

Phylogenetics

SCBT 401 Bioinformatics 3

**Nuttaphon Onparn
3 November 2016**

Outline

- **Phylogenetics**
 - Introduction
 - Anatomy of phylogenetic trees
 - Phylogenetic tree concepts
 - Homology and parsimony
 - Evaluating phylogenetic tree
 - Bootstrapping
 - Grouping in trees (phyletic groups)
 - References

Introduction

- **Phylogenetics**
 - The study of evolutionary relatedness among various groups of organisms.
 - Molecular sequencing data
 - Morphological data matrices
- **Phylogeny** (วงศ์วานวิวัฒนาการ)
 - Greek *phylon* = tribe
 - Greek *genesis* = origin
 - The evolutionary history of a species or group of species.

Introduction

- **What is a phylogenetic tree?**
 - **Graphical representation of relationships**
 - 25+ years ago: morphological data
 - 25 to present day: molecular data
 - **Source of data**
 - **Morphological data: fossils, organisms**
 - **Molecular data: DNA/protein sequences**

Introduction

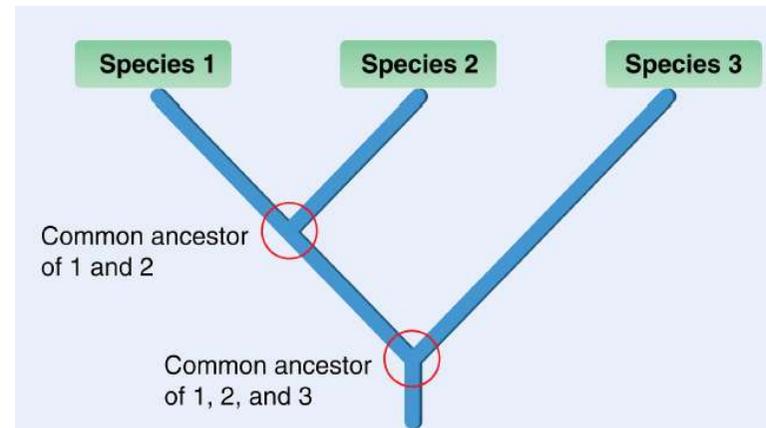
- **Who are using phylogenetic trees?**
 - **Systematist (taxonomist)**
 - Relationships of taxa (often species)
 - **Molecular biologist**
 - Relationships of sequences

Composition of Phylogenetic Tree

- **A simple object of two elements**

- **Phylogenetic tree**

- **Nodes** (ปม)
- **Branches** (กิ่ง)



- **Nodes**

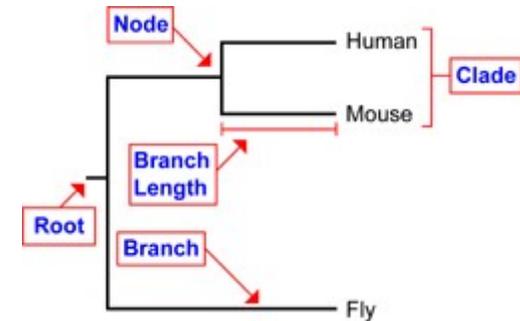
- **Internal node** (a point represent common ancestor)
- **External node** (a point represent taxa/species/seq.)

Branch is a line that intersect and terminate at the node.

Anatomy of Phylogenetic Trees

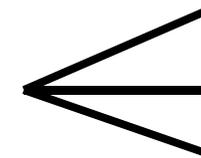
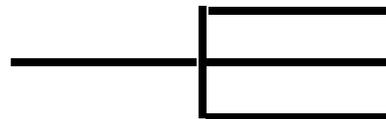
- **Subtree**

- Each subtree is a clade.
- Sister taxa, sister group
 - The two taxa on either side of a split are called sister taxa or sister groups.



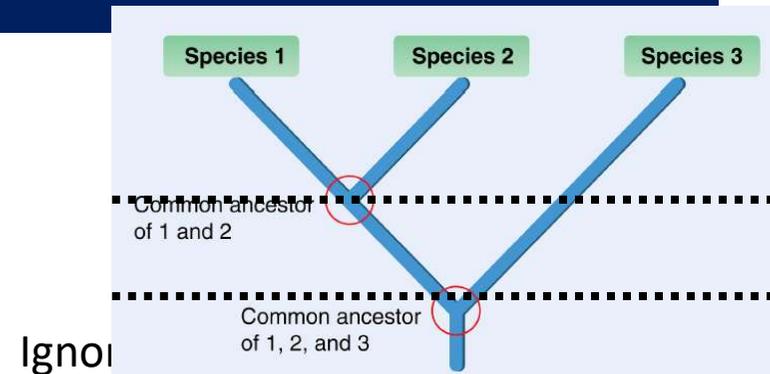
- **Polytomy**

- 2-way or 3-way forks
 - Cladistic uses 2-way fork. 3-way forks appear when no sufficient data to resolve relationship.



Number of Trees and Depth

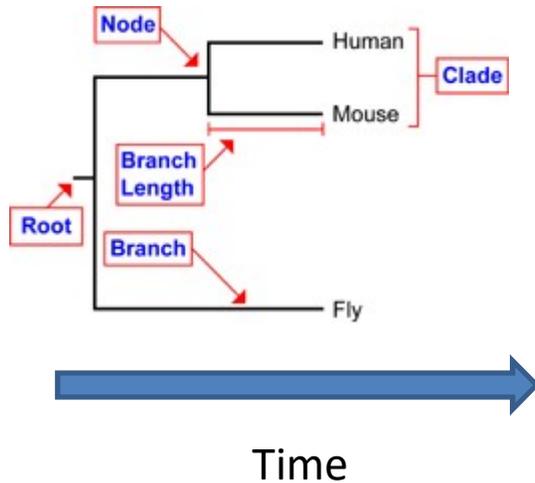
- **Number of possible trees**
 - Given number of species N



Number of species	2	3	4	5	6	7	8	9	10	N
Number of cladograms	1	3	15	105	945	10K	135K	2G	34G	$1 \cdot 3 \cdot 5 \cdot 7 \cdot \dots \cdot (2N-3)$

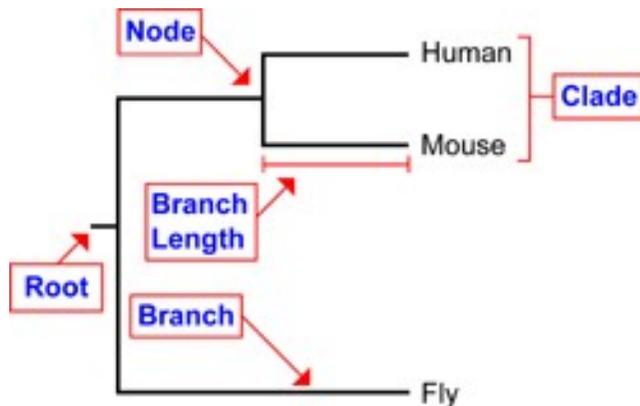
- **Depth of tree**
 - Depth of tree or level is $\log_2 N$
 - If N=32 species, tree will be 5 level deep ($2^5=32$)
 - Underestimate as each depth might not uniform.

