

Lecture 1.1

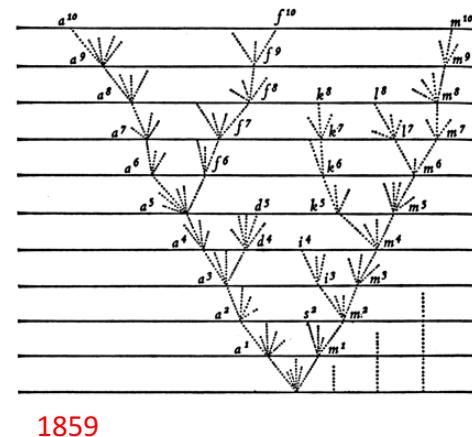
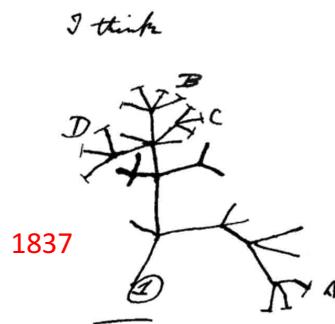
Introduction to Molecular Phylogenetics

Nathan Lo

Phylogenetic Trees

What is a phylogenetic tree?

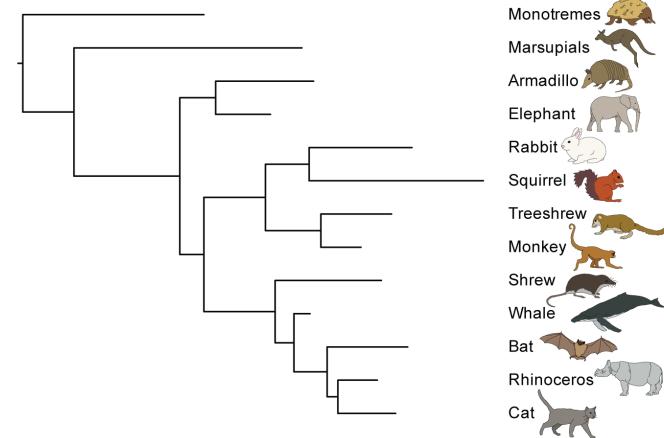
- **Phylogeny:** the true evolutionary relationships among a set of organisms



