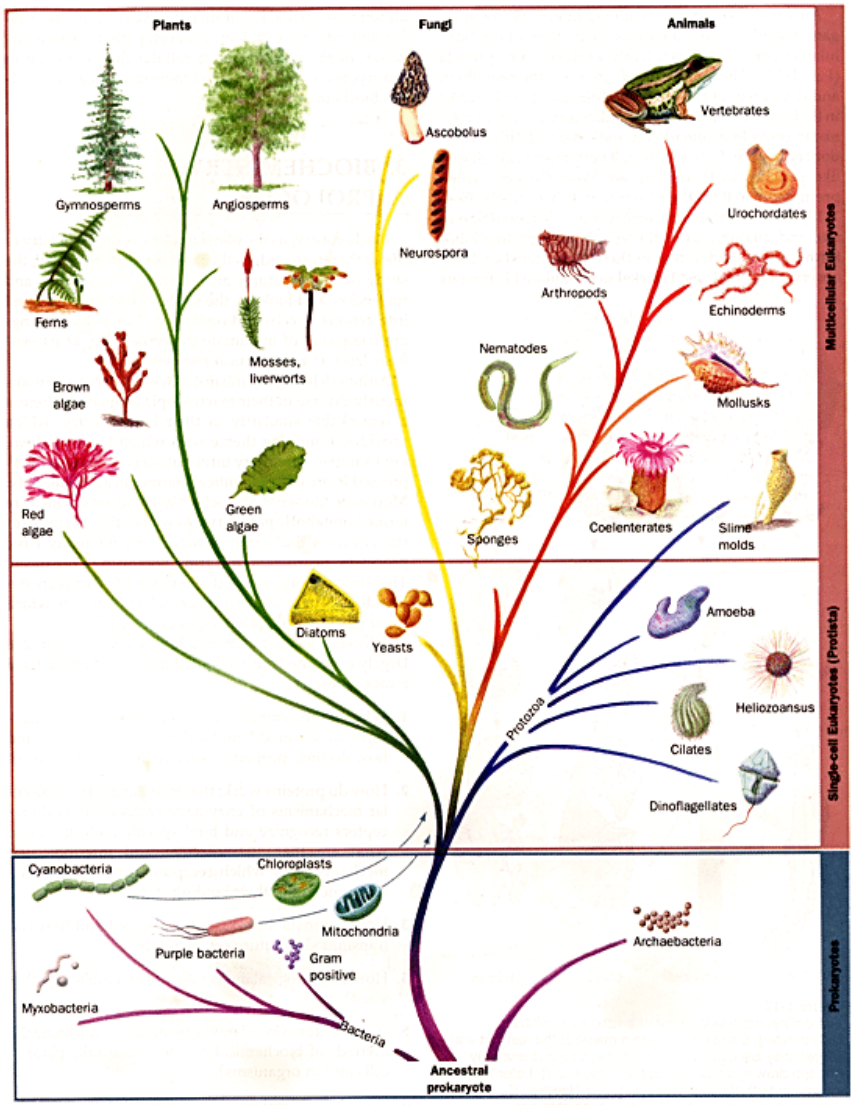


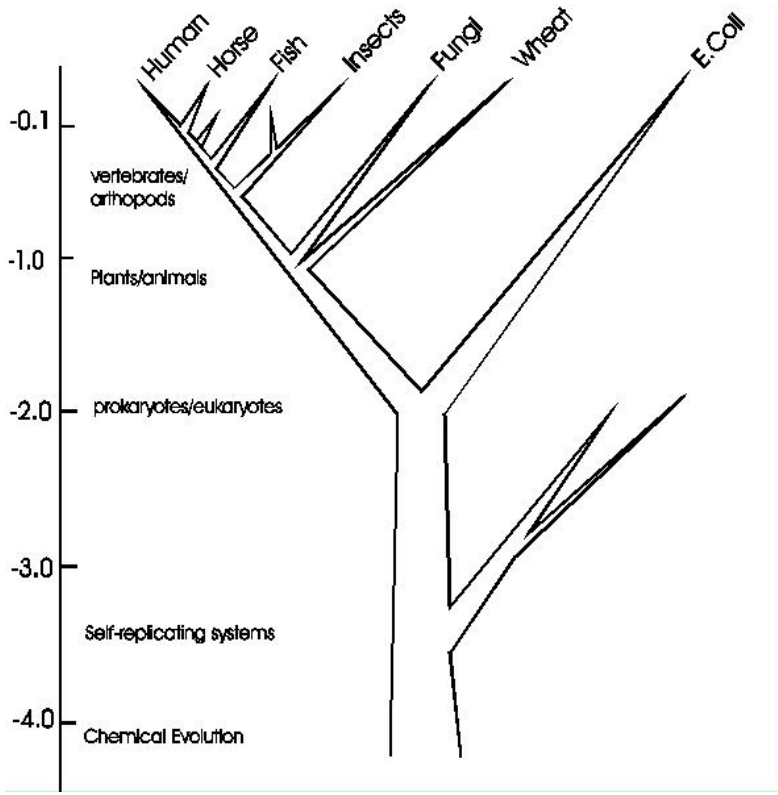
Introduction to Molecular Phylogenetics

Jack A.M. Leunissen
Lab. of Bioinformatics