Time



- **Time scale in cladogram**
 - With time running forward from the base of the tree to the leaves of the tree.
 - Scaled cladogram
 - 1cm = 1 million years
 - Unscale cladogram
 - Cladogram from characteristics with no date.
 - Data from DNA/RNA differences (can be translated to time scale, might not accurate).
 - Unreadable tree if scaled with time.

Structure of Tree



- **Structure of phylogenetic tree** (แผนภูมิวงศ์วานวิวัฒนาการ)

- **Node** (ปม)

- Represents a taxonomic unit. This can be either an existing species or an ancestor.

- **Branch** (กิ่ง)

- Defines the relationship between the taxa in terms of descent and ancestry.

Structure of Tree

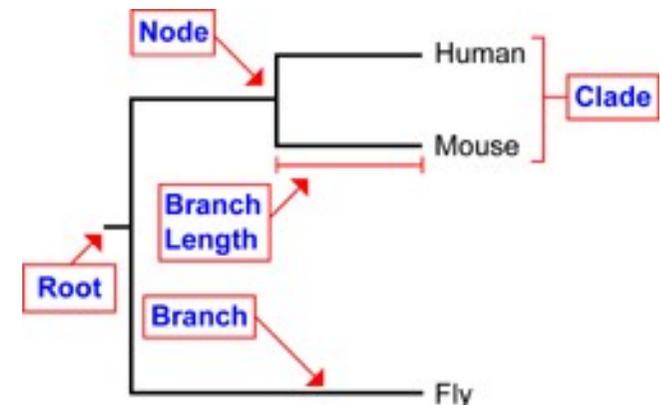
- **Structure of phylogenetic tree** (แผนภูมิตรงศ์วานวิวัฒนาการ)

- **Topology** (รูปทรงของแผนภูมิ)

- The branching patterns of the tree.

- **Branch length** (ความยาวของกิ่ง)

- Represents the number of changes that have occurred in the branch.



Structure of Tree

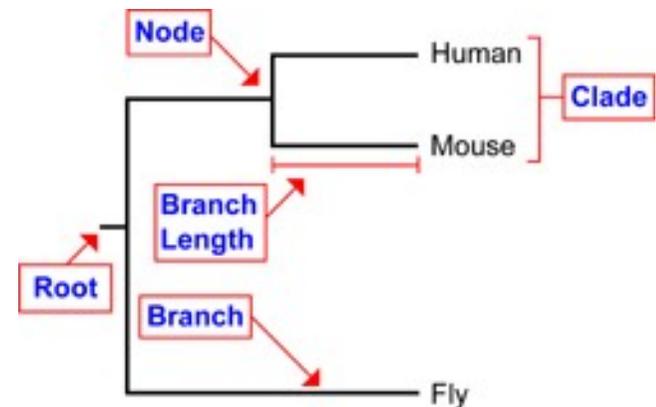
- **Structure of phylogenetic tree (cont.)**

- **Root** (ราก)

- **The common ancestor of all taxa.**
- **Rooted tree** (...ถูกตรึงราก)
- **Unrooted tree** (...ไม่ถูกตรึงราก)

- **Distance scale**

- **Scale that represents the number of differences between organisms or sequences.**



Structure of Tree

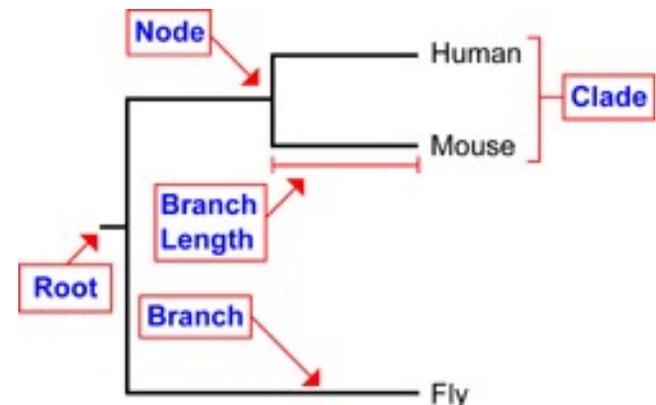
- **Structure of phylogenetic tree (cont.)**

- **Clade**

- A group of two or more taxa or DNA sequences that includes both their common ancestor and all of their descendents.

- **Operational Taxonomic Unit (OTU)**

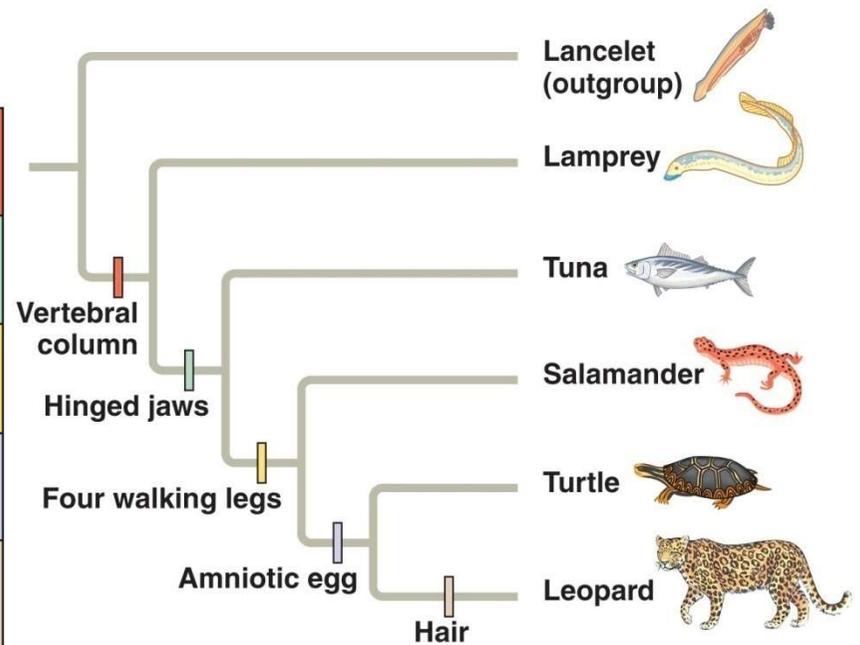
- Taxonomic level of sampling selected by the user to be used in a study, such as individuals, populations, species, genera, or bacterial strains.



