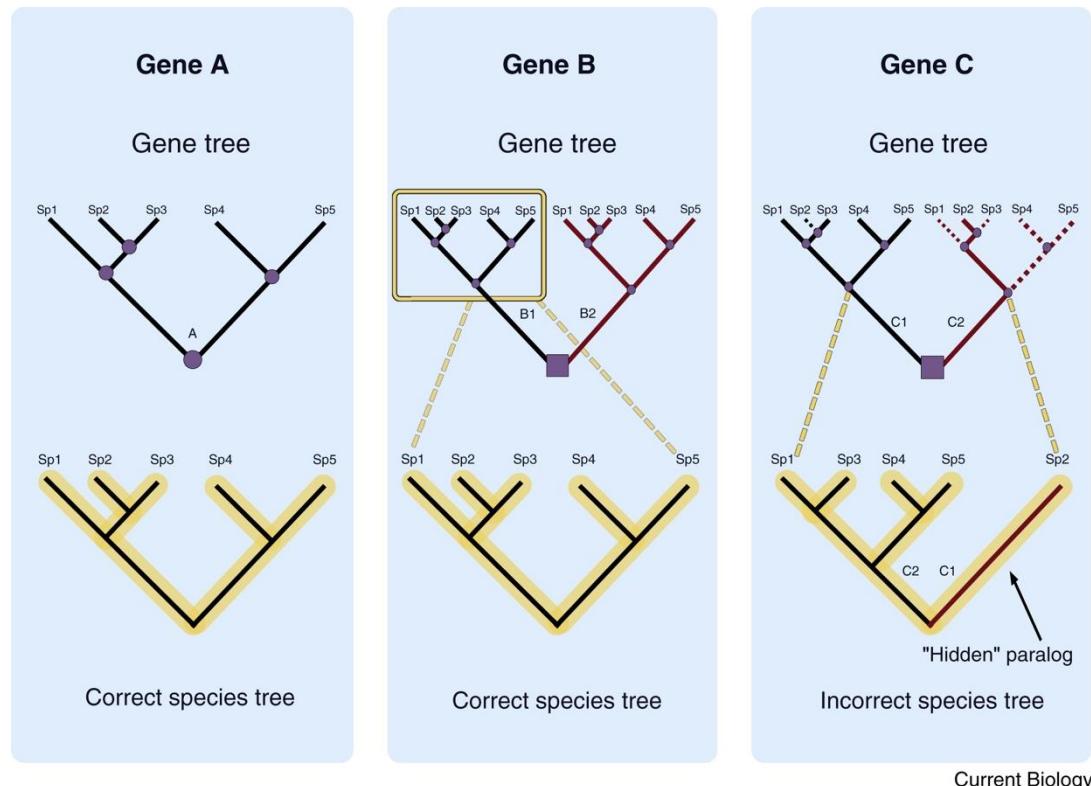
selection for the same or similar traits in distantly related taxa can result in convergent molecular evolution



- exclusion of sequences and sites
- character recoding
- use of models with compositional heterogeneity

# How to do phylogenomics

## Hidden paralogy



# How to do phylogenomics

## Visualizing trees



**Always  
remember  
your trees  
are just  
hypotheses**



# Phylogenomics and comparative genomics

**But what after we have our phylogenetic framework?**

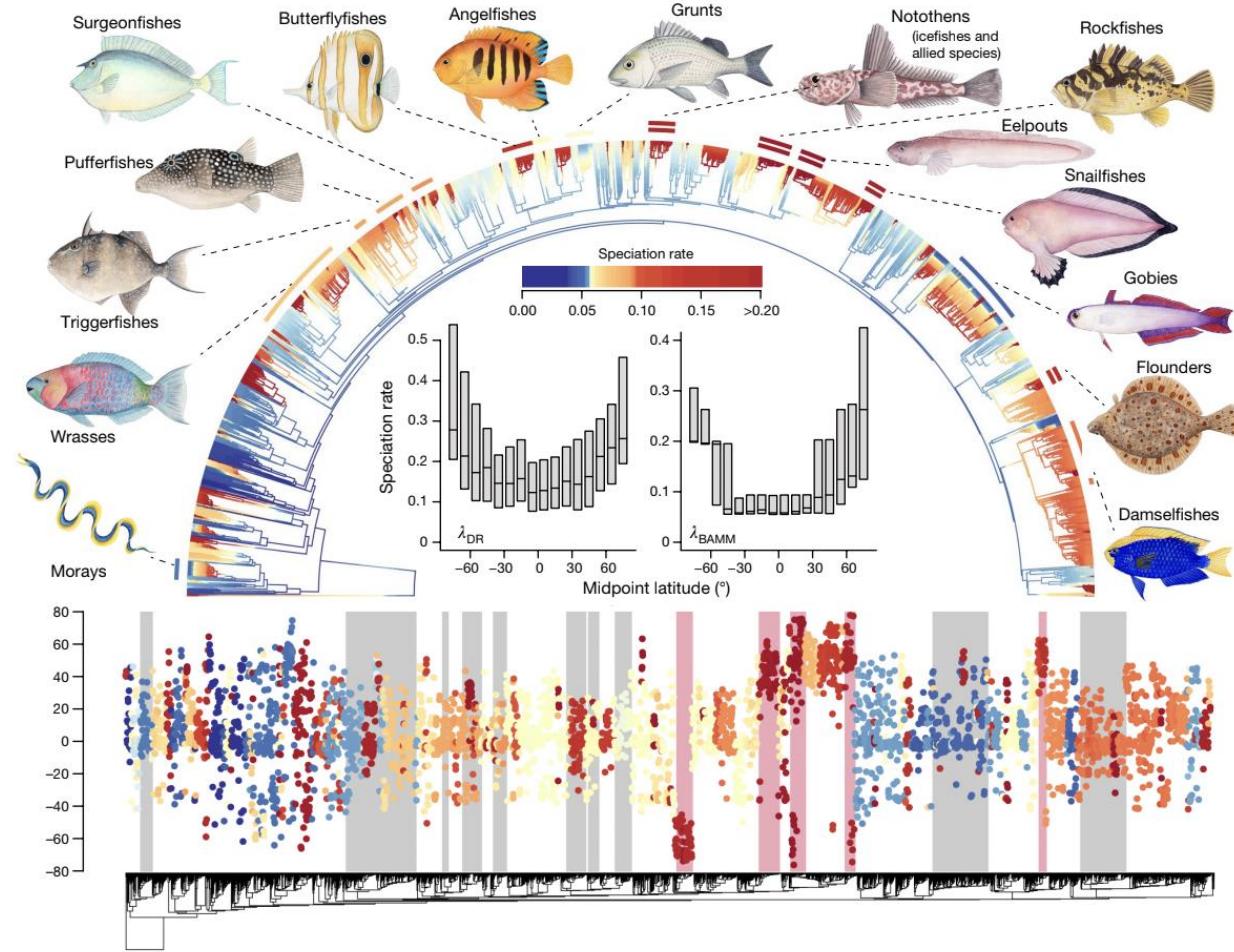
# Phylogenomics and comparative genomics

## Diversification rates

BAMM: <http://bammm-project.org/>



the rates at which new species form (the Speciation rate,  $\lambda$ ) and living species go extinct (the extinction rate,  $\mu$ ). Diversification rates can be estimated from fossils, data on the species diversity of clades and their ages, or phylogenetic trees.