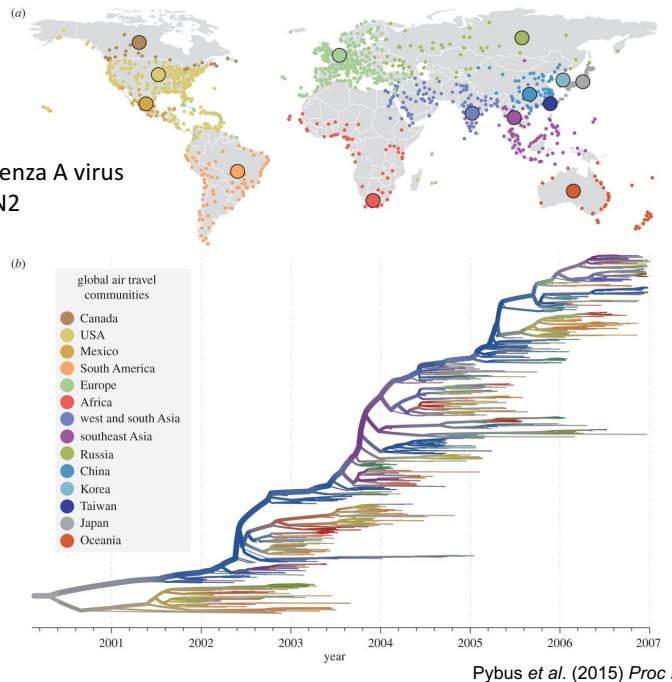
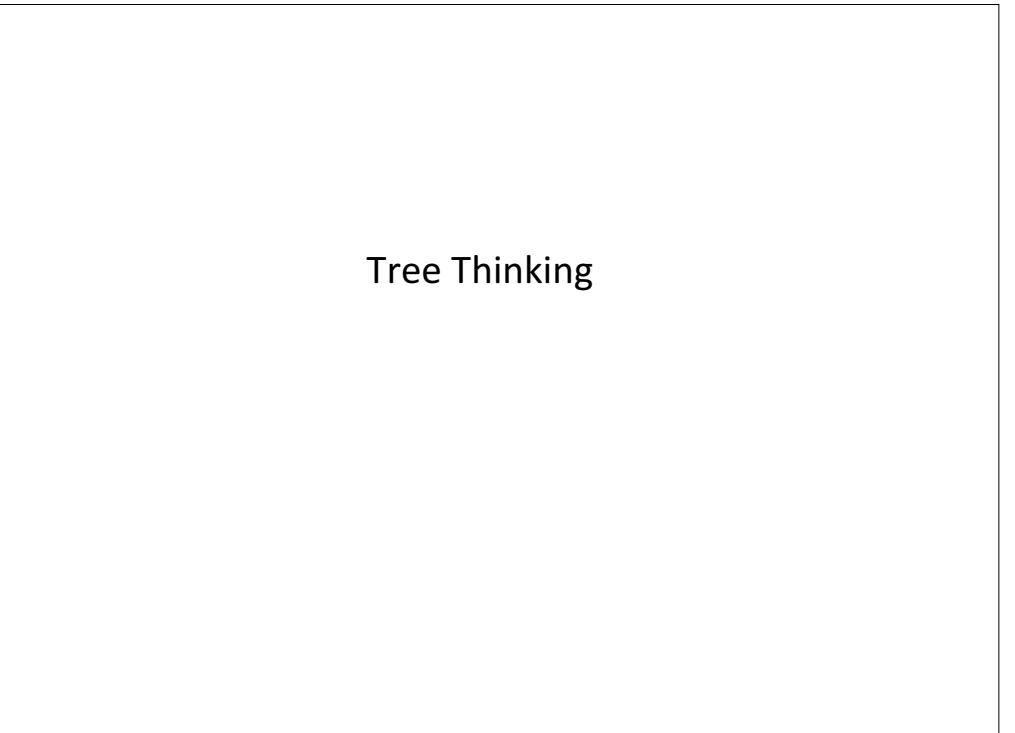
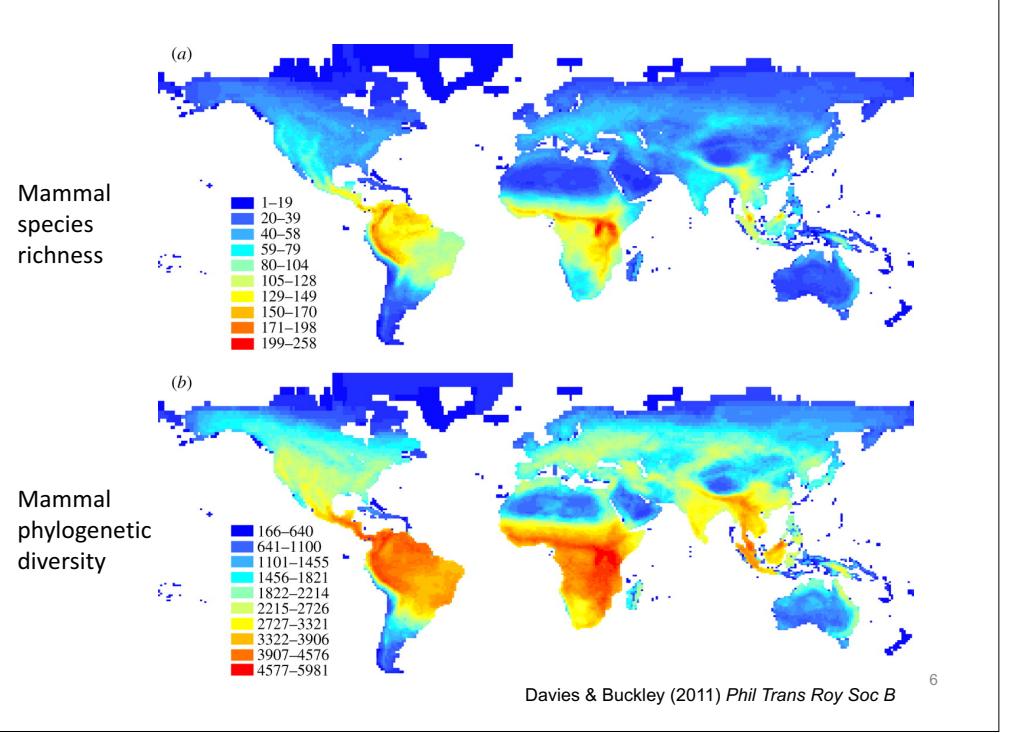
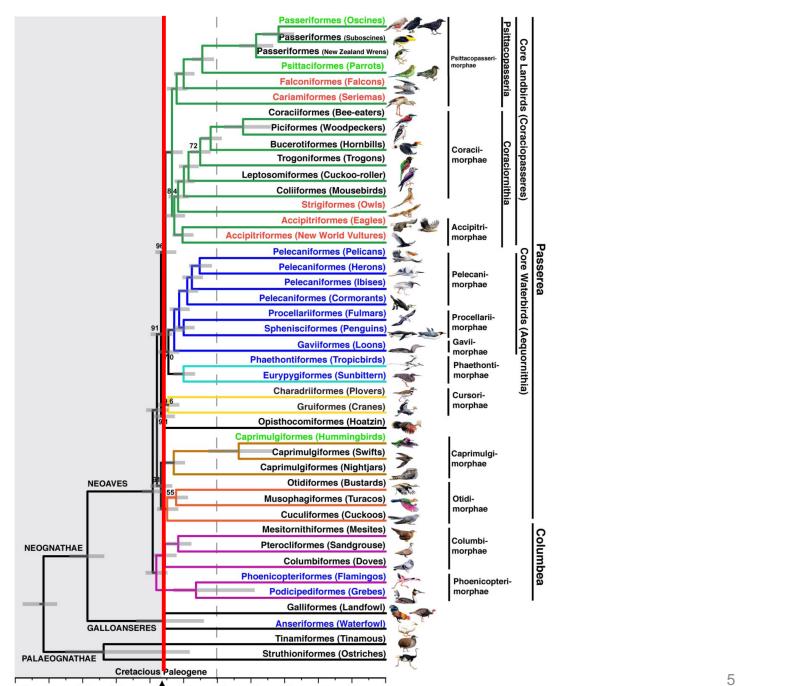
3