		TAXA					
		Lancelet (outgroup)	Lamprey	Tuna	Salamander	Turtle	Leopard
CHARACTERS	Vertebral column (backbone)	0	1	1	1	1	1
	Hinged jaws	0	0	1	1	1	1
	Four walking legs	0	0	0	1	1	1
	Amniotic (shelled) egg	0	0	0	0	1	1
	Hair	0	0	0	0	0	1

(a) Character table

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

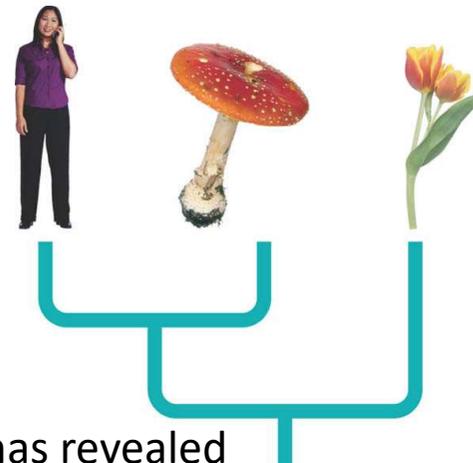


(b) Phylogenetic tree

Copyright © 2008 Pearson Education, Inc., publishing as Pearson Benjamin Cummings.

Molecular Systematics

- **Molecular systematics**
 - Science of inferring evolutionary relationships using DNA, RNA, and other molecules.
 - Gene or genome level of comparison.
 - Information of these molecules lead to “bioinformatics”



Molecular systematics has revealed animal and fungi are closely related.

Homology

- **Homology**

- Similarities due to shared ancestry.

- Homologous structure (human hand, whale flipper)

- Morphological homology

- Molecular homology

- Organisms with similar DNA sequences are more closely related.

- Note: morphological diversity might not correlate with molecular diversity.

Homology vs Analogy

- **Analogy**

- Similarities are due to “convergent evolution”, rather than having shared ancestry.

- Similar characters evolve independently because similar environmental pressure and natural selection.

- **Homology vs analogy**

- Distinguish homology and analogy is critical for phylogenetic reconstruction.

Homoplasy

- **Homoplasy**

- Greek for “to mold in the same way”.
- Analogous structures/sequences are homoplasies.

Convergent evolution of
burrowing characters

Top: marsupial mole

Bottom: eutherian mole



Molecular Homology

- **Molecular homology**
 - Molecular sequences are homologous if there are many shared characters.
 - Same nucleotide sequences among genes \Rightarrow homologous genes (evolve from common ancestor).
 - Molecular homoplasy or molecular homology?
 - Sequences can be similar coincidentally, by chance.

1
 1 C C A T C A G A G T C C
 2 C C A T C A G A G T C C



2
 1 C C A T C A (G) A G T C C
 2 C C A T C A G A G T C C
 (G T A) Insertion
 Deletion



3
 1 C C A T C A A G T C C
 2 C C A T G T A C A G A G T C C



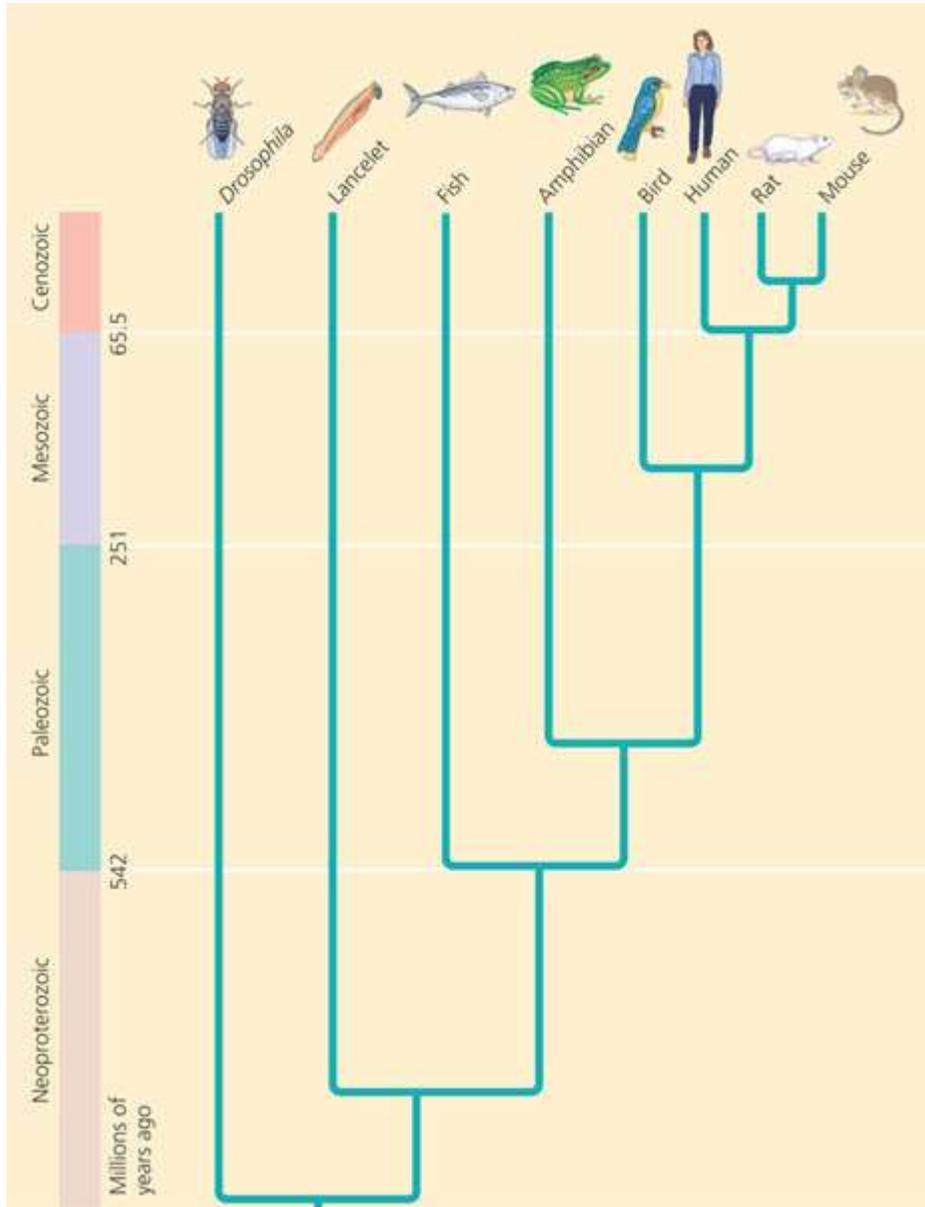
4
 1 C C A T CA A G T C C
 2 C C A T G T A C A G A G T C C

A C G G A T A G T C C A C T A G G C A C T A
 T C A C C G A C A G G T C T T T G A C T A G

Molecular homoplasy

Two distantly related organisms share 25% of sequence similarity.