Dan Rabosky

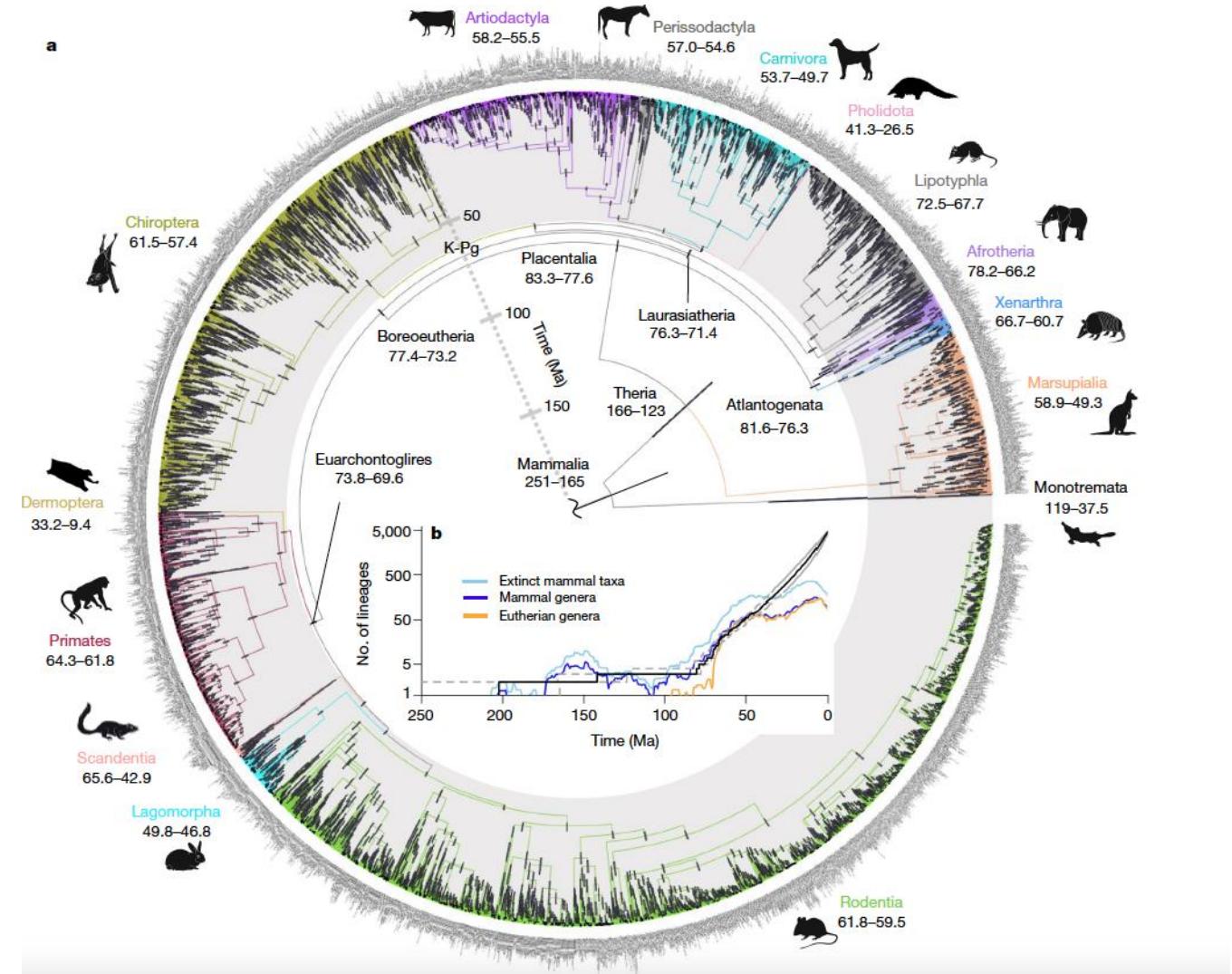
Rabosky et al. (2018). Nature

# Phylogenomics and comparative genomics

## Tree calibration

How?

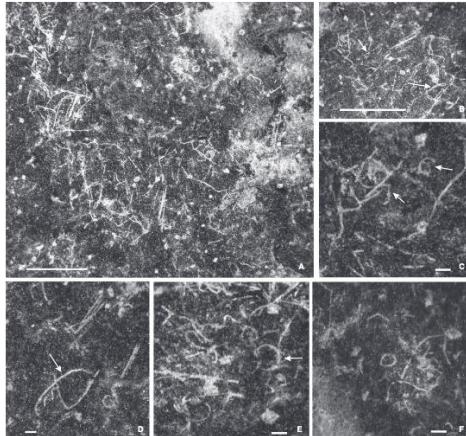
1. geological events
2. estimates from independent molecular dating studies
3. the fossil record
4. Mutation rates



# Phylogenomics and comparative genomics

## Tree calibration

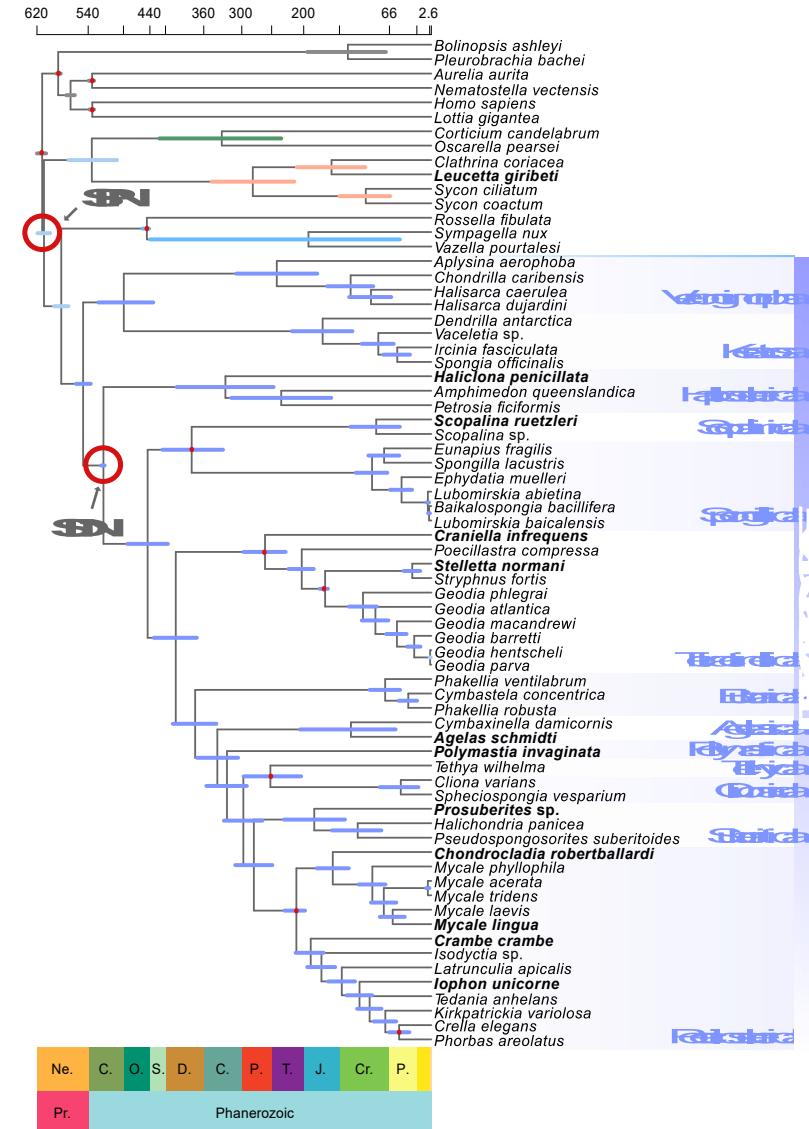
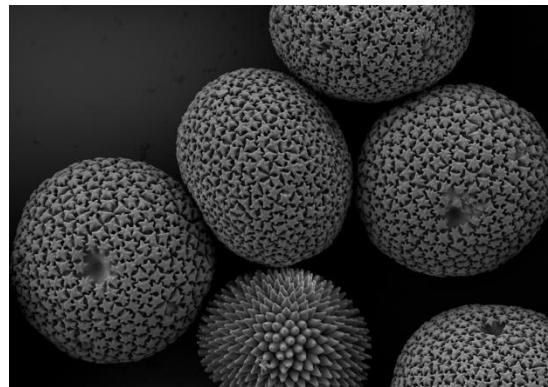
## Limitations



