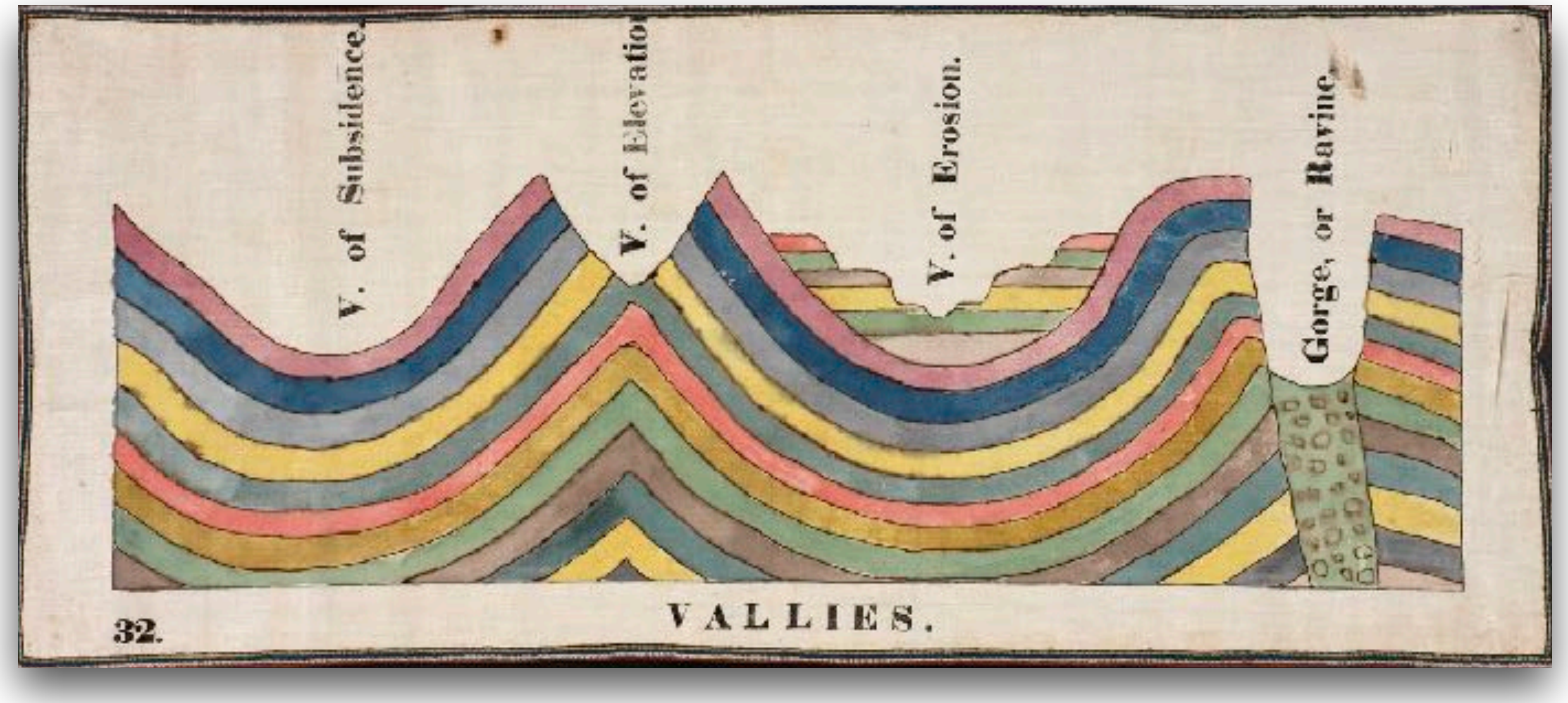
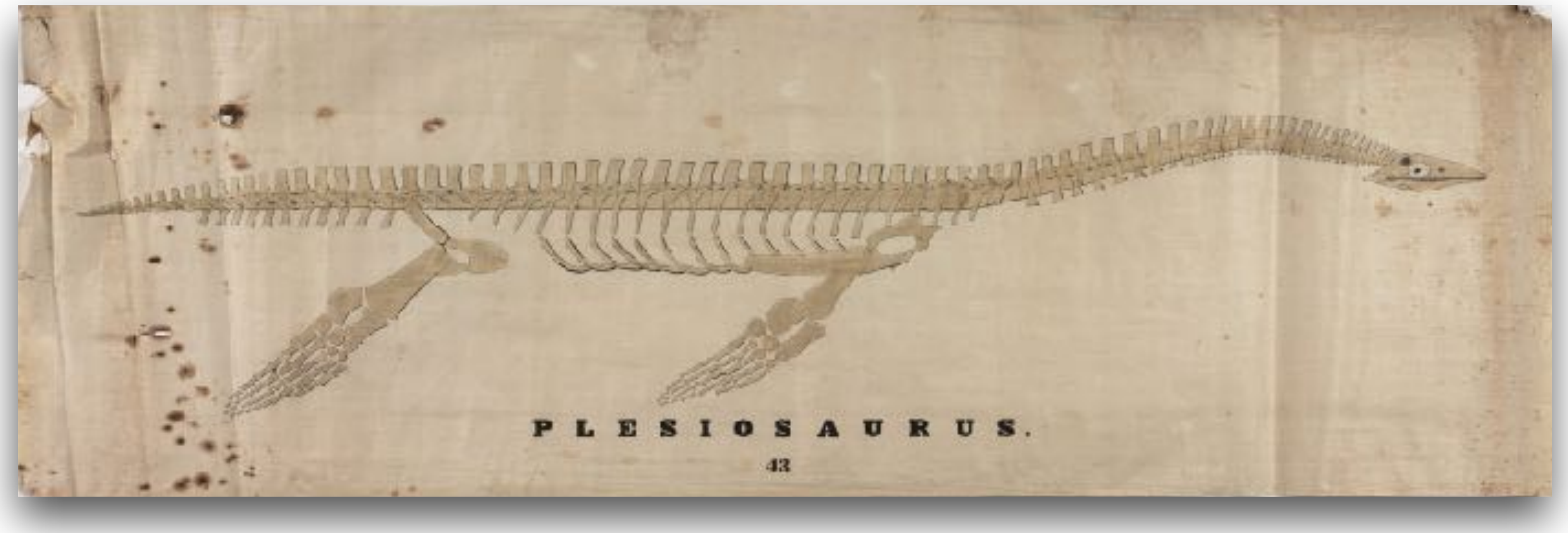




Phylogenetics 101

part A - tree thinking


3
1
2

The diversity of living beeings

- All life shares a common genetic history.
- Phylogeny is the framework for studying its diversity.

The Tree of Life

- A branching diagram that illustrates relationships of biological entities.
- A fundamental tool to investigate evolutionary processes.

Why it matters?

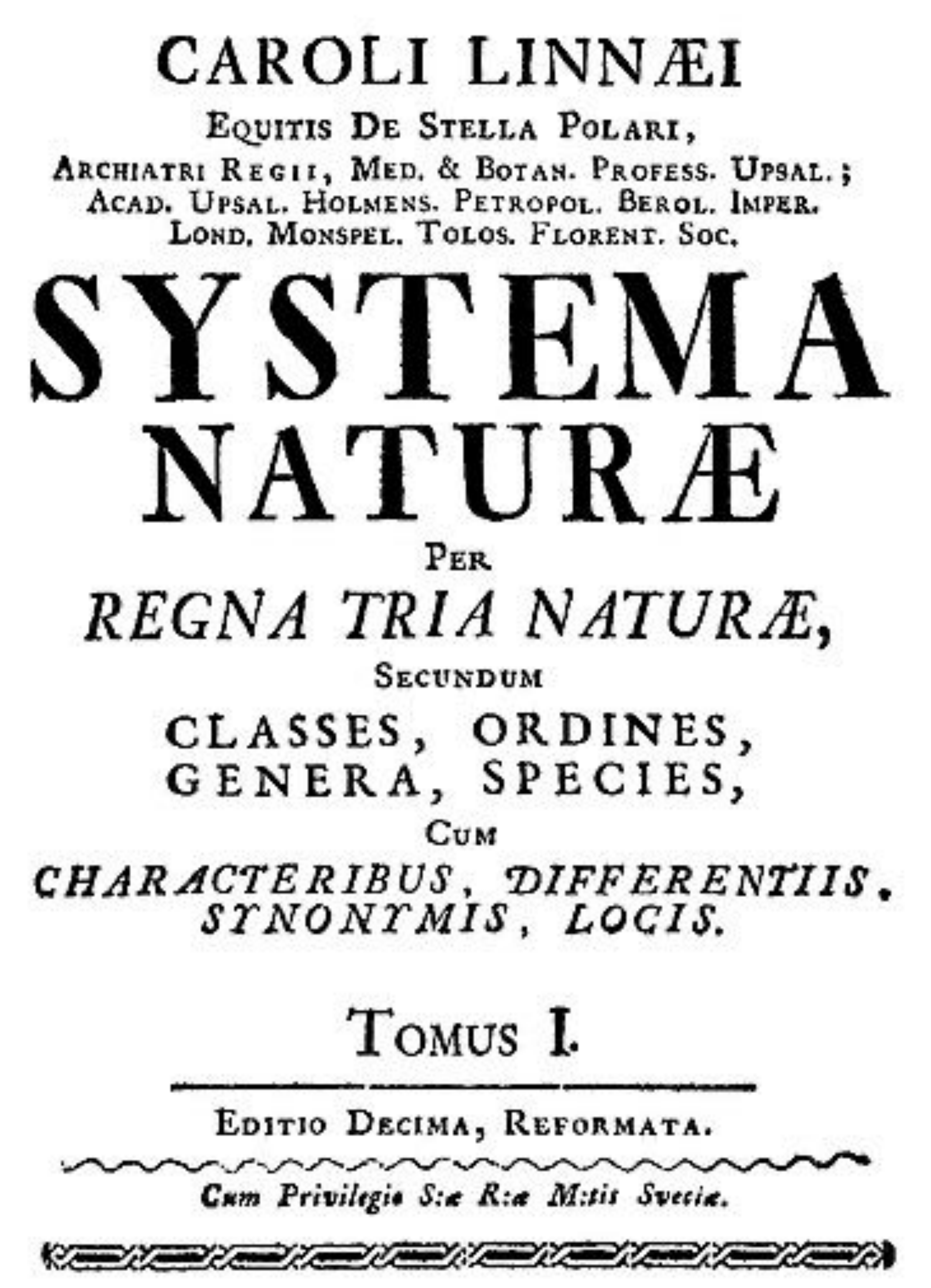
- Helps classify organisms based on shared ancestry.
- Reveals patterns of diversification and adaptation.
- Provides insights into gene and genome evolution.
- Essential for comparative genomics, ecology, and conservation.

Carl Linneus

In 1758 proposed the **binomial nomenclature system** - assigning each species a two-parts name (Genus and species).

Established a **nested hierarchy** (Kingdom, Phylum, Class, Order, Family, Genus, Species) that groups organisms in a structured manner.

Viewed **species as fixed entities**, consistent with the dominant views of his time 😅. Although he noted natural variation, he did not propose any mechanism for evolutionary change.

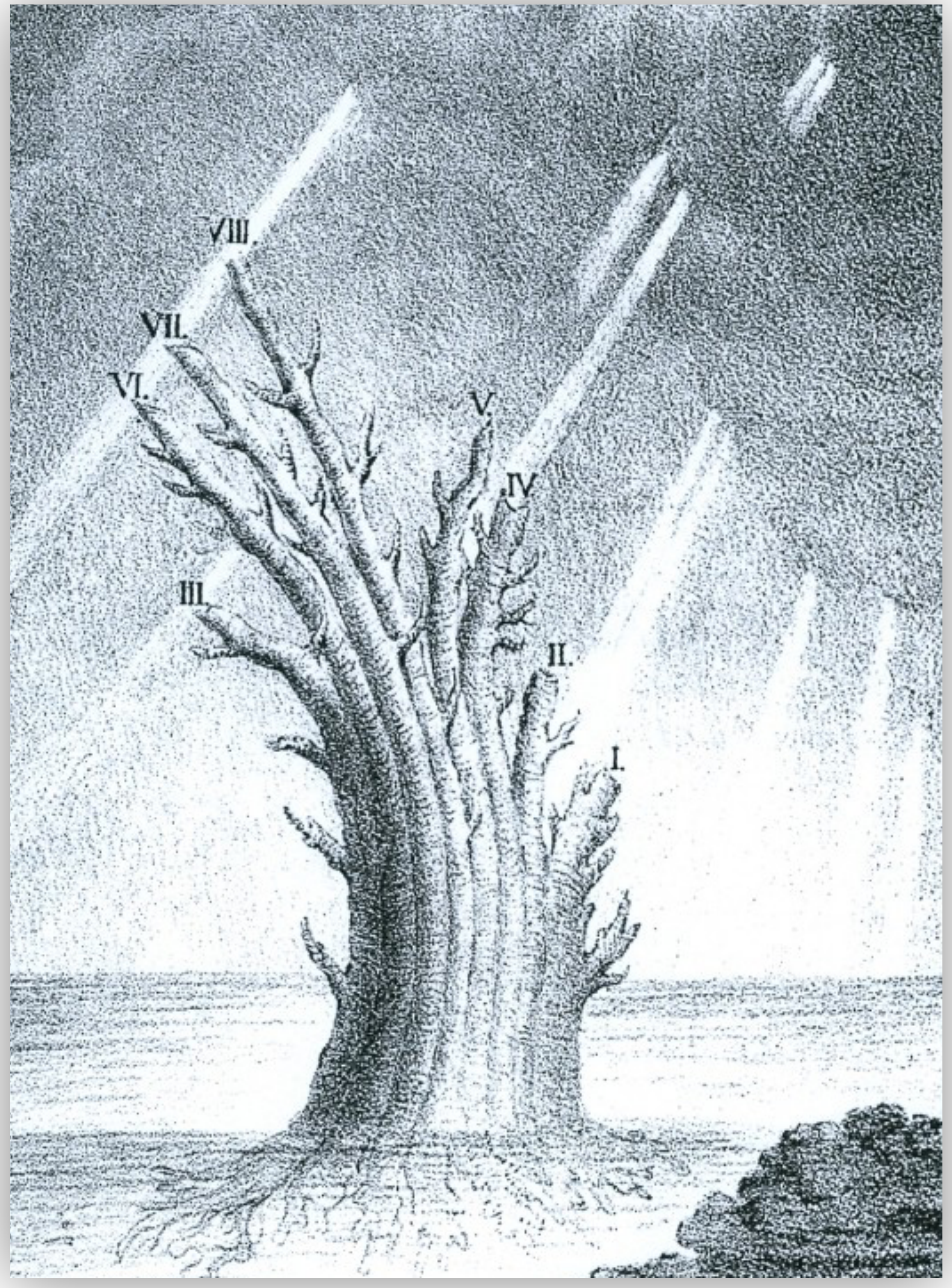


CAROLI LINNÆI
Equitis De Stella Polari,
Archiatri Regii, Med. & Botan. Profess. Upsal.;
Acad. Upsal. Holmens. Petropol. Berol. Imper.
Lond. Monspel. Tolos. Florent. Soc.
SYSTEMA
NATURÆ
Per
REGNA TRIA NATURÆ,
Secundum
CLASSES, ORDINES,
GENERA, SPECIES,
Cum
CHARACTERIBUS, DIFFERENTIIS,
SYNONYMIS, LOCIS.

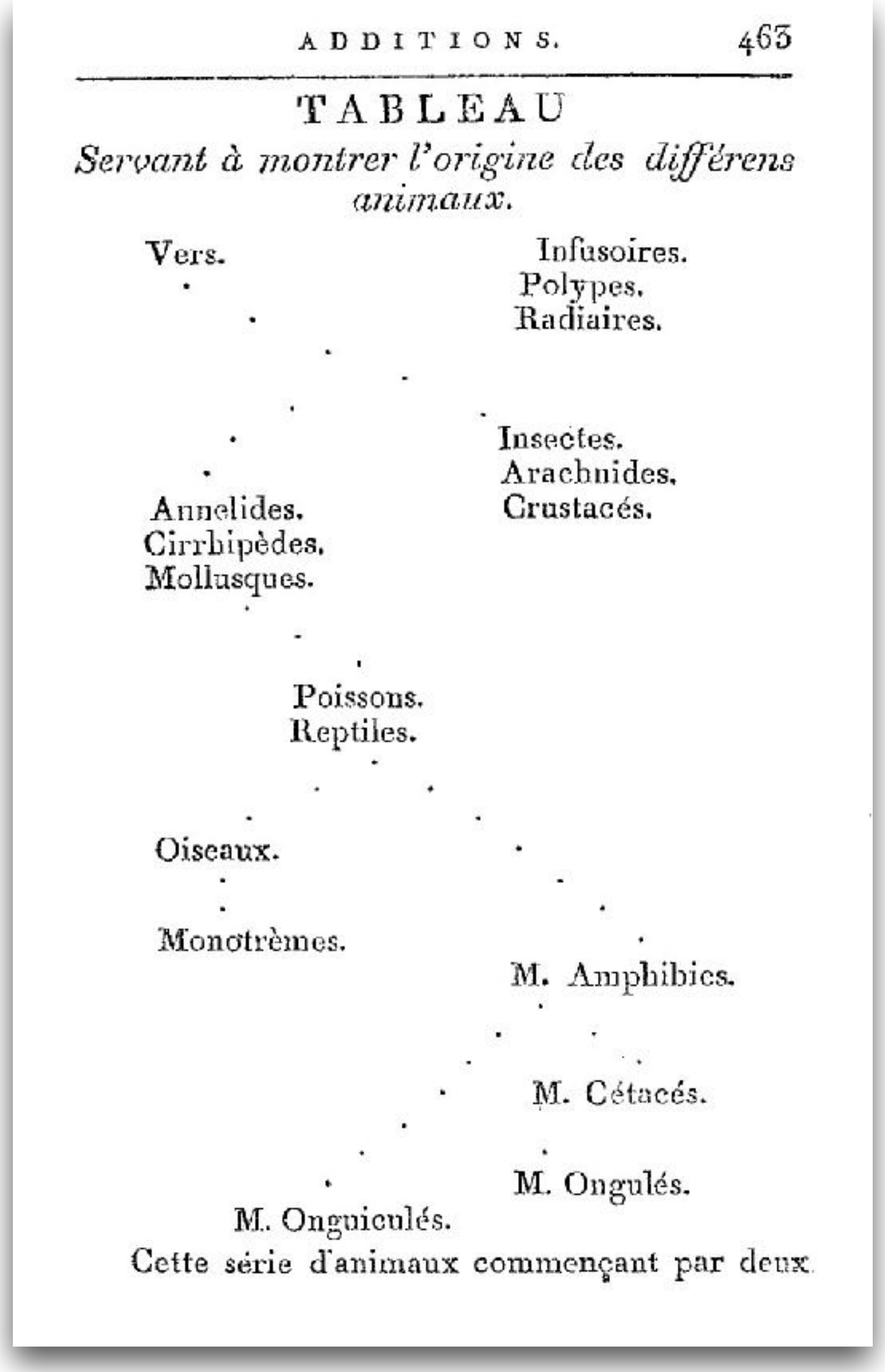
Tomus I.

Editio Decima, Reformata.

Cum Privilegio S:æ R:æ M:tis Sveciæ.



Tree of animal life, from the Zoologia specialis of Carl Edward von Eichwald (1829).



ADDITIONS. 463

TABLEAU

Servant à montrer l'origine des différens animaux.

Vers.

Infusoires.
Polypes.
Radiaires.

Insectes.
Arachnides.
Crustacés.

Annelides.
Cirrhipèdes.
Mollusques.

Poissons.
Reptiles.

Oiseaux.

Monotrèmes.

