

Molecular Phylogenetics Recipe

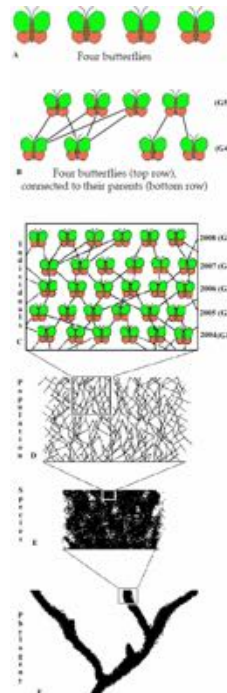


1. Ask a phylogenetic question
2. Collect homologous sequences
3. Do multiple sequence alignment
4. Perform phylogeny estimation
5. Test the reliability of phylogenetic estimate(s)
6. Interpret and apply your results

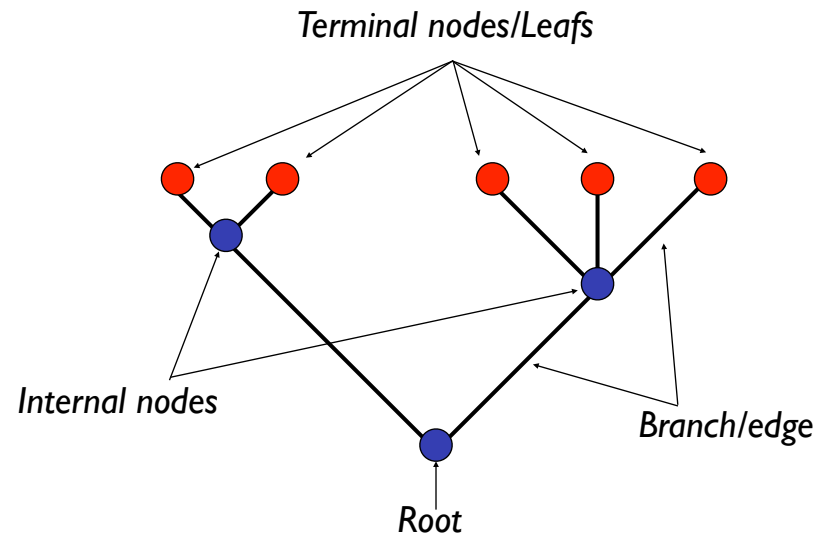
Introduction to Phylogenetics

What is a phylogenetic tree?

- A tree is a mathematical structure (*connected graph with no cycles*), which is used to model the actual evolutionary history of a group of sequences or organisms.
- The actual pattern of historical relationships is an evolutionary tree which we try to estimate



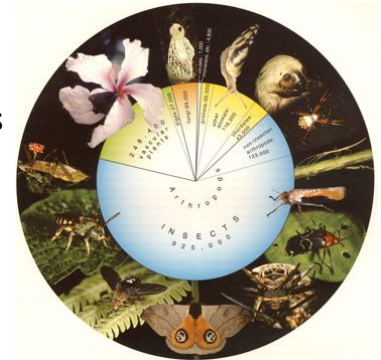
Phylogenetic tree terms



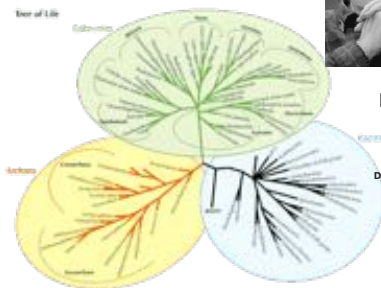


Applications of Phylogenetic Analysis

- Systematics and classification
- Discovering new life forms
- Phylogeography and speciation
- Molecular evolution
- Genomics
- Epidemiology and forensics
- Biotechnology
- Agriculture
- Conservation