## Adding fossils

Is it a **crown group**?

Or a **stem group**?

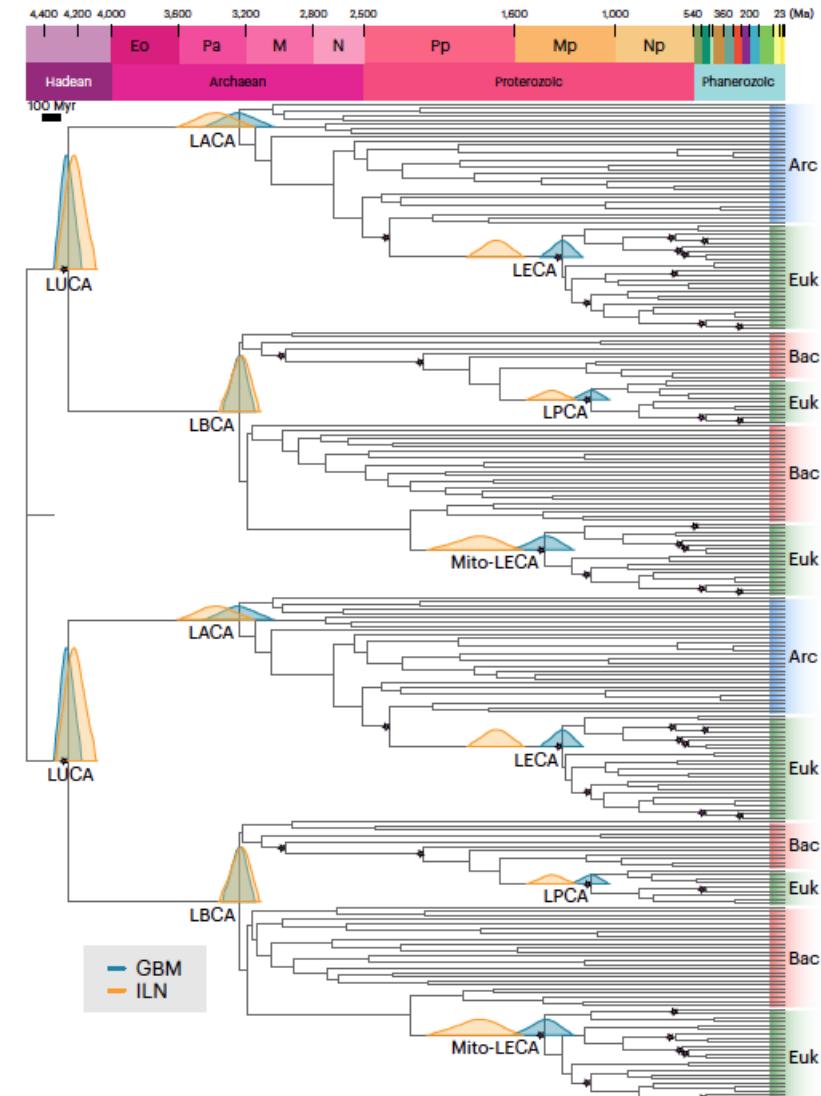


# Phylogenomics and comparative genomics

## Tree calibration

And if I do not have fossils?

- Geological events
- Late Heavy bombardment
- Biomarkers



# Phylogenomics and comparative genomics

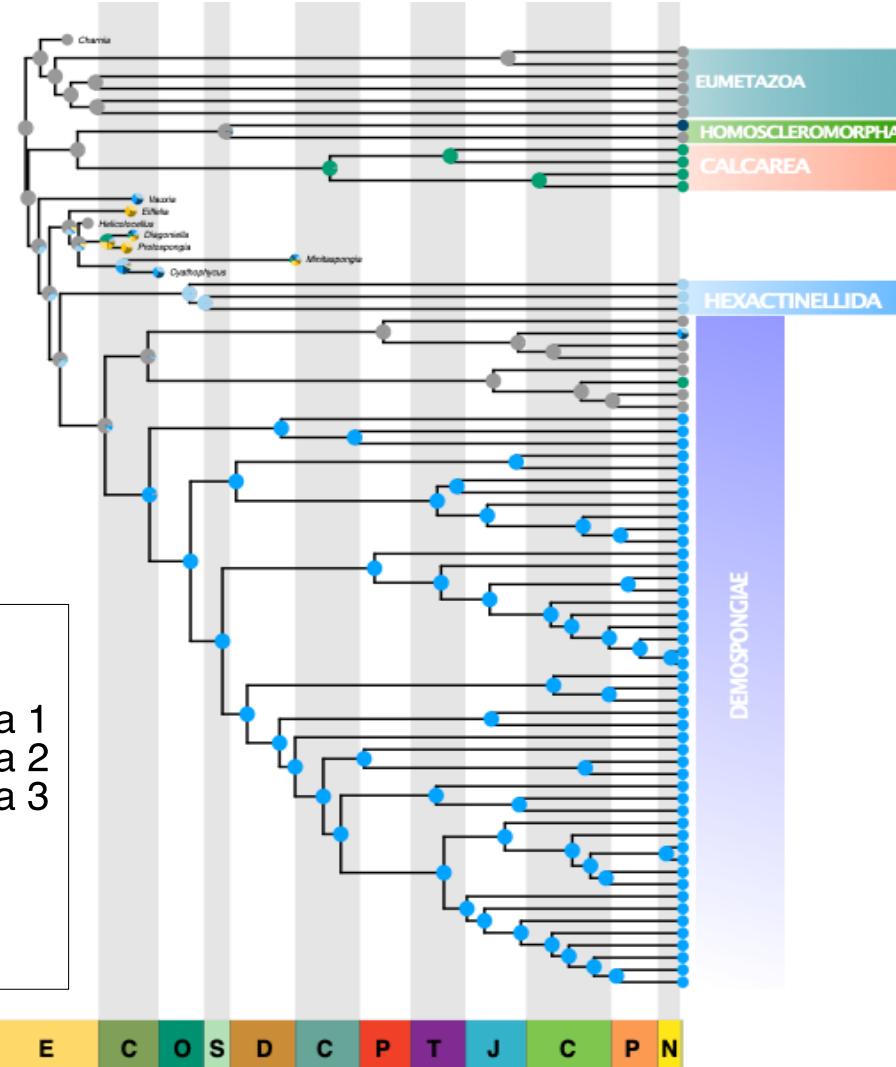
## Ancestral character reconstruction

Combination of information about  
**evolutionary relationships from a phylogenetic tree with the observed state of a character** for each terminal