DNA alignment helps us comparing two homologous sequences, with mutations.

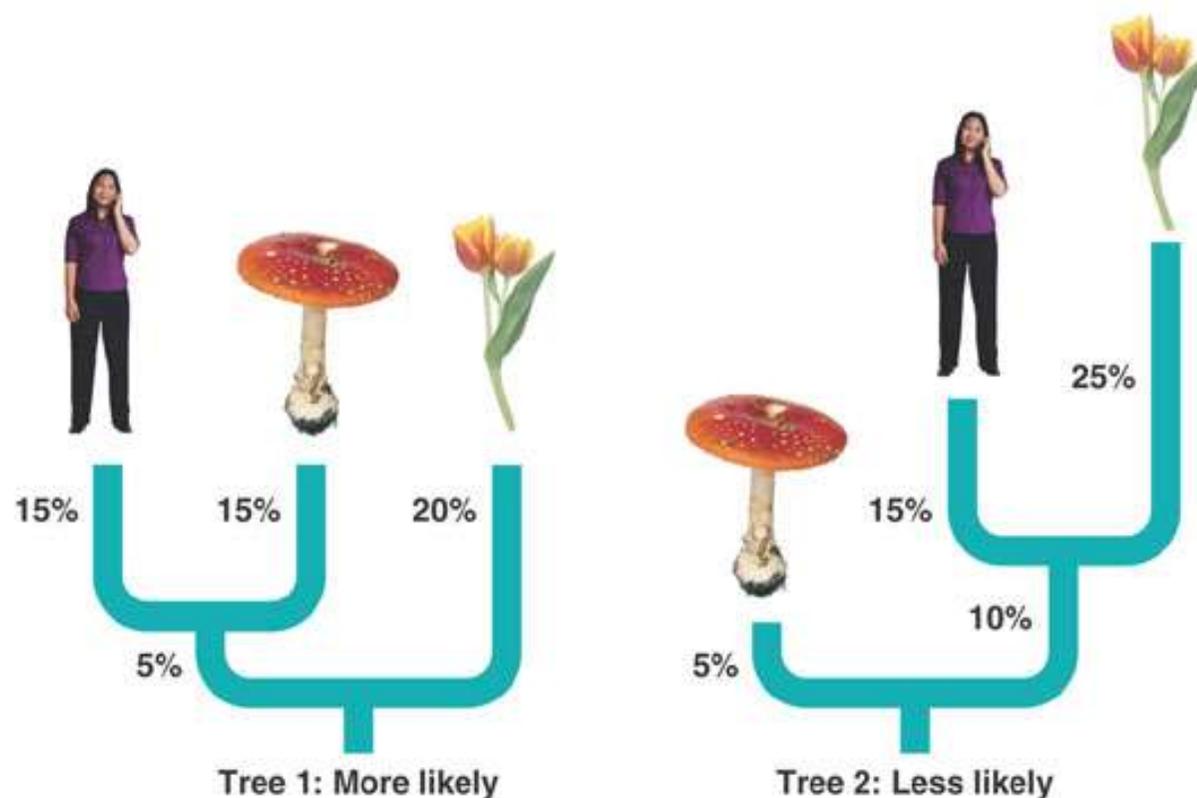


Ultrametric tree

Using molecular data from hedgehog genes, but matching the divergent with fossil records.

	Human	Mushroom	Tulip
Human	0	30%	40%
Mushroom		0	40%
Tulip			0

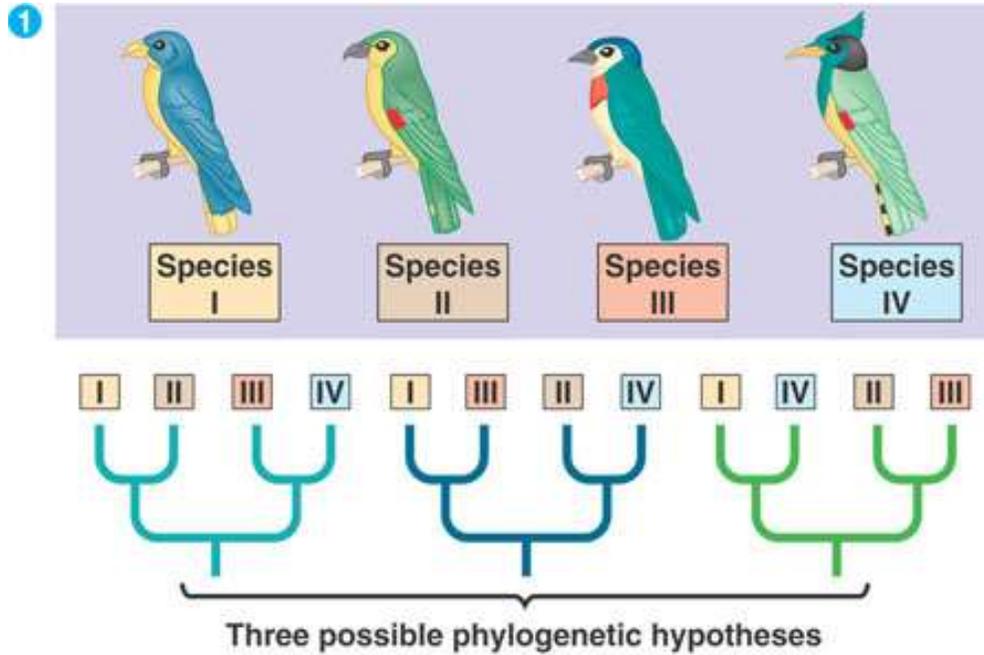
(a) Percentage differences between sequences



(b) Comparison of possible trees

Maximum likelihood

Given the certain rules about DNA changes, a tree can be found to the most likely reflect the evolutionary events.