Phylogenetic trees

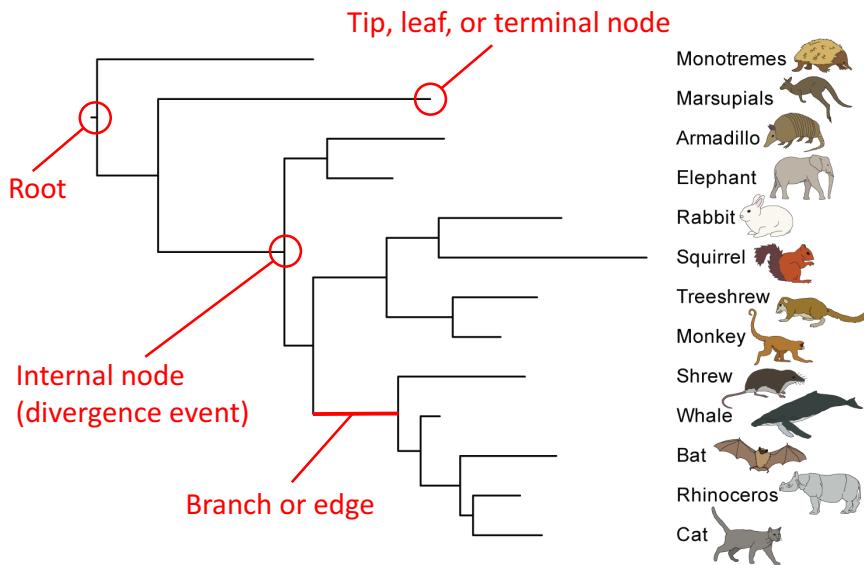
- Topology (relationships)
- Branch lengths (amount of evolutionary change or time)



