

IBIO, Molecular Phylogenetics 2024

# Lecture 1: Introduction

Jadranka Rota and Niklas Wahlberg

Systematic Biology Group

Department of Biology

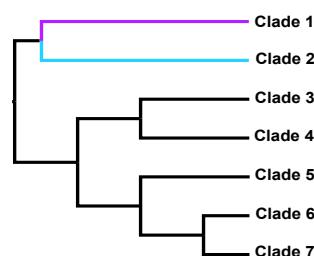
Lund University



1

## Aims

- Cover key concepts in phylogenetics
  - E.g. monophyly, homology, analogy
- Explain why evolutionary history is important in biology
- Understand the basics of statistical phylogenetic inference
- Develop “tree thinking”

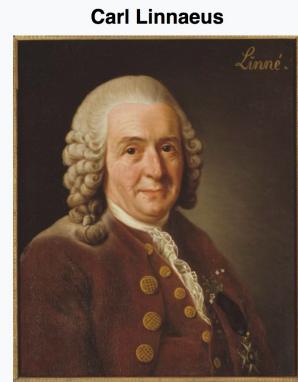


2

1

## Systematics is ....

- The scientific **study of the kinds and diversity of organisms and of the relationships among them**
- Traditionally: taxonomy (naming and classifying organisms, Greek "taxis" = arrangement, and "nomos" = law)
- Since ca. 1980s: largely based on phylogenetics (first morphological, then molecular)
- Most recently: including phylogenomics
- Provides essential framework for recognition and study of biodiversity and evolution



*Carl von Linné* by Alexander Roslin, 1775  
(oil on canvas, Gripsholm Castle)

Known as the “Father of modern taxonomy”  
(Source: Wikipedia)

3

## Human need for taxonomy

- Naming of organisms around us
  - Makes communication a lot easier!
- Sorting different groups into higher categories
  - Helps us organize the living world around us
  - E.g. porcini (Karljohanssvamp) is an edible species of mushrooms, but many other mushrooms are very poisonous, e.g. fly amanita (Röd flugsvamp)



Porcini - *Boletus edulis* Bull. (1782)



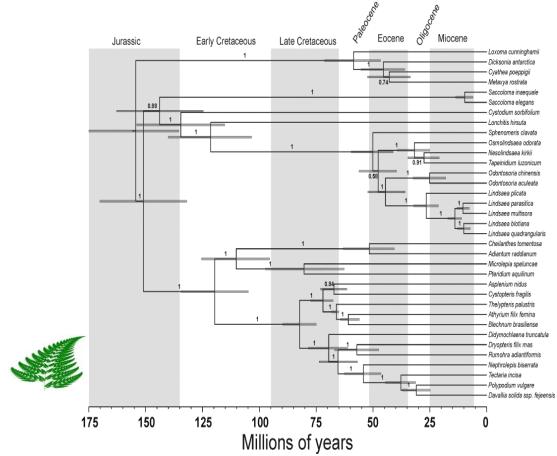
Fly amanita - *Amanita muscaria* (L.) Lam. (1783)

**Note:** one taxon, many taxa (in some Greek words singular ending is –on, plural is –a, e.g. also phenomenon, phenomena)

4

## Systematics also includes...

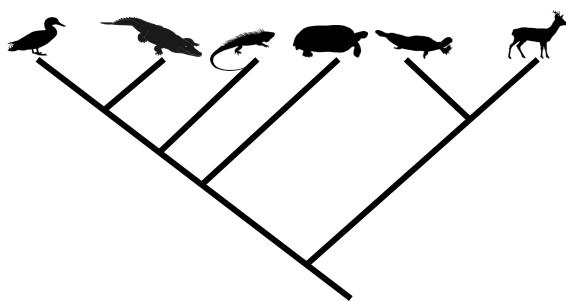
- The study of character evolution
- The study of molecular evolution
- The study of speciation/extinction dynamics
- The study of historical biogeography
- The study of the temporal framework of evolution



5

## Evolutionary History

- How do we learn about the evolutionary history of organisms?
- Why should we care about it?



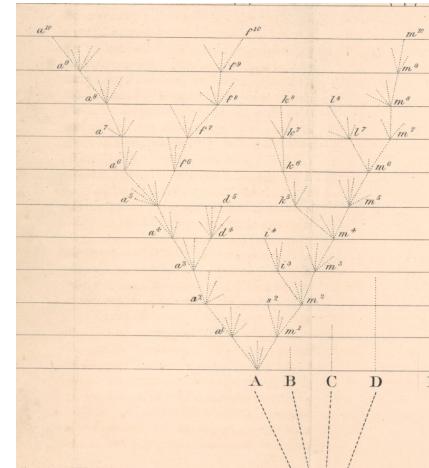
6

**“Nothing in biology makes sense except in the light of evolution”**

- Theodosius Dobzhansky, essay written in 1973

**“Nothing in evolution makes sense except in the light of phylogeny”**

- Quote from Jay Savage, in 1997 Society of Systematic Biology Presidential Address

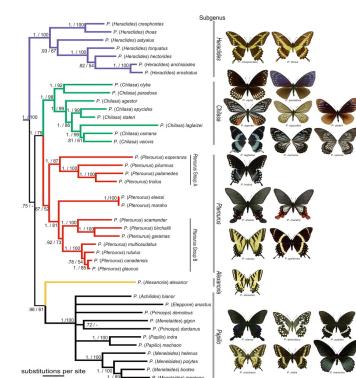


The only figure in Darwin's On the Origin of Species

7

## The very basic facts

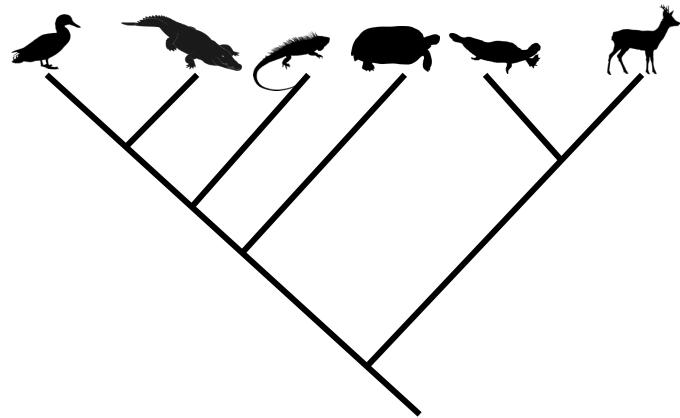
- What we see today in nature is the outcome of what has happened in the past
- Ecology and evolution are inseparable
- “Species” or “genes” are not individual entities without any connections to other species or genes
  - phylogeny



8

## What is a phylogeny?

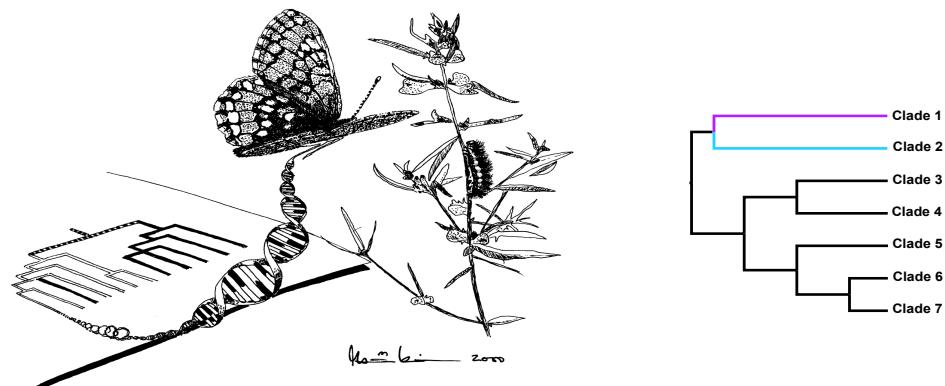
- A phylogeny is the historical genealogy of a group of species



9

## A phylogeny is an inference

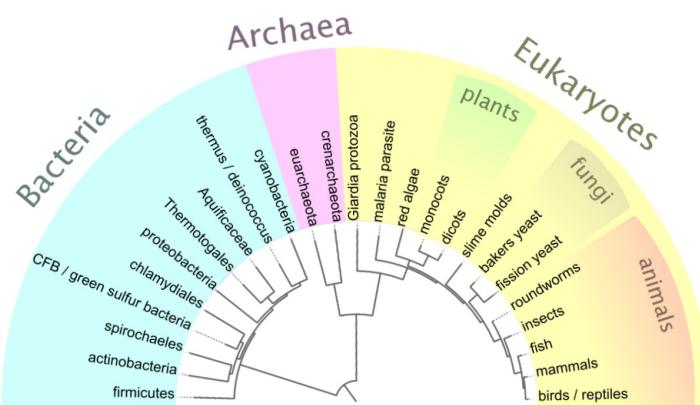
- Envisioned as a dichotomously branching tree
- A phylogeny cannot be observed
- A phylogenetic hypothesis can be inferred from observed data



10

## What we are after

- **Phylogenies – the Tree of Life**
- **With phylogenies we are attempting to get a good working framework for Life**
- **Getting to the root of how evolution has worked**



Source: Wikipedia

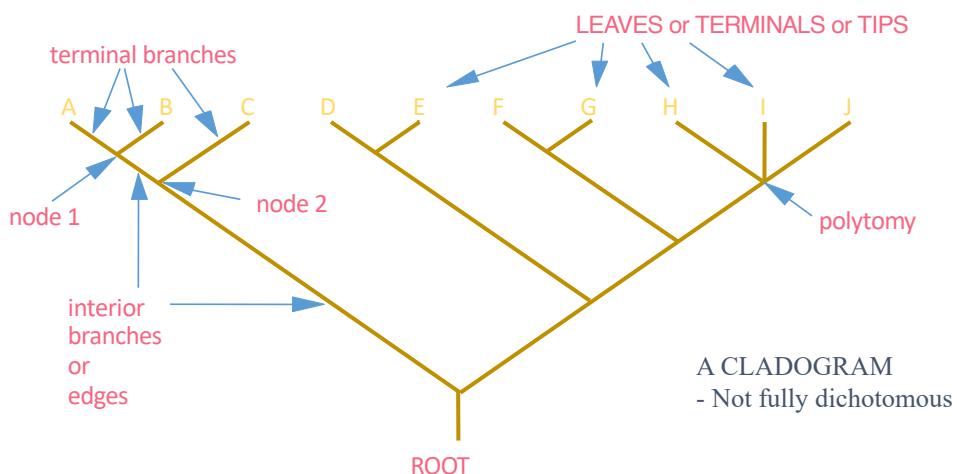
11

## Some basic concepts

- **Cladogram** – a tree diagram which depicts a hypothesised evolutionary history (topology)
- **Phylogram** – a tree which indicates by branch length the degree of change believed to have occurred along each lineage (topology with informative branch lengths)
- **Chronogram** – a tree in which branch lengths are directly in proportion to time (a type of an ultrametric tree – all tips are equidistant from the root)