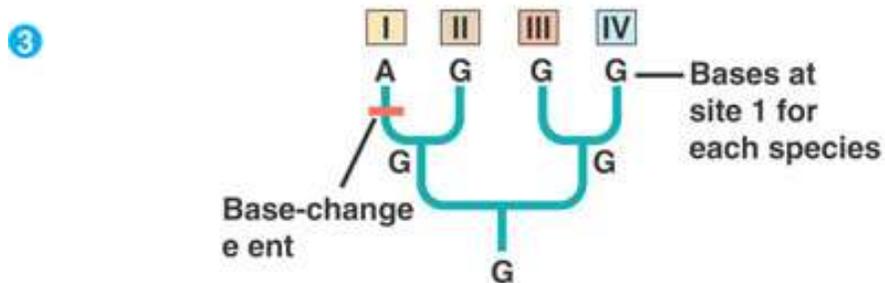
Maximum parsimony

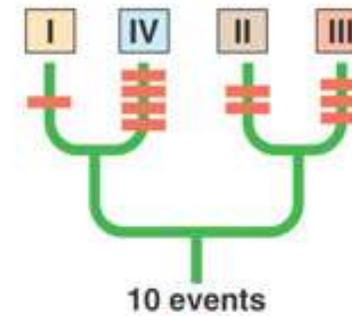
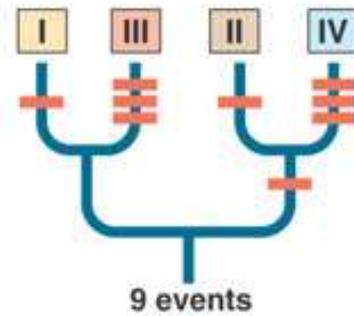
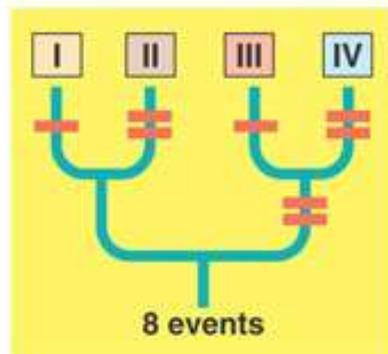
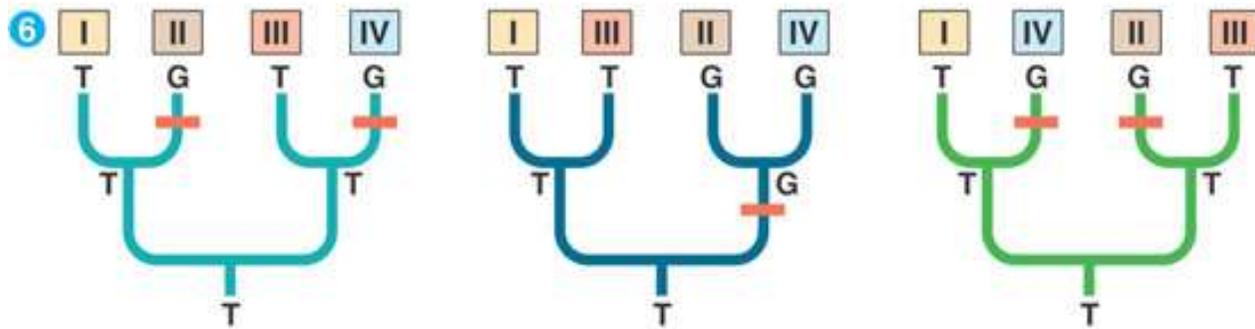
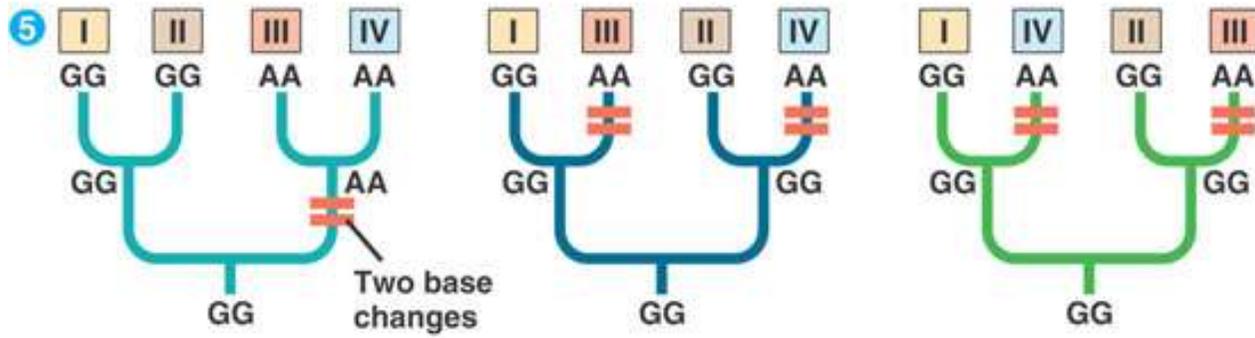
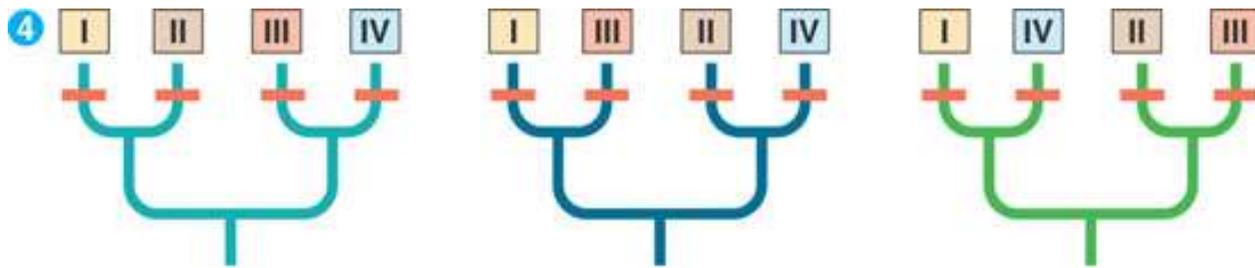
With multiple explanations, the simple explanation is the most consistent with the facts.