M. Amphibies.

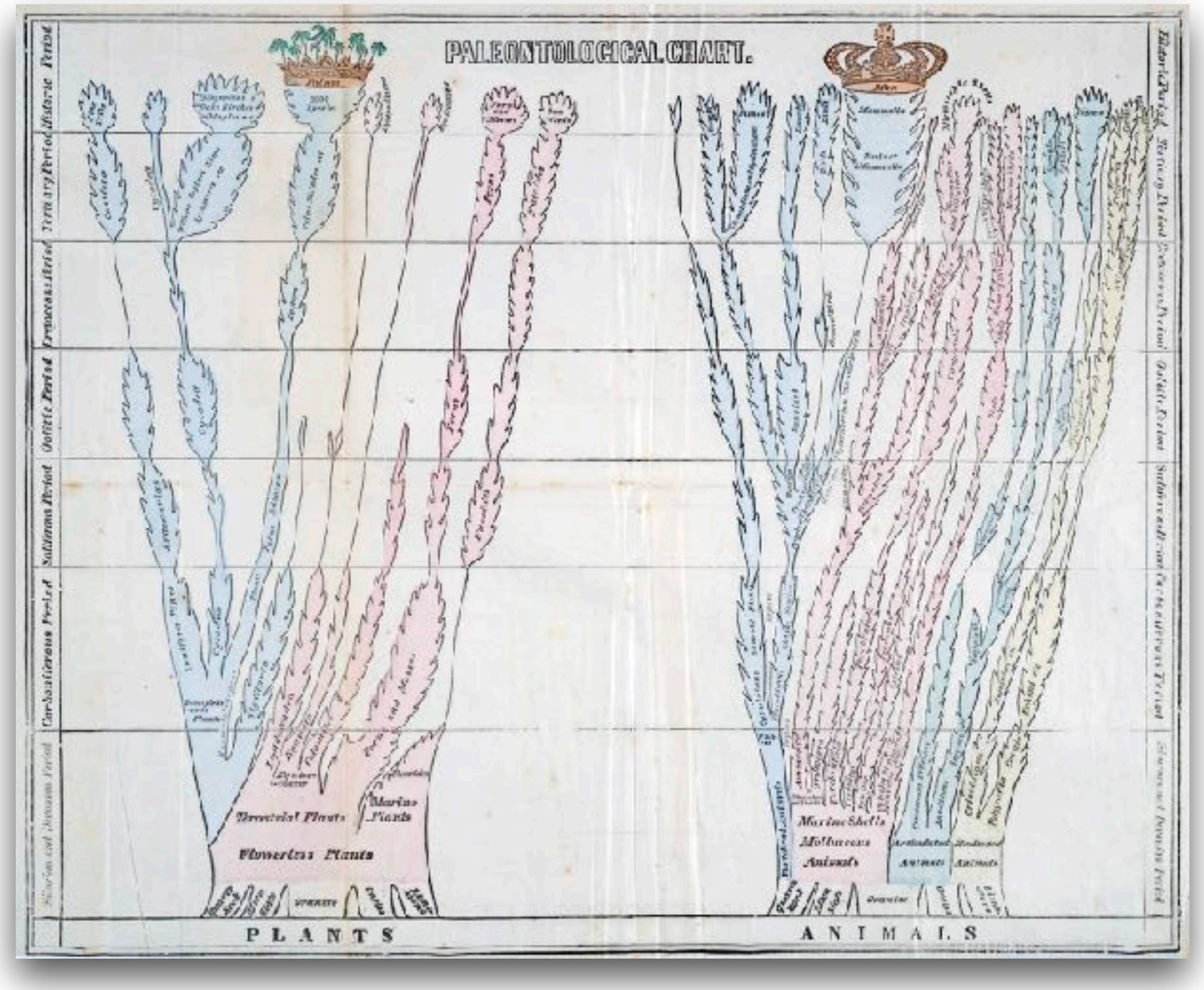
M. Cétacés.

M. Ongulés.

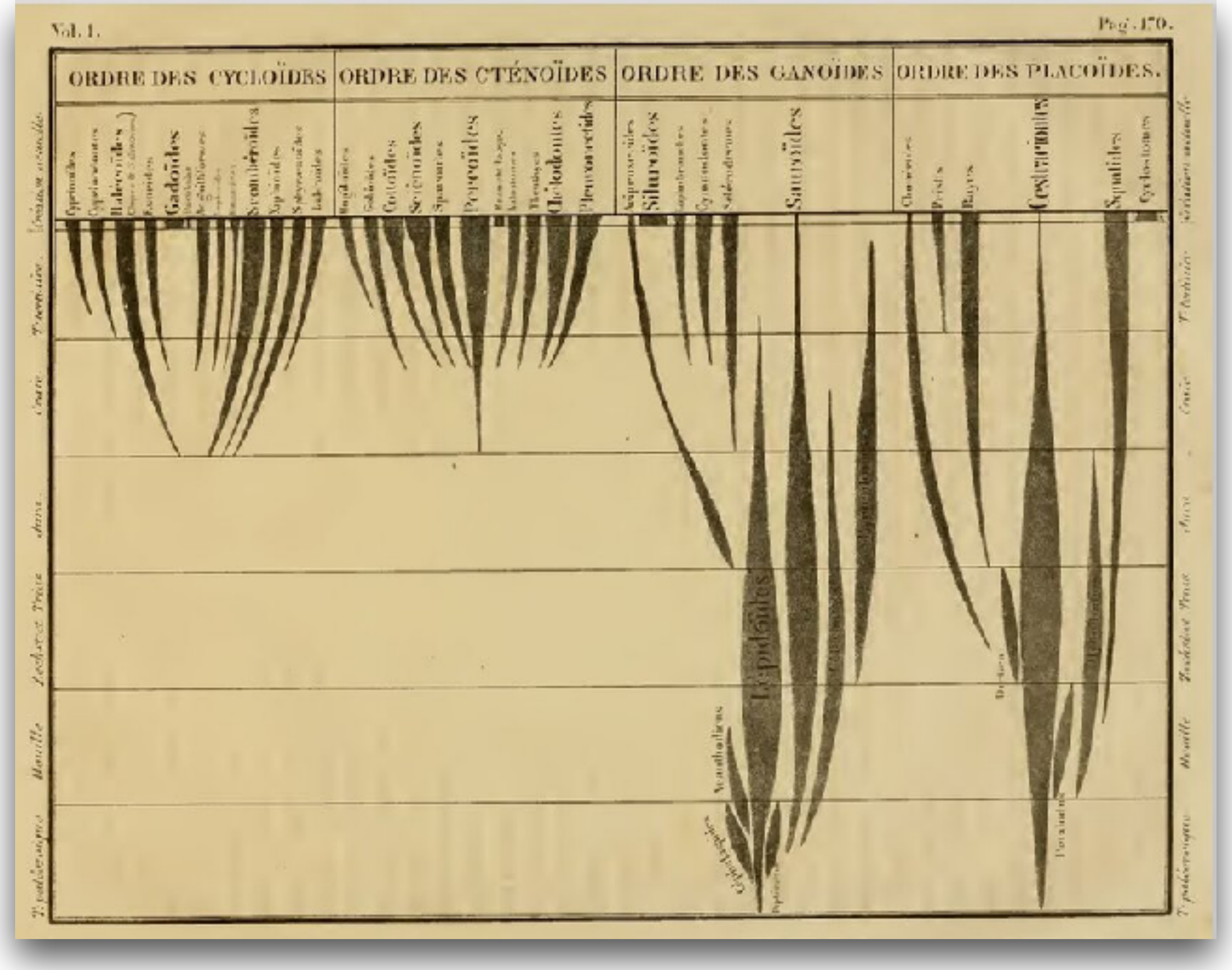
M. Onguiculés.

Cette série d'animaux commençant par deux

Tree depicting the origin of animals, from the Philosophie zoologique of Jean-Baptiste Lamarck (1809).



The Paleontological Chart in the publication 'Elementary Geology' by Edward Hitchcock (1840).



Généalogie de la classe des poissons" in Recherches sur les poissons fossiles by Louis Agassiz (1833).

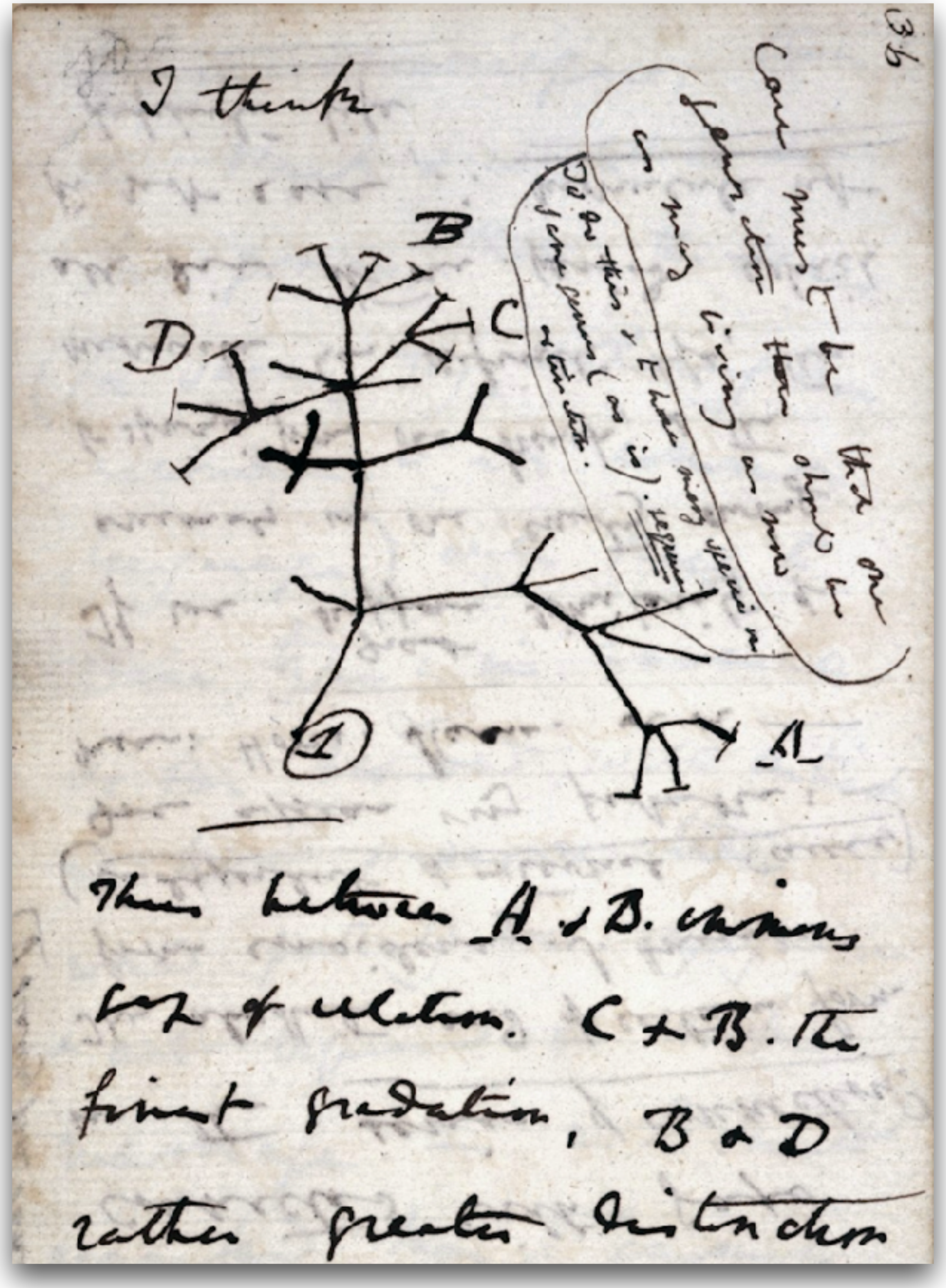
Charles Darwin

In *On the Origin of Species* (1859), Darwin introduced the theory of **evolution by natural selection**, changing our understanding of life's diversity.

- Proposed that **all living organisms** share a **common ancestor**.
- Introduced the **Tree of Life** concept to illustrate evolutionary relationships.

“The affinities of all the beings of the same class have sometimes been represented by a great tree. I believe this simile largely speaks the truth”

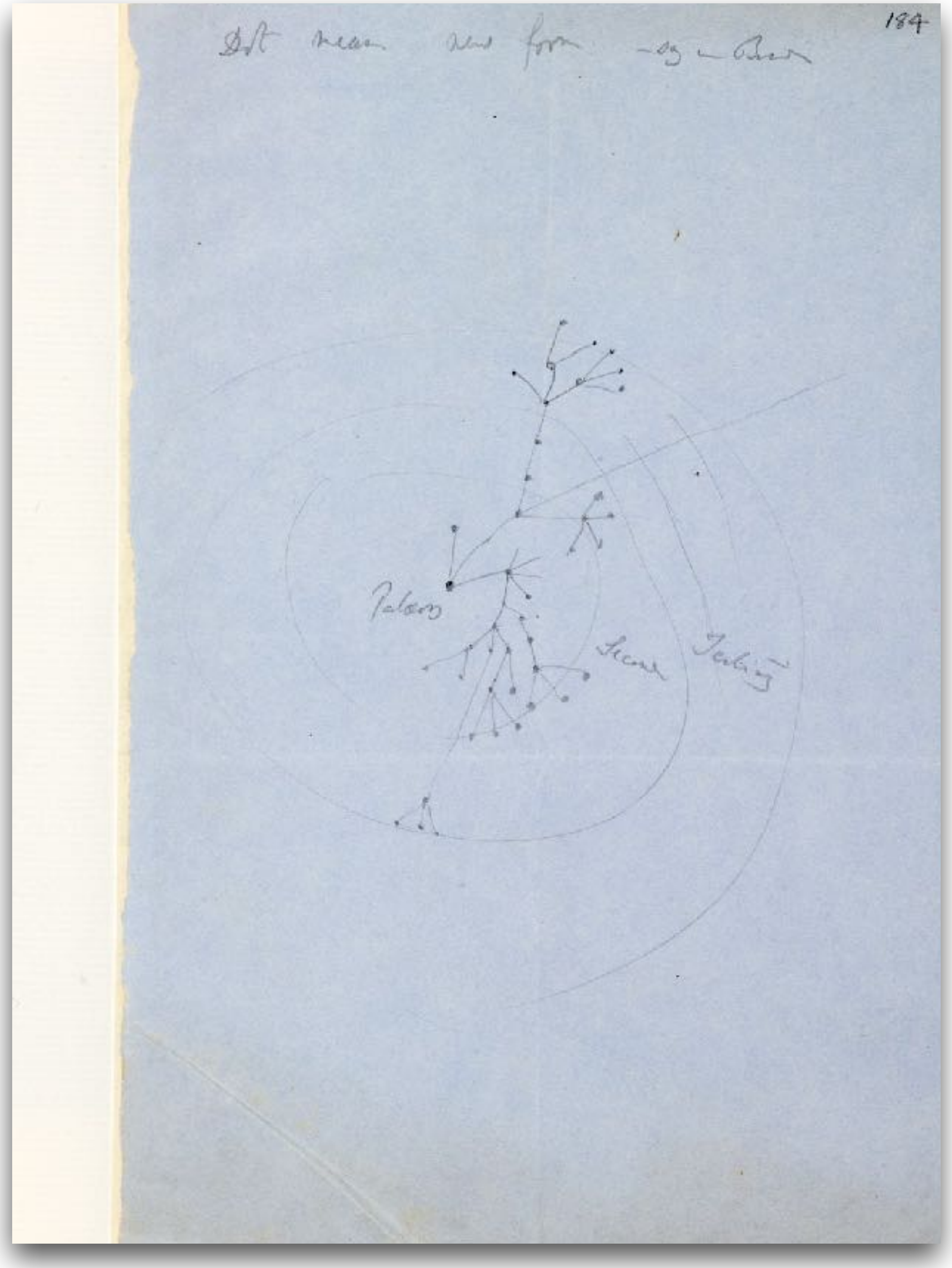
Darwin (1859). On the Origin of Species by Means of Natural Selection.



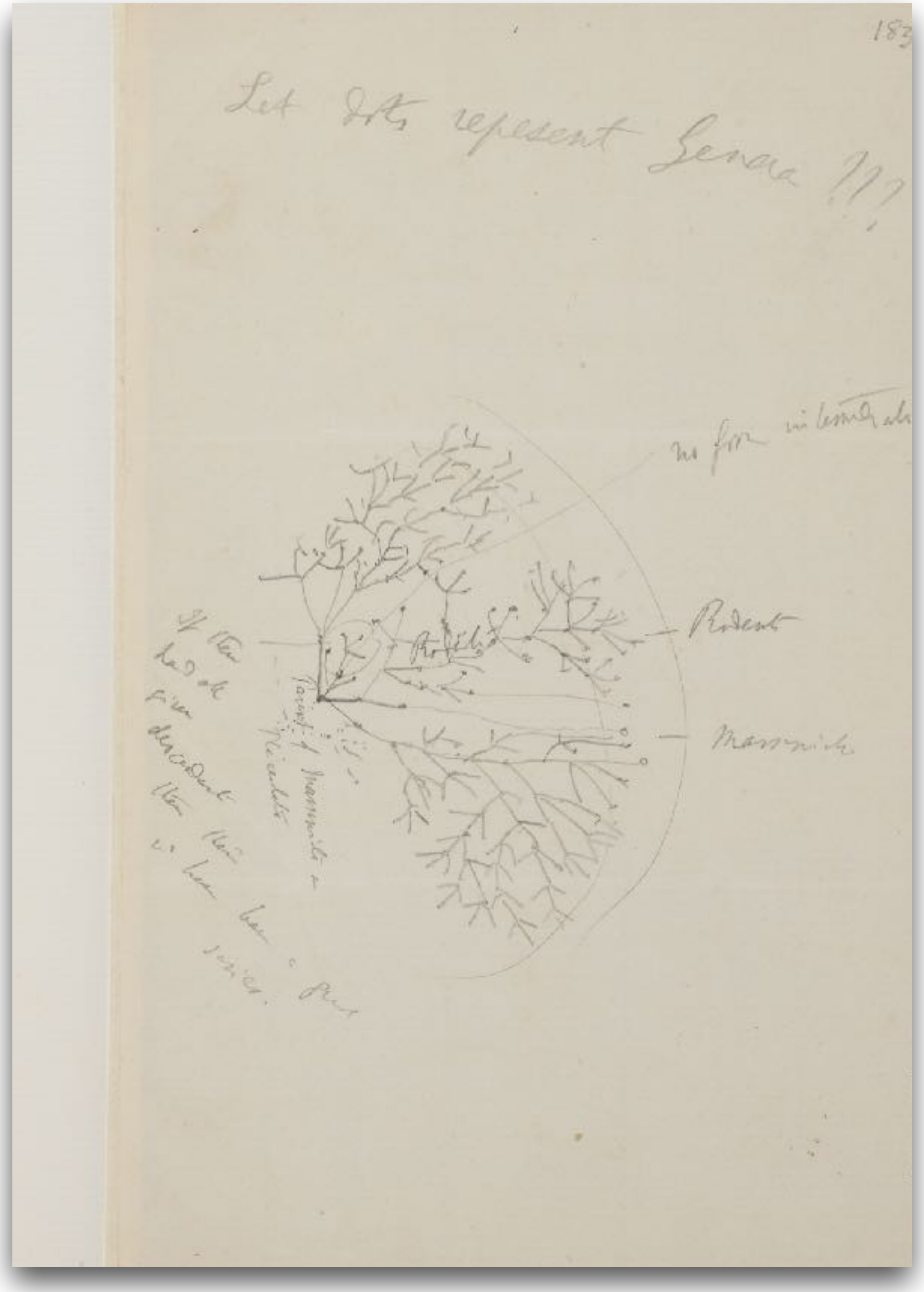
I think

Thus between A & B. immense gap of relation. C & B. the finest gradation, B & D rather greater distinction