Discovering new life forms



Environmental Genome Shotgun Sequencing of the Sargasso Sea

J. Craig Venter,^{1*} Karin Remington,¹ John F. Heidelberg,³ Aaron L. Halpern,² Doug Rusch,² Jonathan A. Eisen,³ Dongying Wu,³ Ian Paulsen,³ Karen E. Nelson,³ William Nelson,³ Derrick E. Fouts,³ Samuel Levy,² Anthony H. Knap,⁴ Michael W. Lomas,⁶ Ken Nealson,⁵ Owen White,⁵ Jeremy Peterson,³ Jeff Hoffman,³ Rachel Parsons,⁶ Holly Baden-Tillson,¹ Cynthia Pfannkuch,¹ Yu-Hui Rogers,⁴ Hamilton O. Smith¹

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 11, pp. 5086-5090, November 1977
Evolution

Phylogenetic structure of the prokaryotic domain: The primary kingdoms

(archaea/bacteria/eukaryote/16S ribosomal RNA/molecular phylogeny)

CARL R. WOESE AND GEORGE E. FOX*

Department of Genetics and Development, University of Illinois, Urbana, Illinois 61801

Font Color Legend
IBEA-SAR samples
Other environmental samples
Cultured species

Proteorhodopsin-like 4

Proteorhodopsin-like 3

Proteorhodopsin-like 2

Proteorhodopsin

Proteorhodopsin-like 1

IBEA SAR Novel Opsins 1-4

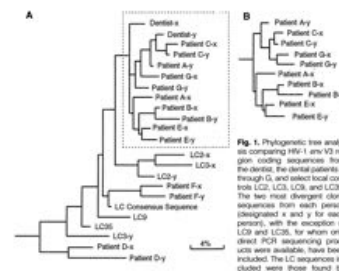
Halorhodopsins

Fungal Opsin

Bacteriorhodopsins

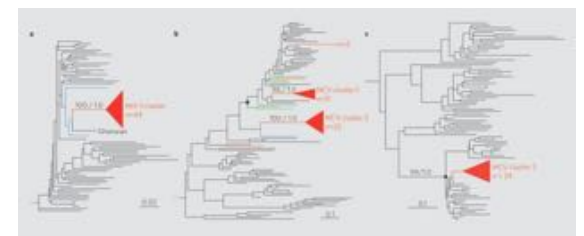
Sensory Opsins

Disease Transmission and Medical Forensics



Molecular Epidemiology of HIV Transmission in a Dental Practice

Chin-Yih Ou, Carol A. Ciesielski, Gerald Myers, Claudius I. Bandea, Chi-Cheng Luo, Bette T. M. Korber, James I. Mullins, Gerald Schochetman, Ruth L. Berkelman, A. Nikki Economou, John J. Witte, Lawrence J. Furman, Glen A. Satten, Kersti A. MacInnes, James W. Curran, Harold W. Jaffe, Laboratory Investigation Group,*



Brief Communications

Nature 444, 838-837 (14 December 2006) | doi:10.1038/444838a | Received 4 November 2006; Accepted 24 November 2006; Published online 6 December 2006

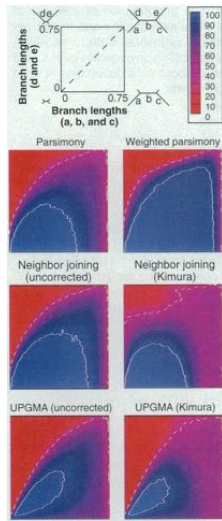
Molecular Epidemiology: HIV-1 and HCV sequences from Libyan outbreak

Tullio de Oliveira¹, Oliver G. Pybus², Andrew Rambaut³, Marco Salemi⁴, Sharon Casanovi⁵, Massimo Ciccozzi⁶, Giovanni Rezza⁷, Guido Castelli Gattinara⁸, Roberta D'Arrigo⁹, Massimo Amicozzi¹⁰, Luc Perrin¹¹, Vittorio Colizzi¹², Carlo Federico Perno¹³ and Benghazi Study Group¹⁴