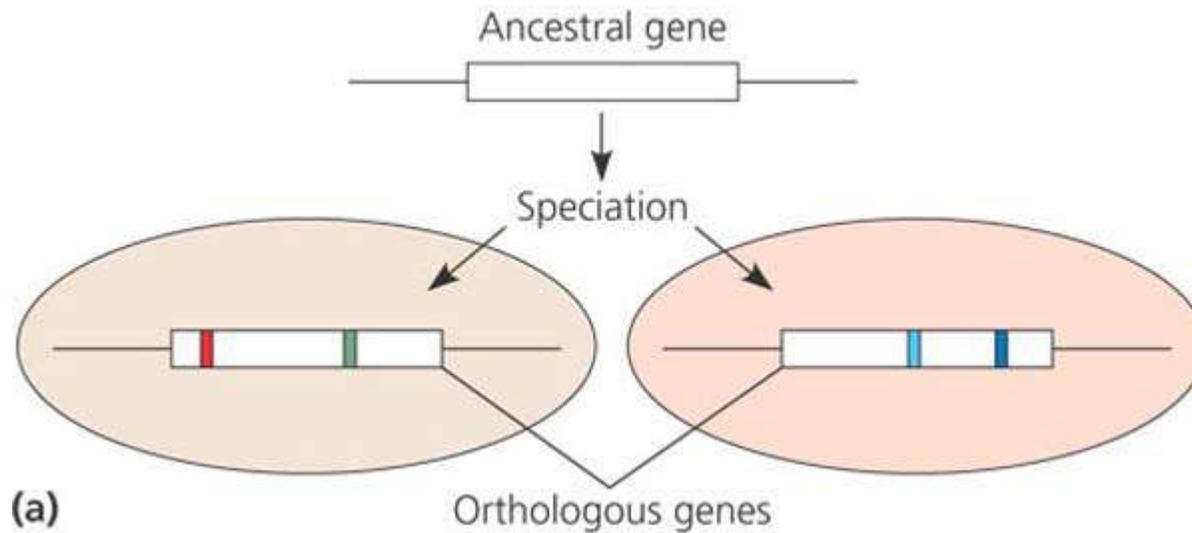
2

Sites in DNA sequence

	1	2	3	4	5	6	7
I	A	G	G	G	G	G	T
II	G	G	G	A	G	G	G
III	G	A	G	G	A	A	T
IV	G	G	A	G	A	A	G



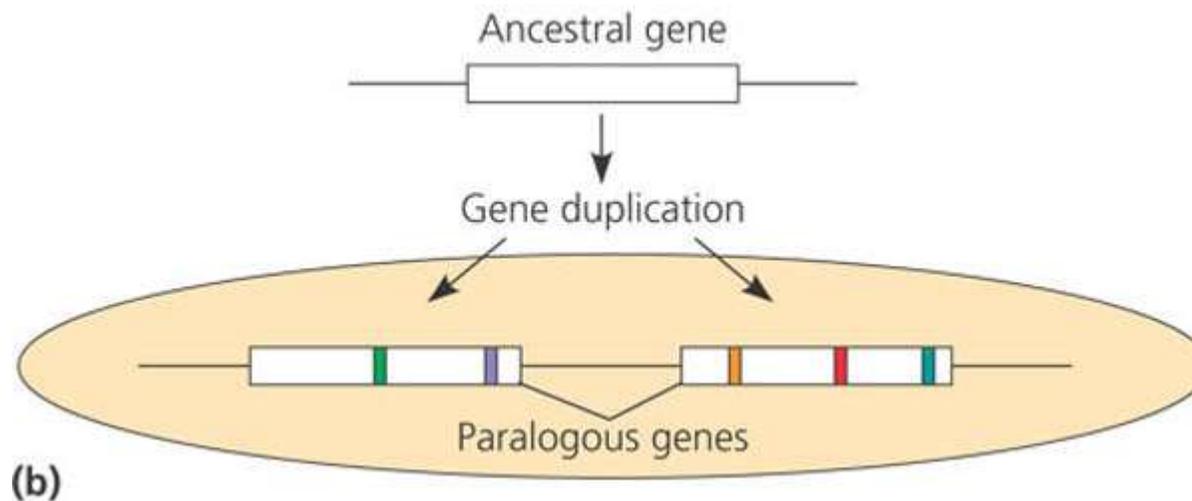




Orthologous genes

Gene that pass straight from generation to generation, but end up in different lineages because of speciation.

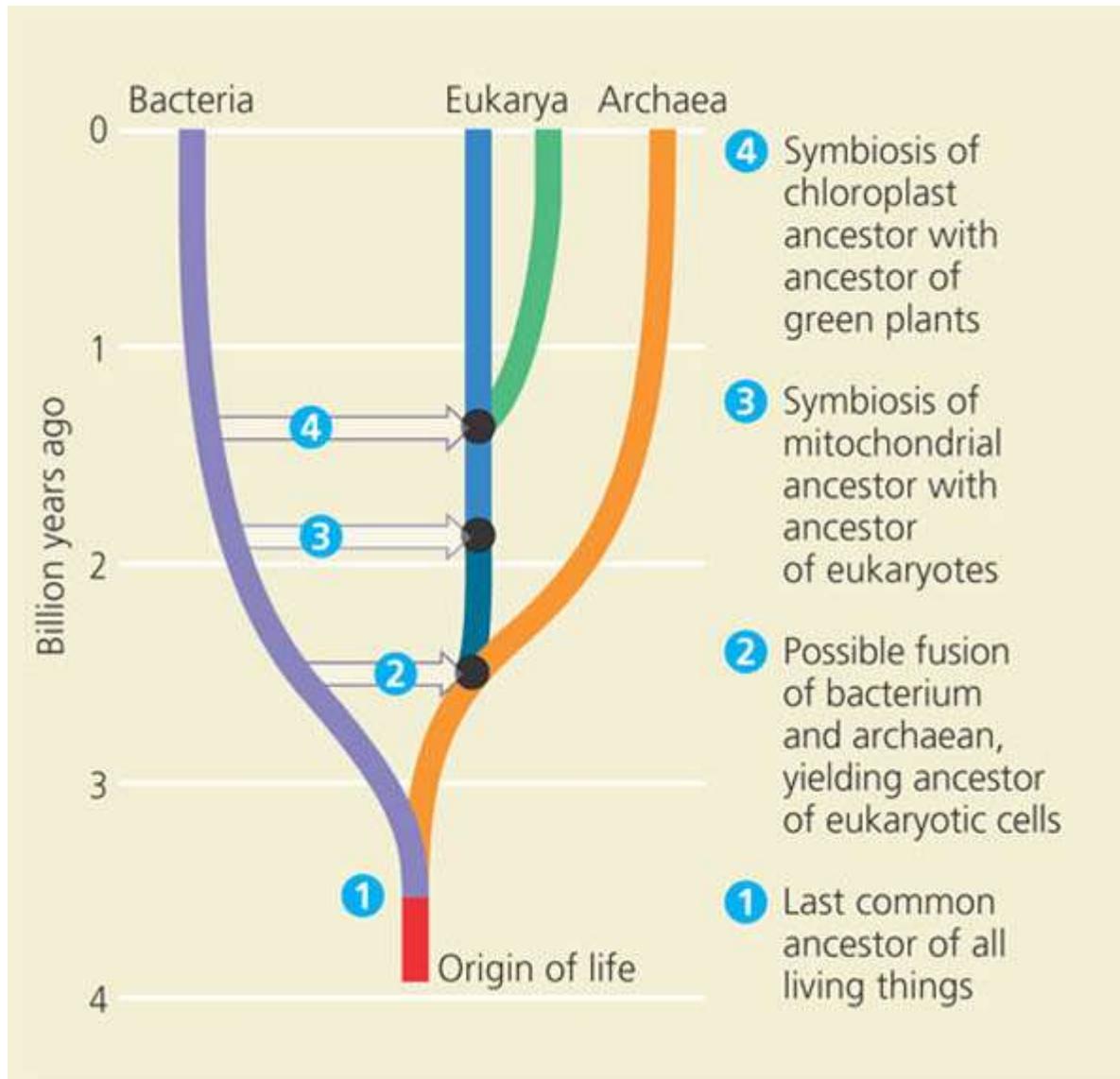
Diverge between lineages.