### Derived vs. Ancestral trait

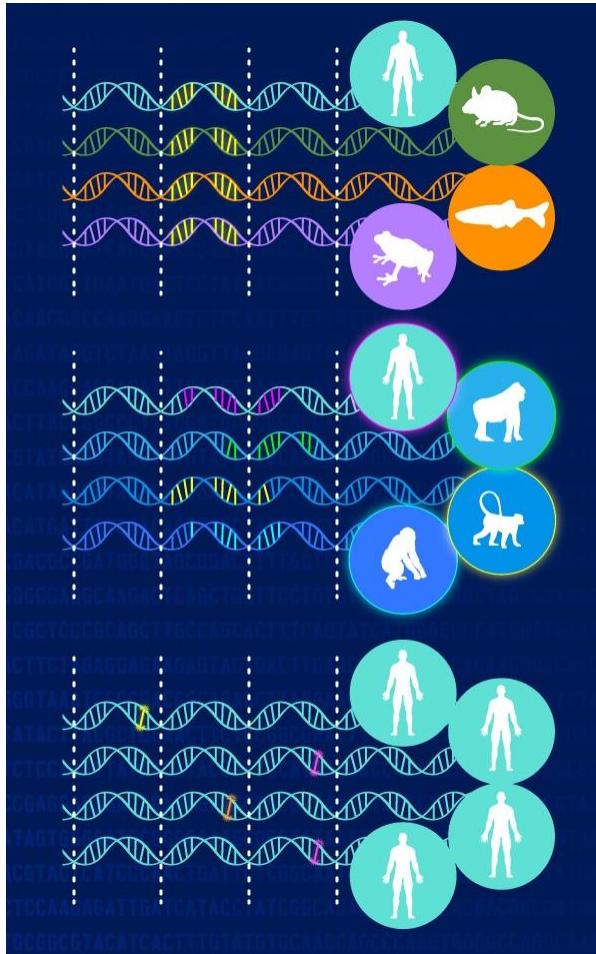


- Spicules absent
- Organic spicules
- Calcite absent silica 1
- Calcite absent silica 2
- Calcite absent silica 3
- Calcite only
- Biminerallitic silica 1
- Biminerallitic silica 2
- Biminerallitic silica 3



# Phylogenomics and comparative genomics

**Comparative genomics:** "is a field of biological research in which researchers use a variety of tools to compare the genome/transcriptome sequences of different species to make sense of evolution and adaptation"



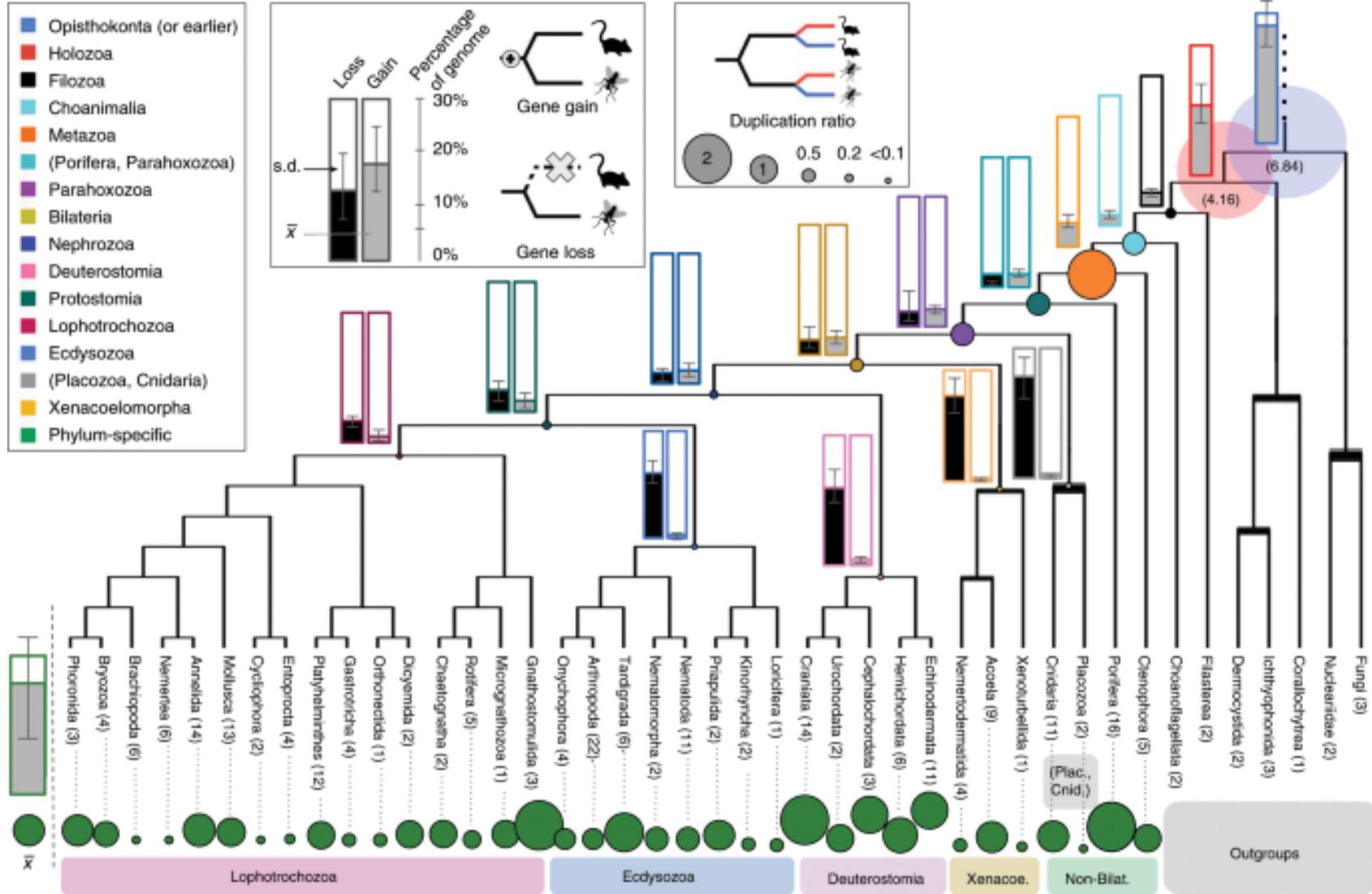
To understand **common features** among distant organisms coded on the DNA and **conserved across evolution**