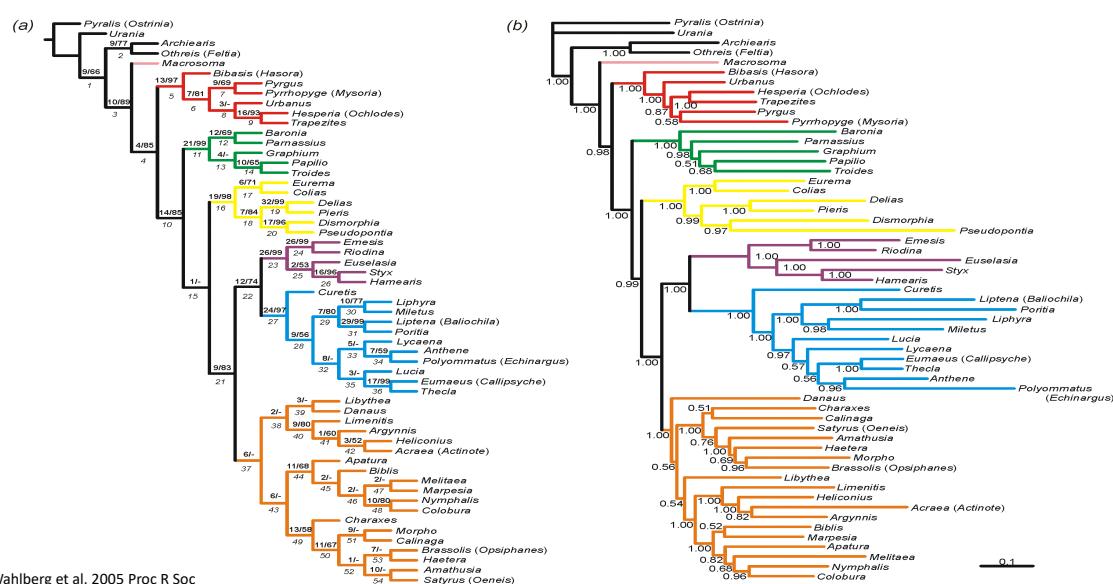
12

## Phylogenetic Trees



13

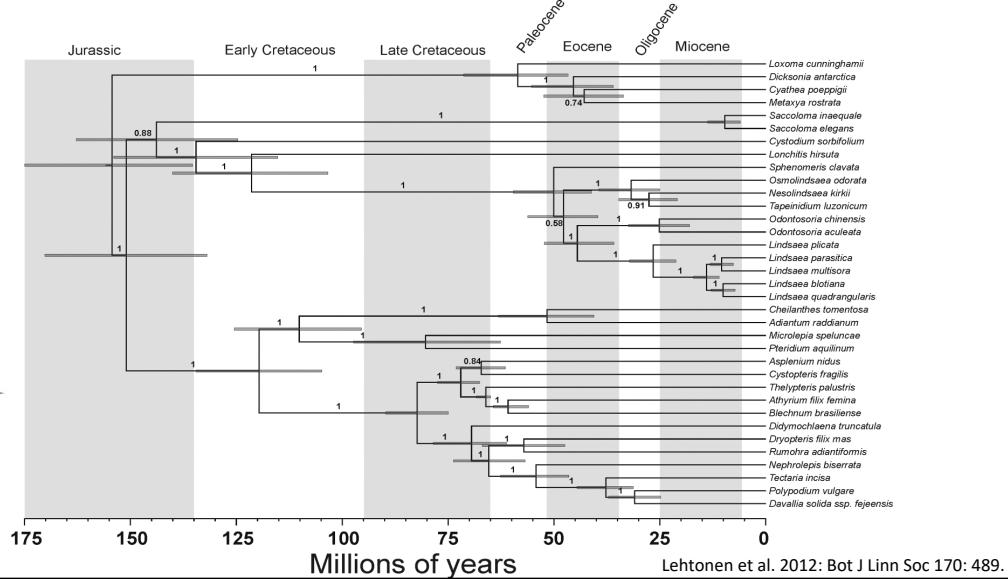
## Cladograms and phylogenograms



14

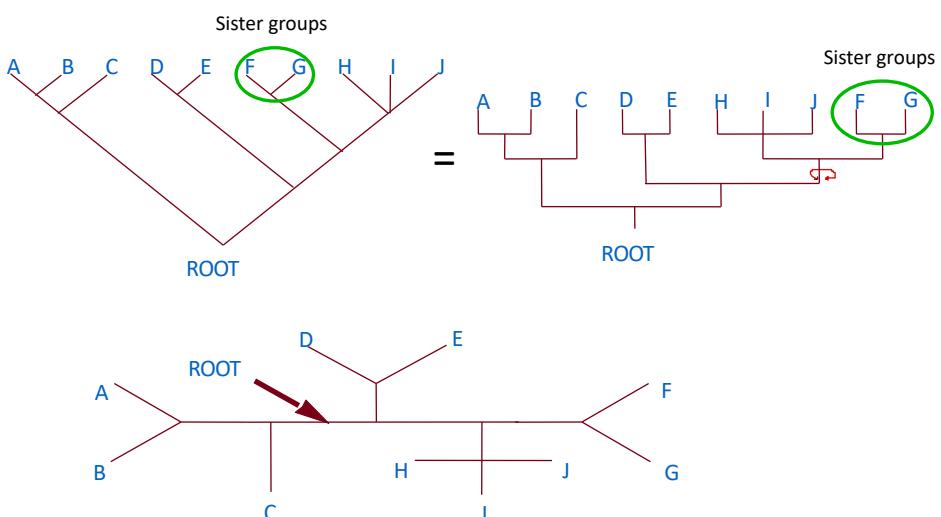
## Chronogram

(an ultrametric tree – all tips equidistant from the root)



15

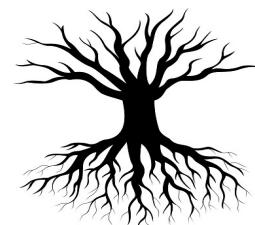
## Trees - Rooted and Unrooted



16

## Rooting a tree

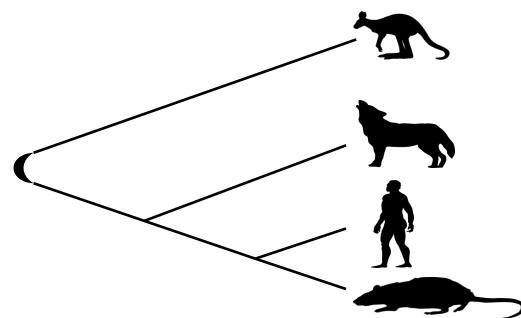
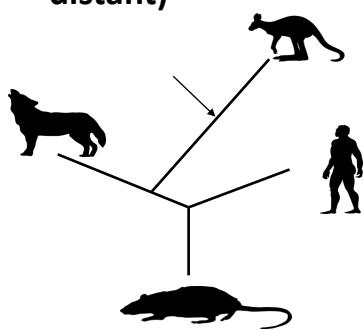
- Rooting a tree using outgroups
  - Commonly we include several outgroups
  - Place the root on the branch leading to the outgroup taxon
- Other ways of rooting a tree
  - Assume a molecular clock
  - Midpoint rooting (root on the longest branch)



17

## Outgroup rooting of unrooted trees

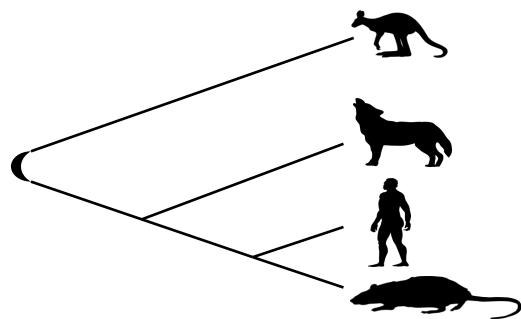
- Outgroup – related group that definitely diverged earlier (palaeontological evidence)
- Not too distantly related (tree method becomes unreliable if it is too distant)



18

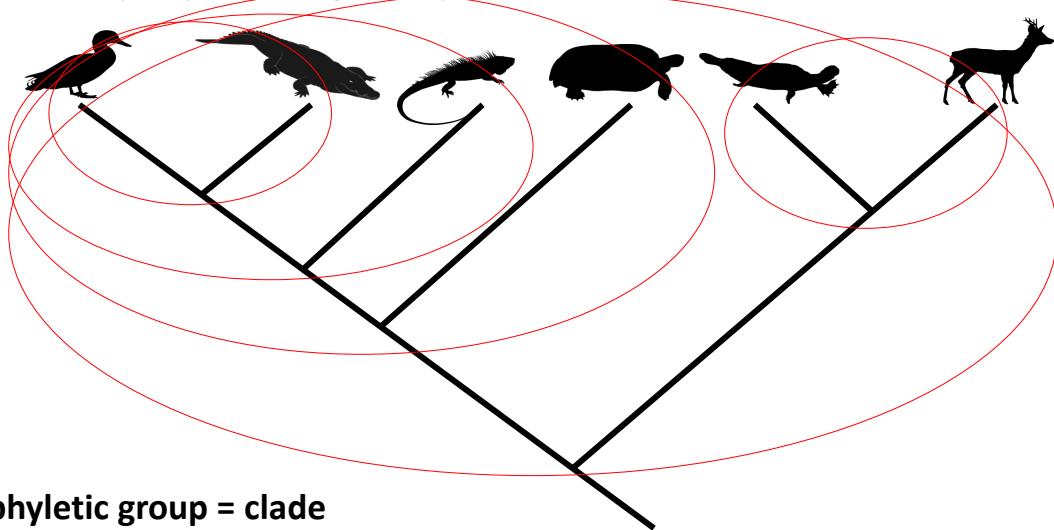
## Phylogenetic systematics

- Uses tree diagrams to portray relationships based upon recency of common ancestry
- **Monophyletic groups (clades)** – contain species which are more closely related to each other than to any outside of the group, including the MRCA (most recent common ancestor)