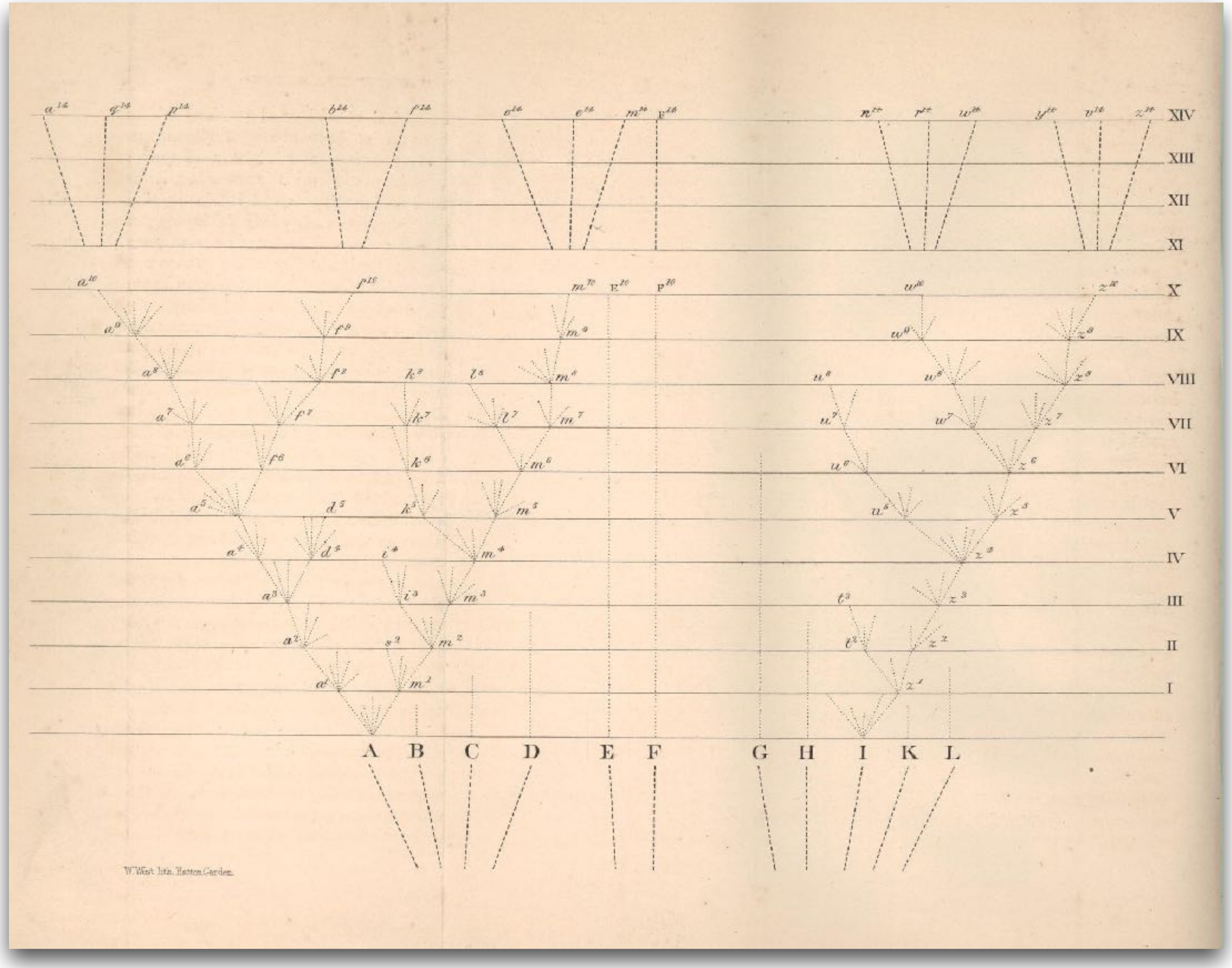
Darwin manuscript ms.dar.00121.38



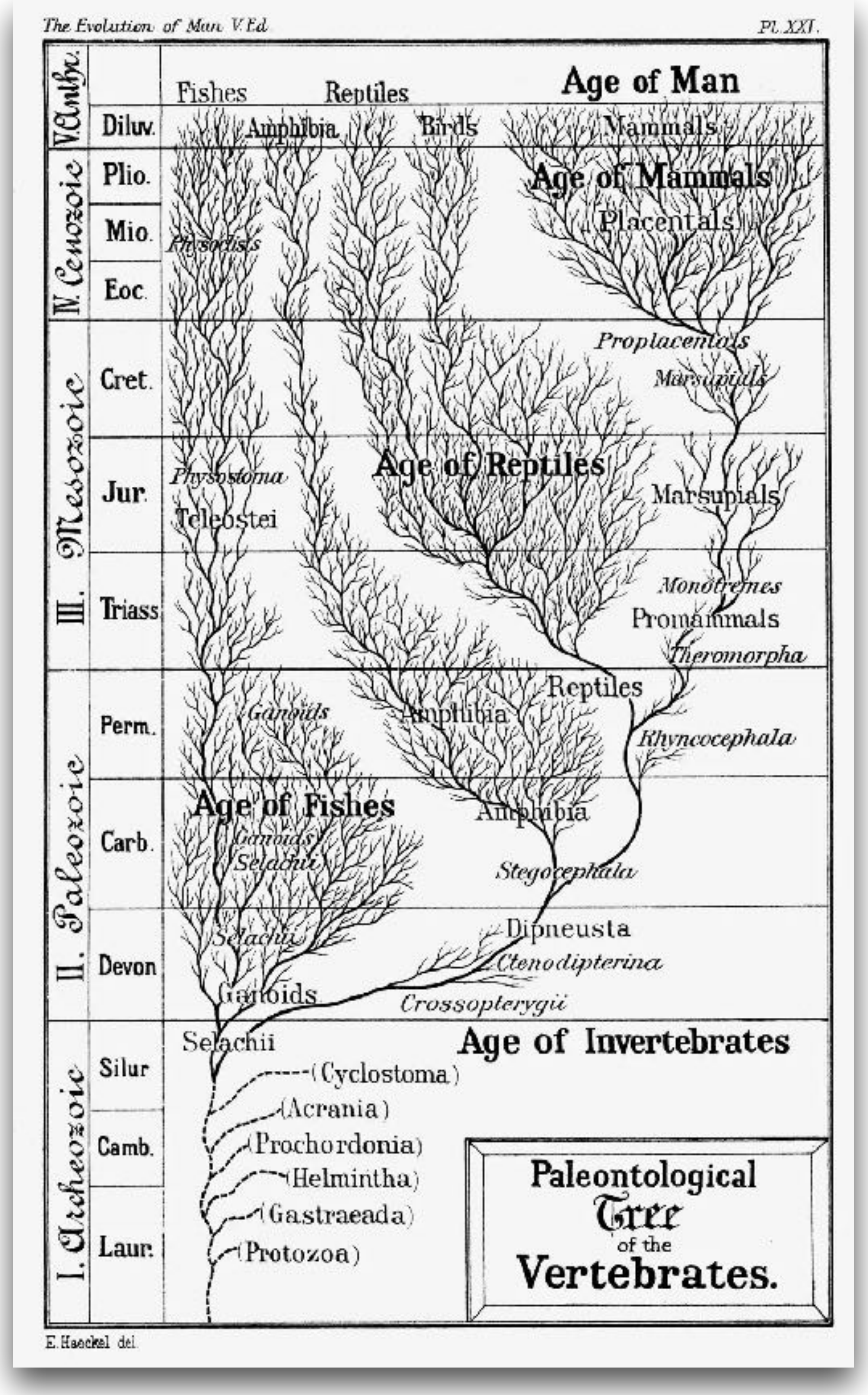
darwin manuscript ms.dar.205.5.184r



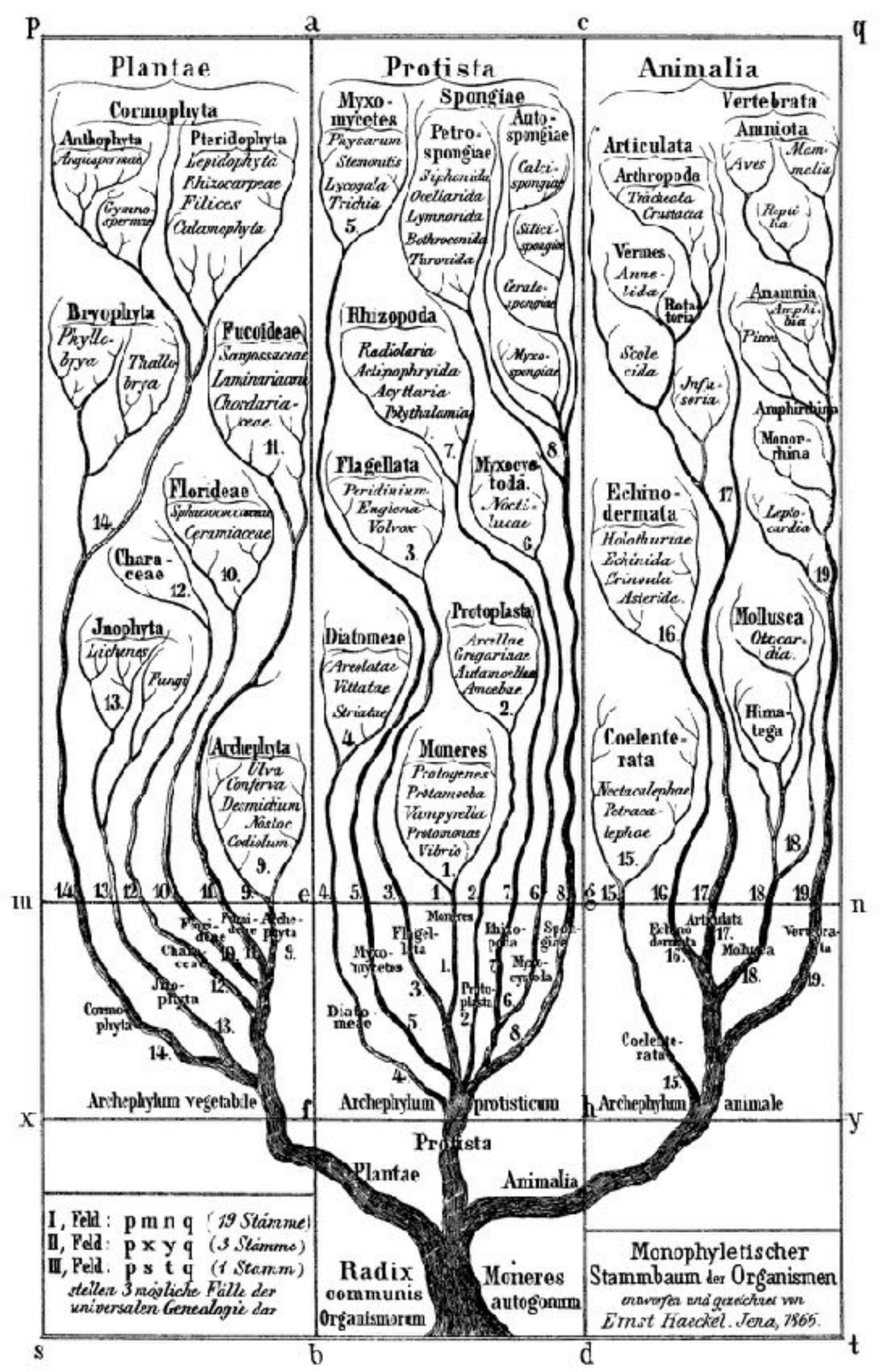
darwin manuscript ms.dar.205.5.183r



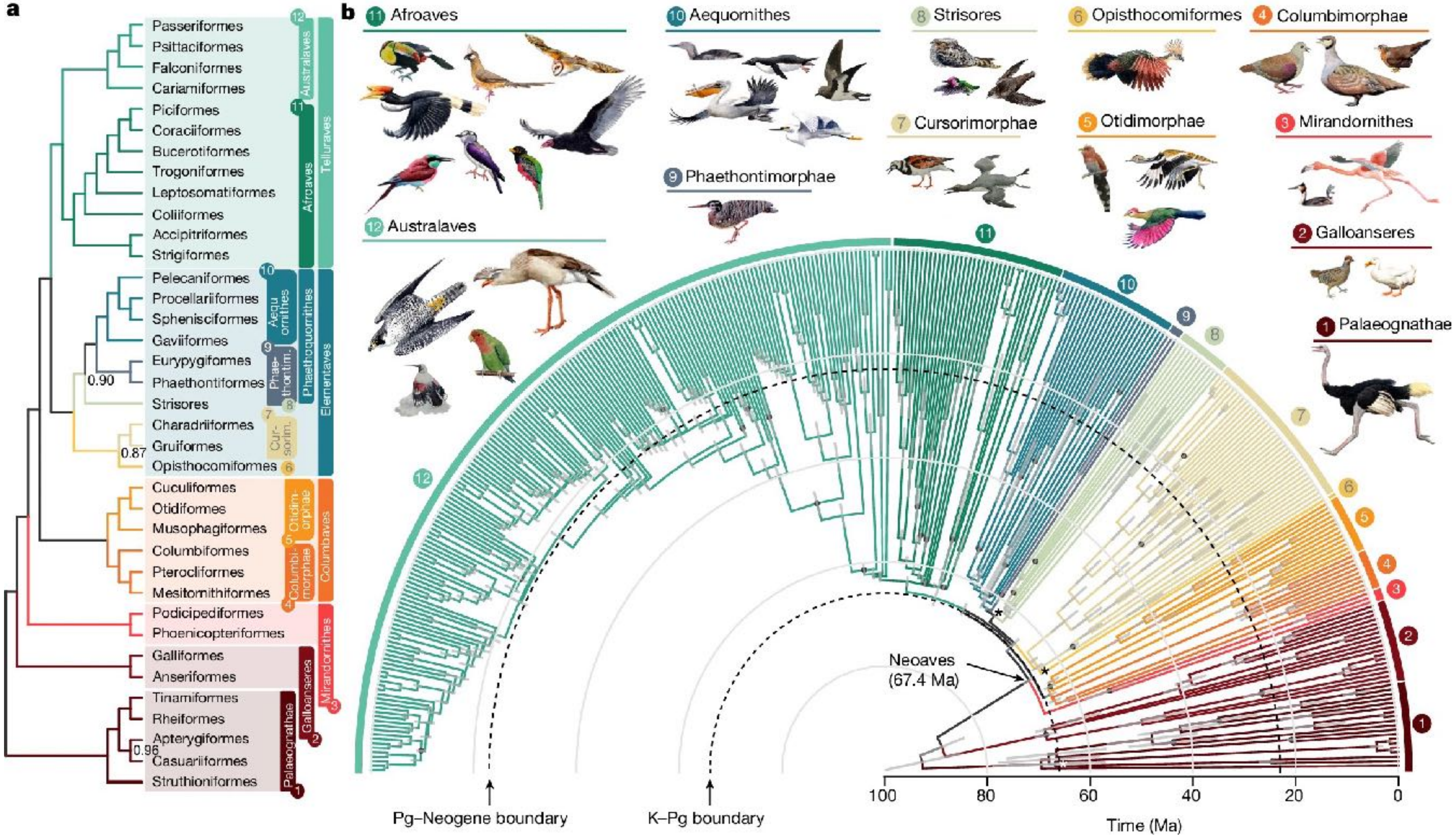
The original phylogenetic tree in Darwin's Origin of Species (1859)



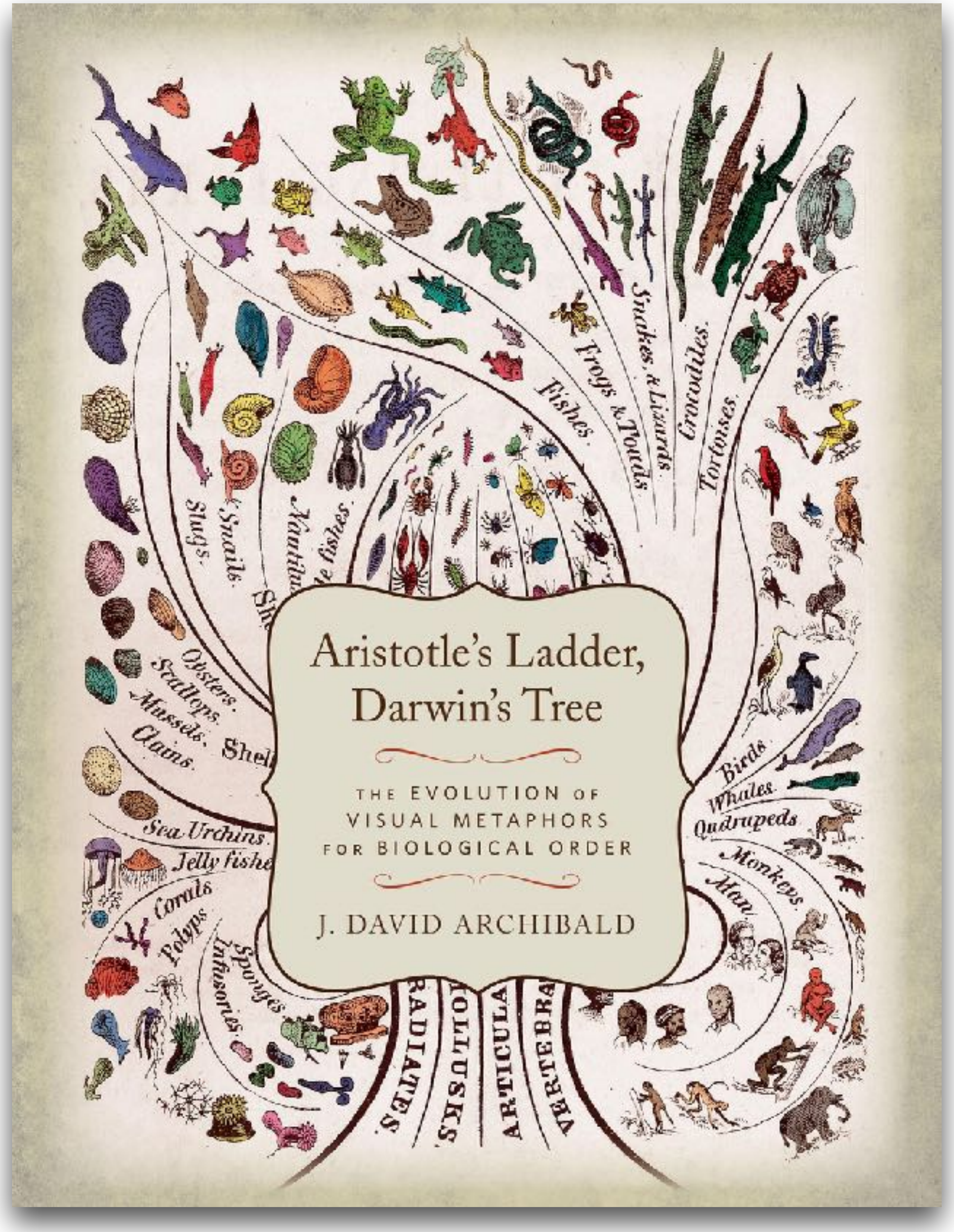
Tree of Vertebrates, from The Evolution of Man fifth edition by Ernst Haeckel (1910)



Genealogical tree depicting three kingdoms of life, from Volume II of Generelle Morphologie by Ernst Heinrich Haeckel (1866).



[Stiller et al. Complexity of avian evolution revealed by family-level genomes. Nature (2024)]



Aristotele's ladder, Darwin's tree by J. David Archibald

Biology Direct

Review

Open Access

Trees and networks before and after Darwin

Mark A Ragan

Address: The University of Queensland, Institute for Molecular Bioscience and Australian Research Council Centre of Excellence in Bioinformatics, 306 Carmody Rd, St Lucia, Brisbane, Queensland 4072, Australia

Email: Mark A Ragan - m.ragan@imb.uq.edu.au

Published: 16 November 2009

Biology Direct 2009, 4:43 doi:10.1186/1745-6150-4-43

Received: 24 October 2009
Accepted: 16 November 2009

This article is available from: http://www.biology-direct.com/content/4/1/43

Abstract

It is well-known that Charles Darwin sketched abstract trees of relationship in his 1837 notebook, and depicted a tree in the *Origin of Species* (1859). Here I attempt to place Darwin's trees in historical context. By the mid-Eighteenth century the Great Chain of Being was increasingly seen to be an inadequate description of order in nature, and by about 1780 it had been largely abandoned without a satisfactory alternative having been agreed upon. In 1750 Donati described aquatic and terrestrial organisms as forming a network, and a few years later Buffon depicted a network of genealogical relationships among breeds of dogs. In 1764 Bonnet asked whether the Chain might actually branch at certain points, and in 1766 Pallas proposed that the gradations among organisms resemble a tree with a compound trunk, perhaps not unlike the tree of animal life later depicted by Eichwald. Other trees were presented by Augier in 1801 and by Lamarck in 1809 and 1815, the latter two assuming a transmutation of species over time. Elaborate networks of affinities among plants and among animals were depicted in the late Eighteenth and very early Nineteenth centuries. In the two decades immediately prior to 1837, so-called affinities and/or analogies among organisms were represented by diverse geometric figures. Series of plant and animal fossils in successive geological strata were represented as trees in a popular textbook from 1840, while in 1858 Bronn presented a system of animals, as evidenced by the fossil record, in a form of a tree. Darwin's 1859 tree and its subsequent elaborations by Haeckel came to be accepted in many but not all areas of biological sciences, while network diagrams were used in others. Beginning in the early 1960s trees were inferred from protein and nucleic acid sequences, but networks were re-introduced in the mid-1990s to represent lateral genetic transfer, increasingly regarded as a fundamental mode of evolution at least for bacteria and archaea. In historical context, then, the Network of Life preceded the Tree of Life and might again supersede it.

Reviewers: This article was reviewed by Eric Bapteste, Patrick Forterre and Dan Graur.