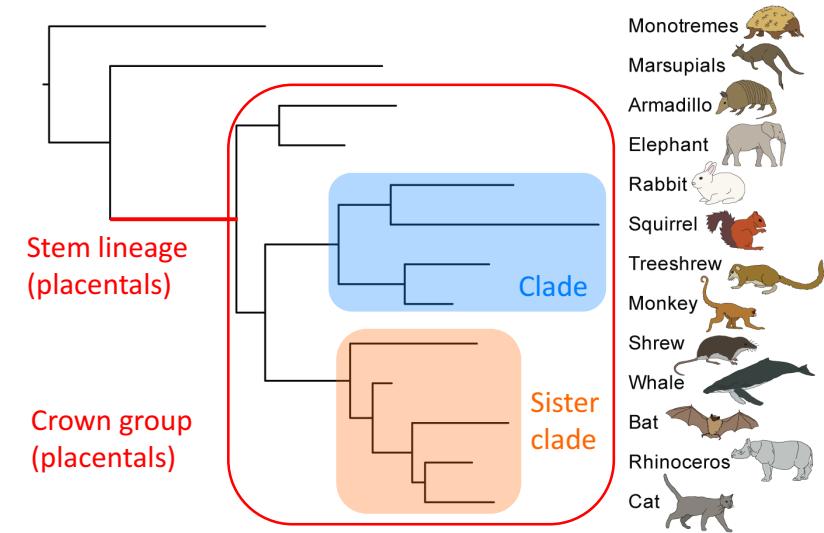
4



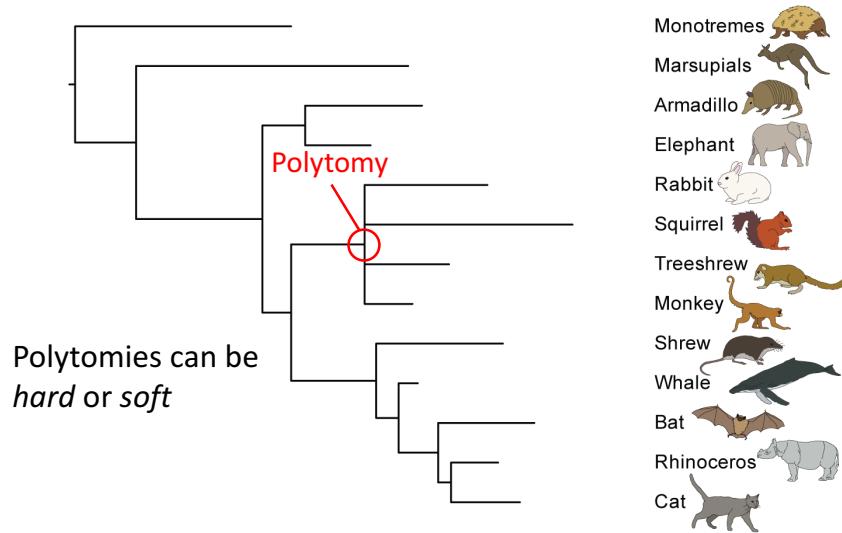
Phylogenetic trees



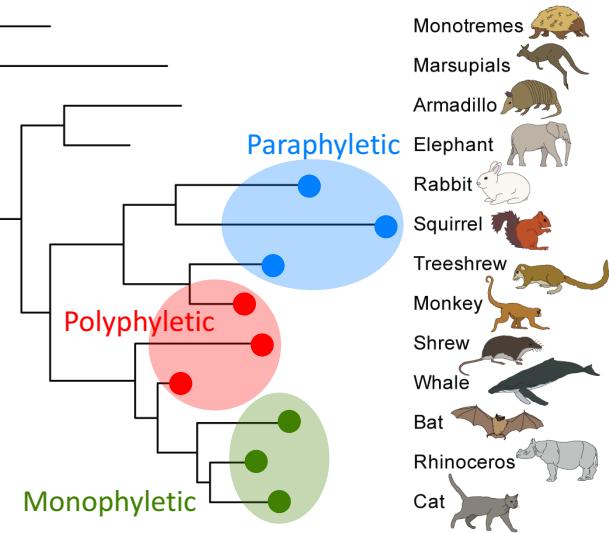
Phylogenetic trees



Phylogenetic trees



Cladistic terms

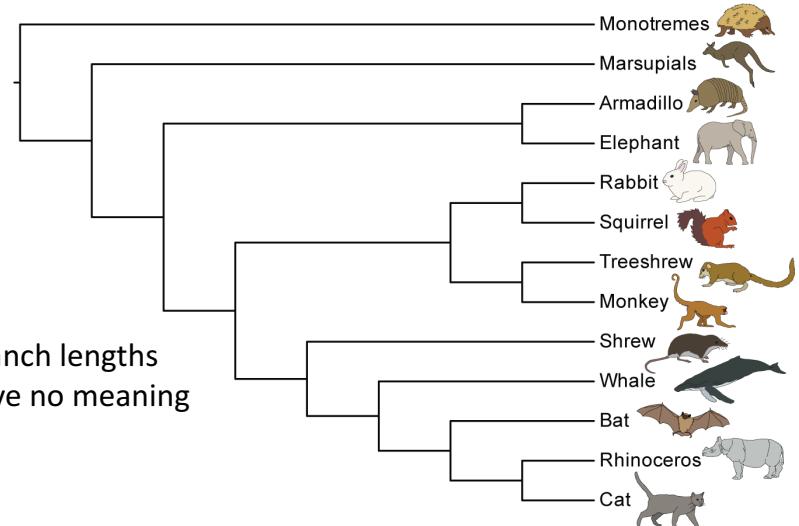


Paraphyletic groups



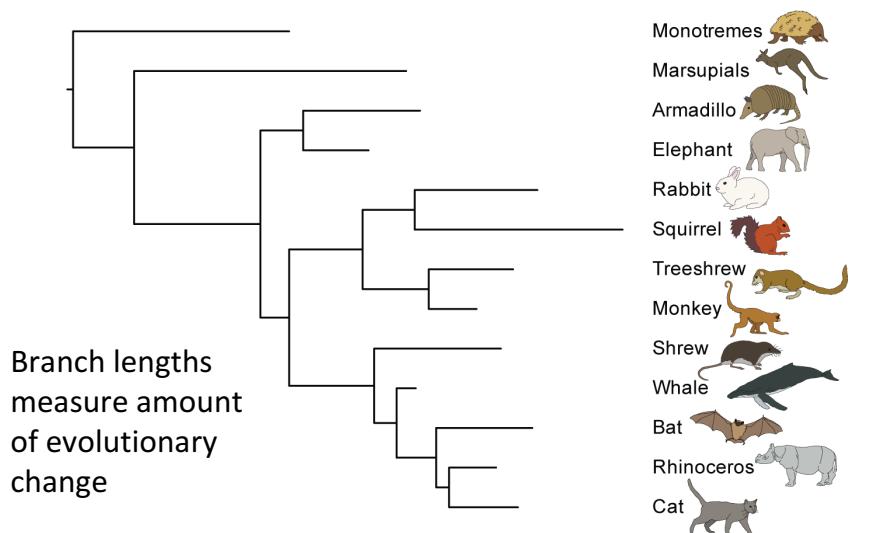
13

Phylogenetic trees: Cladogram



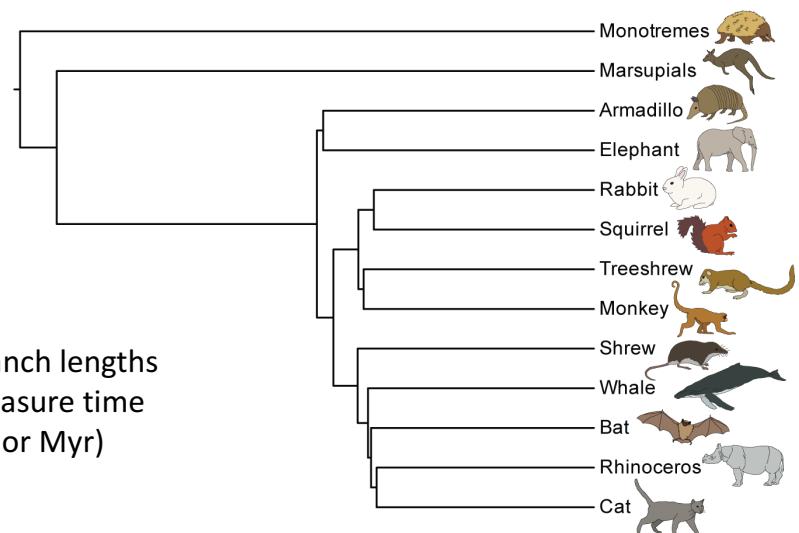
14

Phylogenetic trees: Phylogram



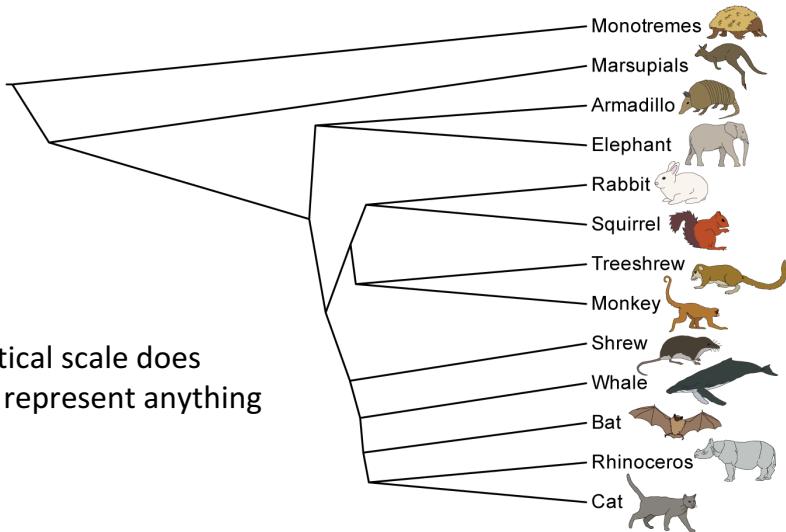
15

Phylogenetic trees: Chronogram



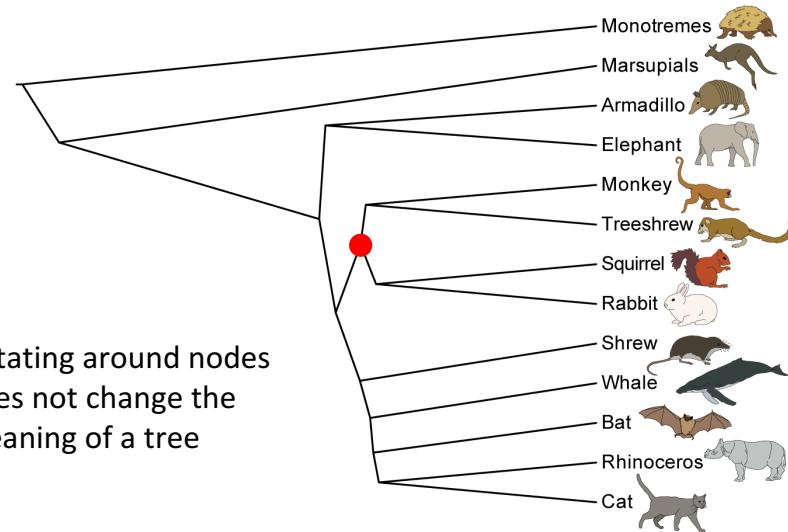
16

Phylogenetic trees



17

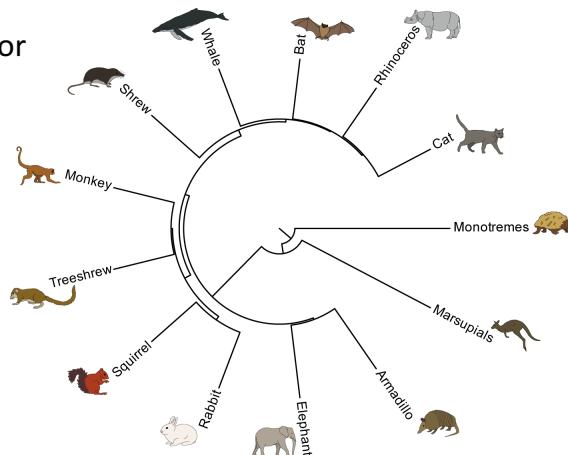
Phylogenetic trees



18

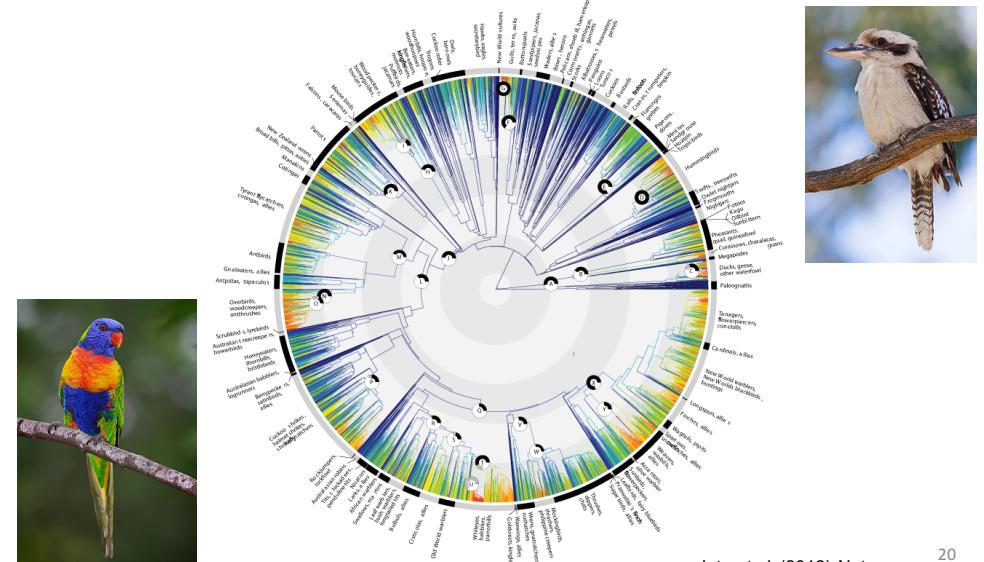