Paralogous genes

Genes that are the result of duplication event, can be found in the same genome, increasing copy number.

Diverge within lineage.



The universal tree of life

Result from different events, symbiosis, horizontal gene transfer, etc.

Evaluating Phylogenetic Trees

- **Bootstrap**

- A computational technique for estimating statistics when distribution is difficult to be analysed.

- **Since phylogenetic study introduction**

- Bootstrap is used to estimate the confidence level of phylogenetic hypotheses.

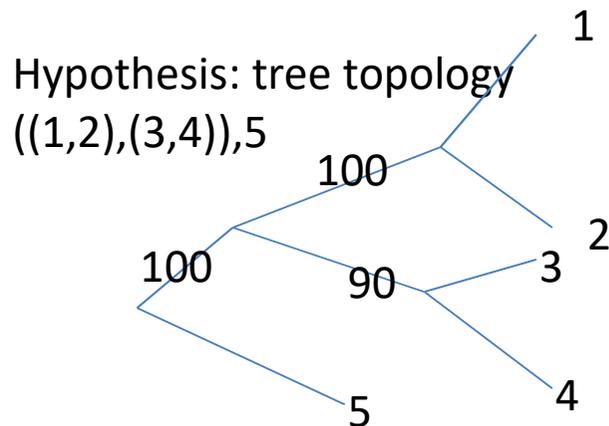
- **Making pseudosamples**

- Original sample represents original sample space.
- Pseudosample represents real sample.

Bootstrapping

- **Original samples**

- OTU 1 GCAGTACT...
- OTU 2 GTAGTACT...
- OTU 3 ACAATACC...
- OTU 4 ACAACACT...
- OTU 5 GCGGCATT...
(outgroup)



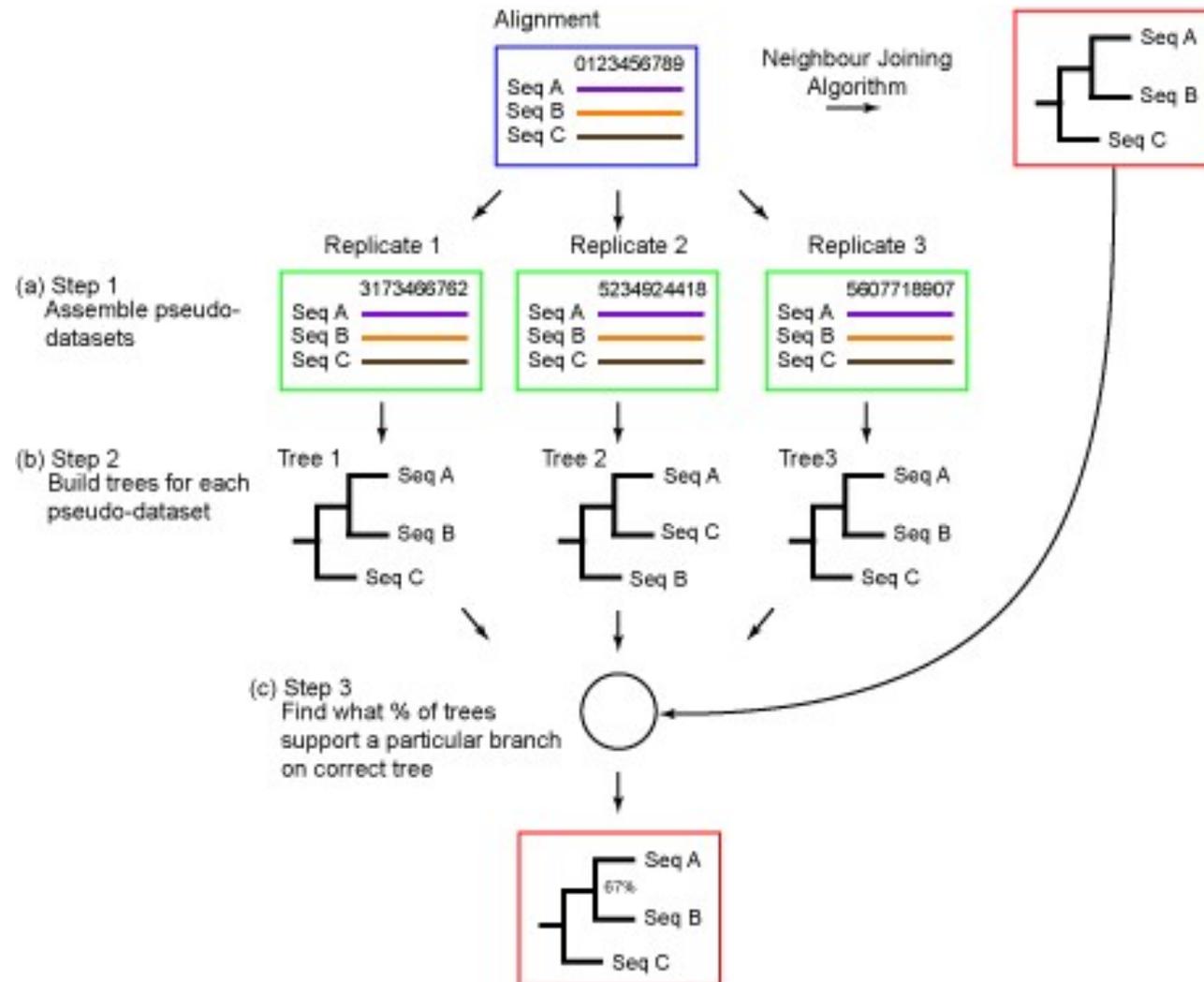
- **Pseudosamples**

- OTU 1 ACATACTC...
- OTU 2 AGATGCTT..
- OTU 3 AAACACTC...
- OTU 4 AAACACCC
- OTU 5 AGACATCC



Construct a tree
(1,2) is counted.
(3,4) is counted.
(1,2),(3,4) is counted.
((1,2),(3,4),5) is counted.
End of 1 bootstrapping.