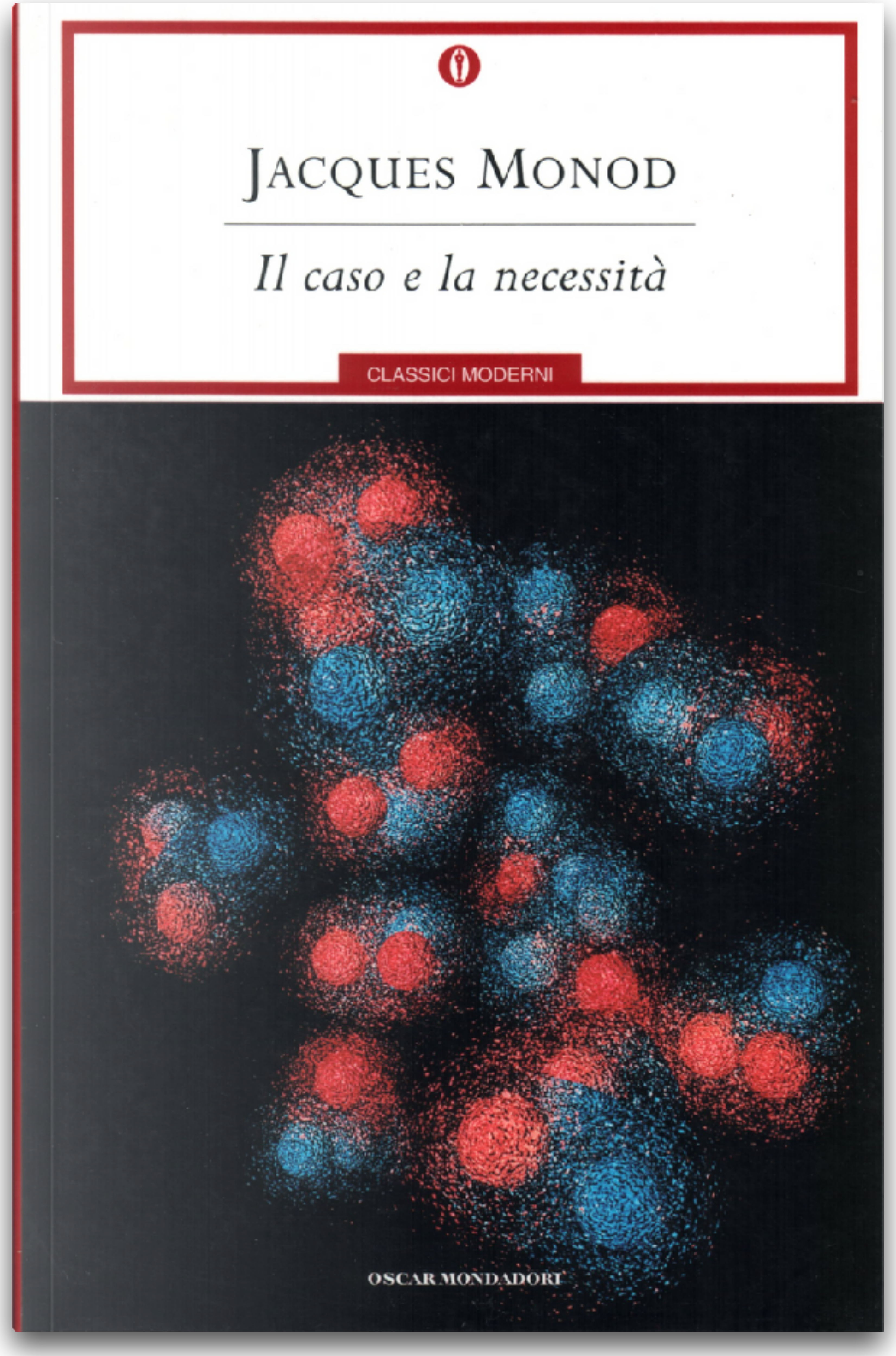
Prefatory quotation

> *(N)ature rises up by connections, little by little and without leaps, as though it proceeds by an unbroken web, it proceeds in a leisurely and placid uninterrupted course. There is no gap, no break, no dispersion of forms: they have, in turn, been connected, ring within ring. That very golden chain is universal in its embrace.* - Juan Eusebio Nieremberg, 1635 [[1], p.29]

The rise and fall of the Great Chain of Being

From very early in the Middle Eastern and European religious and intellectual traditions, chains, cords, ladders and stairways served as metaphors for order in the world, or between earth and heaven [2-6]. The image of a tree sometimes served in the same metaphorical sense [[5], pp.319-329; [6], p.22]. A linear order in nature was com-

Mark A Ragan. Trees and networks before and after Darwin. Biology Direct (2009)

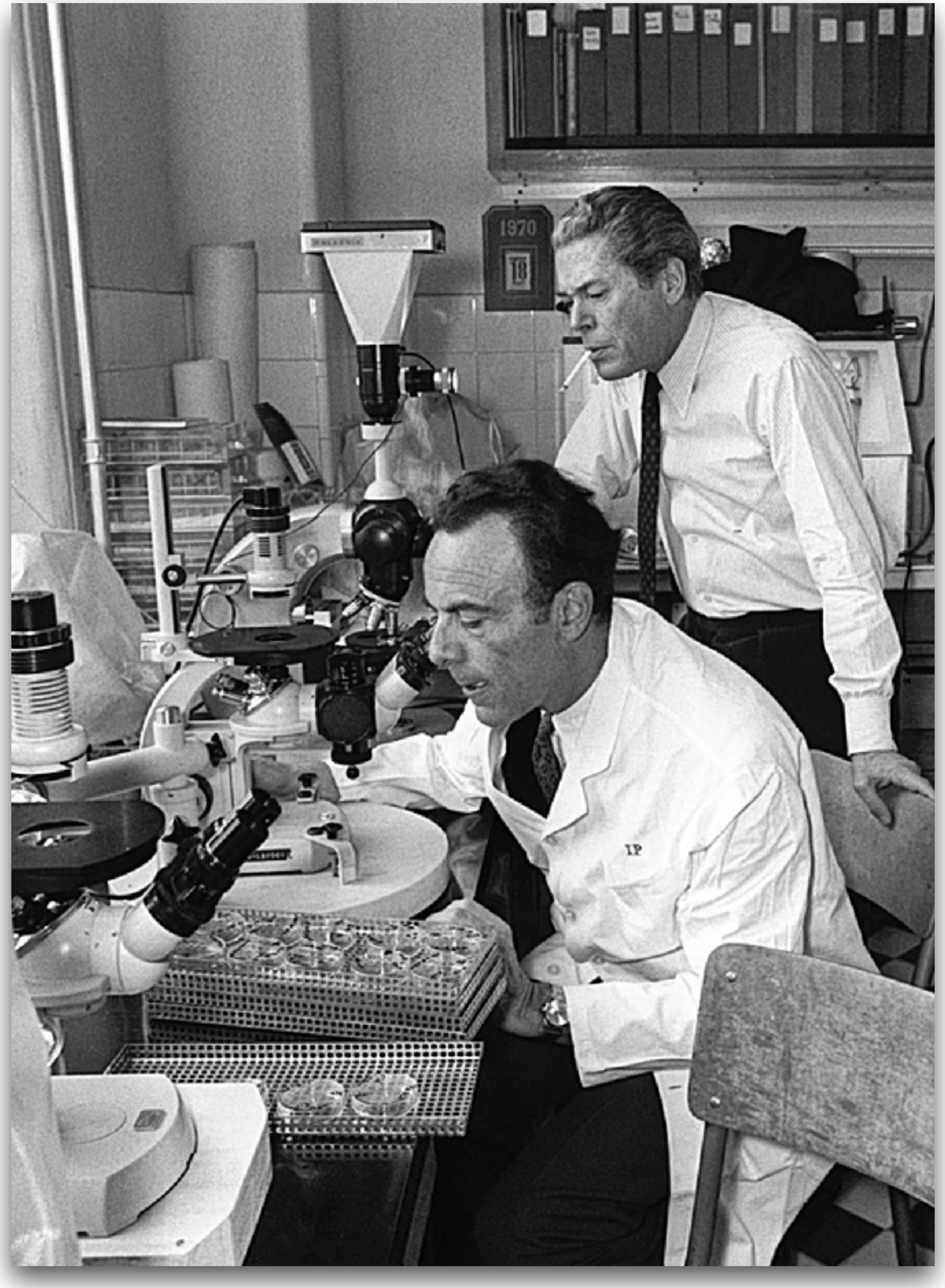


JACQUES MONOD

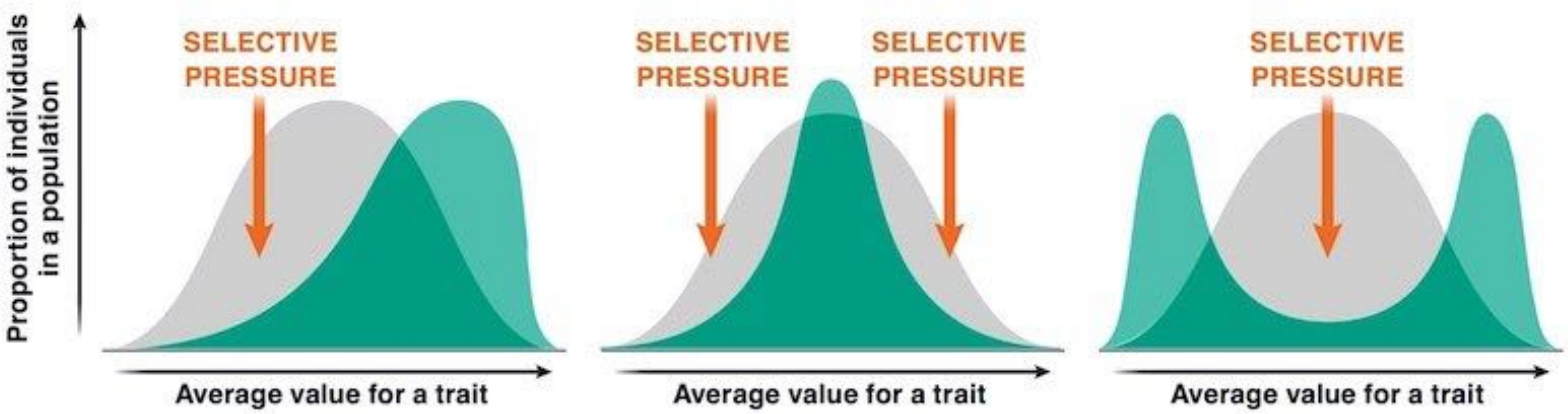
Il caso e la necessità

CLASSICI MODERNI

OSCAR MONDADORI



mutation
&
selection

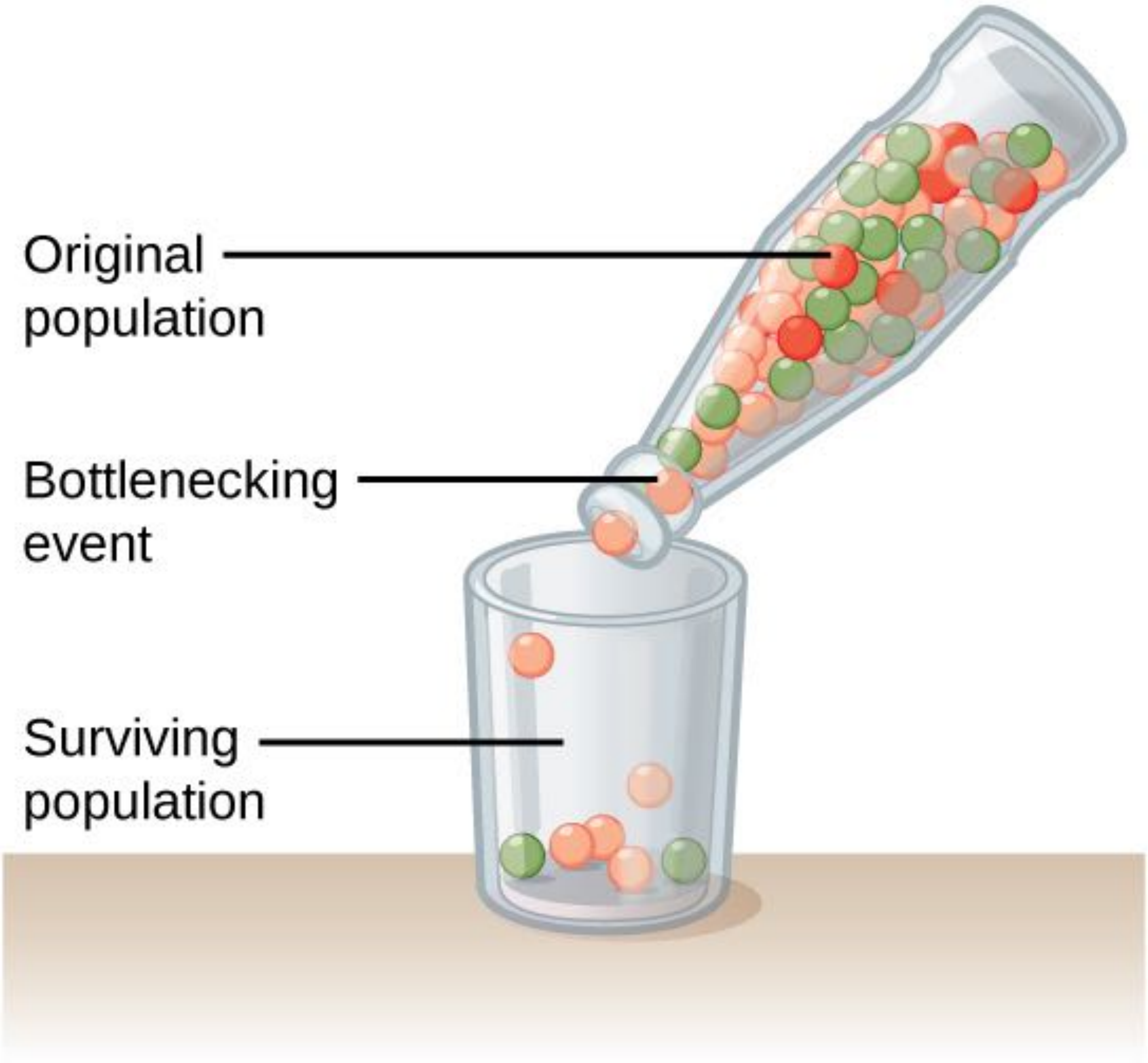


DIRECTIONAL SELECTION
The average value for the trait increases or decreases.

STABILIZING SELECTION
The average value of a trait remains the same while extreme versions of the trait are selected against.

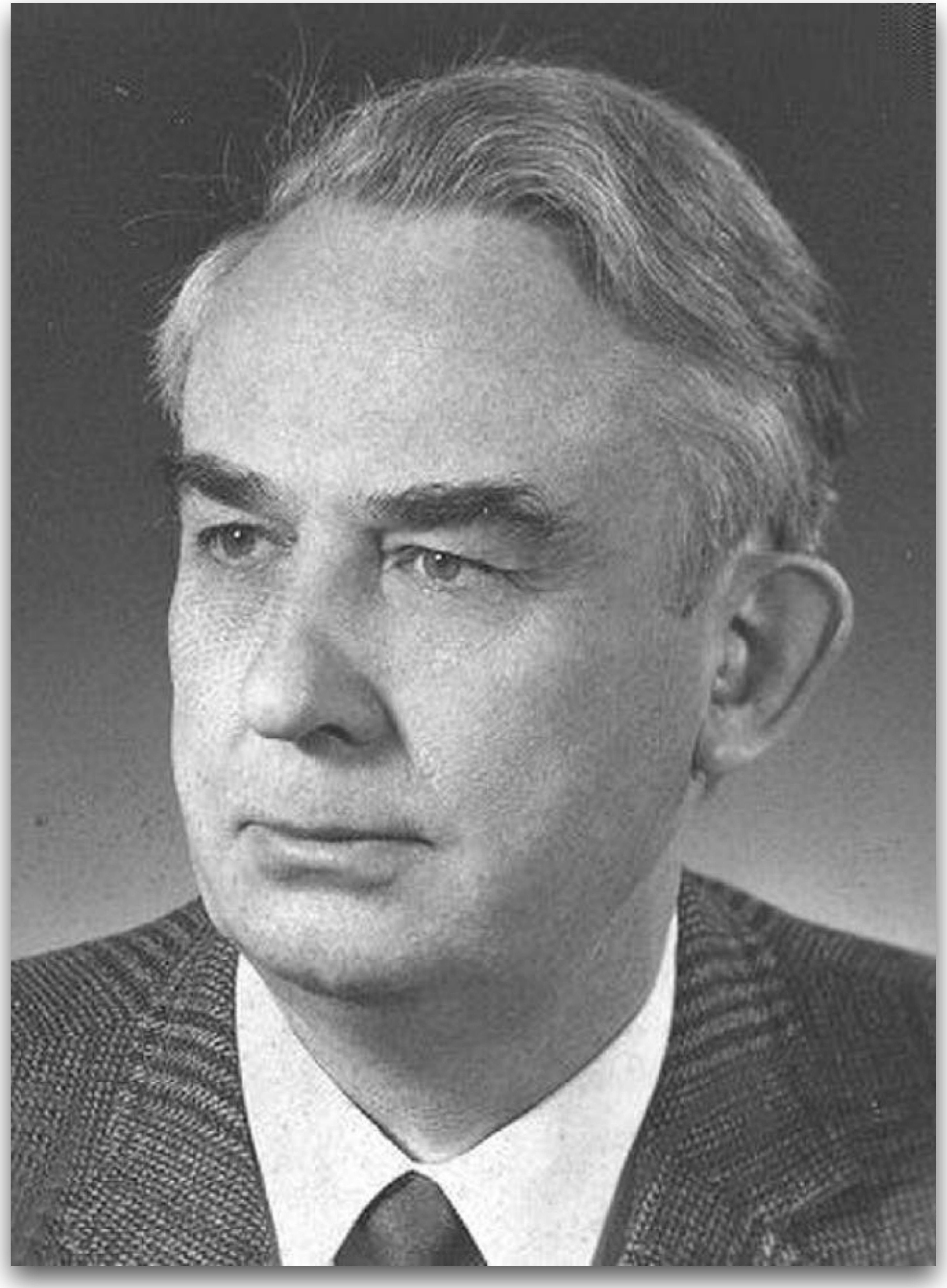
DISRUPTIVE SELECTION
Individuals with extreme phenotypes have the highest fitness.

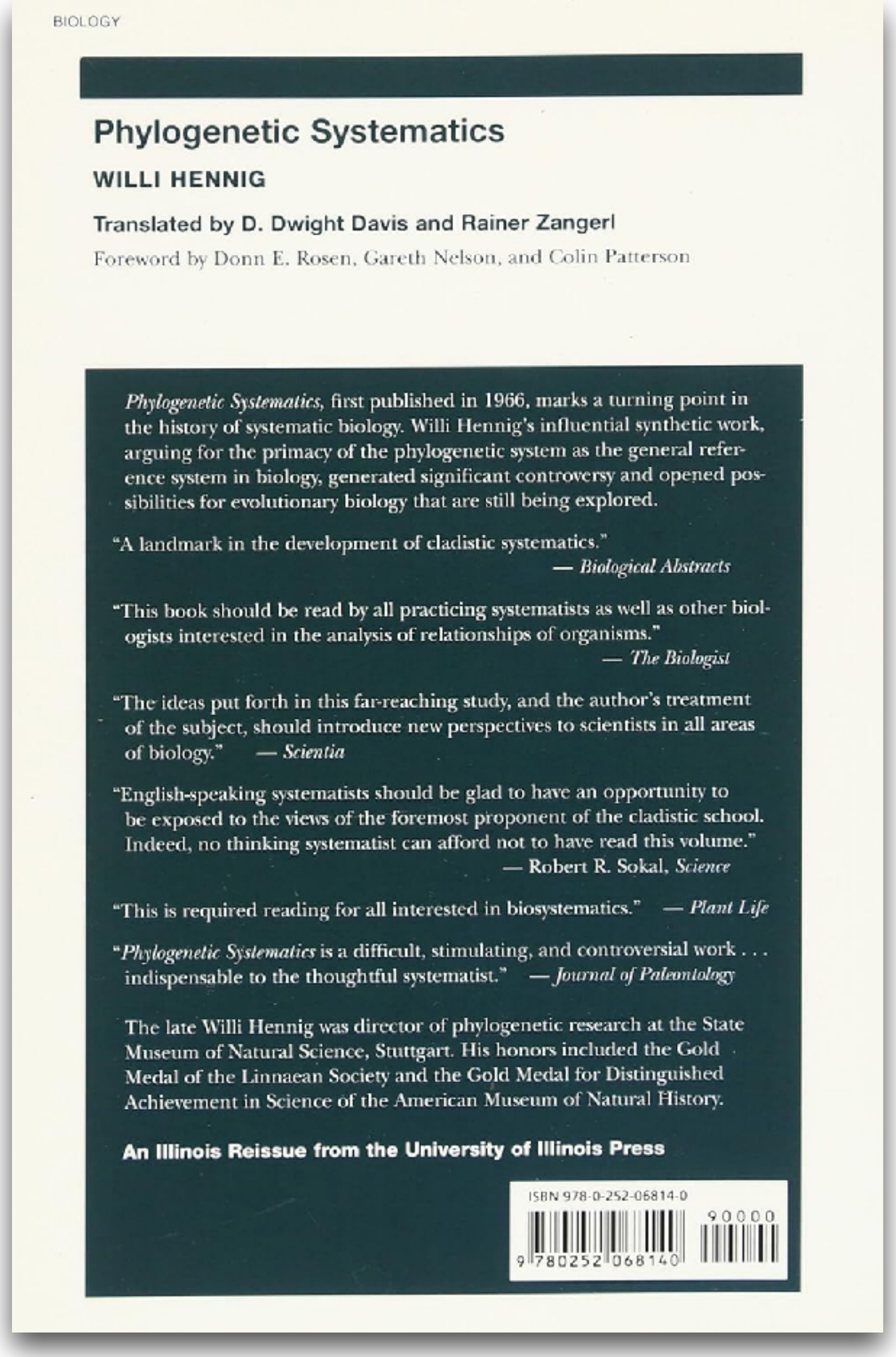
other factors underlie the effect of selection



Phylogenetics

- Phylogenetic systematics, also known as **cladistics**, is a method of biological classification that groups organisms based on their evolutionary history.
- Developed by the German entomologist **Willi Hennig** in 1950, cladistics organizes organisms into groups called clades—lineages that share a most recent common ancestor.
- The term "clade" comes from the Ancient Greek **κλάδος**, meaning "branch."
- The evidence used to infer evolutionary relationships is based on shared derived characteristics (**synapomorphies**)—traits that are unique to a particular group and absent in more distant relatives.





BIOLOGY

Phylogenetic Systematics

WILLI HENNIG

Translated by D. Dwight Davis and Rainer Zangerl

Foreword by Donn E. Rosen, Gareth Nelson, and Colin Patterson

Phylogenetic Systematics, first published in 1966, marks a turning point in the history of systematic biology. Willi Hennig's influential synthetic work, arguing for the primacy of the phylogenetic system as the general reference system in biology, generated significant controversy and opened possibilities for evolutionary biology that are still being explored.

"A landmark in the development of cladistic systematics."
— *Biological Abstracts*

"This book should be read by all practicing systematists as well as other biologists interested in the analysis of relationships of organisms."
— *The Biologist*

"The ideas put forth in this far-reaching study, and the author's treatment of the subject, should introduce new perspectives to scientists in all areas of biology." — *Scientia*

"English-speaking systematists should be glad to have an opportunity to be exposed to the views of the foremost proponent of the cladistic school. Indeed, no thinking systematist can afford not to have read this volume."
— Robert R. Sokal, *Science*

"This is required reading for all interested in biosystematics." — *Plant Life*

"*Phylogenetic Systematics* is a difficult, stimulating, and controversial work . . . indispensable to the thoughtful systematist." — *Journal of Paleontology*

The late Willi Hennig was director of phylogenetic research at the State Museum of Natural Science, Stuttgart. His honors included the Gold Medal of the Linnaean Society and the Gold Medal for Distinguished Achievement in Science of the American Museum of Natural History.