Wageningen University
The Netherlands

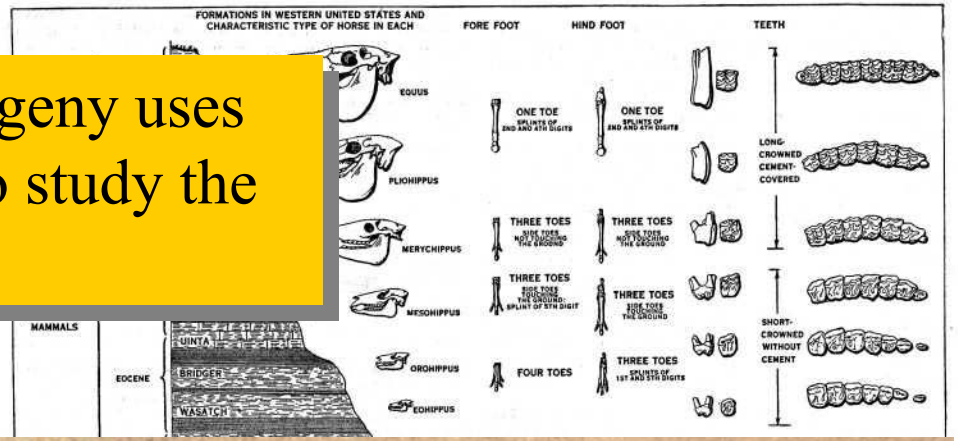
BIOINFORMATICS WAGENINGEN UR

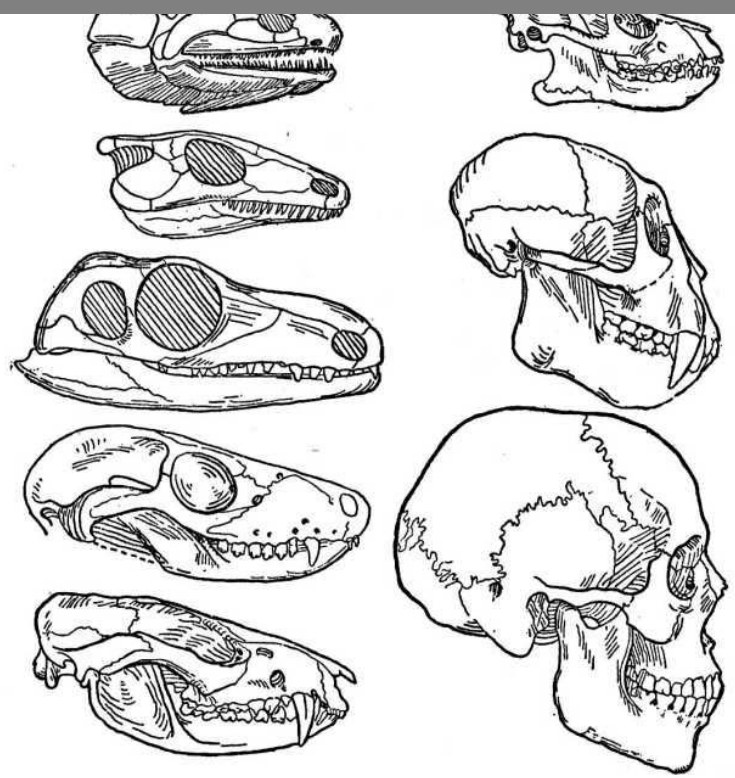
What is Phylogeny?