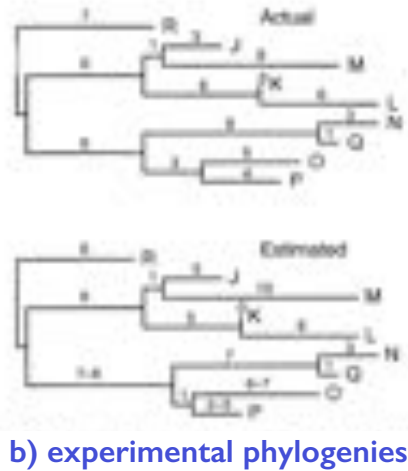
In 1998, outbreaks of human immunodeficiency virus type 1 (HIV-1) and hepatitis C virus (HCV) infection were reported in children attending Al-Fateh Hospital in Benghazi, Libya. Here we use molecular phylogenetic techniques to analyse new virus sequences from these outbreaks. We find that the HIV-1 and HCV strains were already circulating and prevalent in this hospital and its environs before the arrival in March 1998 of the foreign medical staff (five Bulgarian nurses and a Palestinian doctor) who stand accused of transmitting the HIV strain to the children.

Ou et al. 1992

How do we know that phylogenetics work?



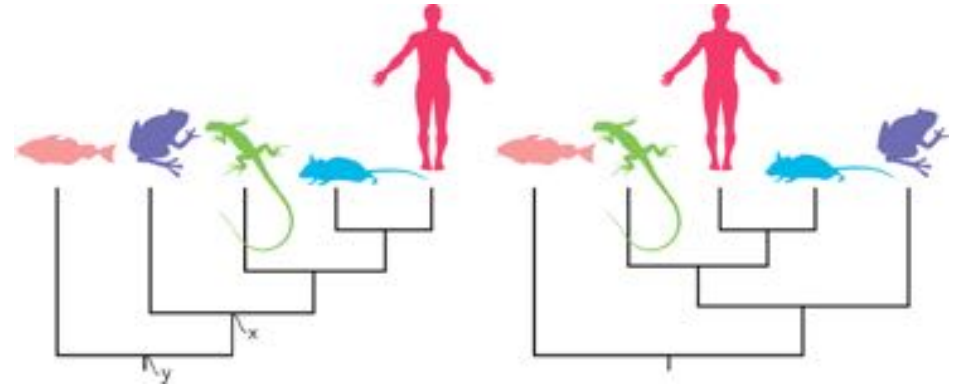
a) simulations



b) experimental phylogenies

ARTICLE
Application and Accuracy of Molecular Phylogenies
David M. Hillis, John P. Huelsenbeck, Clifford W. Cunningham

How to read a phylogenetic tree?



Which phylogenetic trees is more accurate?

Baum et al. 2005



Opinion

TRENDS in Ecology and Evolution Vol.20 No.3 March 2005

Full text provided by www.sciencedirect.com
SCIENCE @ DIRECT®

Do early branching lineages signify ancestral traits?

Michael D. Crisp and Lyn G. Cook

School of Botany and Zoology, The Australian National University, Canberra, ACT 0200, Australia



'To catch a glimpse of the very beginnings of the mammalian dynasty, we must travel to Australia. I'm looking for one of the most ancient of all mammals. It's so ancient, it shares at least one characteristic with the reptiles.' [David Attenborough, *The Life of Mammals*, Episode 1 (BBC, 2003)], which is referring to monotremes (the Australasian platypus and echidnas)



Opinion

TRENDS in Ecology and Evolution Vol.20 No.3 March 2005

Full text provided by www.sciencedirect.com
SCIENCE @ DIRECT®

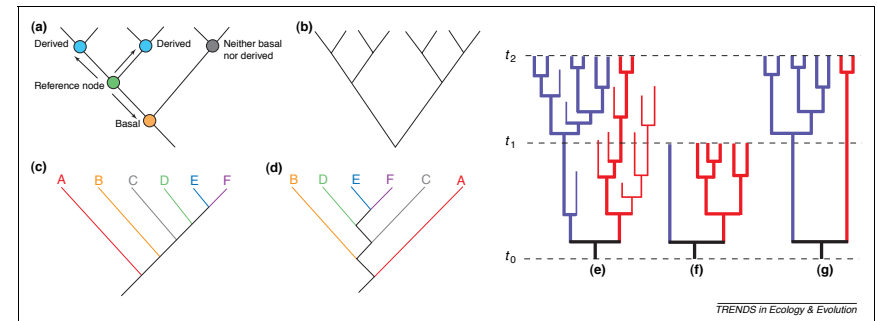
Do early branching lineages signify ancestral traits?