19

## Monophyletic groups

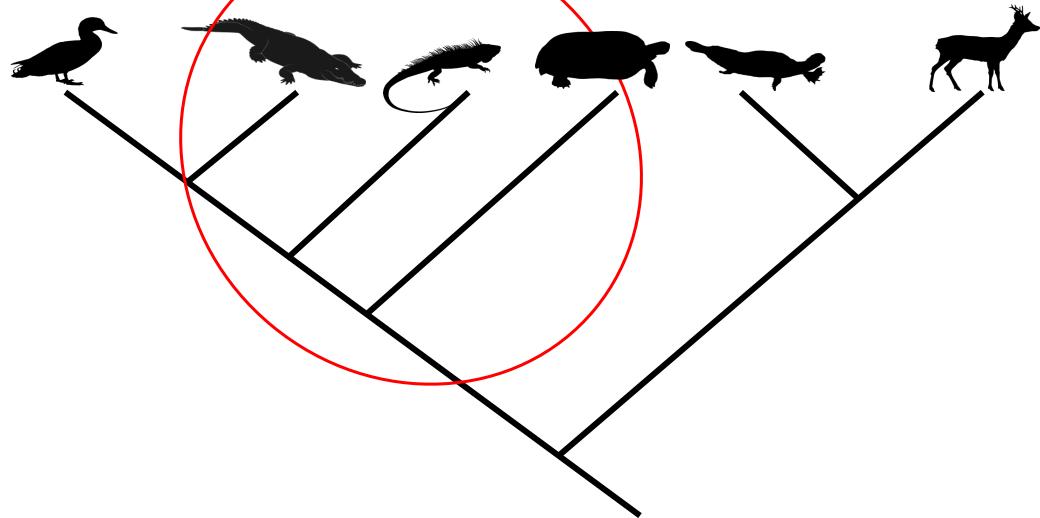


**Monophyletic group = clade**

- Include an ancestor and all of its descendants

20

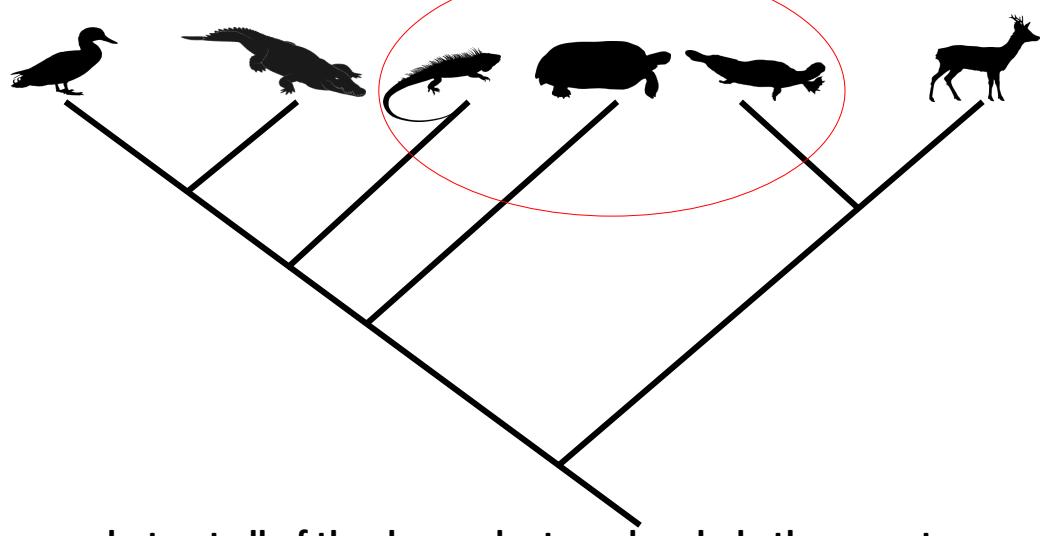
## Paraphyletic groups



- Include ancestor and some but not all of its descendants

21

## Polyphyletic groups



- Include some but not all of the descendants and exclude the ancestor

22

## Sister groups

- By definition, sister groups are of equal age
- Common mistake
  - The sister group that has fewer species is referred to as basal
  - Possible to have nodes that are more basal than other nodes, but not lineages compared to their sister group
  - Rather than saying “this group is basal”, one should say “this group is sister to all other lineages”

This question is actually not yet considered resolved!

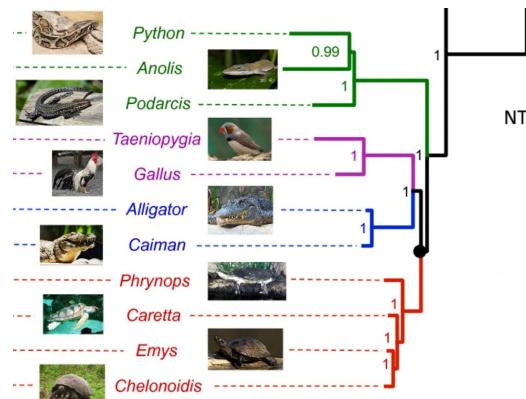
Research article | Open Access | Published: 27 July 2012

**Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria)**

Ylenia Chiari, Vincent Cahais, Nicolas Galtier & Frédéric Delsuc

BMC Biology 10, Article number: 65 (2012) | Cite this article

32K Accesses | 241 Citations | 50 Altmetric | Metrics



23

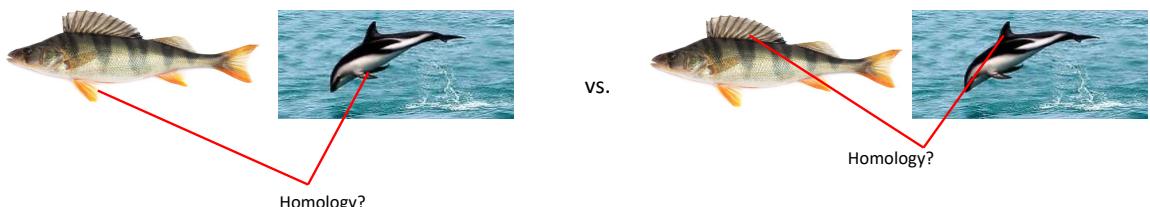
## Some premises underlying phylogenetic inferences

- Phylogenetic inferences are premised on
  - the inheritance of ancestral characters
  - the existence of a shared evolutionary history
- Homology considered as evidence of common ancestry
- A tree-like model of evolution
  - There are evolutionary processes that don't fit this model, e.g. lateral transfer

24

## Homology

- The most fundamental concept in inferring phylogeny is **homology**
- We need to be sure the characters we are studying are homologous, i.e. "the same" character in different organisms
- Otherwise our analyses will be misled



25

## Owen's definition of homology

**Homologue:** the same organ under every variety of form and function (true or essential correspondence)

Richard Owen 1843

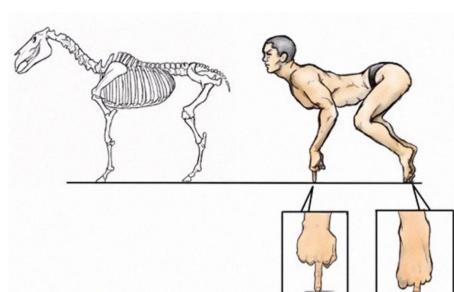
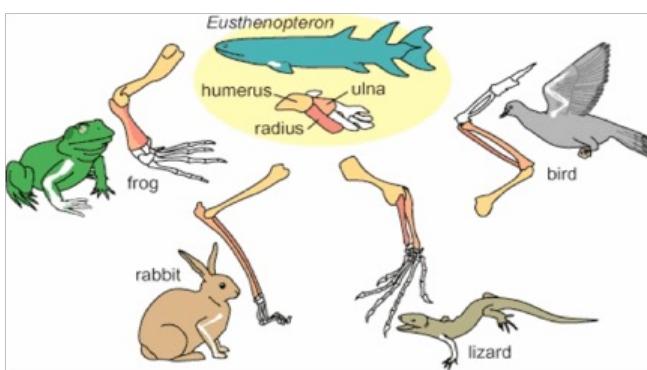


Image source: Satoshi Kawasaki

26

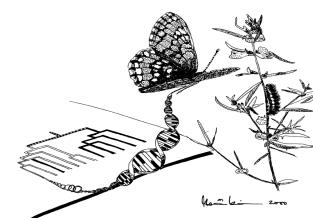
## Main kinds of data in phylogenetic inference

- **Morphological**
  - Traditionally used in phylogenetic inference
  - Still necessary for fossils and when molecular data are lacking
  - Can also help when molecular data are ambiguous
- **Molecular**
  - Most commonly used nowadays
  - Ease of sequencing led to a revolution in molecular phylogenetics

27

## Phylogenetic analysis is an attempt to infer the past

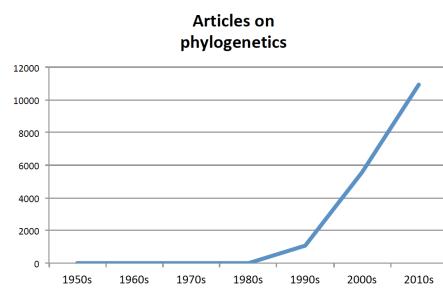
- Inferring a phylogeny is an attempt to produce a best estimate of an evolutionary history based upon incomplete information
- Our direct information about the past is limited
  - Fossil record very incomplete
  - Access to contemporary species and molecules



28

## The rise of systematics

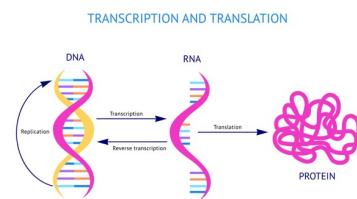
- Within the last 25 years the number of phylogenetic studies has skyrocketed
- Largely due to the advent of easy DNA sequencing methods
- Is helping us understand biodiversity and evolutionary processes better



29

## Why DNA is the Ultimate Source of Information

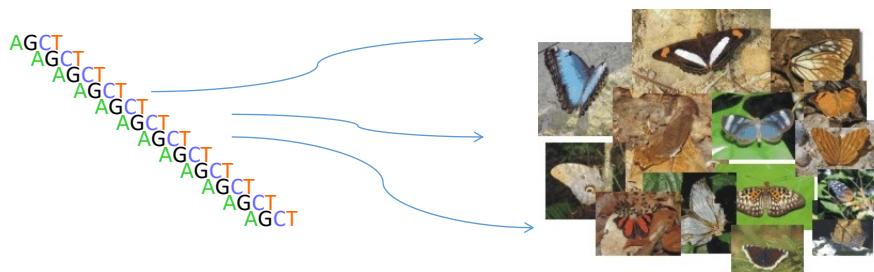
- Higher levels lack information
  - For example, one can infer protein sequence from DNA sequence data, but not complete DNA sequence from protein sequence data
- Lower levels provide no additional useful information
  - For example, sub-atomic structure does not provide information about historical relationships



30

## Why *molecular systematics*?

- Ease of data generation for large numbers of taxa
- Ease of generating a large number of independent data sets for given taxa
- Molecular characters behind the morphological characters we see



31

## Molecular systematics as a part of understanding evolution

- **Biochemistry** — basic low-level processes (e.g., nucleotide substitution, amino acid interactions)
- **Molecular genetics** — fundamental genetic processes (e.g., DNA replication, recombination)
- **Population genetics** — micro-evolutionary processes
- **Systematics** — macro-evolutionary processes

32

## What is “Molecular” in Molecular Systematics?

- Carbohydrates — No
- Lipids — No
- Secondary metabolites — No
- Proteins (amino acids) — Yes
- Nucleic Acids (DNA and RNA) — Yes

Proteins and DNA are sometimes described as “informational macromolecules”



33

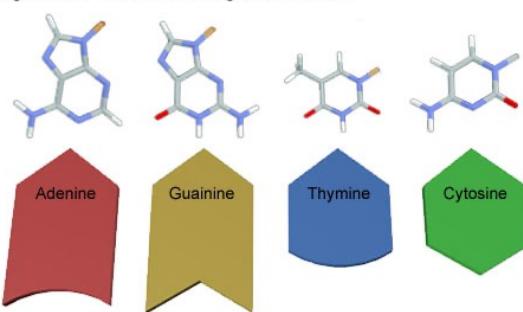
## DNA as a source of information

- ▶ DNA has four characters

Purines

Pyrimidines

Figure B-3: The Four Nitrogenous Bases

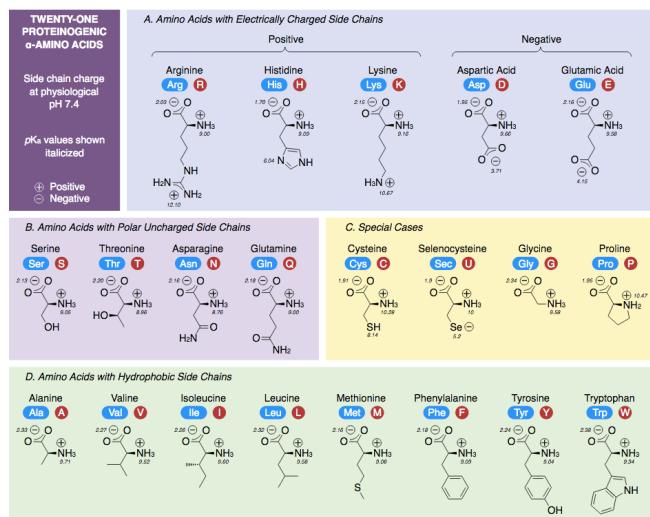


Each base has a distinct shape that can be used to distinguish it from the others. 3D representations of the four bases are shown, with the corresponding chemical structures drawn above.