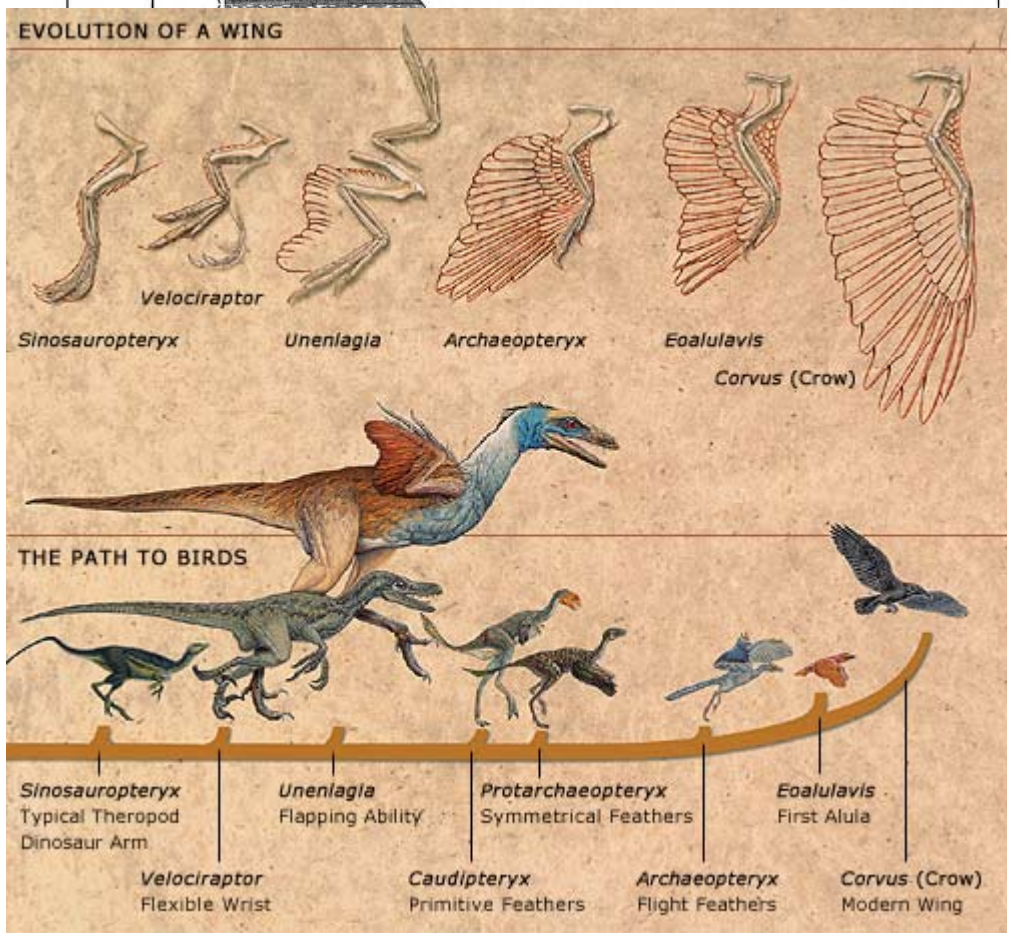
- A description how genes, proteins or species are related within families
- It assumes the objects under investigation are related through *evolution*



The classical approach to phylogeny uses *morphological* characteristics to study the relationship between species