Michael D. Crisp and Lyn G. Cook

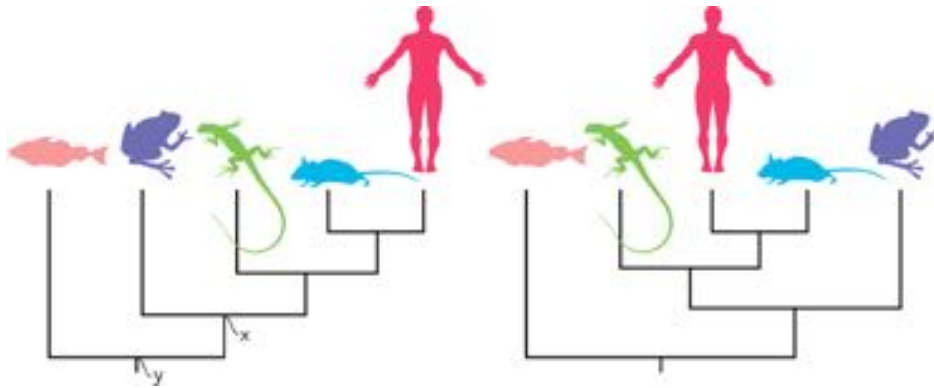
School of Botany and Zoology, The Australian National University, Canberra, ACT 0200, Australia



'To catch a glimpse of the very beginnings of the mammalian dynasty, we must travel to Australia. I'm looking for one of the most ancient of all mammals. It's so ancient, it shares at least one characteristic with the reptiles.' [David Attenborough, *The Life of Mammals*, Episode 1 (BBC, 2003)], which is referring to monotremes (the Australasian platypus and echidnas)

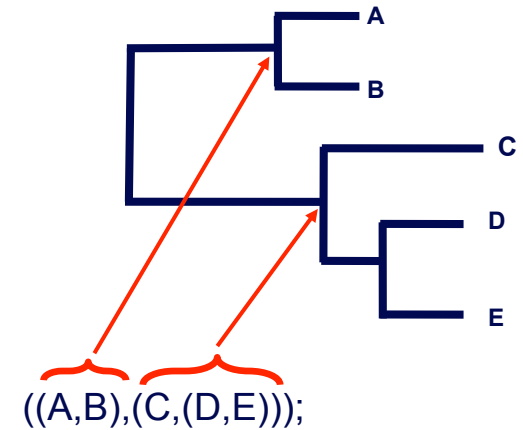


How to write phylogenetic trees?