An Illinois Reissue from the University of Illinois Press

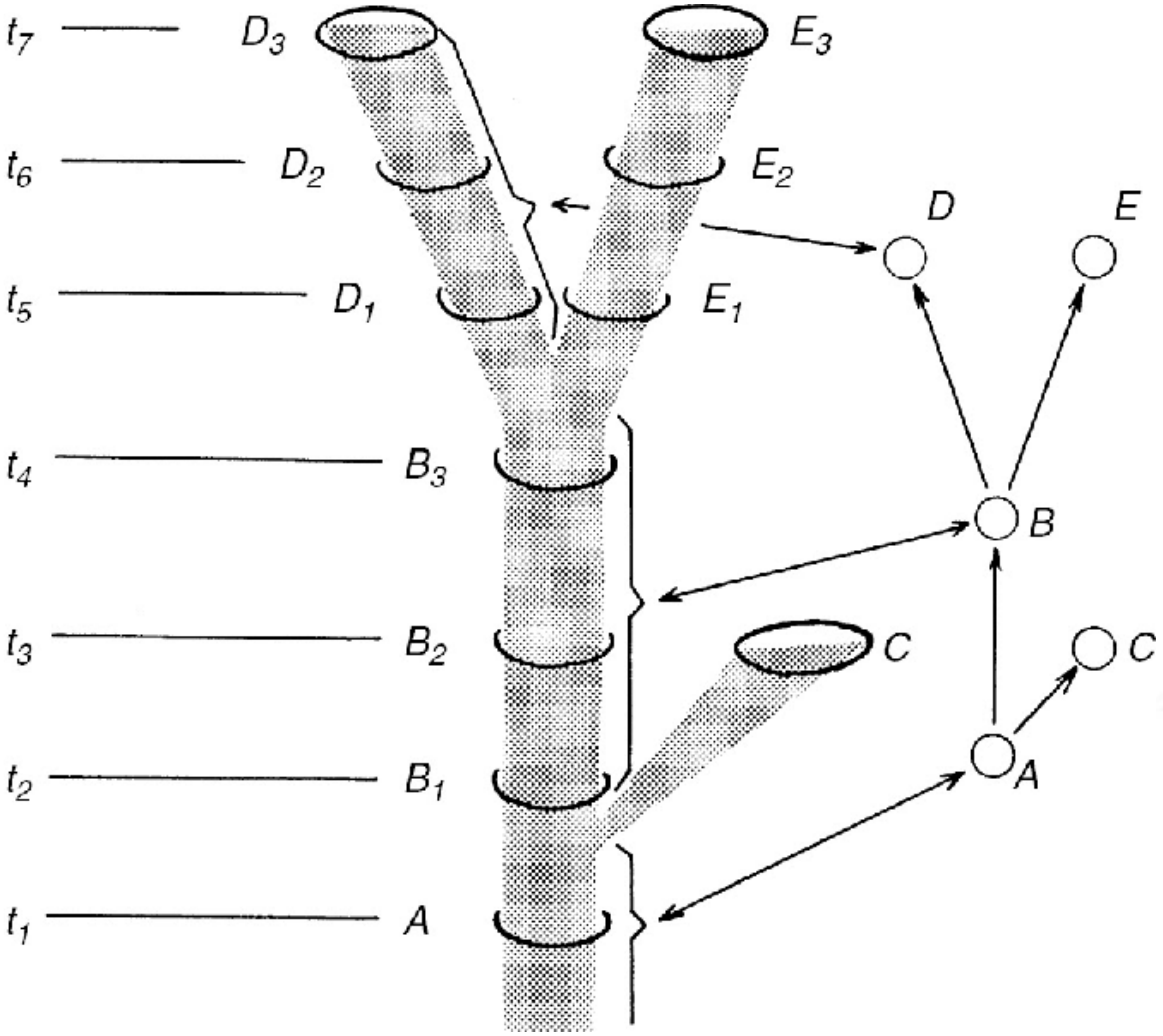
ISBN 978-0-252-06814-0
90000
9 780252 068140

Cladistics

VOLUME 39 • NUMBER 6 • DECEMBER 2023
ISSN 0748-3007

The International Journal of the Willi Hennig Society

WILEY
wileyonlinelibrary.com/journal/cla



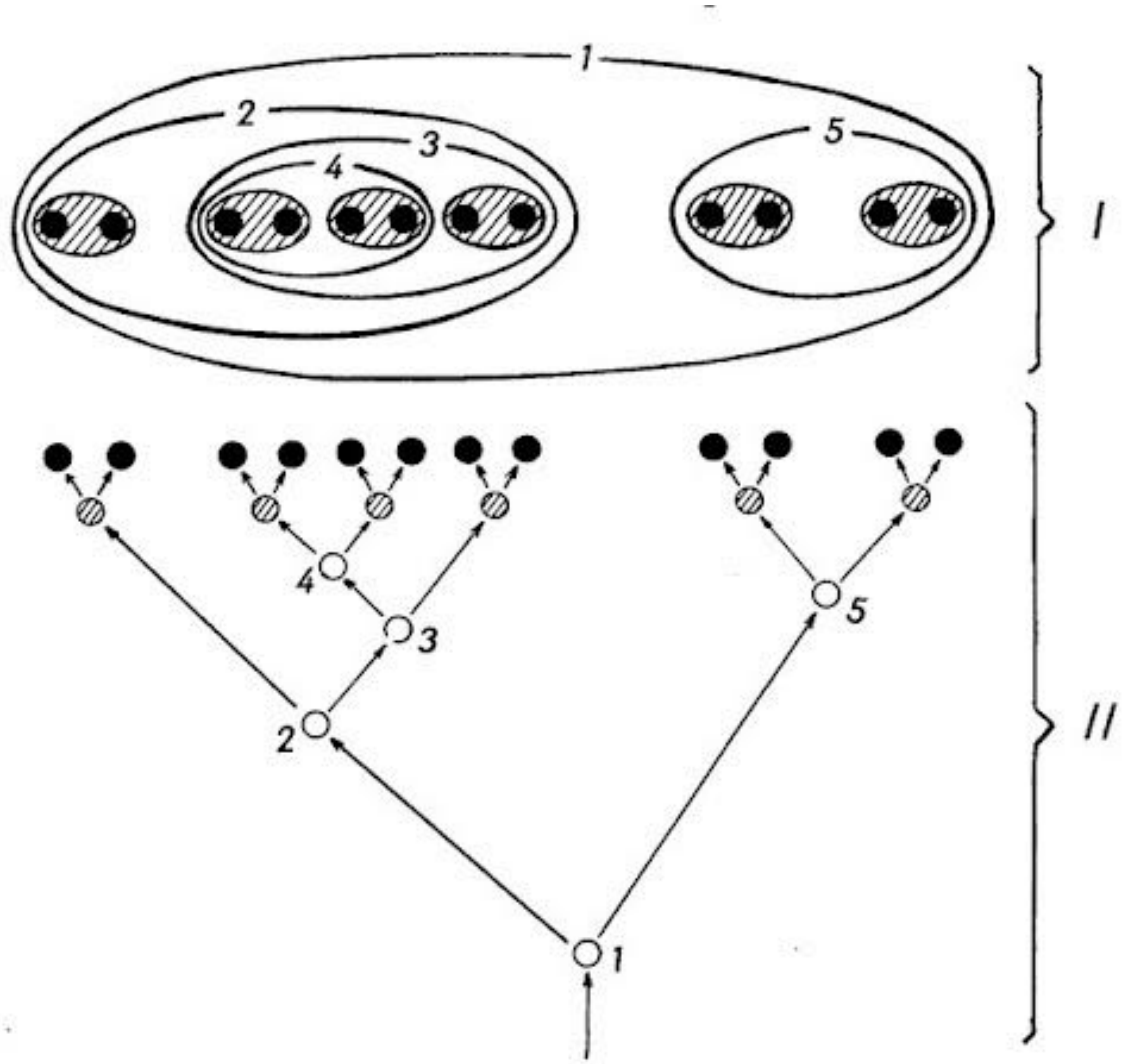


Figure 18. The phylogenetic kinship relations between the species of a monophyletic group, represented in two different ways.

Willi Hennig. Phylogenetic Systematics (1966)

Synapomorphies (Shared Derived Traits)

Synapomorphies are evolutionary traits that are **shared by two or more taxa** and inherited from their most recent common ancestor. These traits define **monophyletic groups (clades)** and provide strong evidence of evolutionary relationships. **Example:** The presence of feathers in birds is a synapomorphy that distinguishes them from other reptiles.

Autapomorphies (Unique Derived Traits)

Autapomorphies traits **unique to a single taxon** and not shared with any other group. While they help in distinguishing individual lineages, they do not provide information about shared ancestry among multiple taxa. **Example:** The elongated neck of giraffes (*Giraffa camelopardalis*) is an autapomorphy, as it is not found in closely related species.

Symplesiomorphies (Shared Ancestral Traits)

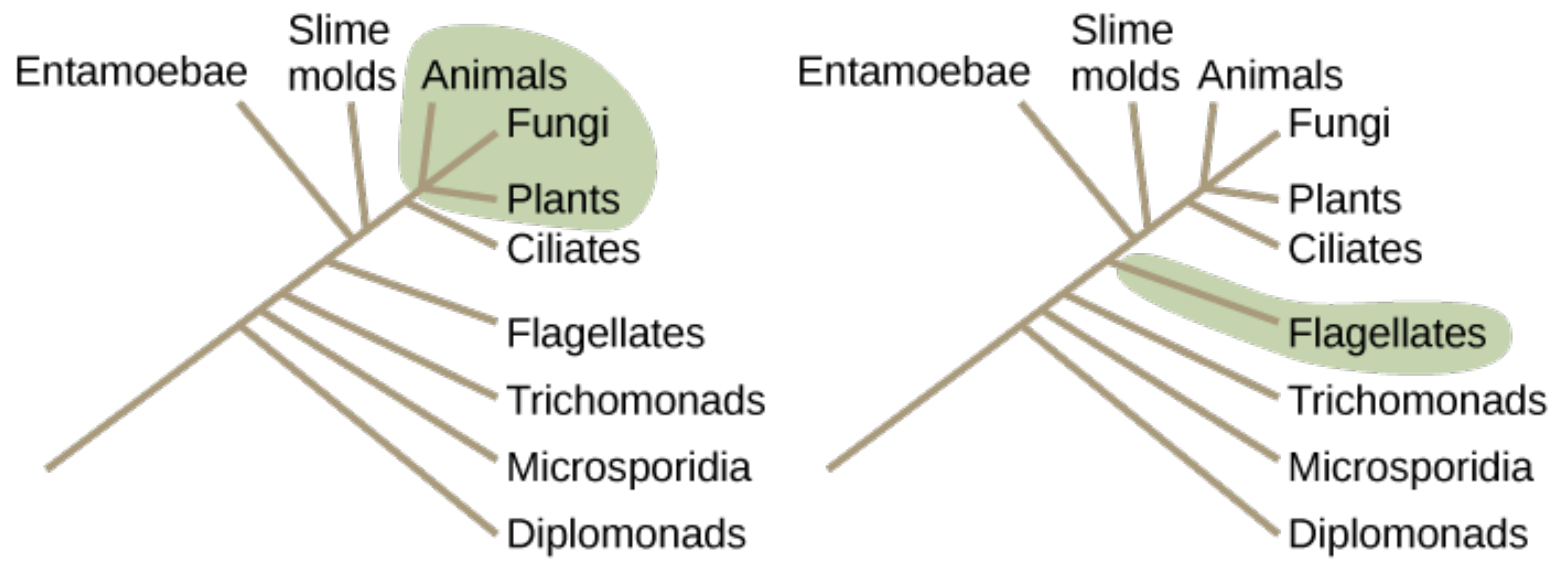
Symplesiomorphies are **ancestral traits shared by multiple taxa** that were inherited from a common ancestor predating the focal clade. These traits are misleading for phylogenetic classification, as they do not indicate close relationships among the taxa that possess them. **Example:** four limbs in mammals, reptiles, and amphibians is a symplesiomorphy inherited from the common ancestor of tetrapods.

Plesiomorphies (Ancestral Traits)

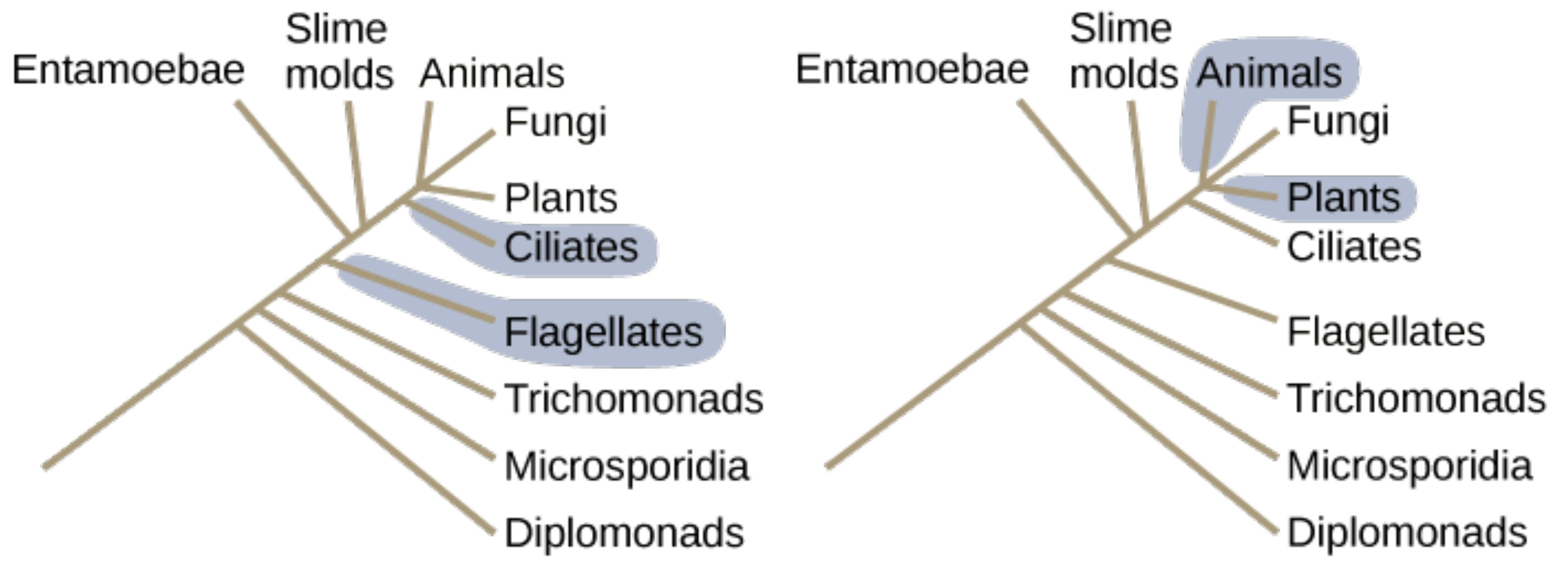
Plesiomorphic traits **originated in a distant ancestor** and are retained by multiple taxa, but they do not define a specific clade. It is a symplesiomorphy discussed in relation to a more derived state. **Example:** vertebrae in mammals is a plesiomorphy because it was inherited from early vertebrates and is also found in fish, amphibians, and reptiles.

Synapomorphies define clades

Clades



Not Clades

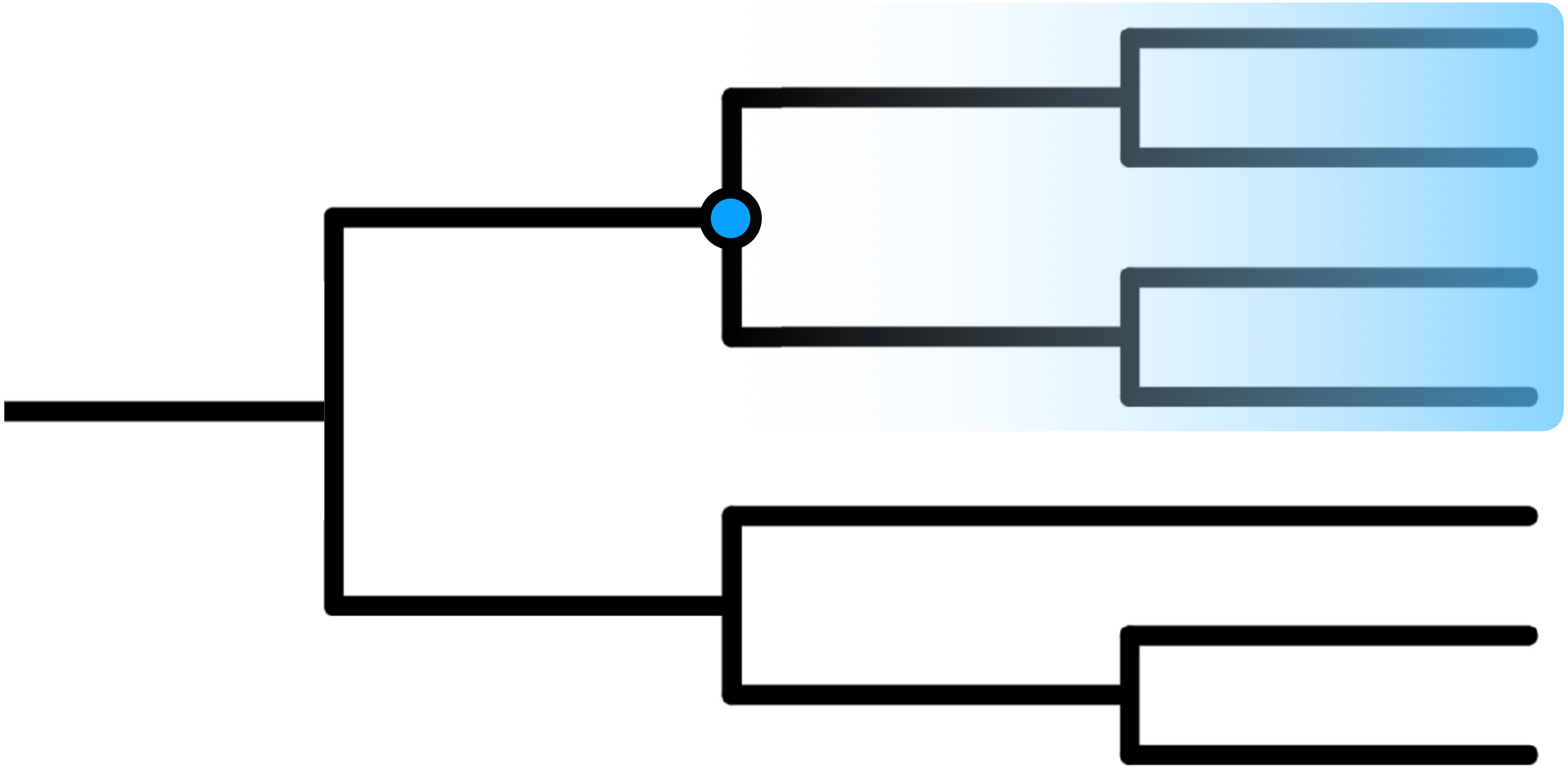


Cladistics? Phylogenetics? Systematics? 😱

- **Systematics** the broadest field ≈ including taxonomy (identification and classification) and phylogenetics (study of evolutionary relationships).
- **Taxonomy** is theory and practice of identifying, describing, naming, and classifying organisms.
- **Phylogenetics** is a subfield within systematics, specifically focused on reconstructing the evolutionary history and relationships among species.
- **Cladistics** is a specific approach within phylogenetics that focuses on common ancestry and shared derived traits.