Phylogenetic trees: Circular

- Root is placed in centre
- Cladogram, phylogram, or chronogram
- Often used to display large trees
- Difficult to interpret



19

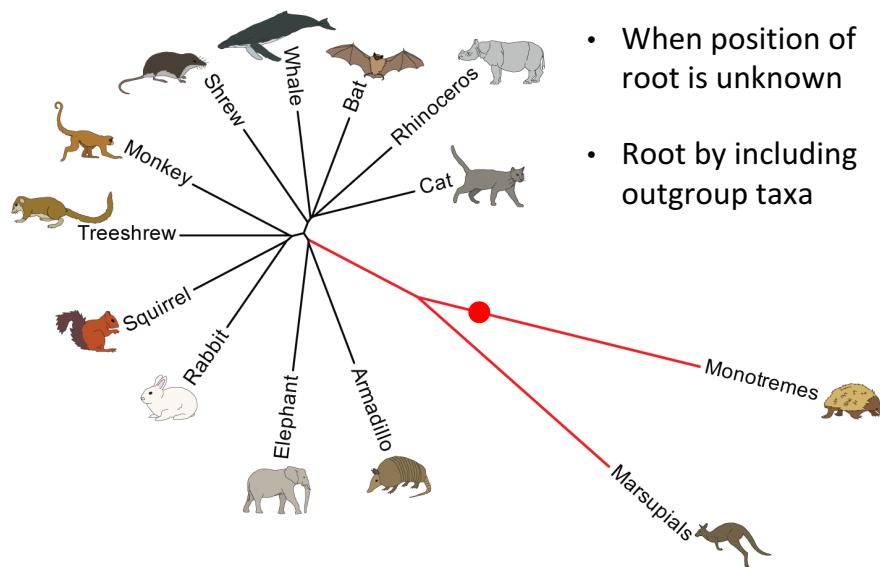
Phylogenetic trees: Circular



Jetz et al. (2012) Nature

20

Phylogenetic trees: Unrooted



21

Rooting

- **Include outgroup taxa**
 - Taxon closely related to ingroup
 - Taxon is not part of ingroup
- **Root at midpoint**
 - Highly unreliable if internal branches are short
- **Use a molecular clock**
 - Automatically estimates position of root

22

Phylogenetic trees: Newick format

- Without branch lengths (cladogram):
 - `(Monotremes,(Marsupials,((Elephant,Armadillo),((Squirrel,Rabbit),(Monkey,Treeshrew)),(Shrew,(Whale,(Bat,Cat,Rhinoceros))))));`
- With branch lengths (phylogram/chronogram):
 - `(Monotremes:12.0,(Marsupials:11.0,((Elephant:1.0,Armadillo:1.0):9.0,((Squirrel:1.0,Rabbit:1.0):2.0,(Monkey:1.0,Treeshrew:1.0):2.0):5.0,(Shrew:4.0,(Whale:3.0,(Bat:2.0,(Cat:1.0,Rhinoceros:1.0):1.0):1.0):1.0):4.0):2.0):1.0):1.0);`

23

Molecular Phylogenetics

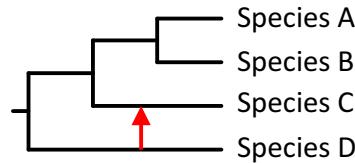
Phylogenetic analysis

- Sometimes we know the phylogeny
 - Viral transmission histories
 - Pedigrees (humans, domesticated animals, lab organisms, etc.)
- Usually we do not know the phylogeny but we can estimate it
 - Morphological data
 - Molecular data