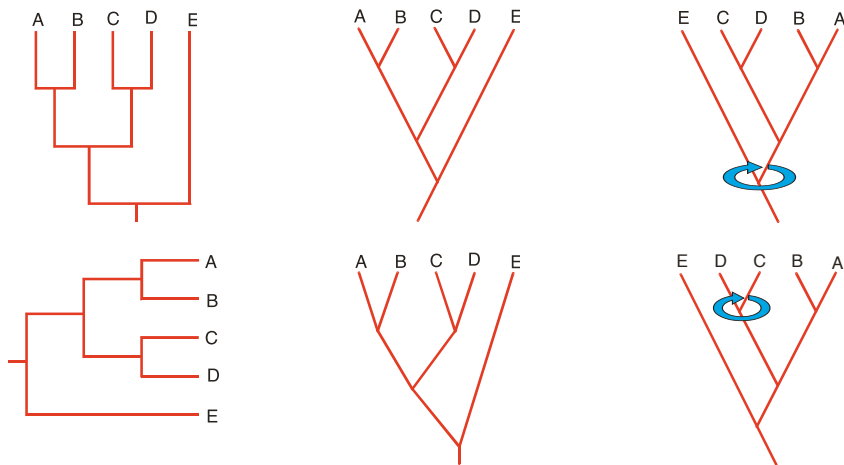
Baum et al. 2005

Newick format



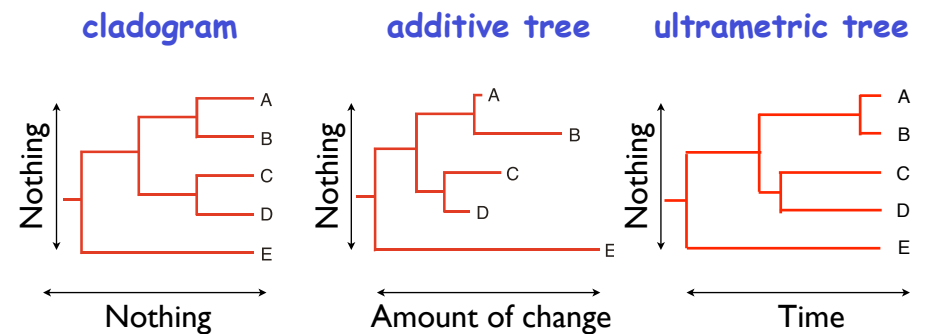
14

Different ways to draw a tree:

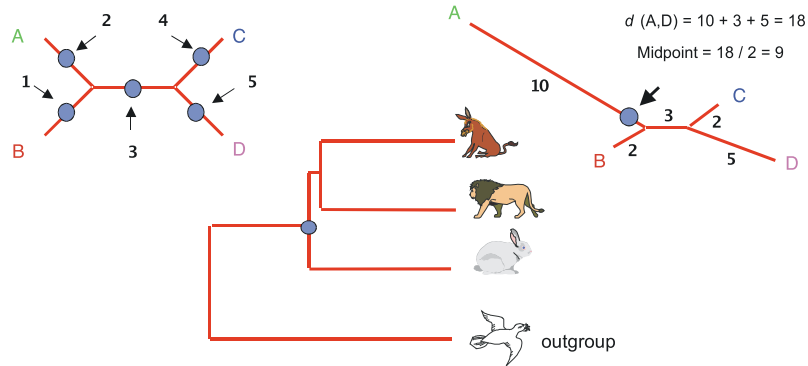


The Newick tree format: $((a,b),(c,d),e)$;

What do different axes of a tree represent?

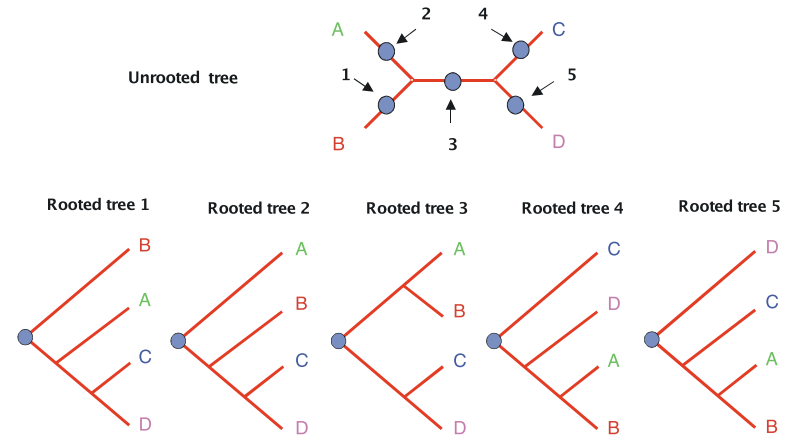


How to root?



Use statistical tools that will root trees automatically (e.g. mid-point rooting) **(BEWARE!)** or outgroup rooting

Number of places for a root



...equals to the number of branches on a tree
...equals to $(2n-3)$ for n -species tree
(each new species adds 2 branches to a tree)

Why molecular phylogenetics?