To assess what is shared and what is novel to understand how **evolutionary innovations** appear

To understand how intraspecific genomic variation is linked to disease and **adaptation**

# Phylogenomics and comparative genomics

## Genome content comparisons



**Gene content:** similar gene complements across lineages

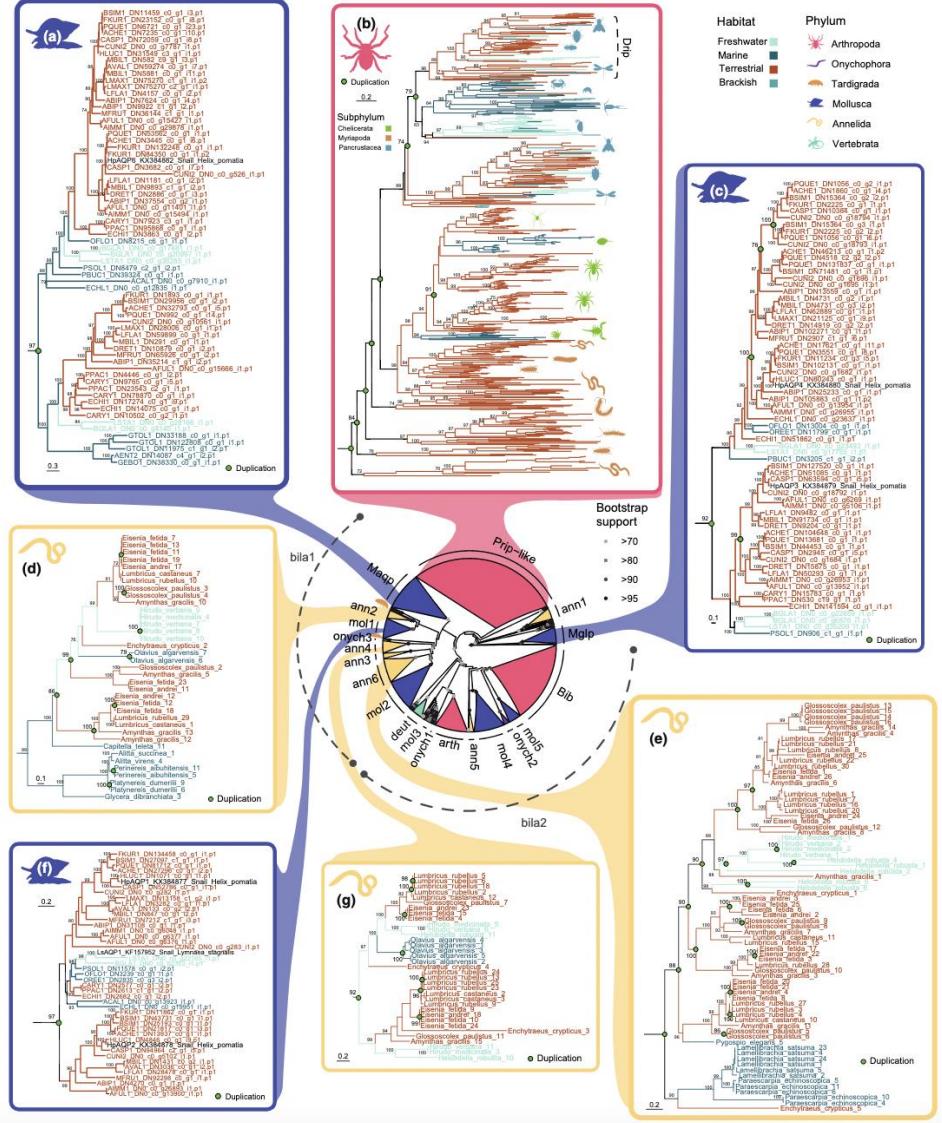
Fernández & Gabaldón (2020).  
Nature Eco Evo

# Phylogenomics and comparative genomics

## Genome content comparisons

**Gene evolution:** evolutionary trajectories of genes, expansions, losses, divergences

**Phylostratigraphy:** the process of determining the phylogenetic origin of every gene in a genome

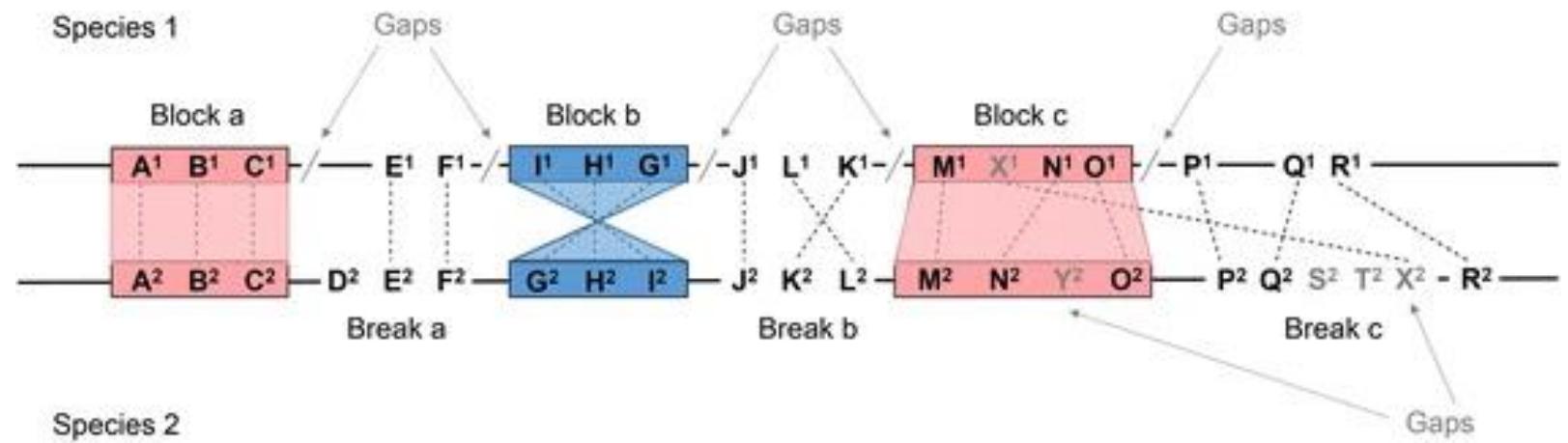


# Phylogenomics and comparative genomics

## Genome content comparisons

### Synteny

The term synteny was introduced in 1971 by John H. Renwick as a definition of **two (or more) genes residing on the same chromosome** of a given species even if genetic linkage cannot be demonstrated or tested for