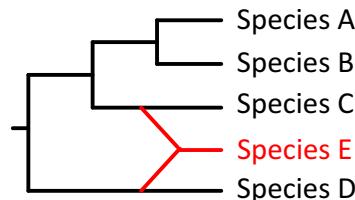
25

Non-treelike evolution

Horizontal gene transfer



Hybrid speciation

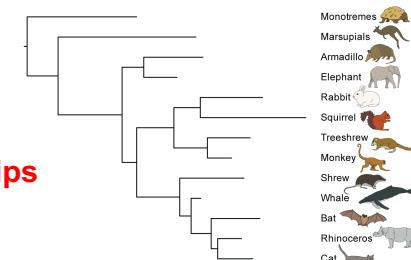


27

Fundamental assumptions

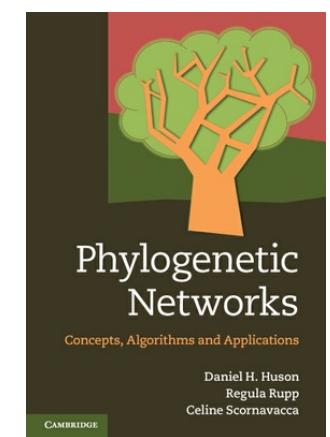
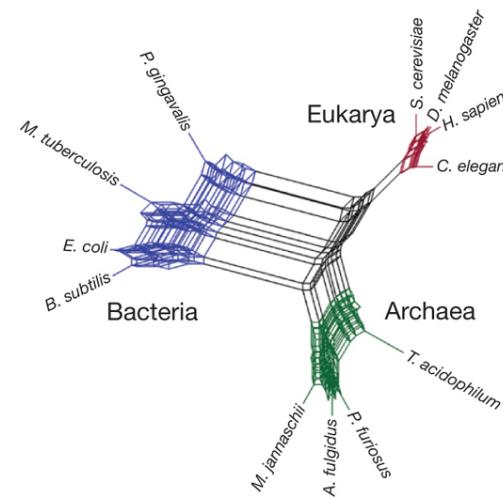
- Phylogenetic methods make several fundamental assumptions:
 - Relationships among taxa can be represented by a tree
 - Homologous characters are being compared
 - Characters are mutually independent
 - Lineages evolve independently

When might relationships
not be treelike?



26

Phylogenetic networks

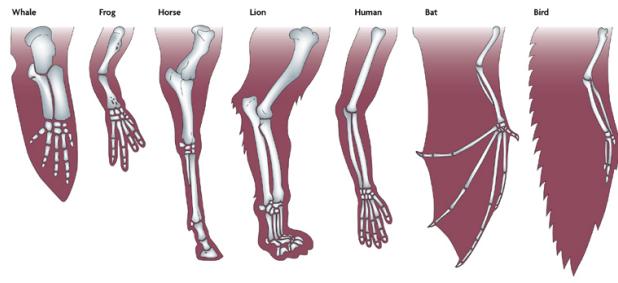


Theobald (2010) *Nature*

28

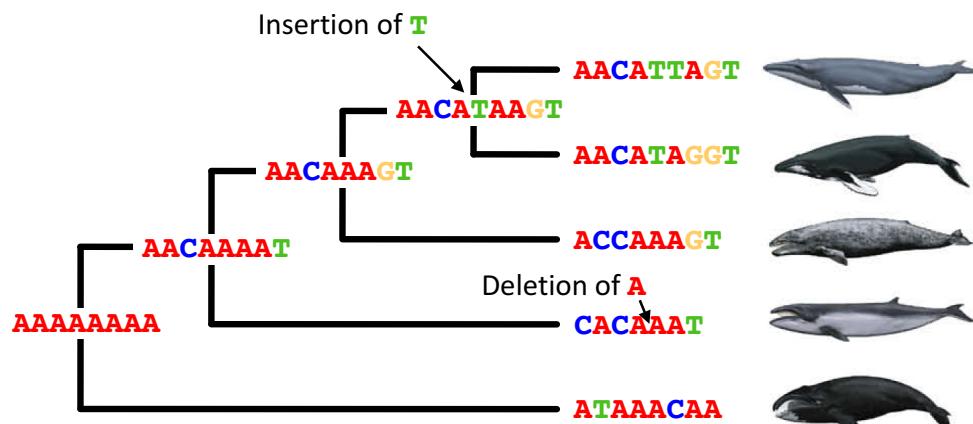
Fundamental assumptions

- Phylogenetic methods make several fundamental assumptions:
 - Relationships among taxa can be represented by a tree
 - Homologous characters are being compared
 - Characters are mutually independent
 - Lineages evolve independently



29

Example: Whales



31

Character homology

- Comparing strings of nucleotides
- Each nucleotide site is a character
- But DNA sequences can vary in length

blue whale CGTTAGTACACT
humpback whale CGATAGTTCACT
gray whale CGTTAGTTTACC
fin whale CATTGGATTACT
right whale CATTGGTTTACT



30

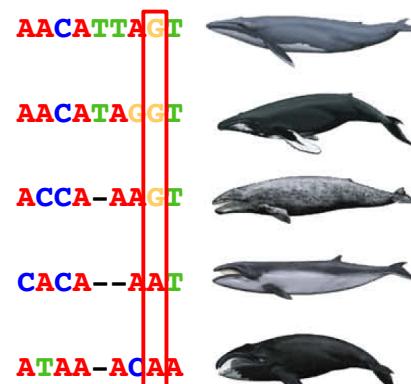
DNA sequence alignment

Whale Species	DNA Sequence
blue whale	AACATTAGT
humpback whale	AACATAGGT
gray whale	ACCA-AAGT
fin whale	CACA--AAT
right whale	ATAA-ACAA