- The stream of heredity makes phylogeny: in a sense, it is phylogeny. Complete genetic analysis would provide the most priceless data for the mapping of this stream. *George G. Simpson, 1945*



- "I do not fully understand why we are not proclaiming the message from the housetops ... We finally have a method that can sort homology from analogy." *Stephen J. Gould, 1985*

Molecular Phylogenetics Recipe

1. Ask a phylogenetic question
2. Collect homologous sequences
3. Do multiple sequence alignment
4. Perform phylogeny estimation
5. Test the reliability of phylogenetic estimate(s)
6. Interpret and apply your results



Definitions of homology

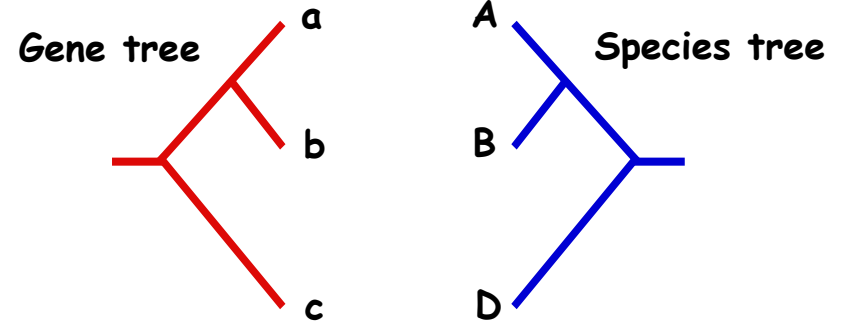
- **Homology**: is the relationship of two characters that have descended, usually with divergence, from a common ancestral character.
- **Analogy** is the relationship of two characters which, although similar, have descended convergently from unrelated characters.

Walter Fitch, 2000

Homology

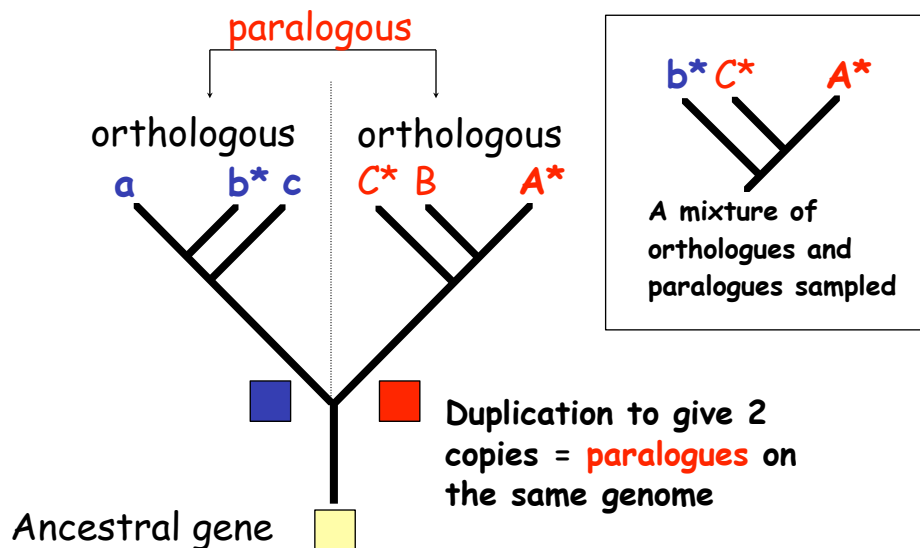
a personal view on some of the problems

Gene trees and species trees



We often assume that gene trees give us species trees

Orthologues and paralogues



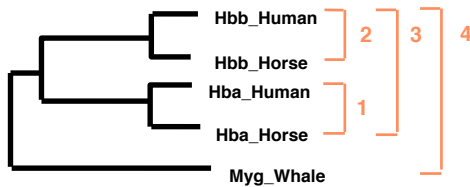
Molecular Phylogenetics Recipe

1. Ask a phylogenetic question
2. Collect homologous sequences
3. Do multiple sequence alignment
4. Perform phylogeny estimation
5. Test the reliability of phylogenetic estimate(s)
6. Interpret and apply your results