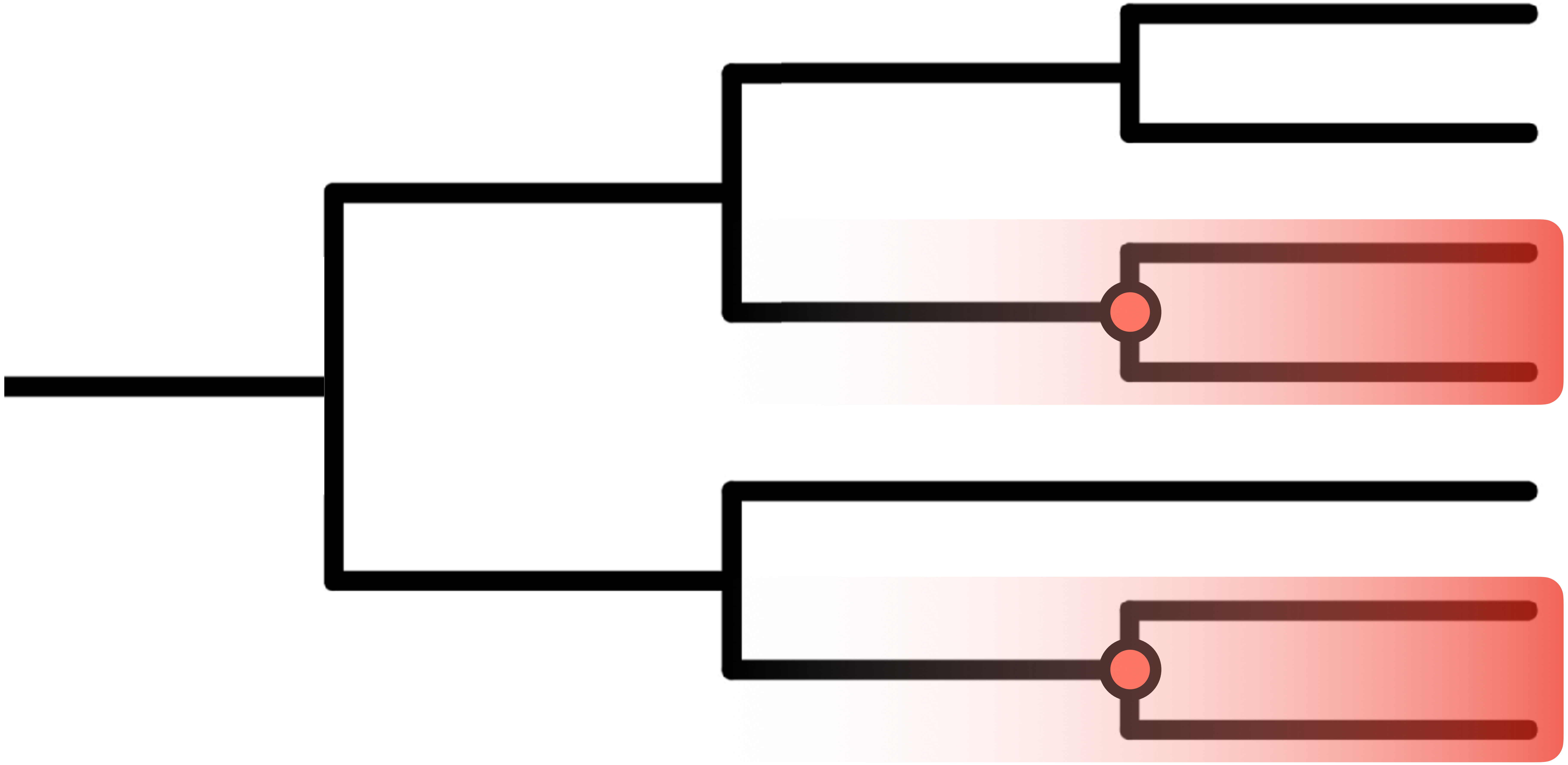
So, in a way … **Systematics > Phylogenetics > Cladistics**

Cladistics is an approch within phylogenetics, which is part of systematics.
All cladistics is phylogenetics, but not all phylogenetics is cladistics.



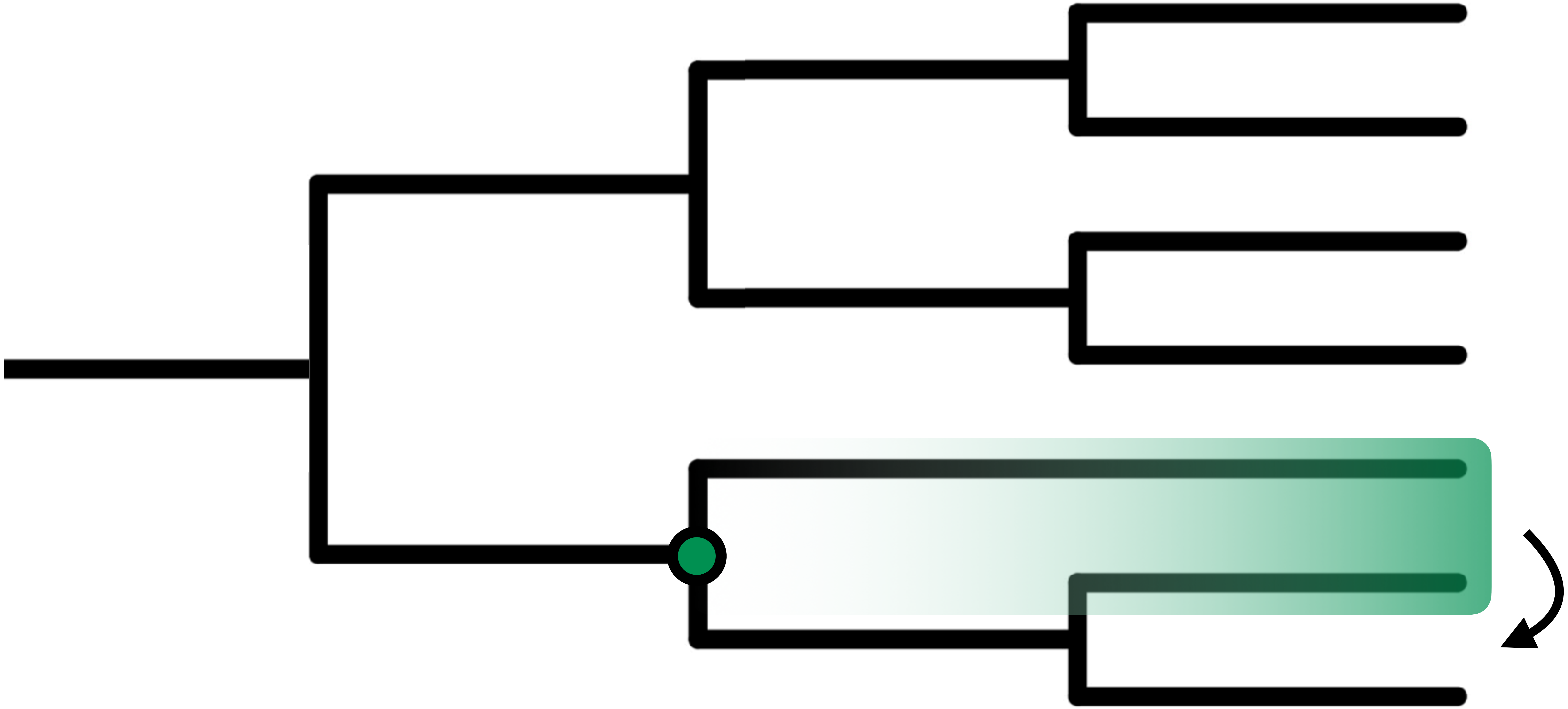
monophily

A grouping of organisms which contains: **(1)** its own most recent common ancestor, i.e. excludes non-descendants of that common ancestor; **(2)** all the descendants of that common ancestor, without exception. The condition of a taxonomic grouping being a **clade**.



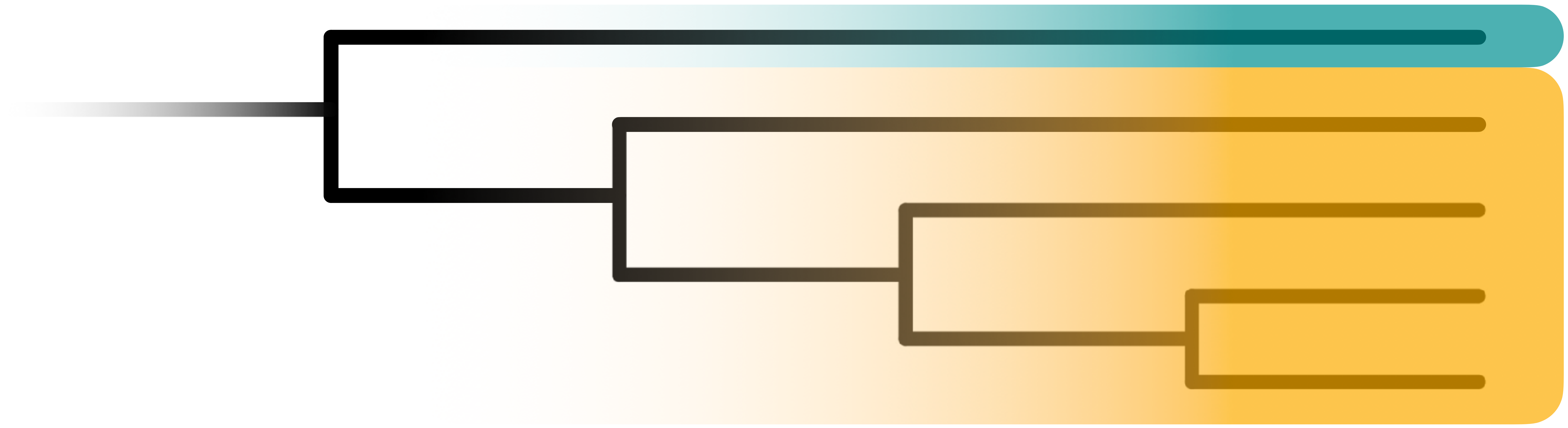
polyfily

A polyphyletic group is an assemblage that includes two or more **separate groups**, each with a **separate common ancestor**. The most recent common ancestor of the species in the polyphyletic assemblage would be the ancestor also of species not included in them.



parafyly

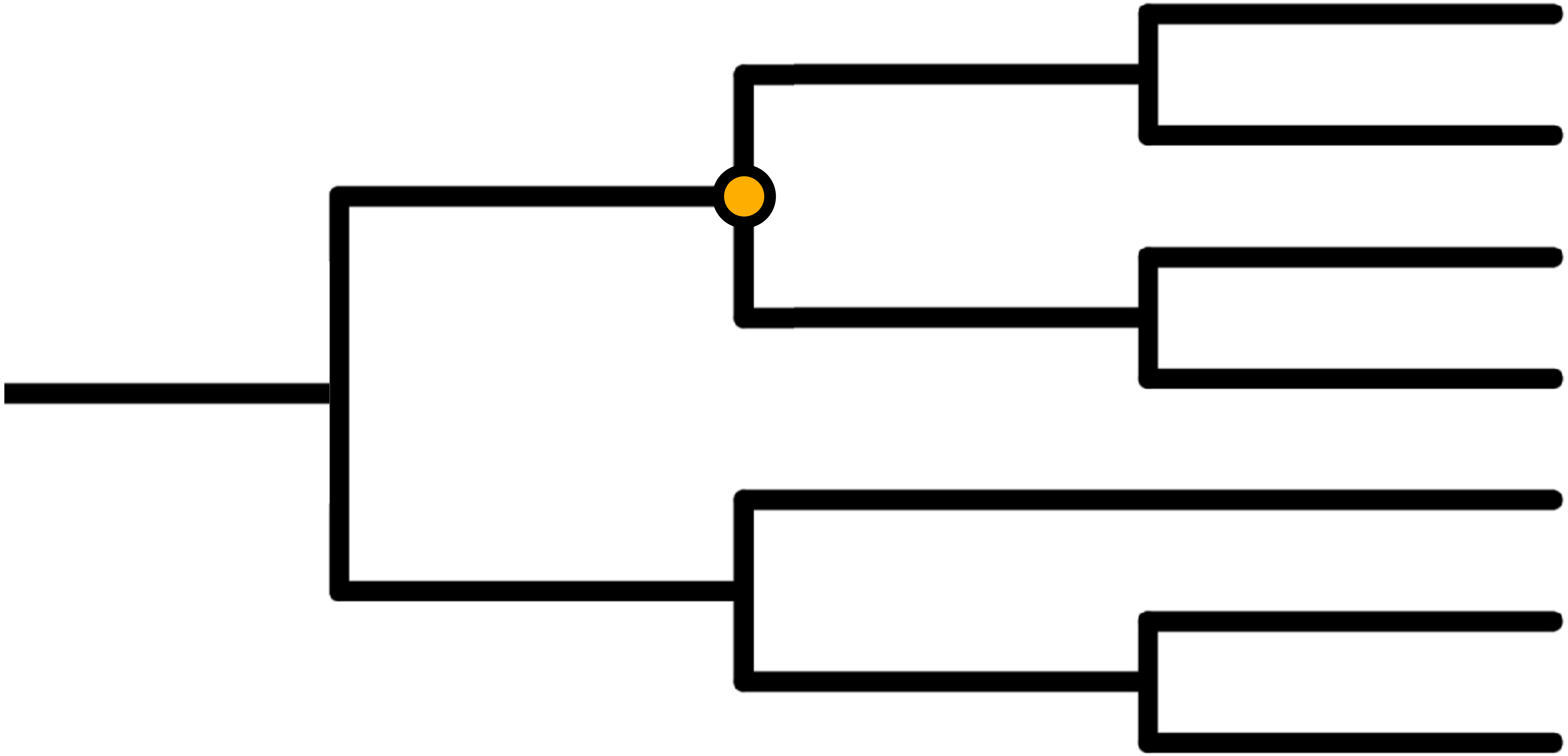
A group that consists of the grouping's **last common ancestor and some but not all of its descendant** lineages. The grouping is said to be paraphyletic with respect to the excluded subgroups.



ingroup and outgroup

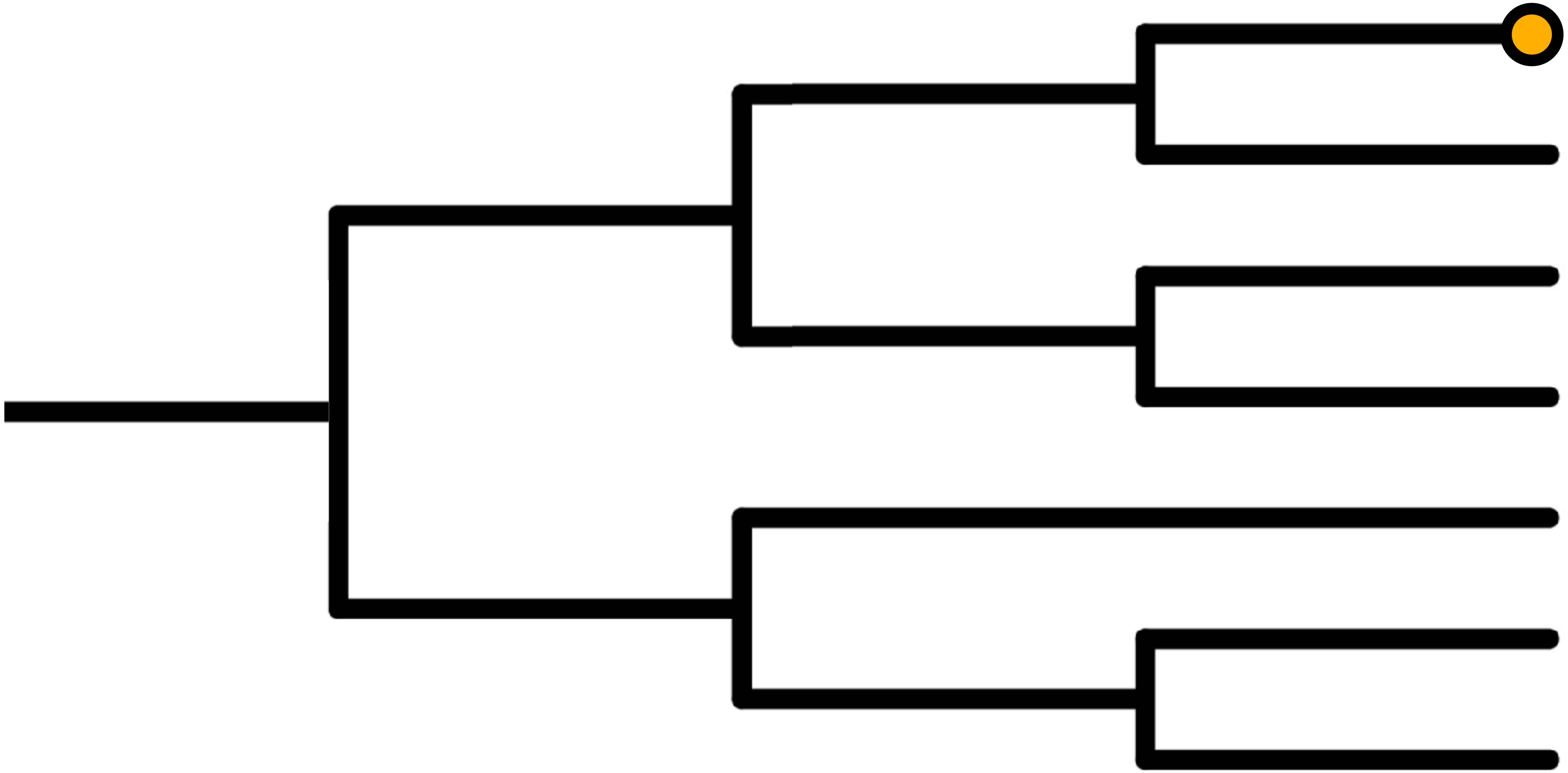
- ingroup: a group of tips assumed a priori to represent a clade, and the focus of our phylogenetic analyses.

- outgroup: a group of tips assumed a priori to lie outside the monophyly of the focal clade. It also serves to give a direction to our tree.



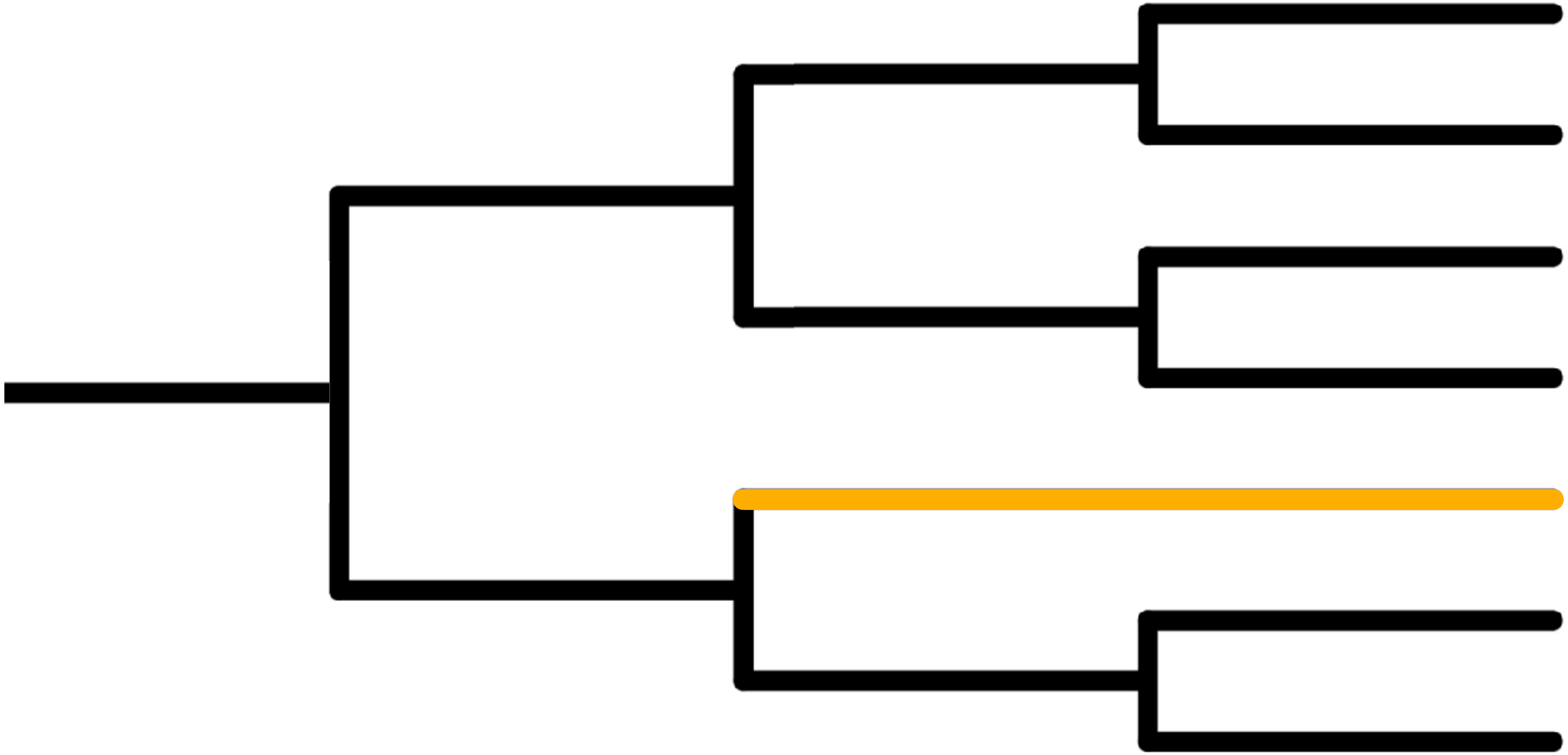
node

Nodes are the points at the ends of branches which represent **real sequences** or **hypothetical sequences** at various points in evolutionary history. In a tree of species an internal node in a tree is, biologically, an **ancestor**. You can also thin to nodes as **bipartitions or splits**.