34

# Proteins as a source of information

Protein sequences have 20 characters

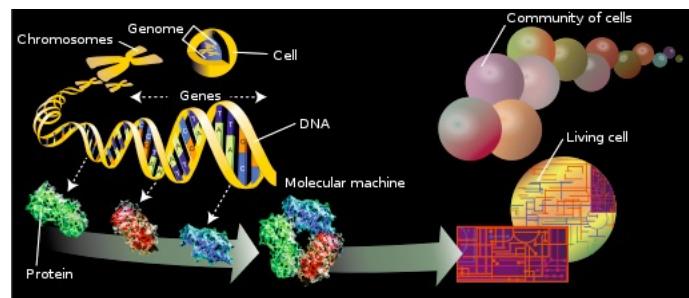


Source: Wikipedia

35

# DNA found as various entities in the genome

- Protein-coding genes
  - introns and exons
- Ribosomal DNA
- Repetitive elements
- Regulatory regions
- Junk DNA
- etc



36

## Homology in DNA sequences!

- Two steps:
- Are we looking at the same region of the genome in our species of interest?
  - Orthology vs paralogy
- Are we looking at the same site within our chosen marker in our species of interest?
  - Alignment

Sequence alignment of DNA markers from 1030 samples. The markers are color-coded by nucleotide: A (green), T (red), C (blue), and G (black). The sequence is highly conserved across all samples, with most positions showing the same nucleotide across all samples.

37

## Orthology or paralogy?

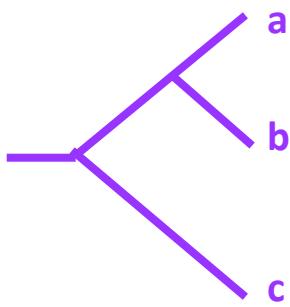
- Are the genome regions sequenced from different species the same (homologous)?
- Gene duplication
  - 1) duplicate gene degenerates - pseudogene
  - 2) duplicate gene acquires new function
- A problem particularly acute currently as we analyze phylogenomic data

38

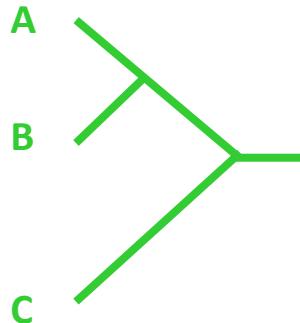
19

## Orthology: gene trees and species trees

Gene phylogeny



Organism phylogeny



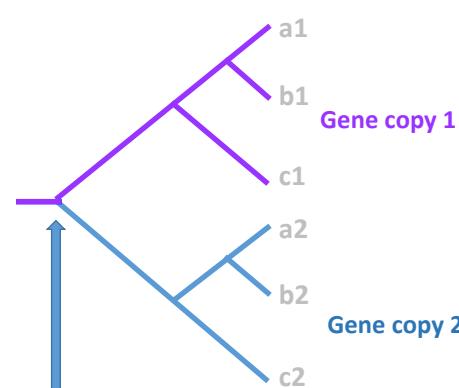
### ORTHOLOGY

Orthologs: genes that arose due to speciation

39

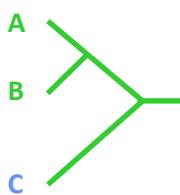
## Paralogy: can produce misleading trees