Overview of ClustalW Procedure

Hbb_Human	1	-			
Hbb_Horse	2	.17	-		
Hba_Human	3	.59	.60	-	
Hba_Horse	4	.59	.59	.13	-
Myg_Whale	5	.77	.77	.75	.75



alpha-helices

1	PEEKSAVTALWGKVN--VDEVGG	2	3	4
2	GEEKAAVLALWDKVN--EEEVGG			
3	PADKTNVKAANGKVGAHAGEYGA	1		
4	AADKTNVKAAWSKVGGHAGEYGA			
5	EHEWQLVLHVWAKVEADVAGHGQ			

CLUSTAL W

Quick pairwise alignment:
calculate distance matrix

Neighbor-joining tree
(guide tree)

Progressive alignment
following guide tree

Molecular Phylogenetics Recipe

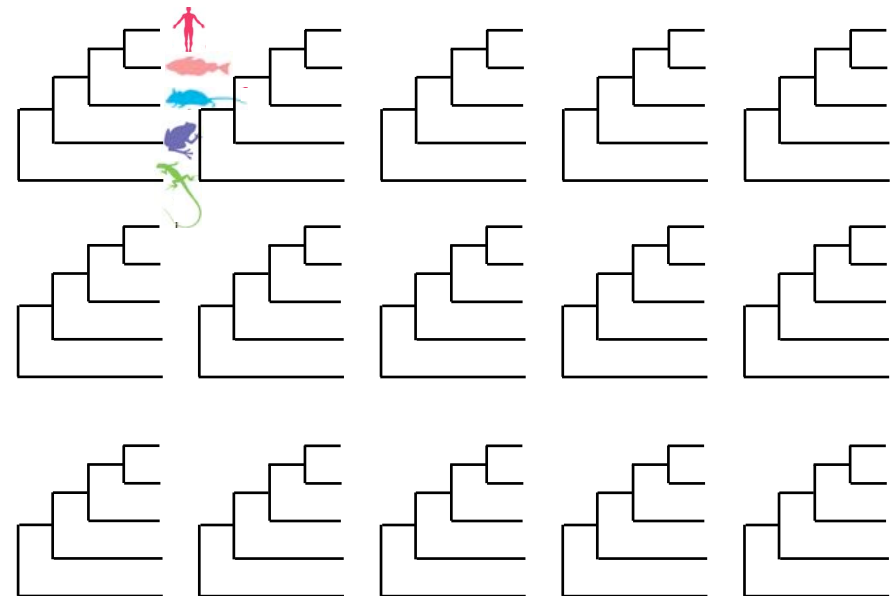
1. Ask a phylogenetic question
2. Collect homologous sequences
3. Do multiple sequence alignment
4. Perform phylogeny estimation
5. Test the reliability of phylogenetic estimate(s)
6. Interpret and apply your results



How to build a phylogenetic tree?



How to build a phylogenetic tree?



Number of taxa	Number of unrooted trees
1	-
2	1
3	1
4	3
5	15
6	105
7	945
8	10,395
9	135,135
10	2,027,025
11	34,459,425
12	654,729,075
13	13,749,310,575
14	7,905,853,580,625
15	213,458,046,676,875
20	221,643,095,476,699,771,875
50	27529213532835651545259729751524430639300973035816196098326553772152587890625
53	> the number of atoms in the universe
67	> volume of the universe in cubic Ångströms
Tree of life	~10 ^{70,000,000}

**(2n-3) places
per tree to
add a new
taxon**

Methods for phylogeny estimation

- Four main methods:
 - Maximum parsimony
 - Distance
 - Maximum likelihood
 - Bayesian

		Type of data	
		Distances	Nucleotide sites
Tree-building method	Clustering algorithm	UPGMA Neighbour joining	
	Optimality criterion	Minimum evolution	Maximum parsimony Maximum likelihood

Distance Method

Cavalli-Sforza and Edwards (1967)

Fitch and Margoliash (1967)

Influenced by Sokal and Sneath (1963)

Distance Methods

General idea:

- Calculate a measure of the distance between each pair in a group of sequences
- Find a tree that predicts the observed set of distances as closely as possible