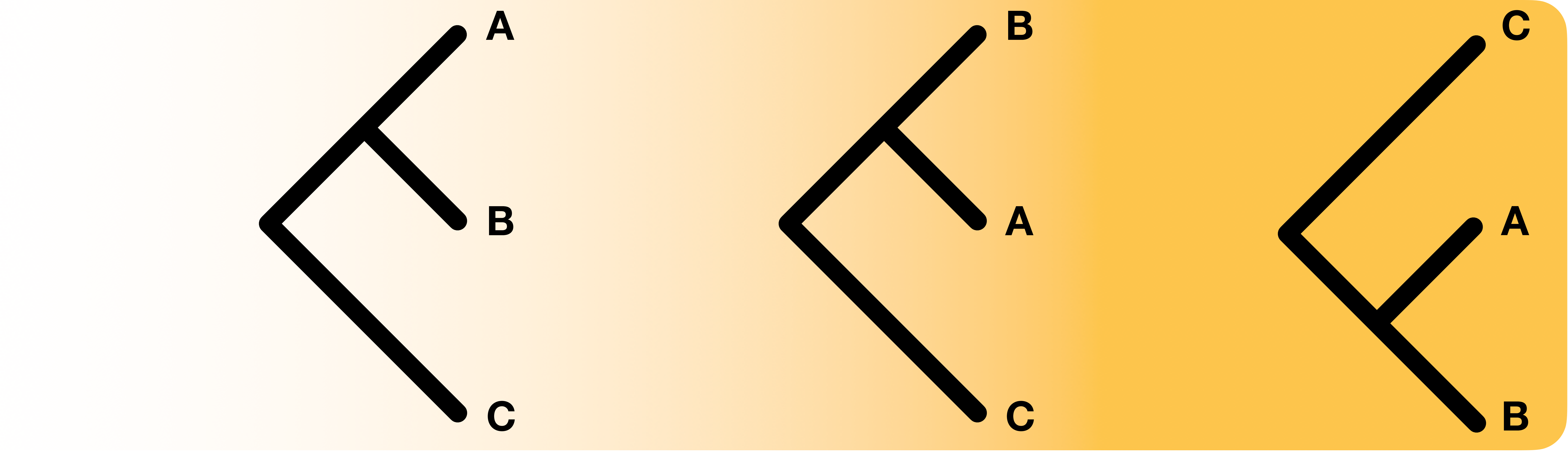
nodes (terminal or external)

The sequence that we sampled and used to construct our phylogeny occur on single terminal branches, known as **tips or leafs**. Often referred to as **OTU**, which stands for Operational Taxonomic Unit, a noncommittal term used for the objects of study (be they species, populations or individuals).

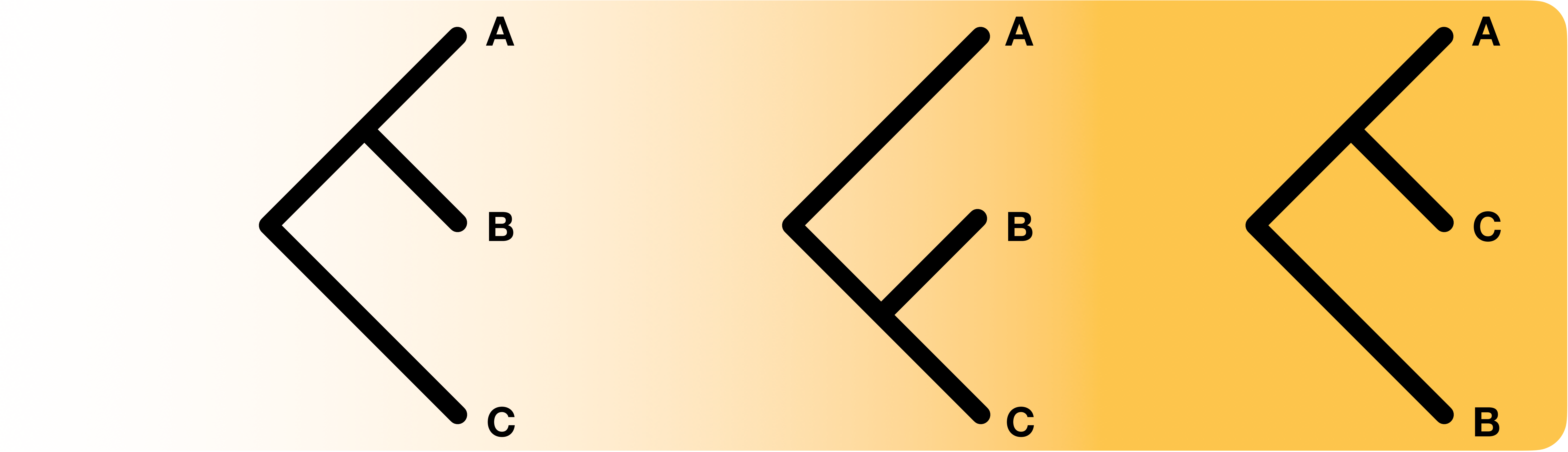


branch

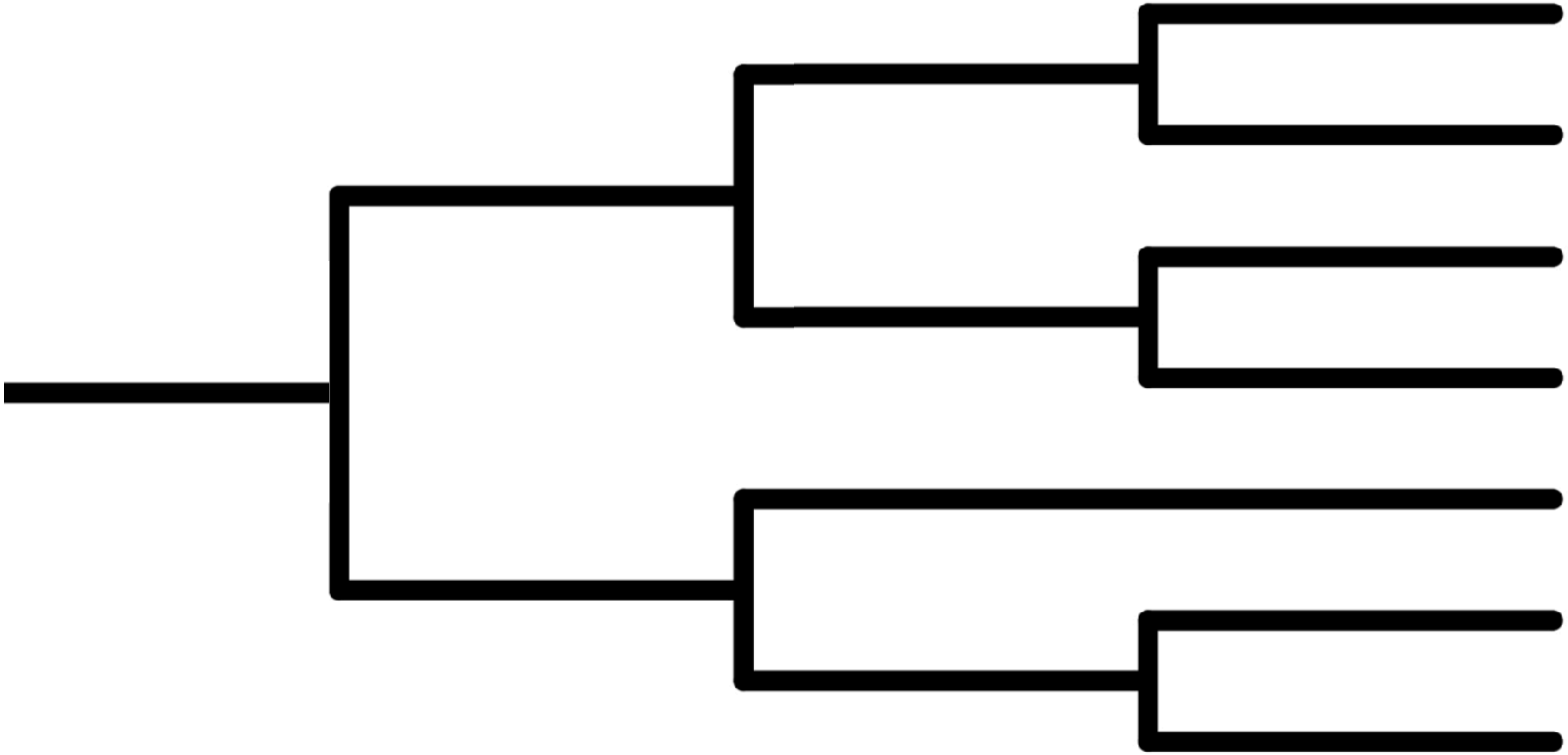
The branching structure of the tree is its **topology**. It is of particular significance because it indicates **patterns of relatedness** among taxa. Beware: any internal node can be rotated and the tree is the same.



SAME TREES

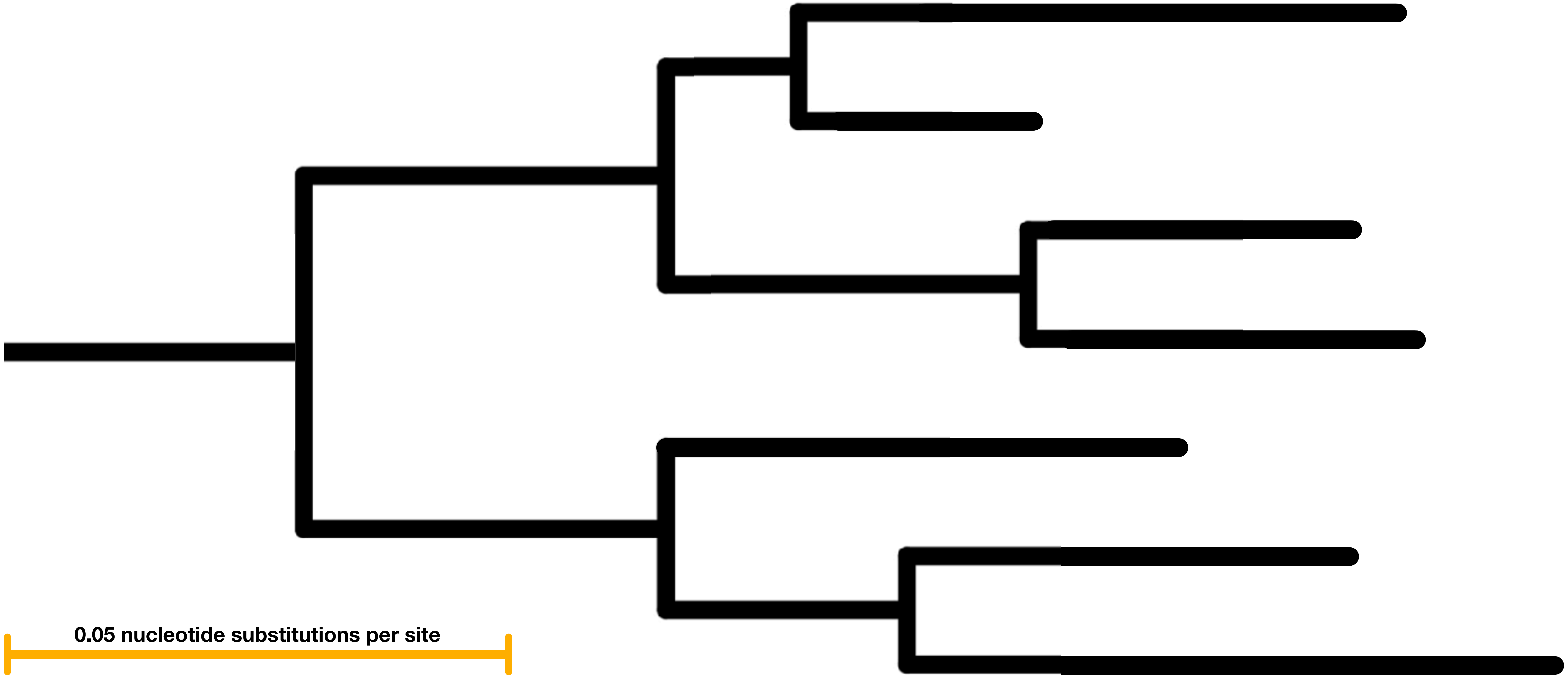


DIFFERENT TREES



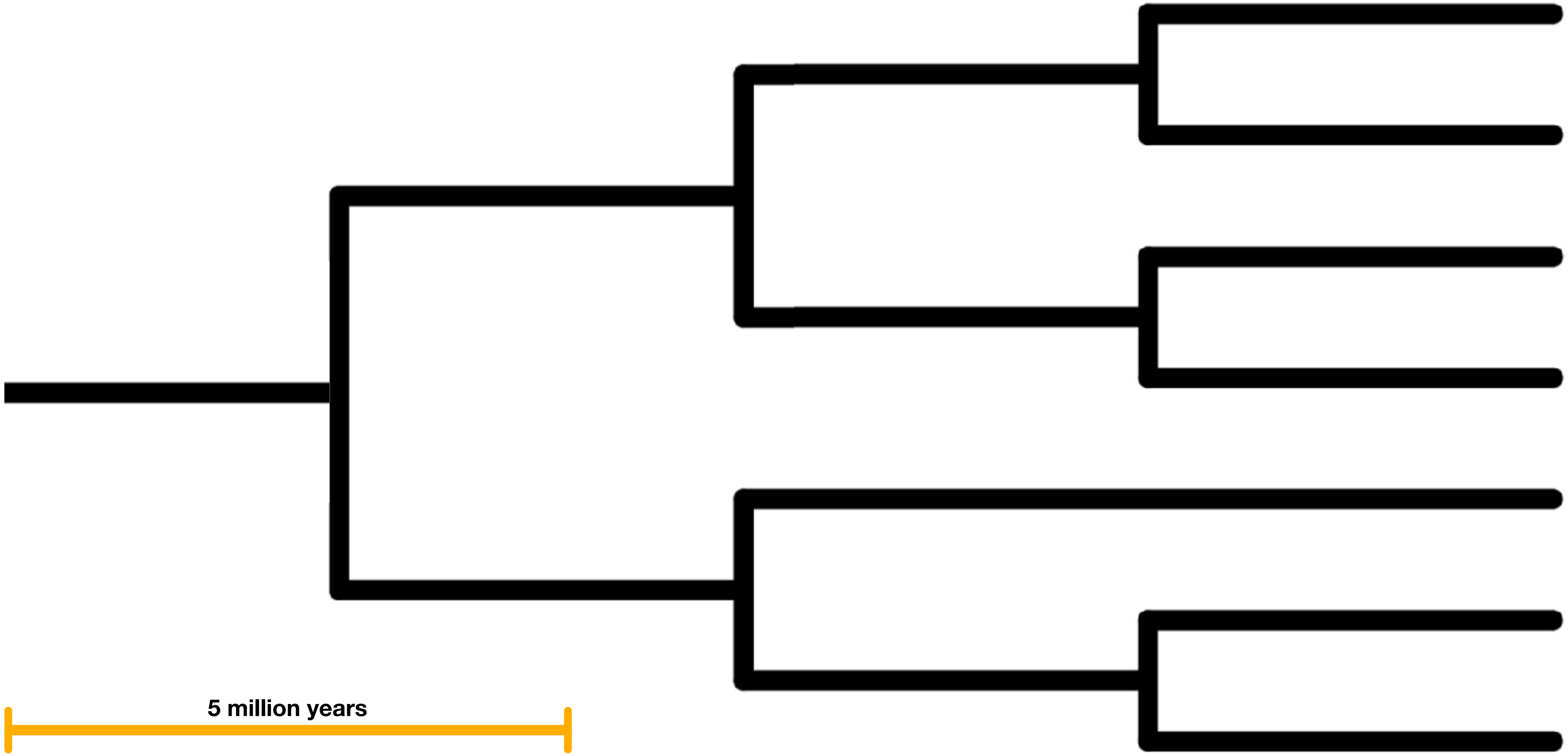
cladogram

A cladogram only represents a branching pattern; **branch lengths do not have a meaning** and do not represent either time or relative amount of character change. They are just a function of the order of branching on the tree.



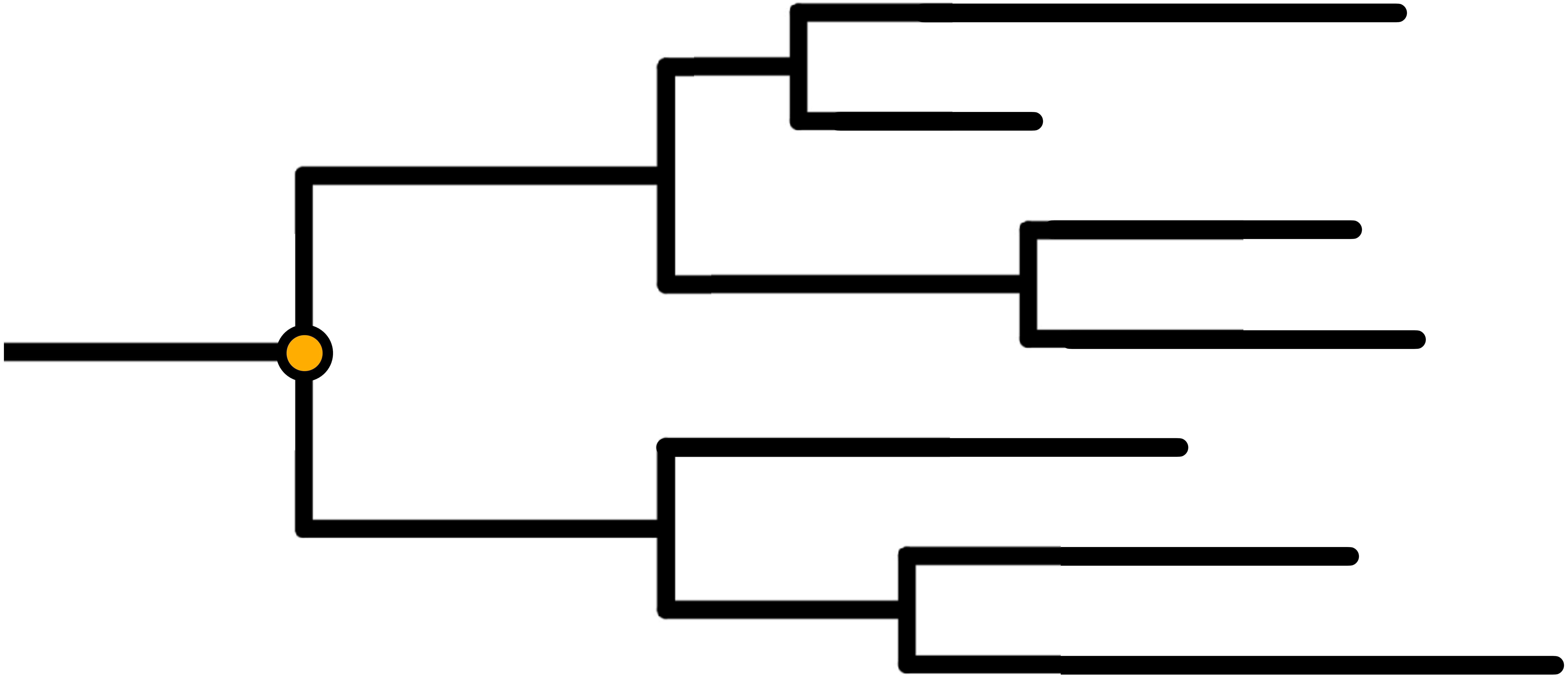
scale: evolutionary change

The line segment shows the length of **branch that represents an amount of change** of 0.05. The vertical dimension in this figure has no meaning and is used simply to lay out the tree visually with the labels evenly spaced vertically. This tree is a **phylogram**.



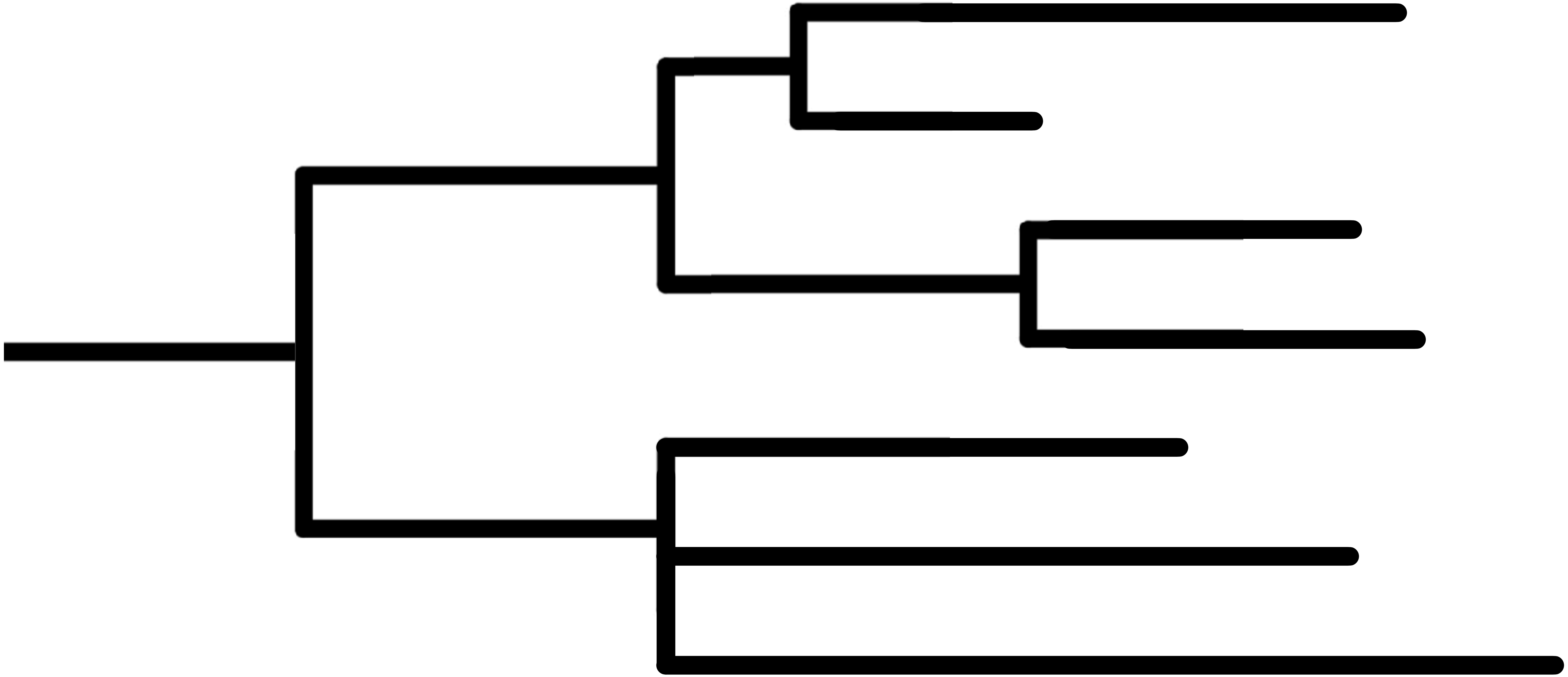
scale: time

A chronogram is a phylogenetic tree that explicitly represents **time through branch lengths**. This tree is a **chronogram** or **timetree**. The process of obtaining such a phylogeny is often called a **divergence time analysis**.