32

DNA sequence alignment

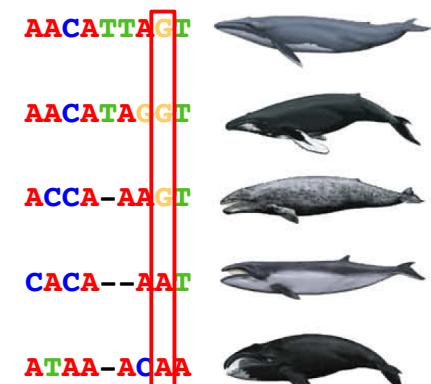
- Homologous site
- Inherited from the common ancestor of all sequences in the alignment
- The aim of sequence alignment is to maximise the number of sites for which you can infer homology



33

DNA sequence alignment

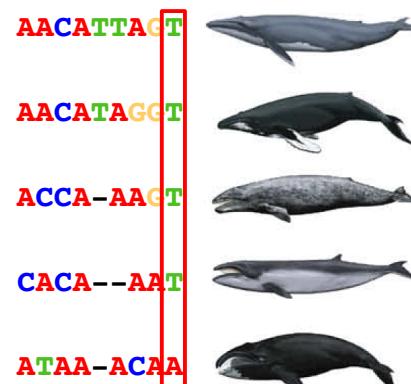
- Groups together the first 3 sequences
- Groups together the last 2 sequences
- Informative for all phylogenetic methods



34

DNA sequence alignment

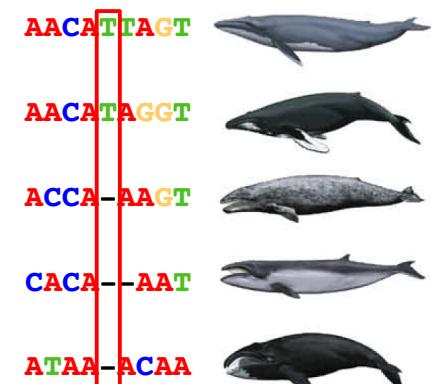
- Does not group any sequences
 - Not useful for maximum parsimony
- But informative for estimating amount of evolutionary change
 - Useful for other methods



35

DNA sequence alignment

- Indel – insertion or deletion
- Potentially informative
- Most phylogenetic methods do not really use indel data



36

A practical approach

Align sequences using automated methods

CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice

Julie D.Thompson, Desmond G.Higgins* and Toby J.Gibson*

Software

MUSCLE: a multiple sequence alignment method with reduced time and space complexity

Robert C Edgar*

Open Access

MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform

Kazutaka Katoh, Kazuharu Misawa¹, Kei-ichi Kuma and Takashi Miyata*

37

A practical approach

Align sequences using automated methods

↓

Adjust alignments by eye

CTATGTGGCACCCAGCCCATGCA--AGC
ATATGTGGCA-----CCCAGGCA--AG--
ATATGTGGCACCCAGCCCATGCATTT--

38

A practical approach

Align sequences using automated methods



Adjust alignments by eye

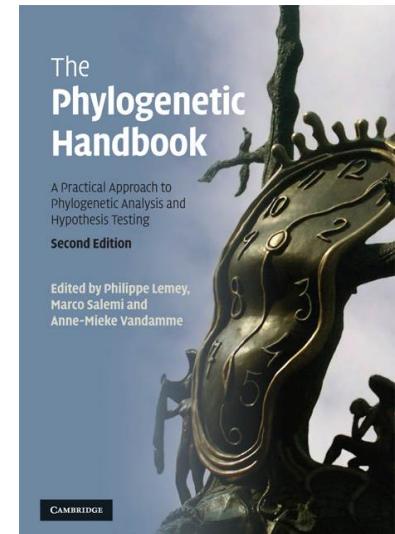
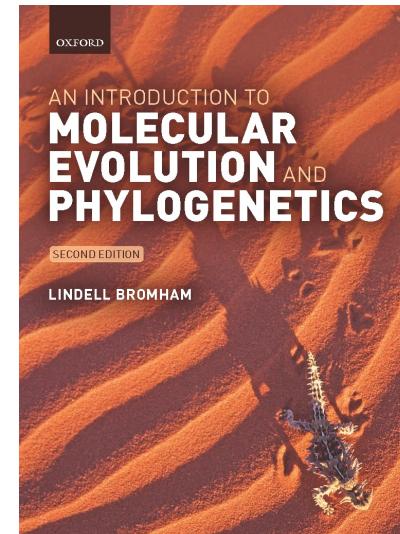


Delete sites with uncertain homology

CTATGTGGCACCCAGCCCATGCA--AGC
ATATGTGGCA-----CCCAGGCA--AG--
ATATGTGGCACCCAGCCCATGCATTT--

39

Useful references



40