Gene phylogenies



gene duplication

Organism phylogeny

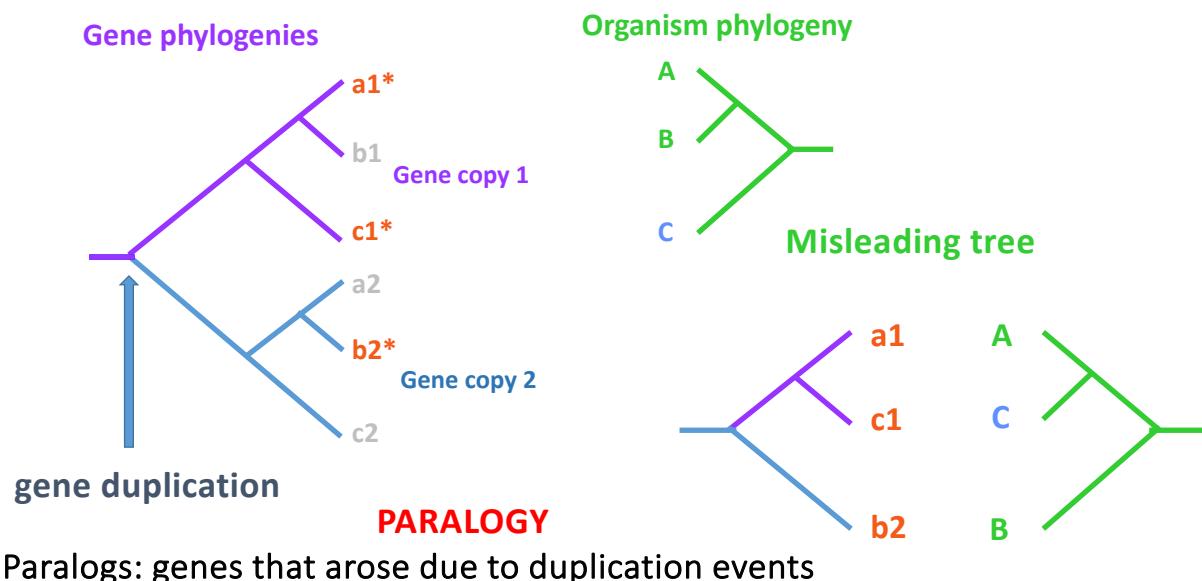


### PARALOGY

Paralogs: genes that arose due to duplication events

40

## Paralogy: can produce misleading trees

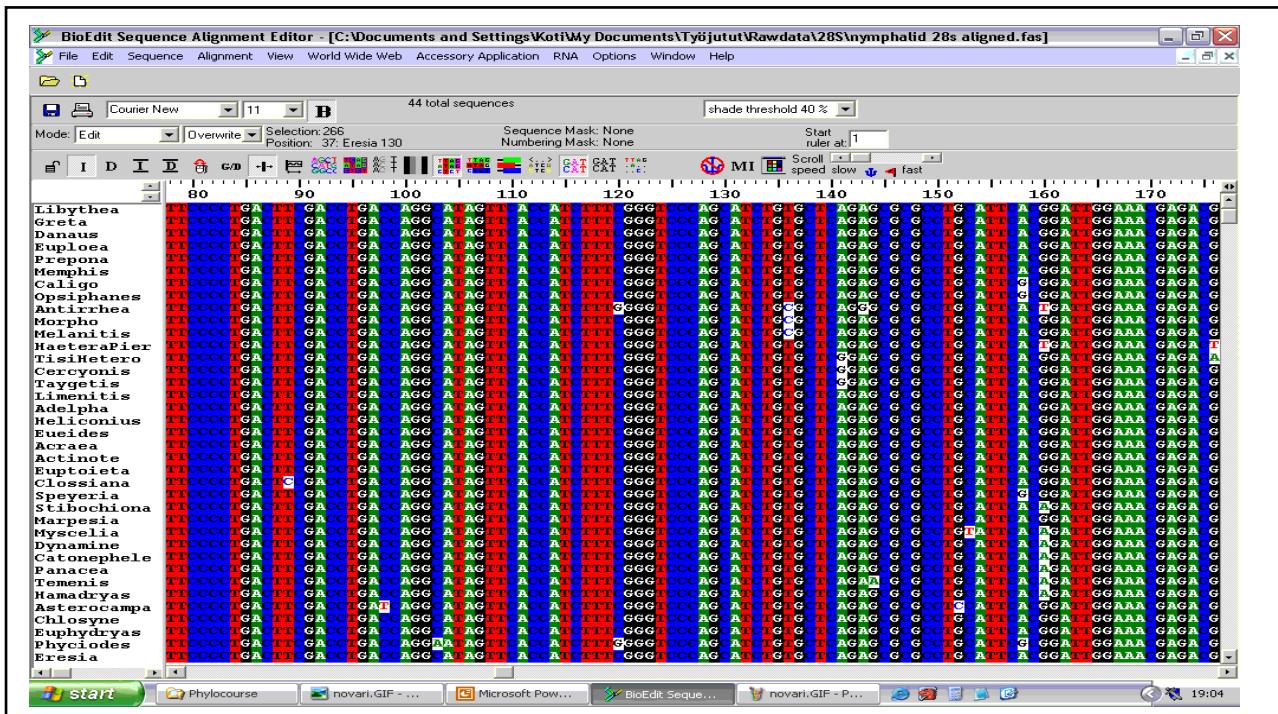


41

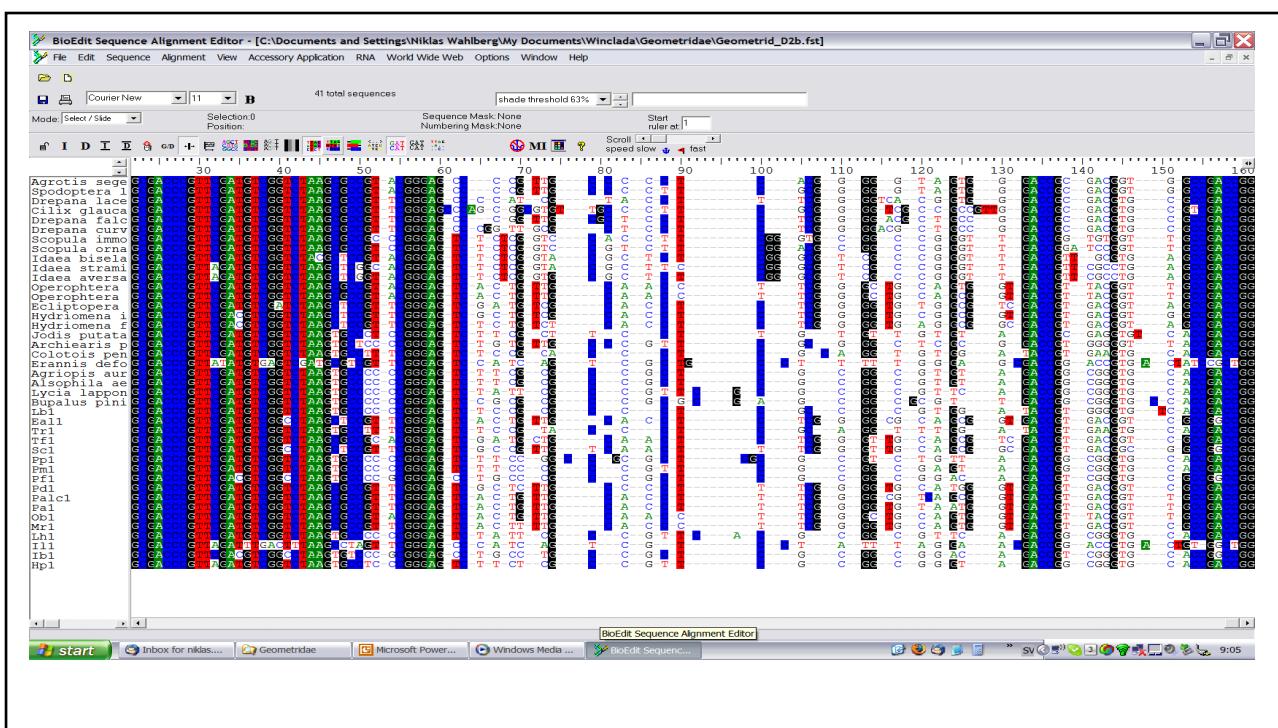
## What sequences should we use?

- **Choice of sequence** - appropriate for question (fast or slow evolving - close or distant relationships).
- Many sequences are a **mosaic** of different rates
  - **rRNA** different structural regions evolve at different rates
  - **Proteins** - synonymous (silent) rate (codon position 3) is often faster than nonsynonymous (positions 1 & 2 - changes aa) rate of change
  - Transitions occur more readily than transversions

42



43



44



45

## Data

- For long, the field of systematics was restricted by the amount of data
  - 10 years ago, datasets comprising 3-5 genes were the norm
  - 5 years ago the genomic revolution swung into full effect
  - We are now faced with an abundance of potential data, but what can we do with it?

46

## The Era of Phylogenomics

- Genomes can now be sequenced relatively easily
- Whole genomes contain a lot of information that is irrelevant for systematics, especially at deep levels
- The field of systematics is still trying to figure out how best to utilize genomic level data
  - What parts of the genome should be used?
  - How can we get at those parts in the most efficient way?
  - Where can we access specimens for our chosen methods?

47

## Phylogenomic data

- RADseq
  - Restriction-site Associated DNA sequencing
- Transcriptomes
  - All genes that are being expressed in a certain tissue at a certain time
- Ultra-Conserved Elements
  - Probes to pull out UCEs and flanking regions
- Anchored Hybrid Enrichment
  - Probes for e.g. exons of protein coding sequences
- Whole Genome Sequencing

48

## Some useful resources

- **Phylogenetics in the Genomic Era (2021)**
  - <https://hal.science/hal-02535070v3>
- **Phylogenomics: An Introduction (2017)**
  - <https://link.springer.com/book/10.1007/978-3-319-54064-1>

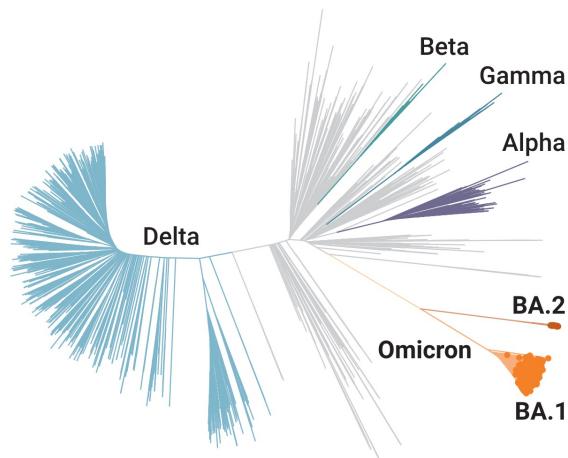
49

## Evolutionary History and Its Importance in Biology

50

## Why do we need phylogenies?

- What is shown here?



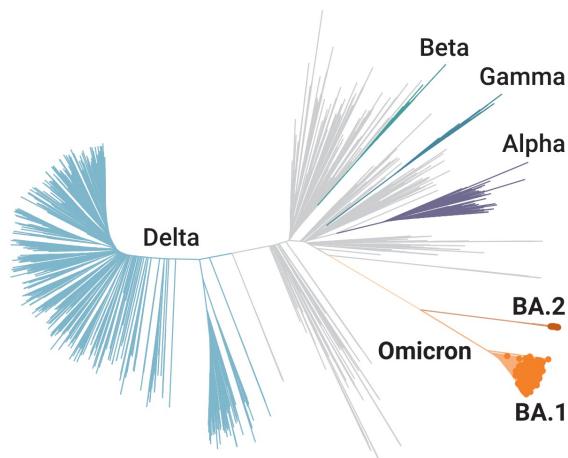
[https://www.science.org/content/article/sudden-rise-more-transmissible-form-omicron-caught-scientists-surprise?utm\\_campaign=SciMag&utm\\_source=Social&utm\\_medium=Twitter](https://www.science.org/content/article/sudden-rise-more-transmissible-form-omicron-caught-scientists-surprise?utm_campaign=SciMag&utm_source=Social&utm_medium=Twitter)

51

## Why do we need phylogenies?

- What is shown here?

Phylogeny of covid-19 strains



[https://www.science.org/content/article/sudden-rise-more-transmissible-form-omicron-caught-scientists-surprise?utm\\_campaign=SciMag&utm\\_source=Social&utm\\_medium=Twitter](https://www.science.org/content/article/sudden-rise-more-transmissible-form-omicron-caught-scientists-surprise?utm_campaign=SciMag&utm_source=Social&utm_medium=Twitter)

52

Go to the website: [menti.com](https://menti.com)  
Code: 1771 4629

- Quiz on the importance of evolutionary history
- 6 questions
- Back to lecture to have a look at the answers together



53

1. Which statement is correct?

- a) Praying mantids, cockroaches, and termites are distantly related insects.
- b) Termites are highly modified cockroaches.
- c) Termites are more closely related to ants than to cockroaches.



54

## Systematics:

praying mantids are the sister group of cockroaches and termites are derived cockroaches



**Phylogenomics resolves the timing and pattern of insect evolution**  
Bernhard Misof et al.  
*Science* **346**, 763 (2014);  
DOI: 10.1126/science.1257570

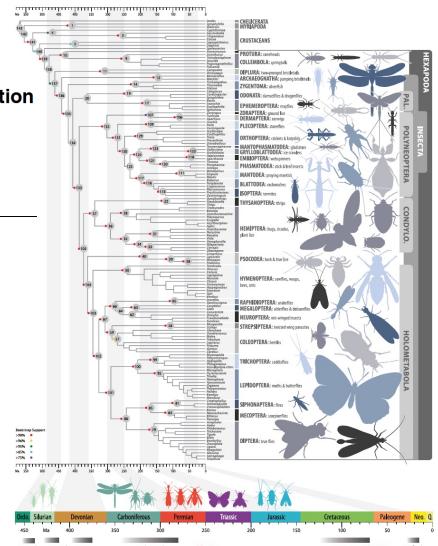
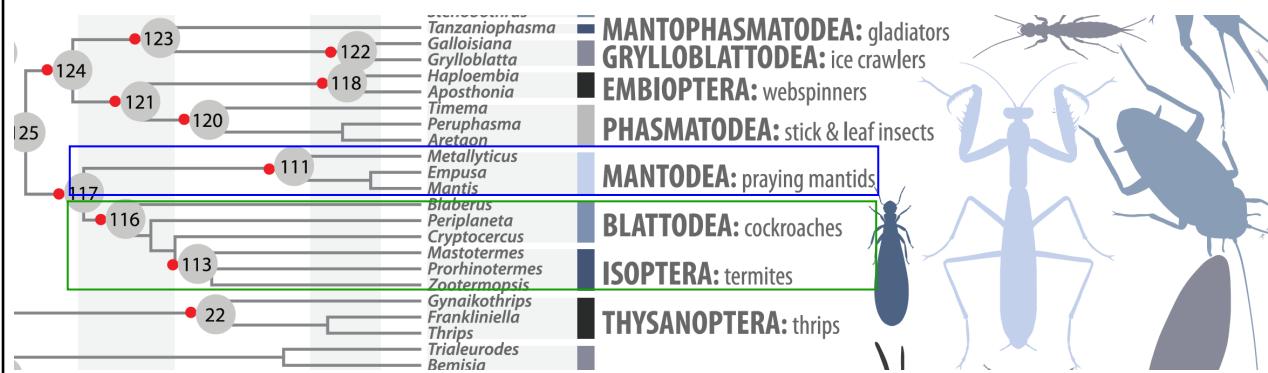


Image source: Wikipedia

55

## Systematics:

praying mantids are the sister group of cockroaches and termites are derived cockroaches



- Cockroaches are paraphyletic with respect to termites
- Termites have to be included within cockroaches to have a natural classification because termites are nested within cockroaches

56

## 2. Which statement is true?

- a) Snakes are the sister group of lizards.
- b) Some lizards are more closely related to the snakes than to other lizards.
- c) Snakes are reptiles that are not closely related to lizards.