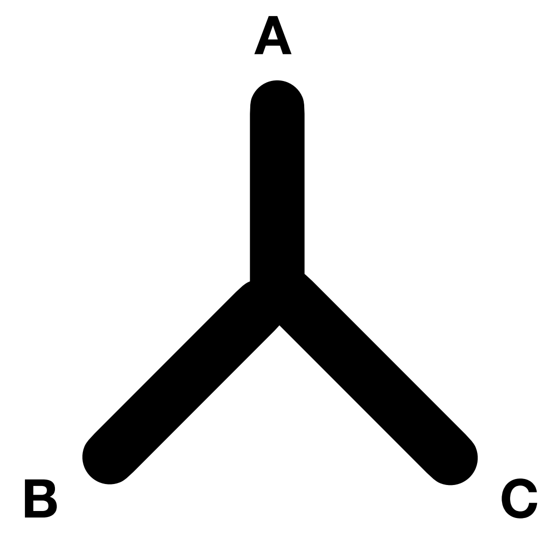
root

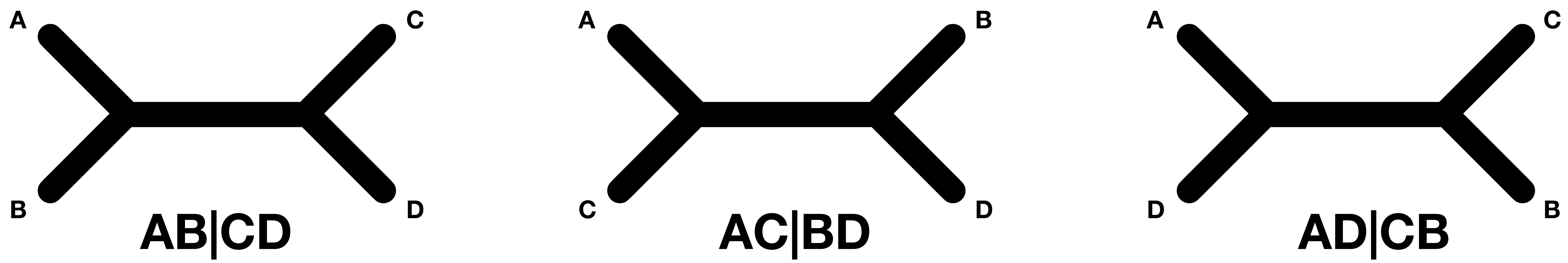
The root is a specific internal node representing **the most recent common ancestor** of all tips in the tree. It is therefore the oldest part of the tree and tells us the direction of evolution. There are two main approaches that we can use to root a tree: **outgroup rooting** and **midpoint rooting**.

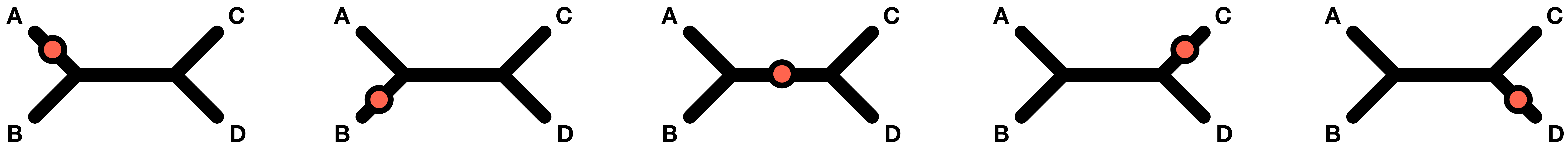


polytomy

Trees can either be **bifurcating** (fully dicotomous) or **multifurcating** (containing a politomy). A polytomy is a node in a phylogenetic tree where **more than two descendant lineages** emerge, indicating an unresolved evolutionary relationship.

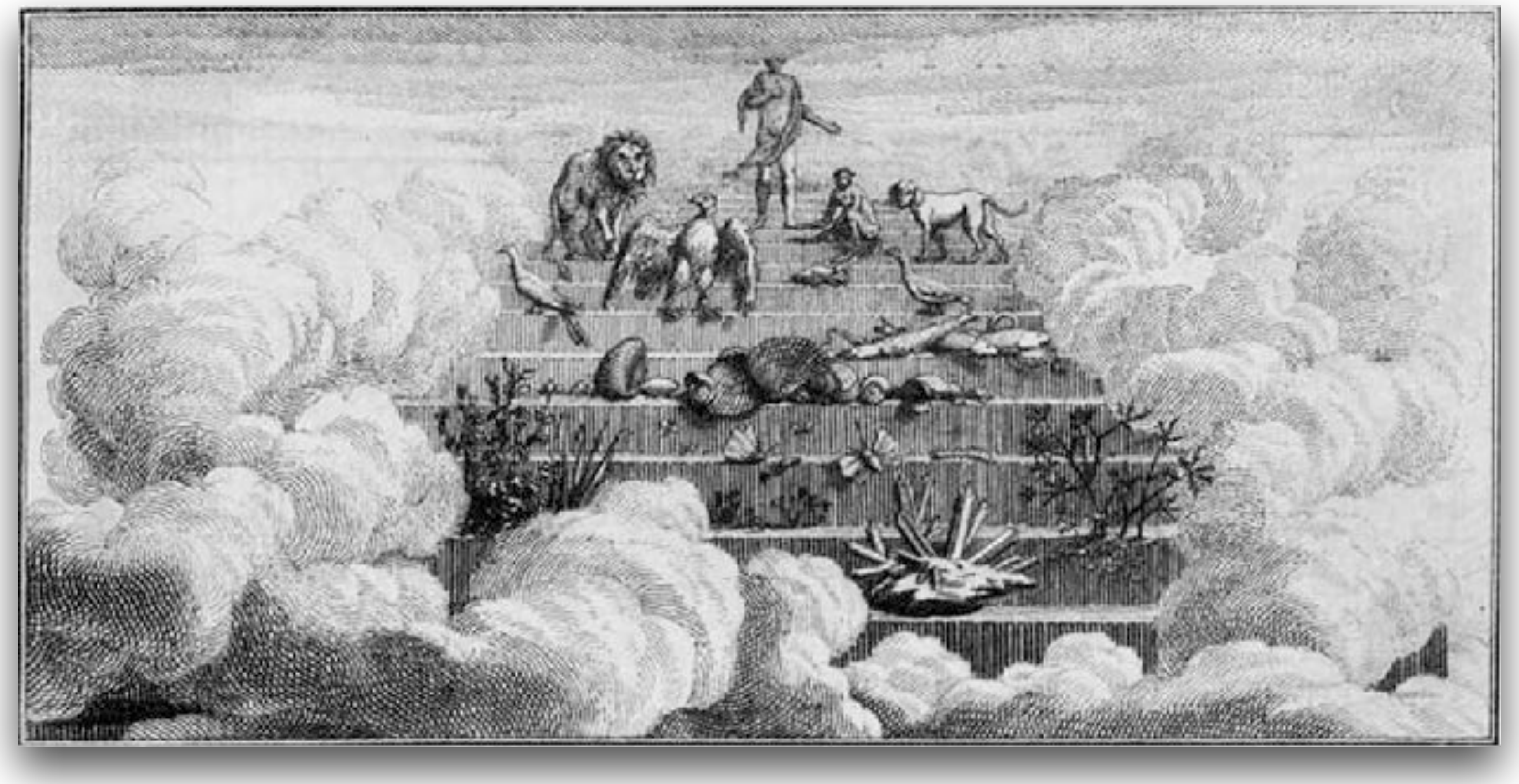






root?

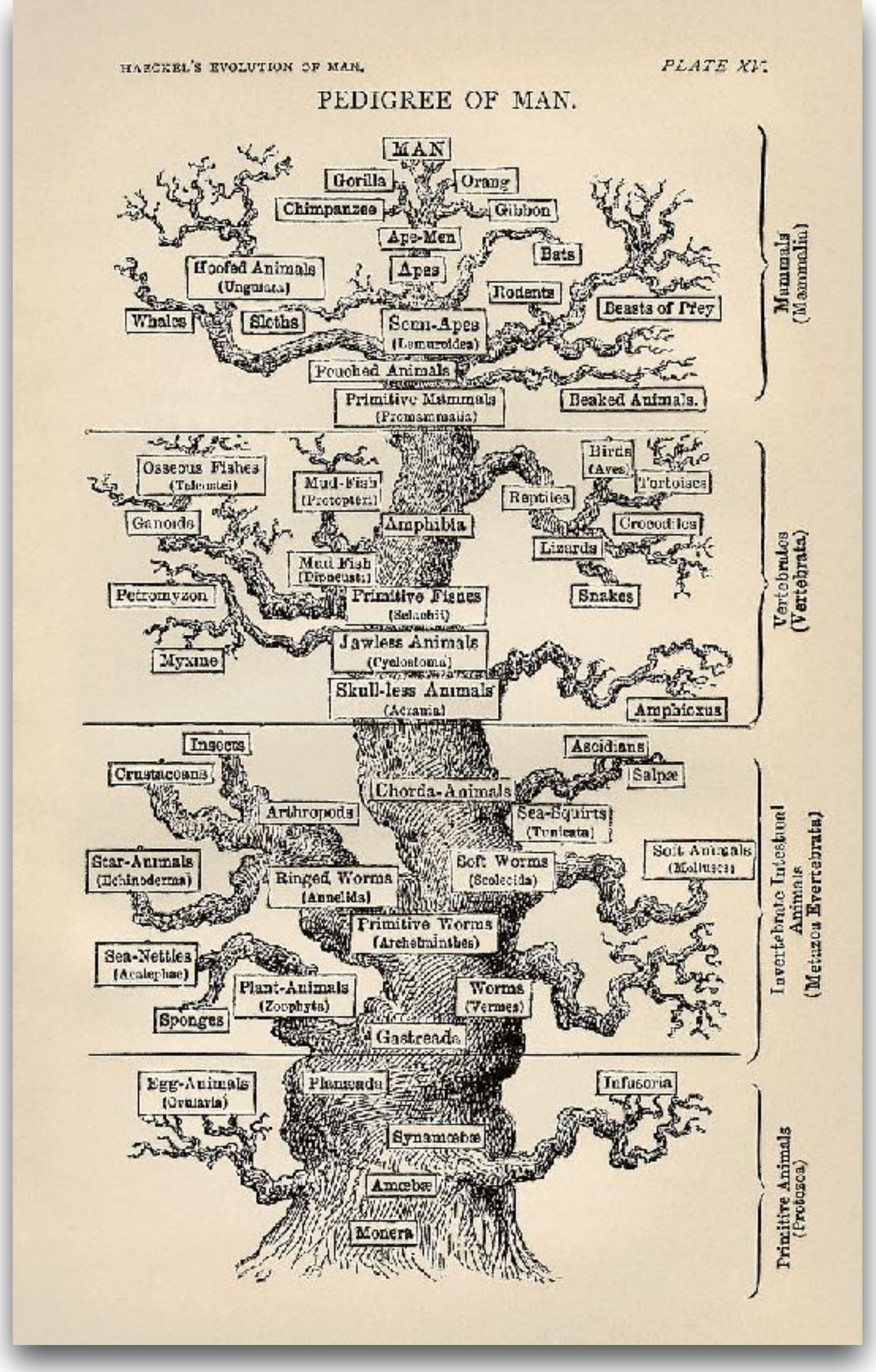
The simplest ambiguous tree is a **quartet**, the unambiguous one a triplet. Trees may be rooted or unrooted. In a rooted tree, each node with some descendants represents the most recent common ancestor of those descendants. In the practical we will see that complexity grows quickly 😱.



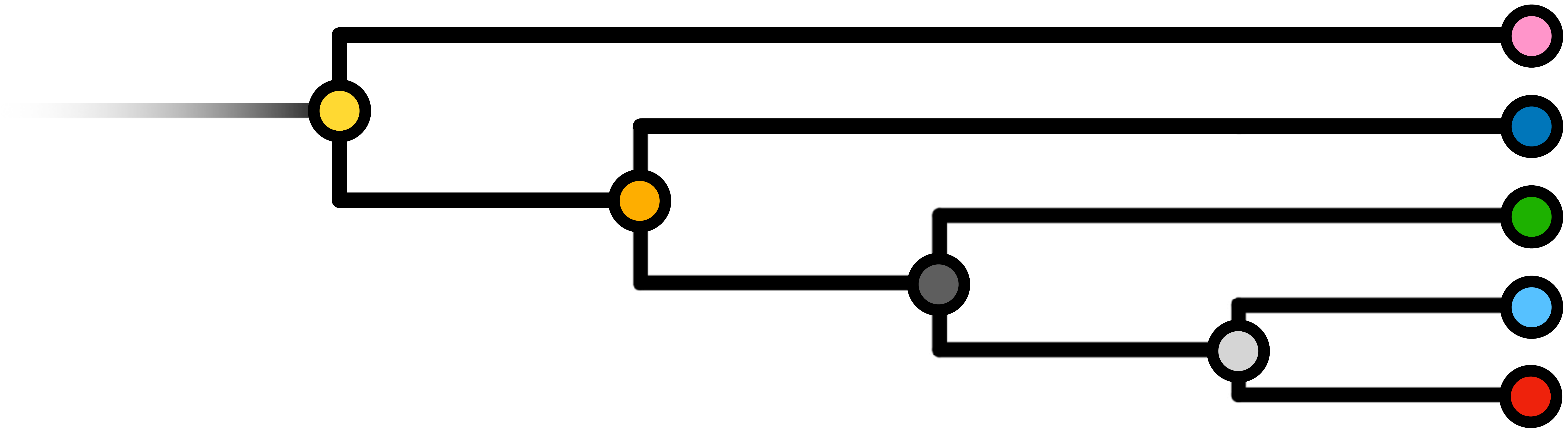
Kette der Wesen, Scala Naturae by Charles Bonnet (1781).

trees are not ladders

The ladder concept (Scala Naturae) is an old view that tend to organize ancestry based on complexity or some putative evolutionary superiority. Terms as “primitive”, “ancient” or “lower” referred to a lineage are misleading as they imply ladder-thinking 😑.



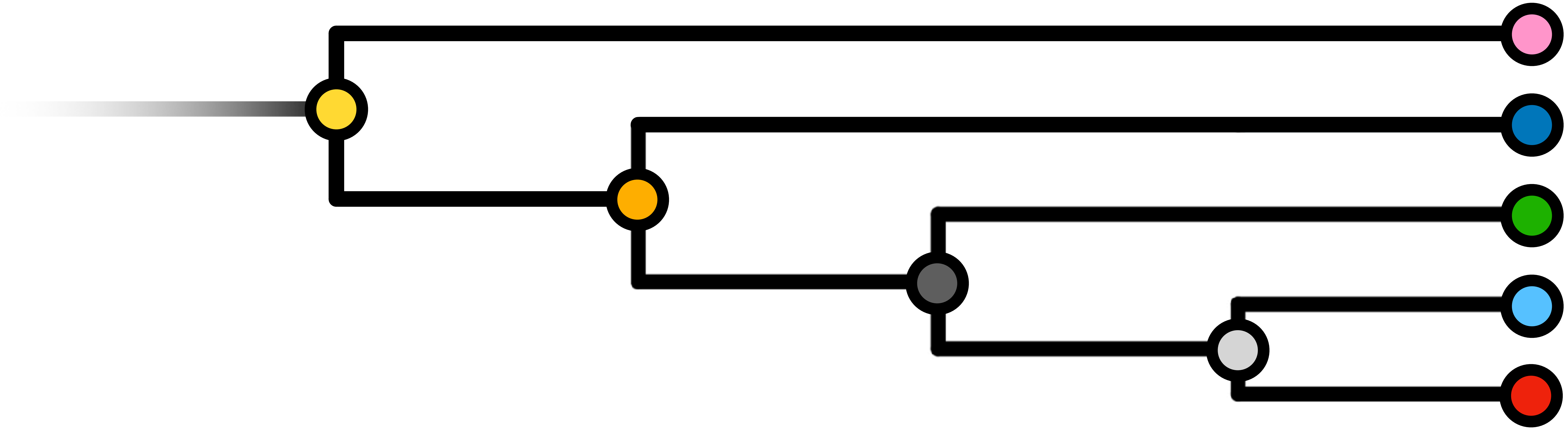
Ernst Haeckel's tree from the The Evolution of Man (1879).



The **red tip** is more closely related tot he **blue tip** than to the **green tip** because it share a common ancestor more recently with the **blue tip** (ancestor = **light gray node**) than they do with the **green tip** (ancestor = **dark gray node**).

The **purple tip** is more closely related to the **green tip** then they are to the **pink tip** because the **purple tip** shares a common ancestor with the **green tip** more recently (ancestor = **orange node**) than they do with the **pink tip** (ancestor = **yellow node**).

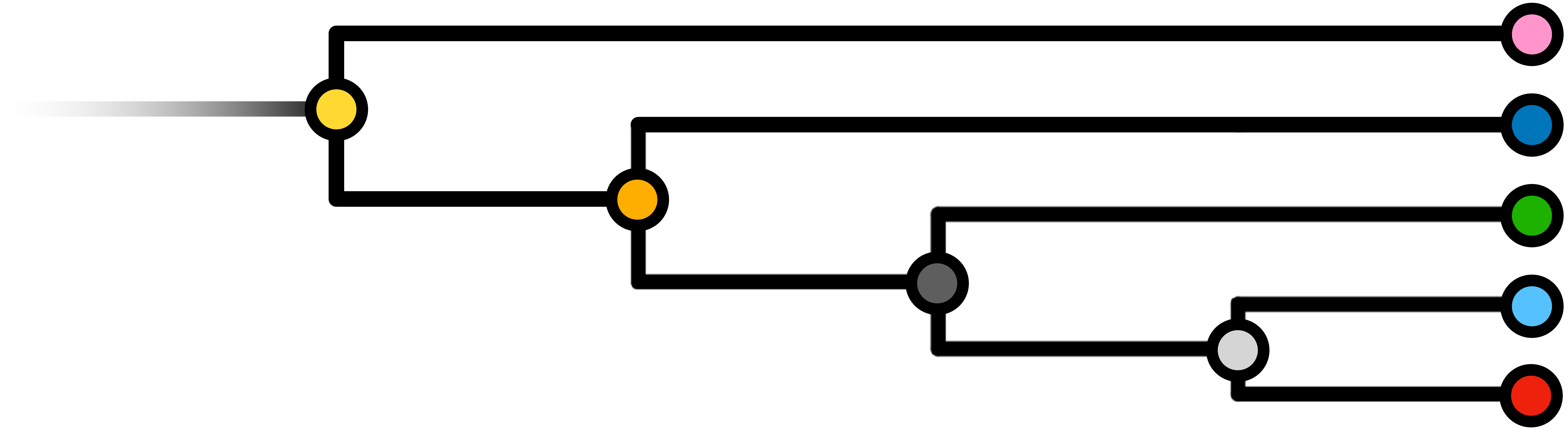
The **pink tip** is equally related to blue as they are to the **purple tip**. This is less intuitive, but if you trace back to the MRCAs you will see why: the **blue tip** and **purple tip** share the same common ancestor (**yellow node**) with **pink tip**, so neither species is more closely related to **pink tip**.



Common misconceptions pt. 1

Taxa that are adjacent on the tips of phylogeny are more closely related to one another than they are to taxa on more distant tips of the phylogeny.

In a phylogeny, information about relatedness is conveyed by the pattern of branching, not by the order of taxa at the tips of the tree. Organisms that share a more recent branching point (i.e., a more recent common ancestor) are more closely related than are organisms connected by a more ancient branching point (i.e., one that is closer to the root of the tree).



Common misconceptions pt. 2

Taxa that appear near the “bottom” of a phylogeny are more “advanced” than other organisms on the tree.

Taxa that are nearer the “top” of a phylogeny represent the ancestors of the other organisms on the tree.

IMPORTANT

- A phylogenetic tree is always an **hypothesis** about how species are related through evolution.
- Evolution describes events occurring in the **past** and that cannot be directly observed.
- Phylogenetic trees are **inferred** from data and are subject to revision as new evidence emerges.

SOME LEXICON:

rooted *versus* unrooted tree
clade
polytomy
monophily
poliphily
parafyly
internal node
terminal node
branch
bipartitions
cladogram
phylogram
chronogram or timetree
dicotomous *versus* non-dicotomous
automorphy
sinapomorphy

FINISH