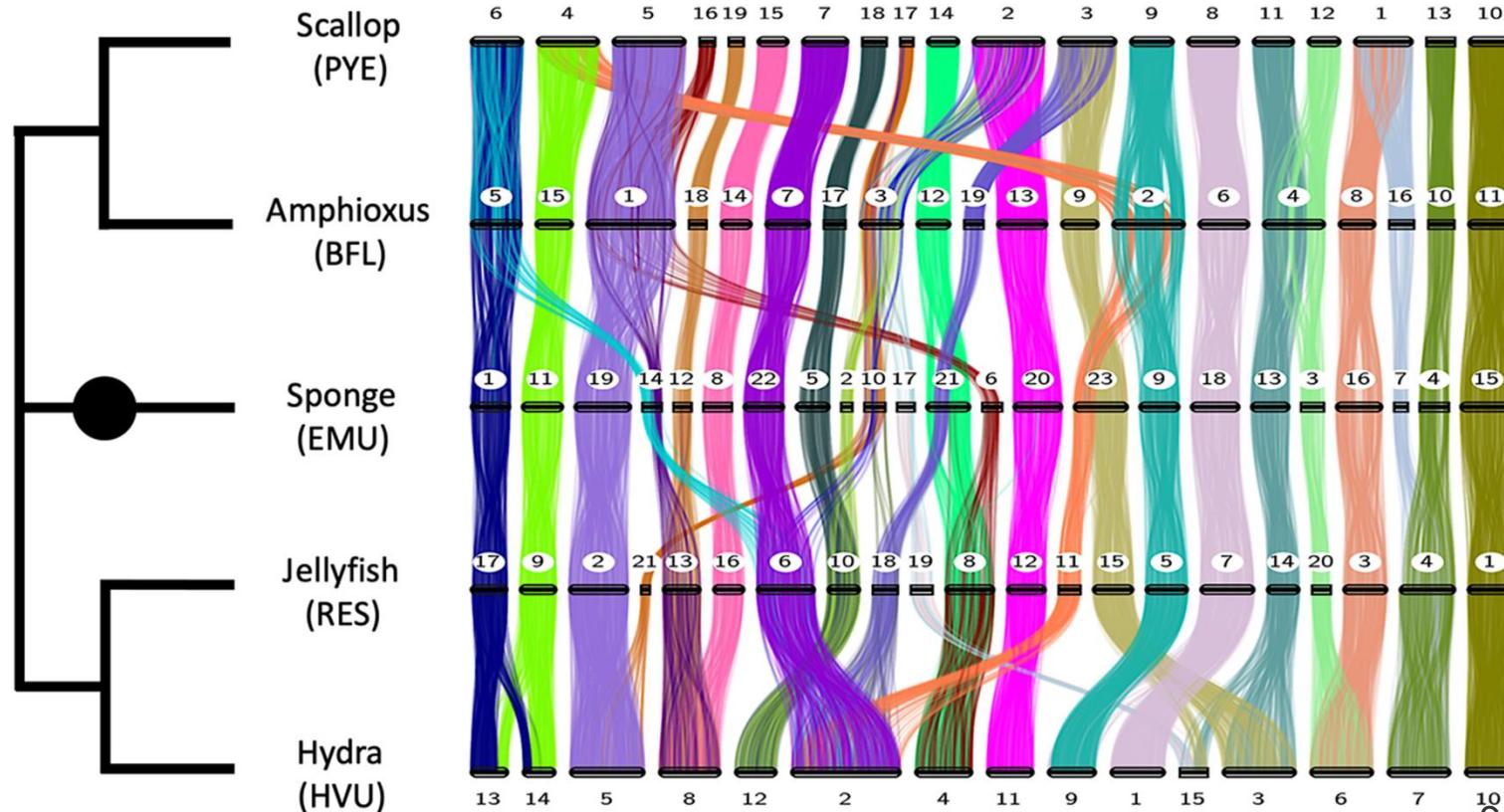


**Blocks = Ancient Linkage Groups  
macrosynteny patterns**

# Phylogenomics and comparative genomics

## Genome content comparisons

Synteny across distant Metazoa: surprisingly conserved



29 groups of genes (colours)  
Metazoan ALGs

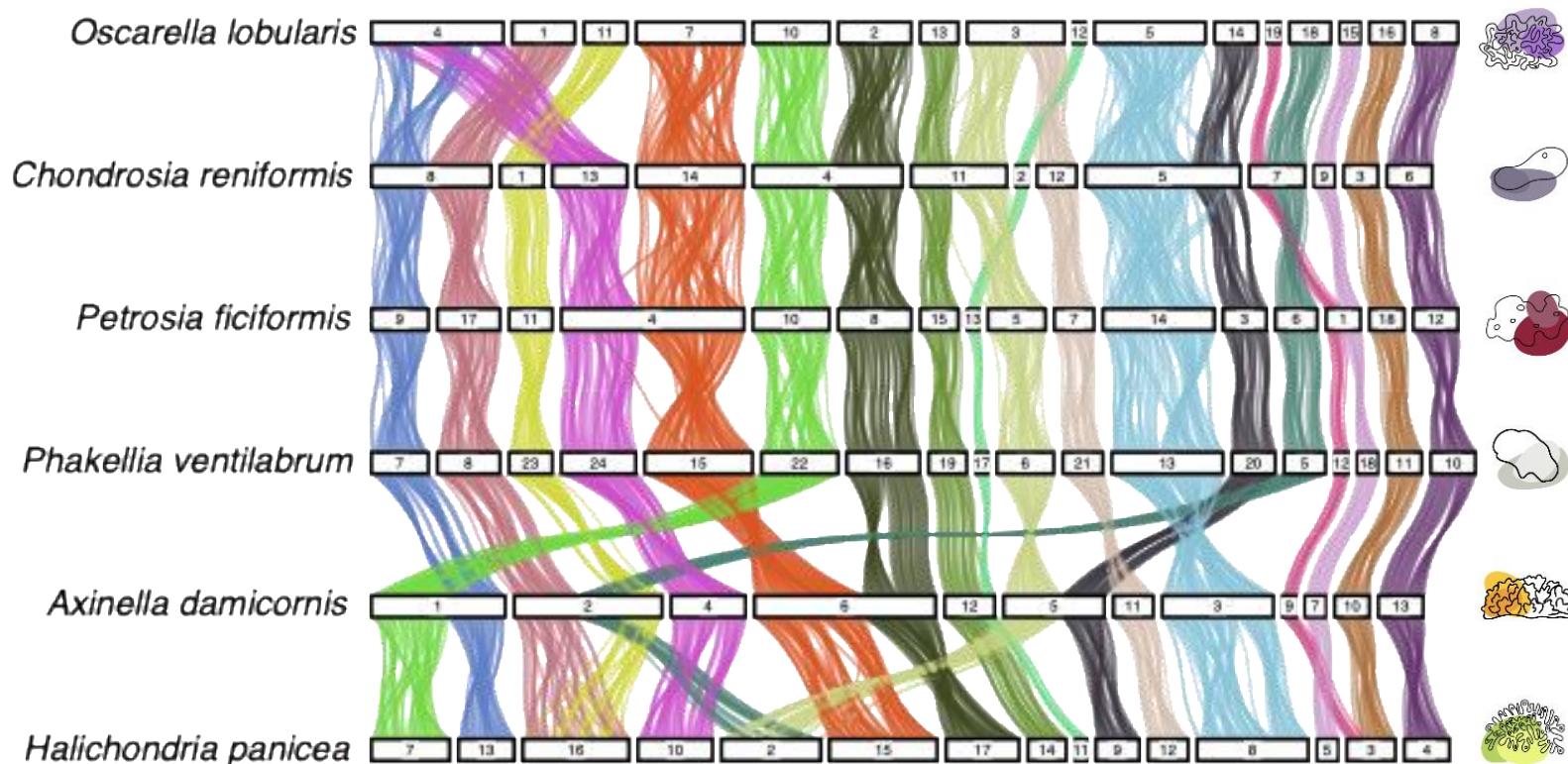
More than 2000 orthologs