57

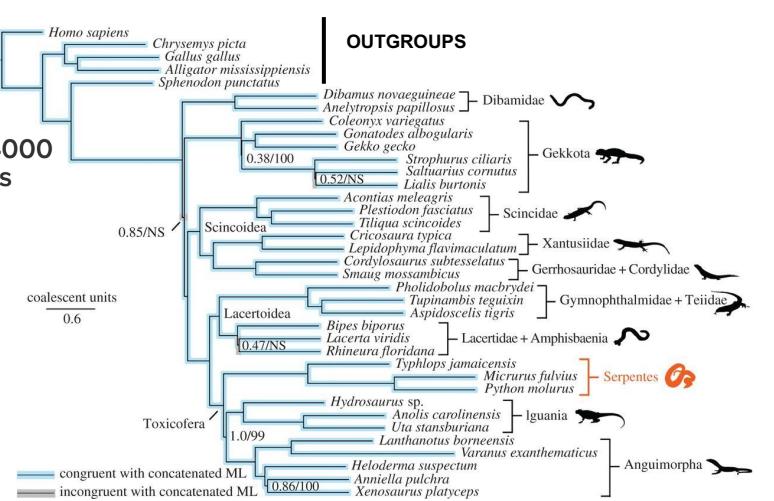
## Systematics:

Snakes are one of several lizard lineages that have lost limbs

**Phylogenomic analyses of more than 4000 nuclear loci resolve the origin of snakes among lizard families**

Jeffrey W. Streicher and John J. Wiens

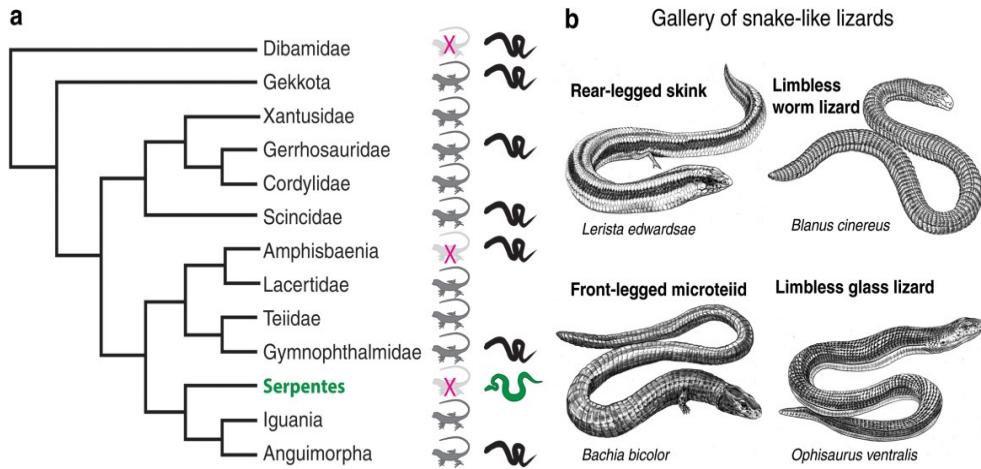
Published: 13 September 2017 | <https://doi.org/10.1098/rsbl.2017.0393>



58

## Systematics:

Snakes are one of several lizard lineages that have lost limbs



Leal & Cohn (2018) Developmental, genetic, and genomic insights into the evolutionary loss of limbs in snakes. *Genesis*, Volume: 56, Issue: 1, First published: 02 November 2017, DOI: (10.1002/dvg.23077)

59

## 3. Dinosaurs...

- a) ... belong to reptiles but are not closely related to birds.
- b) ... are the sister group of birds.
- c) ... are paraphyletic with respect to birds.

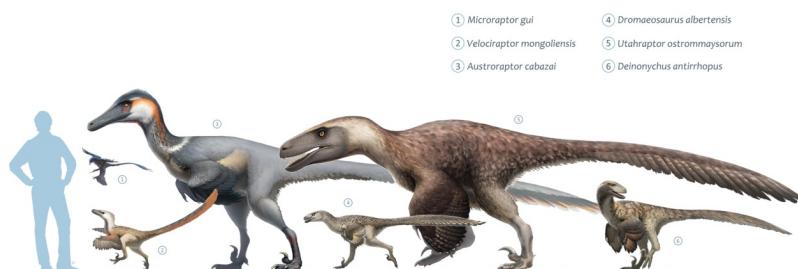
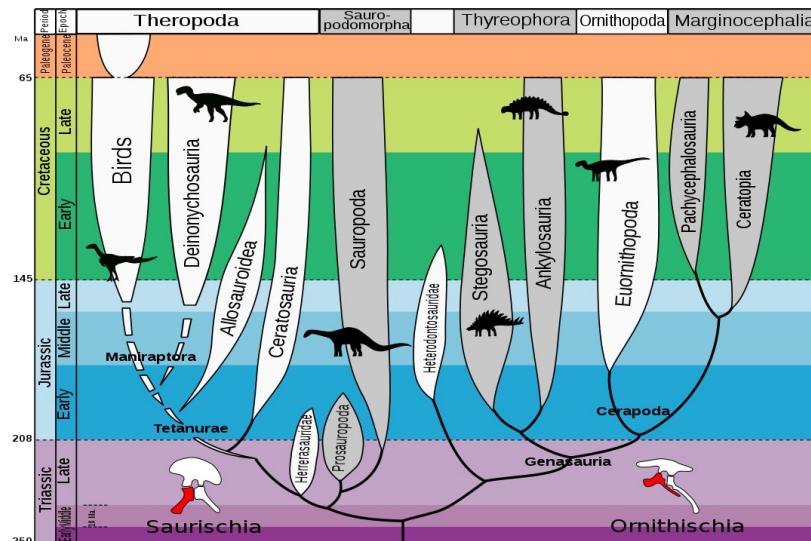


Image source: Wikipedia

60

# Are dinosaurs extinct?

- Avian dinosaurs
- Non-avian dinosaurs
- Dinosaurs are paraphyletic with respect to birds
- Pseudoextinction



61

## 4. Carnivorous plants ...

- ... evolved only once.
- ... evolved independently at least 5 times.
- ... are a myth. There are no plants that eat animals.



Image source: Wikipedia

62

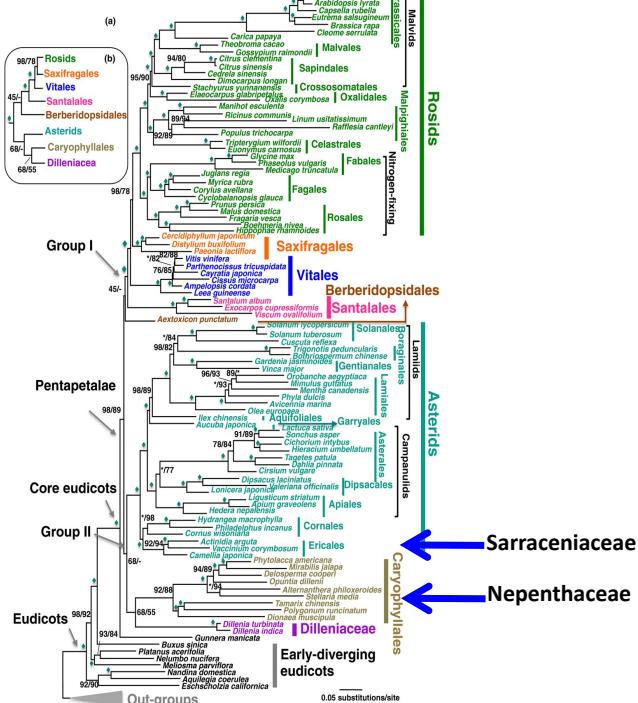
## Convergent evolution: pitcher plants



Nepenthaceae  
Order Caryophyllales



Sarraceniaceae  
Order Ericales



Zeng et al. (2017) <https://doi.org/10.1111/nph.14503>

63

## Classification of carnivorous plants

- Order: Caryophyllales
  - Families: Nepenthaceae, Drosophyllaceae, Dioncophyllaceae, Droseraceae
- Order: Oxalidales
  - Family: Cephalotaceae
- Order: Ericales
  - Family: Sarraceniaceae
- Order: Lamiales
  - Families: Lentibulariaceae, Byblidaceae
- Order: Poales
  - Family: Bromeliaceae

64

## 5. Swim bladder in fishes ...

- a) ... evolved from the precursor of lungs.
- b) Swim bladder and lungs are two organs that are not homologous.
- c) Lungs evolved from the swim bladder.

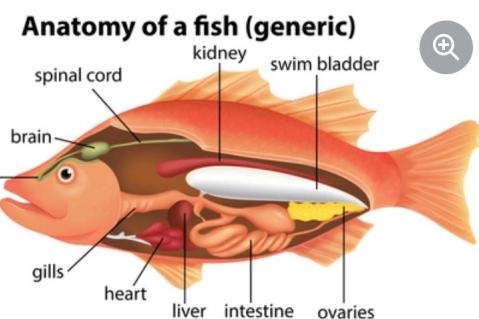


Image source: Shutterstock.com

65

## Understanding directionality of character evolution

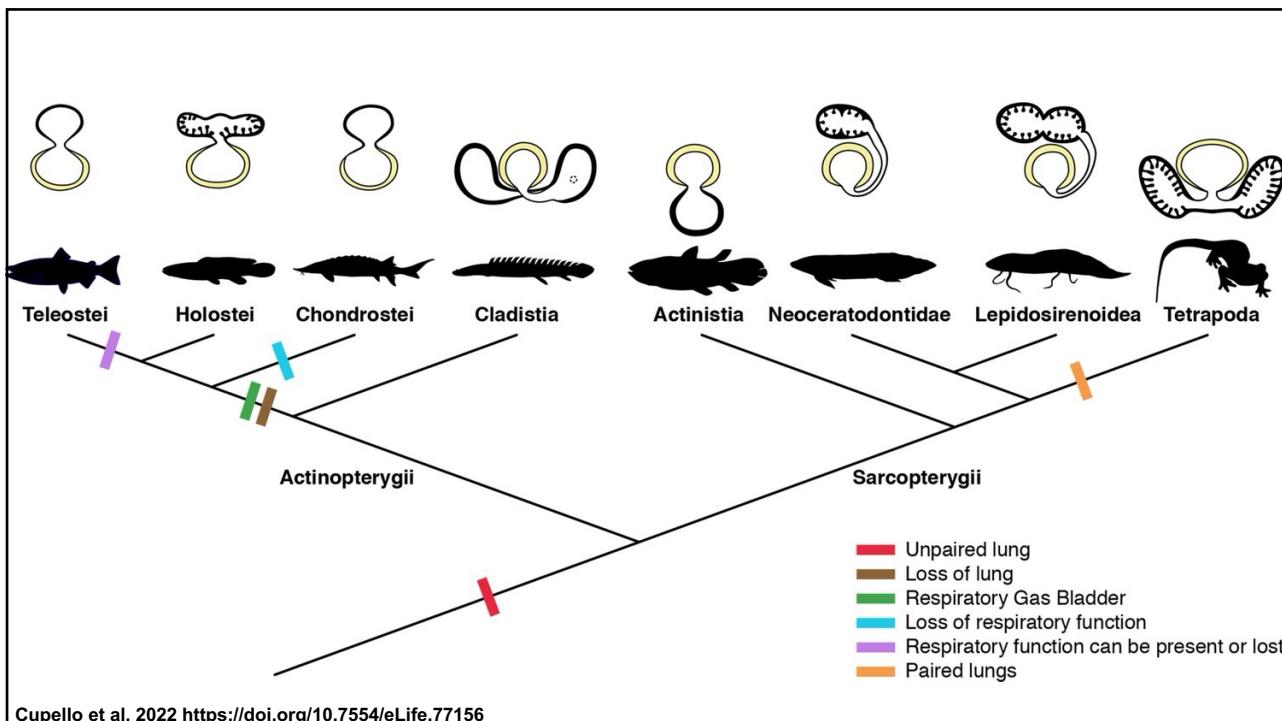
- **Swim bladder in fishes and lungs in tetrapods are homologous**
- **Darwin erroneously believed that lungs evolved from the swim bladder**
- **He was right thinking that an organ with one function can evolve into an organ with a different function BUT the swim bladder actually evolved from the lung precursor!**

The illustration of the swimbladder in fishes is a good one, because it shows us clearly the highly important fact that an organ originally constructed for one purpose, namely flotation, may be converted into one for a wholly different purpose, namely respiration.