Bootstrapping



Phenetic Approach

- **Phenetic approach**
 - Classify organisms based on overall similarity.
 - Regardless of evolutionary relationship.
 - Clustering and ordination method
 - Distance matrix
 - Reduction of variations to manageable level.
 - 2D, 3D graph
 - Loss of information vs simplicity
 - Assumed to be approximate phylogenetic.

Cladistic Approach

- **Cladistic approach**
 - A form of biological systematics which classifies living organisms on the basis of shared ancestry.
 - Construct a tree representing the ancestry of organisms and species, objective and quantitative.
 - **Cladogram**
 - A phylogenetic tree inferred by clustering synapomorphies.

Shared Traits

- **Shared traits**

- **Synapomorphy**

- A shared trait found among two or more taxa and their most recent common ancestor, whose ancestor in turn does not possess the trait.

- E.g. halteres (modify hindwing) found in order Diptera, but not in other insects.

- **Symplesiomorphy**

- A shared trait found among two or more taxa, but which is also found in taxa with an earlier common ancestor.

- E.g. five digits appendages of rats and apes originated very early in tetrapod evolution.

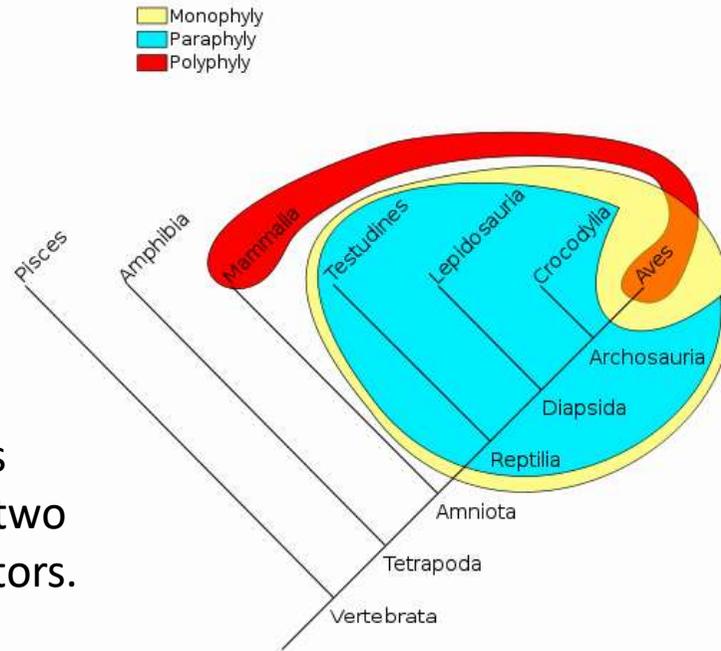
Grouping of organisms

- **Grouping of organisms in phylogenetic trees**
 - **Monophyletic group**
 - Group of all descendants from a common ancestor.
 - **Paraphyletic group**
 - Group of some descendants from a common ancestor.
 - **Polyphyletic group**
 - Group of descendants from more than one common ancestor.

Paraphyletic group:

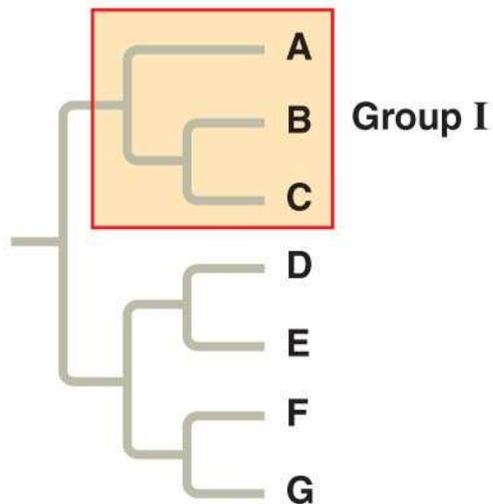
modern reptiles (cyan)
contain not all
descendants from a
common ancestor.

Polyphyletic group: warm-
blood animals (mammals
and birds; red) are from two
different common ancestors.

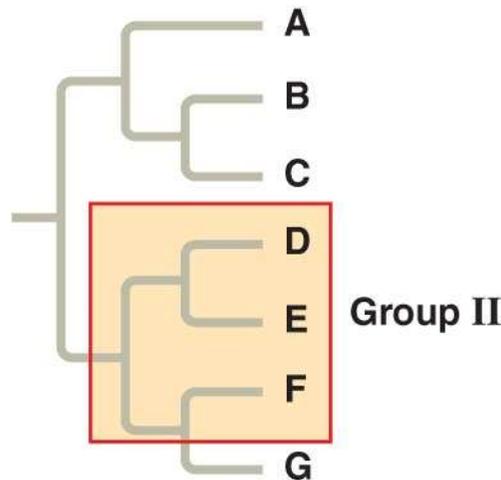


Monophyletic group

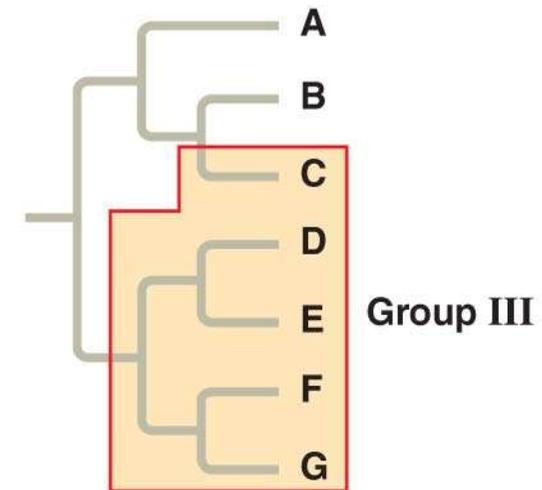
(clade): birds and
reptiles evolved
from one common
ancestor. (yellow)



(a) Monophyletic group (clade)



(b) Paraphyletic group



(c) Polyphyletic group

References

- **Textbooks**

- Freeman, S., & Herron, J. C. (2004). *Evolutionary analysis* (3rd ed.). Upper Saddle River, NJ: Pearson Education.
- Campbell, N. A. (2008). *Biology*. San Francisco, Pearson Benjamin Cummings.
- Hall, B.G. (2004), *Phylogenetic Trees Made Easy. A How-To Manual*. 2nd Edition. Sinauer Associates, Inc. USA.

- **Publicaitons**

- Dudas G, Rambaut A. Phylogenetic Analysis of Guinea 2014 EBOV Ebolavirus Outbreak. *PLOS Currents Outbreaks*. 2014 May 2. Edition 1. doi: 10.1371/currents.outbreaks.84eefe5ce43ec9dc0bf0670f7b8b417d.

- **Internet**

- Biology Homepage: <http://biology.sc.mahidol.ac.th>
- Email: nuttaphon.onp@mahidol.ac.th
- Office: B408