# Phylogenomics and comparative genomics

## Genome content comparisons

Synteny across more related groups even more striking

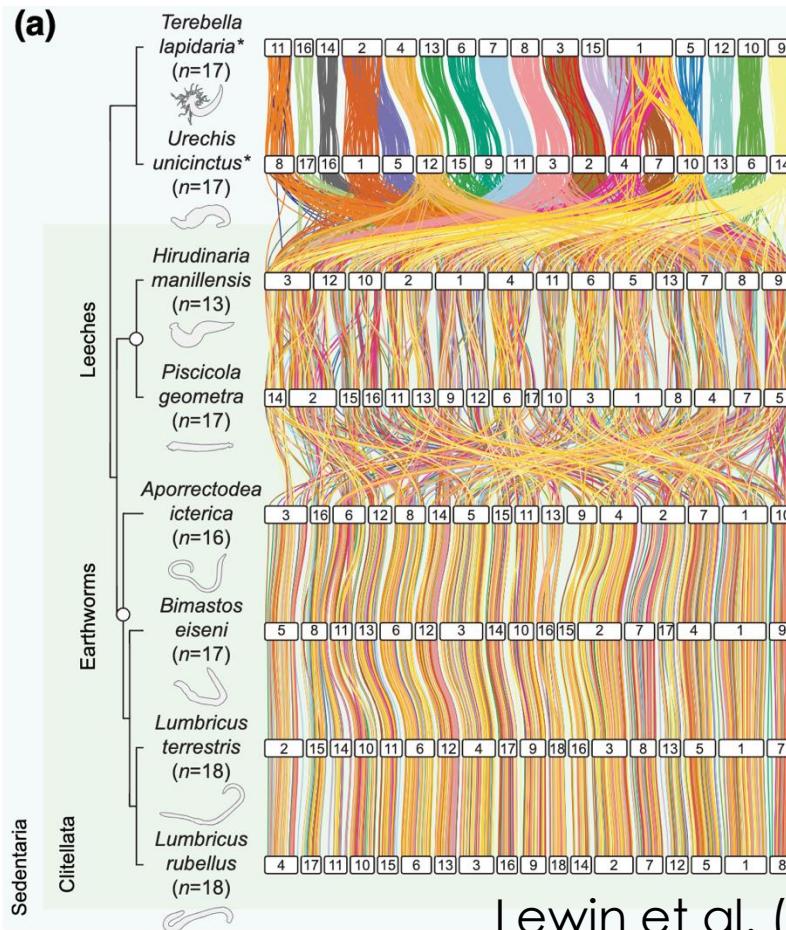
### Porifera



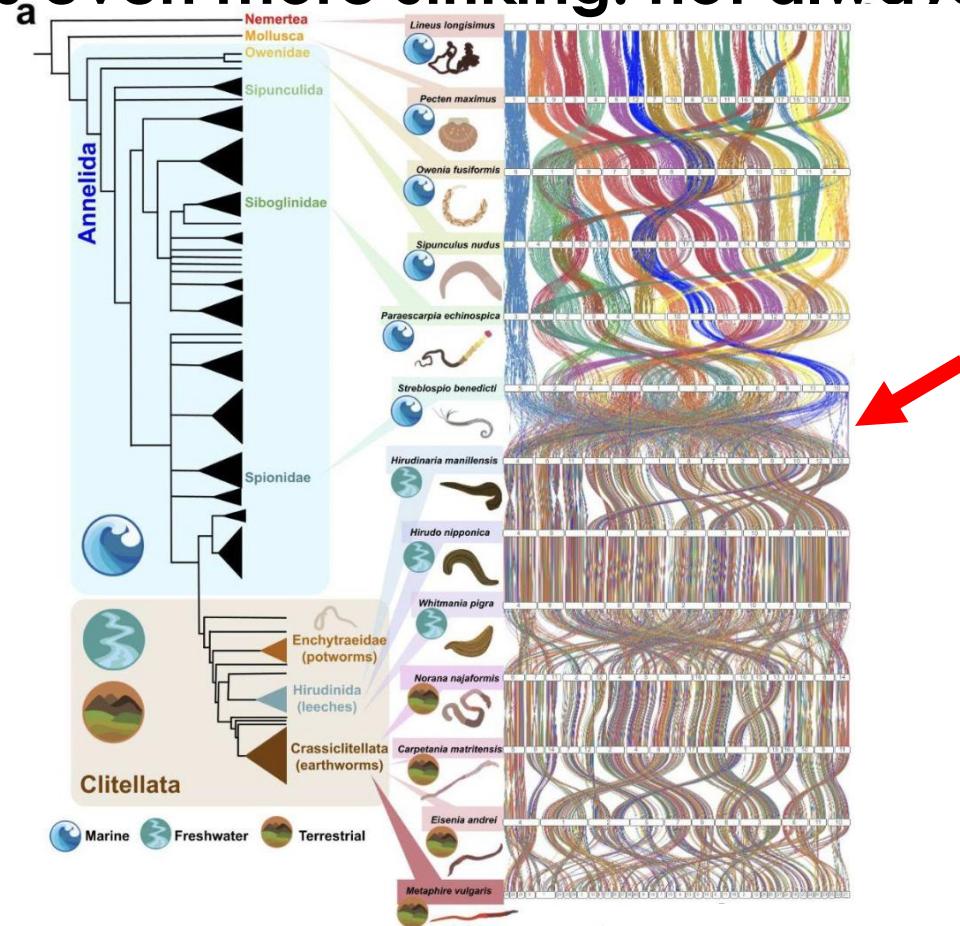
# Phylogenomics and comparative genomics

## Genome content comparisons

Synteny across more related groups <sup>a</sup> even more striking: not always!