All physiologists admit that the swim-bladder is homologous, or "ideally similar" in position and structure with the lungs of the higher vertebrate animals: hence there seems to me to be no extreme difficulty in believing that natural selection has actually converted a swimbladder into a lung, or organ used exclusively for respiration.

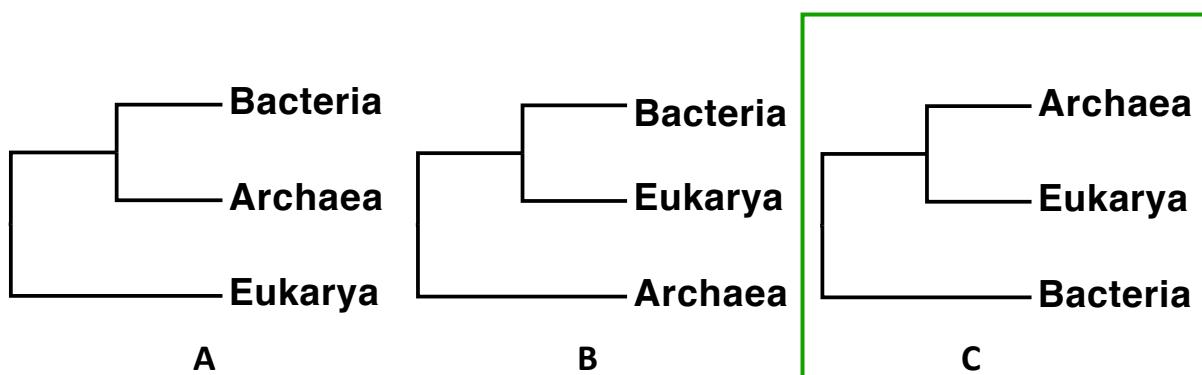
C. Darwin: On the Origin of Species (1961 ed.), p. 210

66



67

6. Best hypothesis for the relationships among the three domains of life is:



68

## Origin of Eukarya

- **Eukaryotes**

- Organisms whose cells have a nucleus enclosed within a nuclear envelope
- Plants, animals, fungi + a lot of single-cell lineages
- **Which lineage is the sister lineage to Eukarya?**
- **Why is this interesting?**
  - Where does the complex cell of eukaryotes originate from?

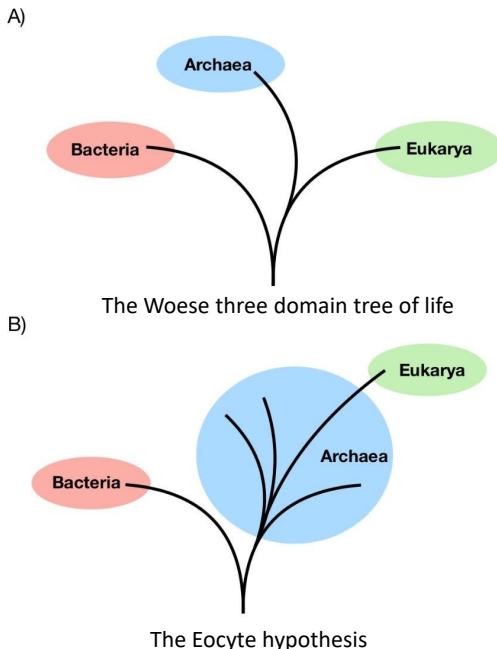


Fig. 4 from MacLeod et al. 2019 (doi: 10.3934/microbiol.2019.1.48)

69

## Origin of Eukarya

- Which lineage is the sister lineage to Eukarya?
- Discovery of Asgard Archaea is complicating the picture

Figure 5. (A) The tree of life resulting from a range of phylogenetic analyses of **conserved markers, ribosomal RNA genes and ribosomal proteins**, placing Asgard archaea as the closest archaeal relatives of eukaryotes. (B) The tree produced from phylogenetic analyses of **RNA polymerase genes** supports the three-domain topology of the tree of life, with Asgard archaea as a sister group to Euryarchaeota.

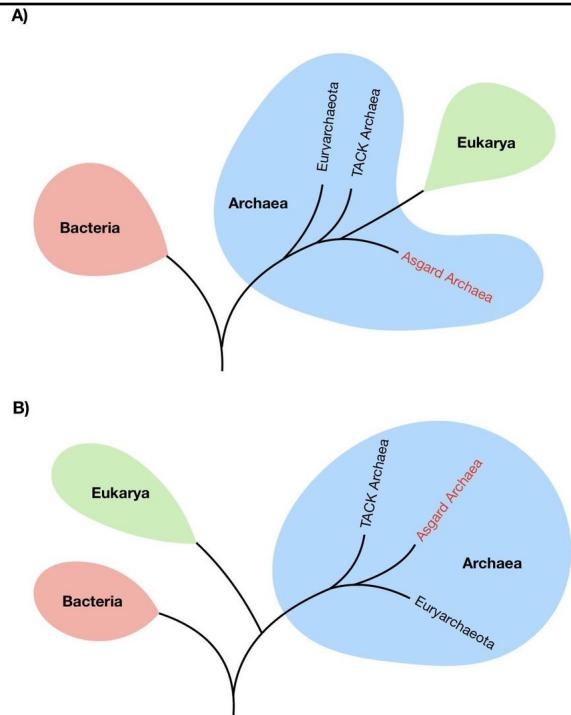


Fig. 5. MacLeod et al. 2019 (doi: 10.3934/microbiol.2019.1.48)

70

## Molecular evolution: arms race between caterpillars and their host plants

- Plants have secondary metabolites
- One of the functions: protection from herbivory
- Insects can evolve the ability to metabolize plant secondary metabolites
- Then plants can evolve new chemicals, and we have an arms race
- Example: Edger et al. 2015 PNAS study



### The butterfly plant arms-race escalated by gene and genome duplications

Patrick P. Edger<sup>a,b,c,1</sup>, Hanna M. Heidel-Fischer<sup>d,1</sup>, Michaël Bekaert<sup>e</sup>, Jadranka Rota<sup>f</sup>, Gernot Glöckner<sup>g,h</sup>, Adrian E. Platts<sup>i</sup>, David G. Heckel<sup>i</sup>, Joshua P. Dier<sup>j,k</sup>, Eric K. Wafulai<sup>l</sup>, Michelle Tang<sup>o</sup>, Johannes A. Hofberger<sup>l</sup>, Ann Smithson<sup>m,n</sup>, Jocelyn C. Hall<sup>o</sup>, Matthieu Blanchette<sup>l</sup>, Thomas E. Bureau<sup>o</sup>, Stephen I. Wright<sup>q</sup>, Claude W. dePamphilis<sup>j</sup>, M. Eric Schranz<sup>l</sup>, Michael S. Barker<sup>b</sup>, Gavin C. Conant<sup>r,s</sup>, Niklas Wahlberg<sup>f</sup>, Heiko Vogel<sup>d</sup>, J. Chris Pires<sup>s,t,u,2</sup>, and Christopher W. Wheat<sup>t,2</sup>

<sup>a</sup>Division of Biological Sciences, University of Missouri, Columbia, MO 65211; <sup>b</sup>Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721; <sup>c</sup>Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720; <sup>d</sup>Department of Entomology, Max Planck

71

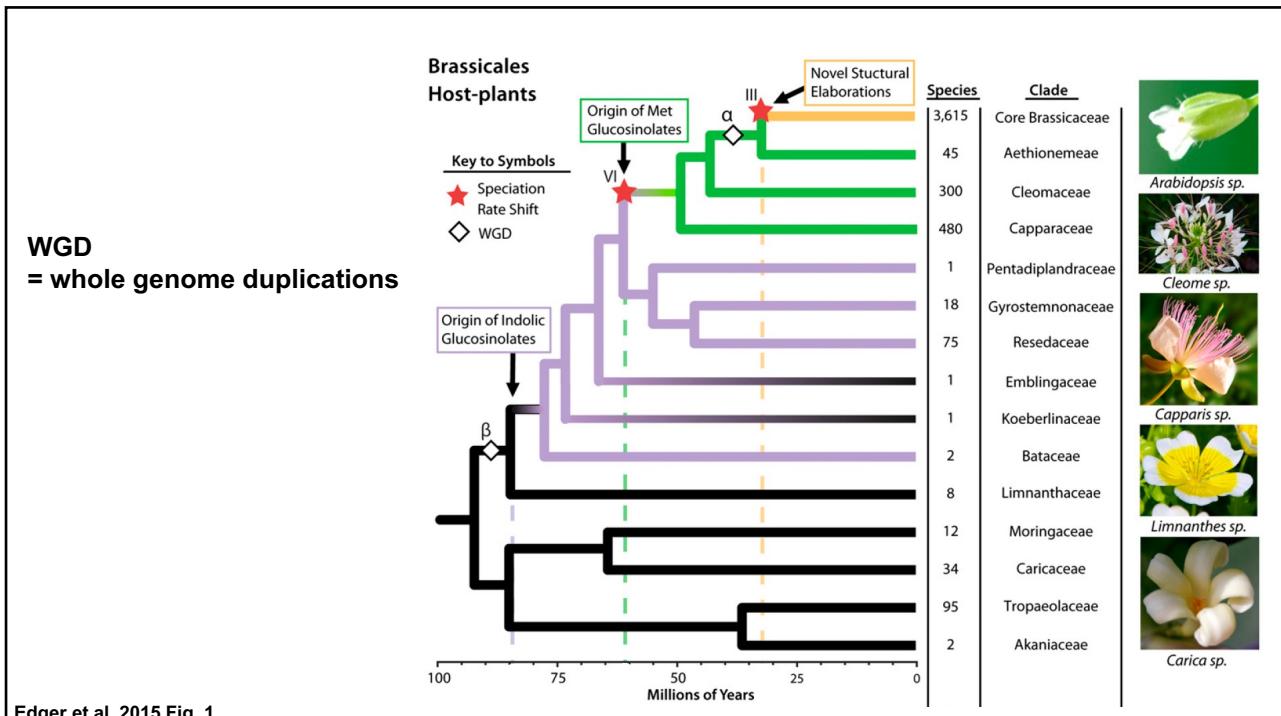
## Edger et al. 2015. The butterfly plant arms-race escalated by gene and genome duplications

- Butterflies: family Pieridae
- Plants: order Brassicales
- A number of gradual changes
- BUT also gene and genome duplications
- The story: “Nearly 90 million years ago, the ancestors of *Brassica* (mustards, cabbage) and related plants developed a chemical defense called glucosinolates.”
  - Glucosinolates – toxic to most insects, for humans source of the sharp taste in wasabi, horseradish, and mustard