Lewin et al. (2024). MBE

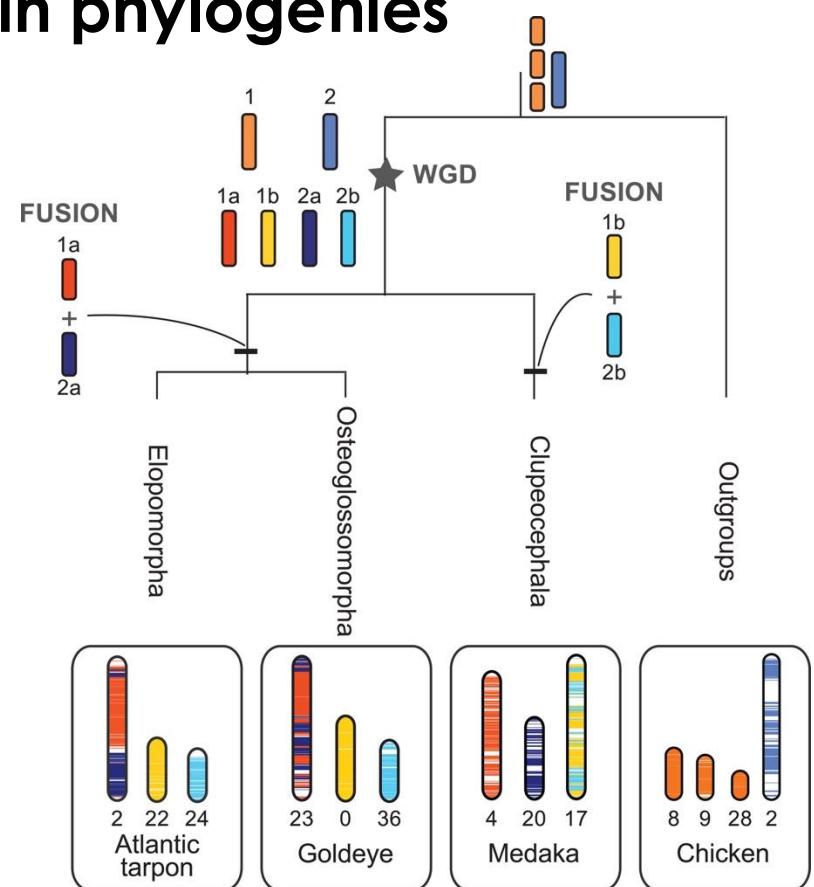
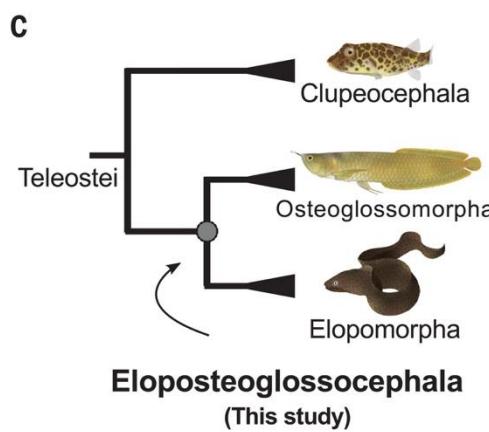
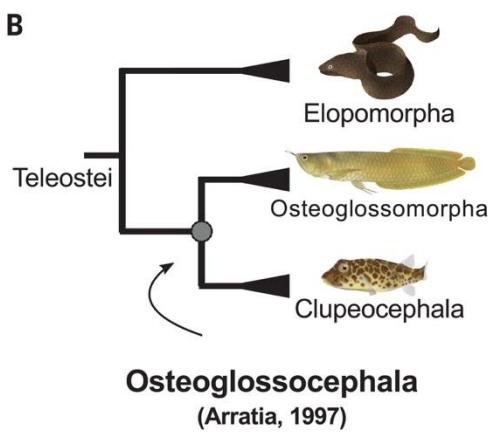
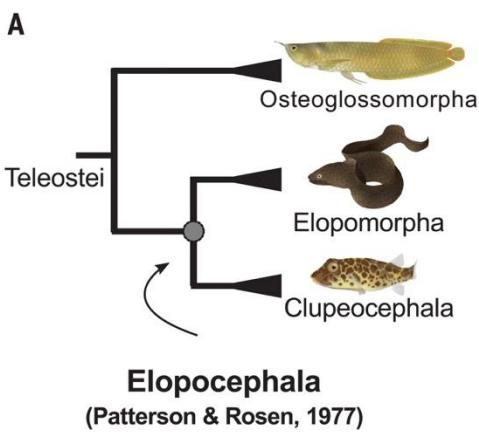


Vargas-Chávez et al. (2024). bioarXiv

# Phylogenomics and comparative genomics

## Genome content comparisons

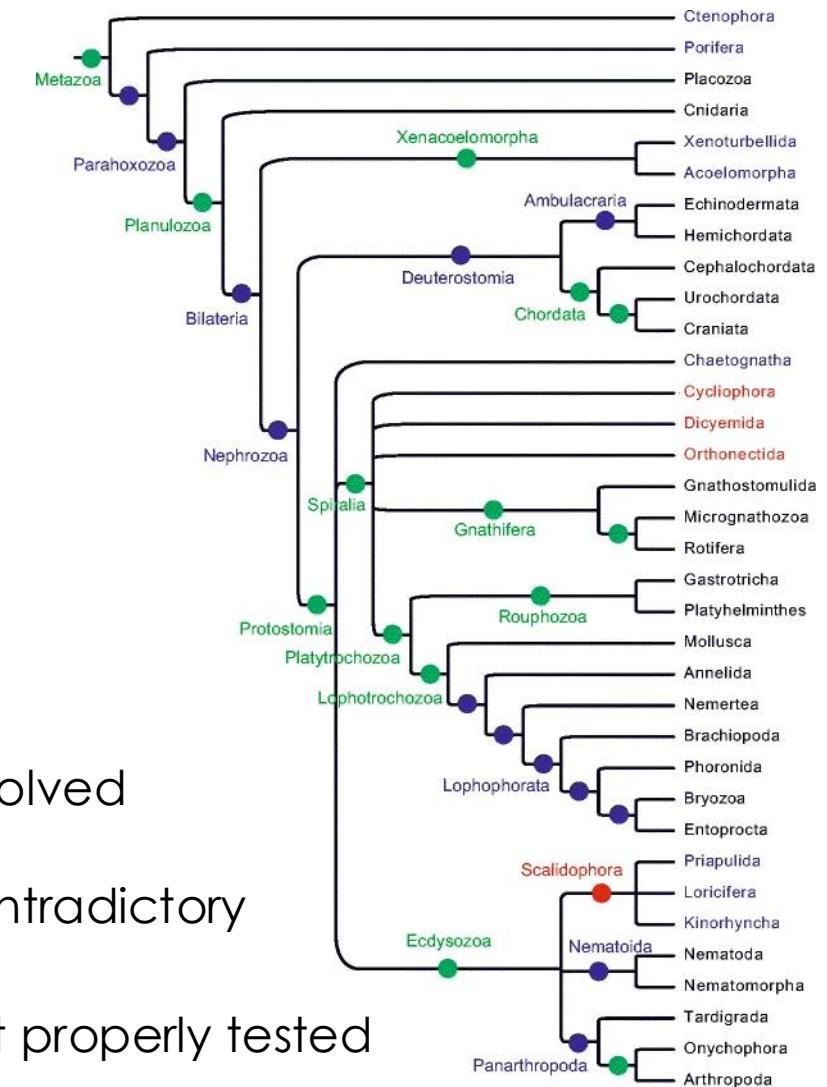
Synteny can help resolve recalcitrant nodes in phylogenies



# Phylogenomics and comparative genomics

What is the **next frontier** for phylogenomics?

1. Resolving **recalcitrant nodes**
2. Improve **taxon sampling** (rare and small animals)
3. Improve **orthology assignment**
4. Focus on **algorithmic developments**
5. Solve efficient **data storage**
6. Figure out better ways to integrate information from genomes, development, morphology, ecology, etc.



- resolved
- contradictory
- not properly tested