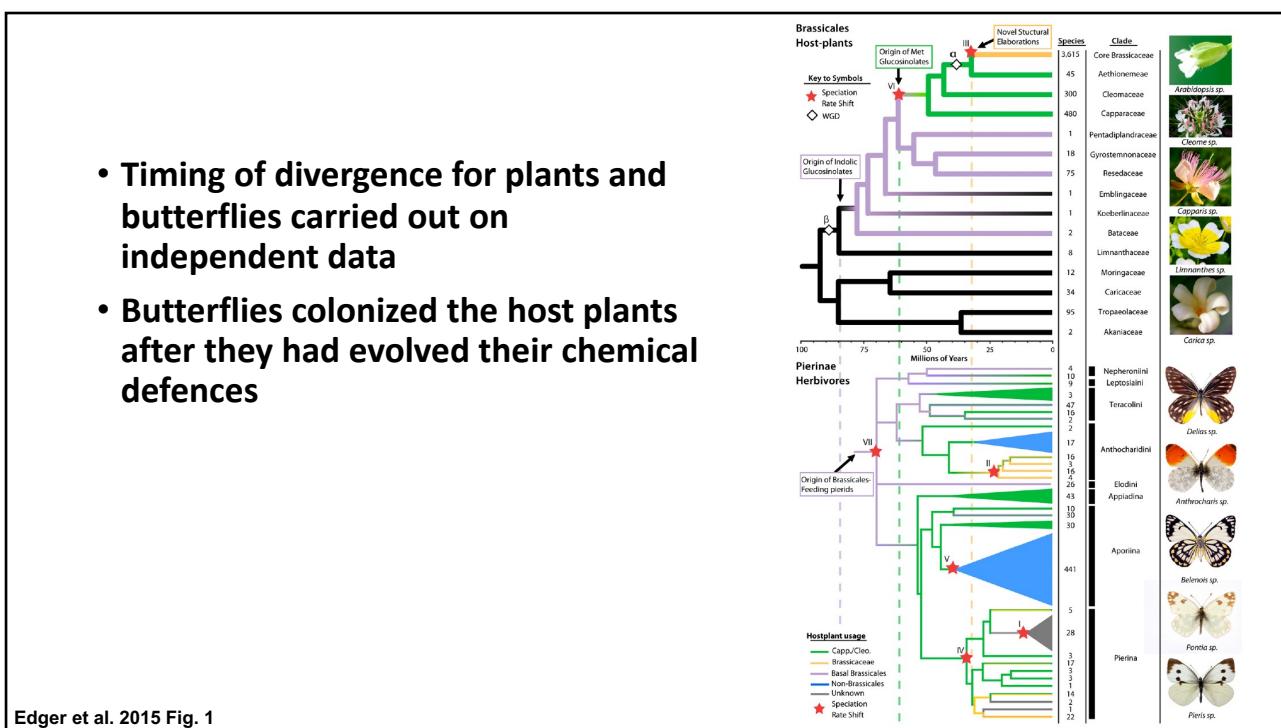


Cabbage Butterfly *Pieris brassicae* (Linnaeus, 1758)  
By Didier Descouens - Own work, CC BY-SA 4.0

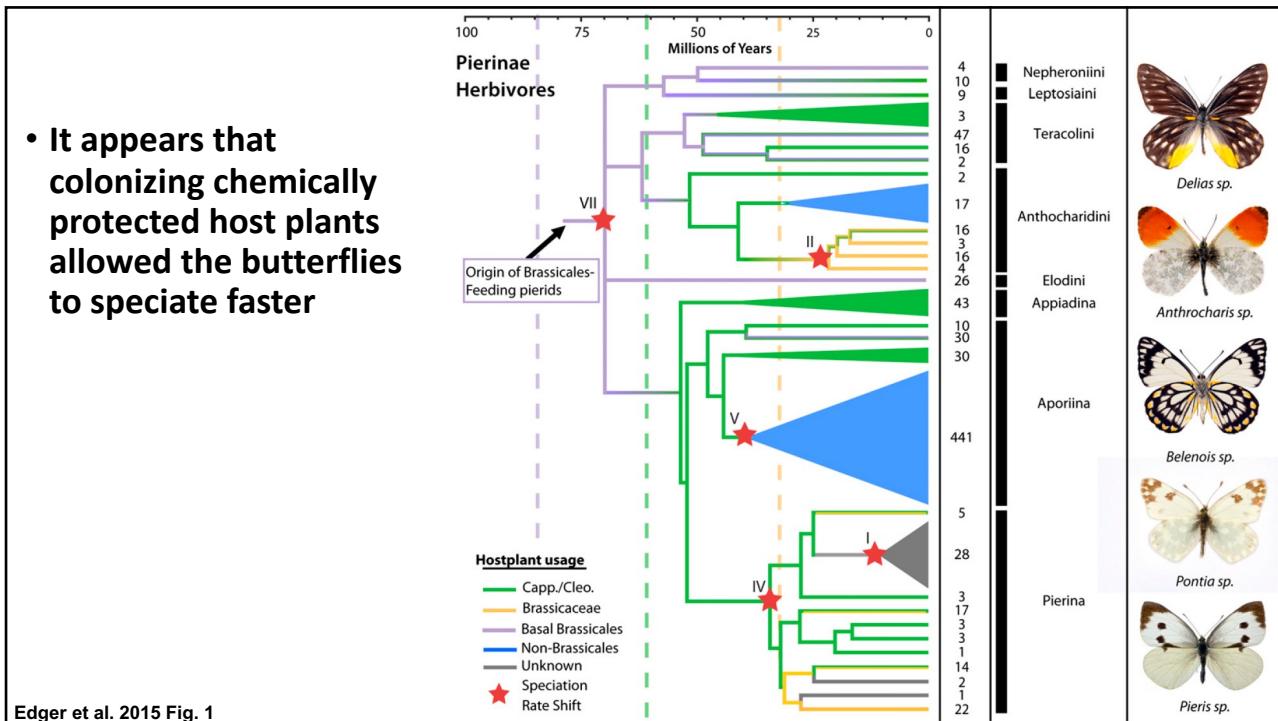
72



73

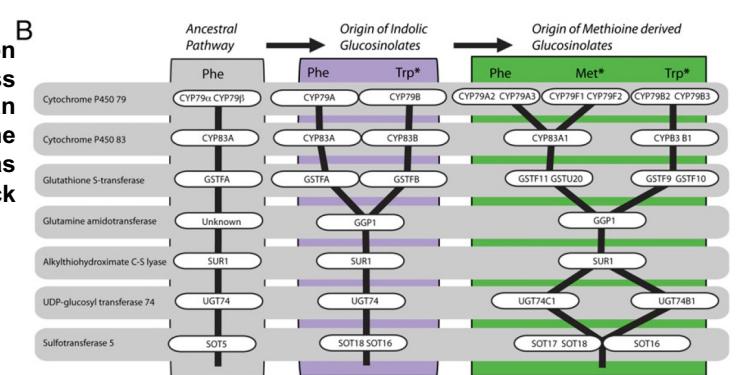


74

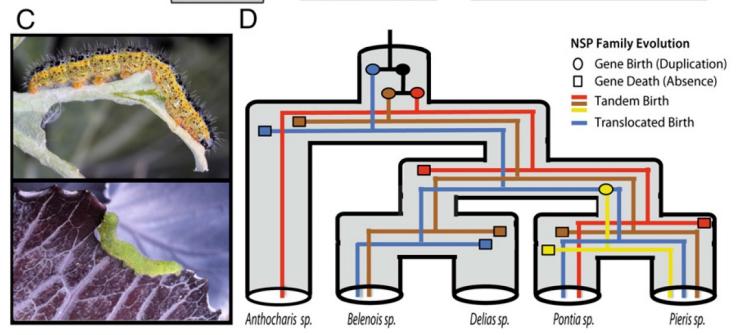


75

**B**  
Figure 2. (B) An illustration of the evolution of core glucosinolate pathways across Brassicales; with substrates tryptophan (Trp), phenylalanine (Phe), and methionine (Met) shown at the top, enzymes depicted as white ovals, and each pathway as black vertical lines.



**C**  
Figure 2. (D) Evolution of the NSP gene family is shown across select Pierinae genera, indicating the birth and death dynamics of four paralogous clades.



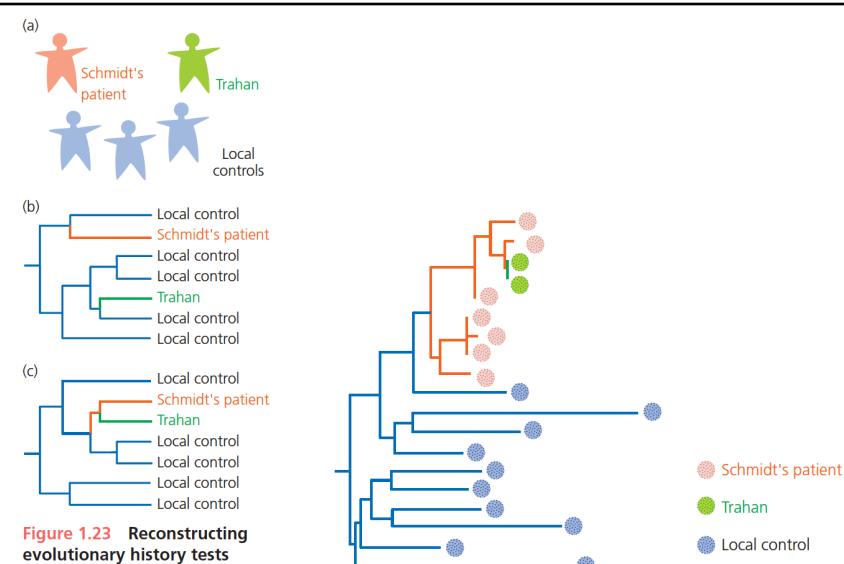
76

\*Nitrile Specifier Protein (NSP): primary detoxification mechanism used by the butterflies to break down this chemical defense system

## Criminal investigations: evolutionary history of a deliberate HIV infection

- Source: Herron & Freeman “Evolutionary Analysis” 5<sup>th</sup> edition
- Players in this true story:
  - Janet Trahan, nurse – the accuser
  - Richard Schmidt, physician – the accused
  - David Hillis, evolutionary biologist – providing crucial evidence
- Legal case:
  - Trahan became sick with HIV and accused Schmidt of deliberately injecting her with blood from a HIV-positive patient as revenge for breaking off their affair
- Analysis:
  - Evolutionary history of viral samples from various people to establish from where Trahan contracted her virus

77



**Figure 1.23** Reconstructing evolutionary history tests the prosecution's theory of a crime (a) The individuals from whom HIV samples were collected. (b) The viral evolutionary tree predicted if the suspect is innocent. (c) The evolutionary tree predicted if the suspect is guilty.

**Figure 1.24** Incriminating evidence This reconstructed evolutionary history is consistent with the prosecution's charges. Redrawn from Metzker et al. (2002).

Source: Herron & Freeman “Evolutionary Analysis” 5<sup>th</sup> edition

78

## Importance of Understanding Evolutionary History: examples we talked about (there could be many, many more)

- Viral evolution and its impact on human health
- Systematics – understanding the relationships among different groups of organisms and how they acquired their current form and function
  - E.g. snakes and lizards, birds and dinosaurs, the main domains of life
- Molecular evolution
  - Genome and gene duplications as weapons in an arms race between plants and caterpillars
- Even criminal investigations!