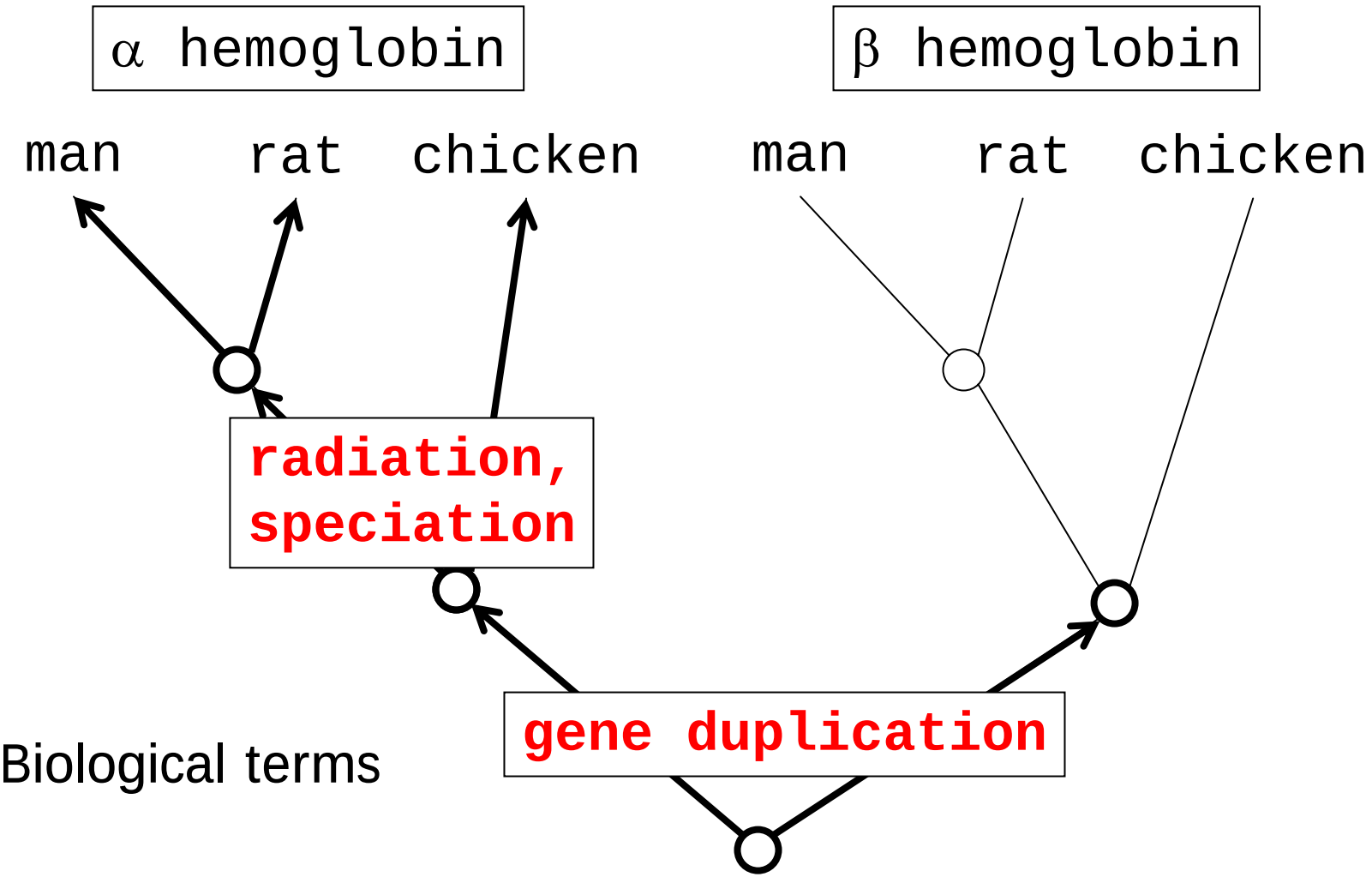


BIOINFORMATICS WAGENINGEN UR

What is Molecular Phylogeny?

- The use of molecular data of establish the relationship between species, organisms or gene families
- Data can be a variety of characteristics, such as:
  - Protein sequences
  - DNA hybridisation
  - Gene frequencies
  - Codon usage
  - DNA sequences
  - Immunological data
  - Restriction patterns
  - Gaps in sequences

Technical Terms

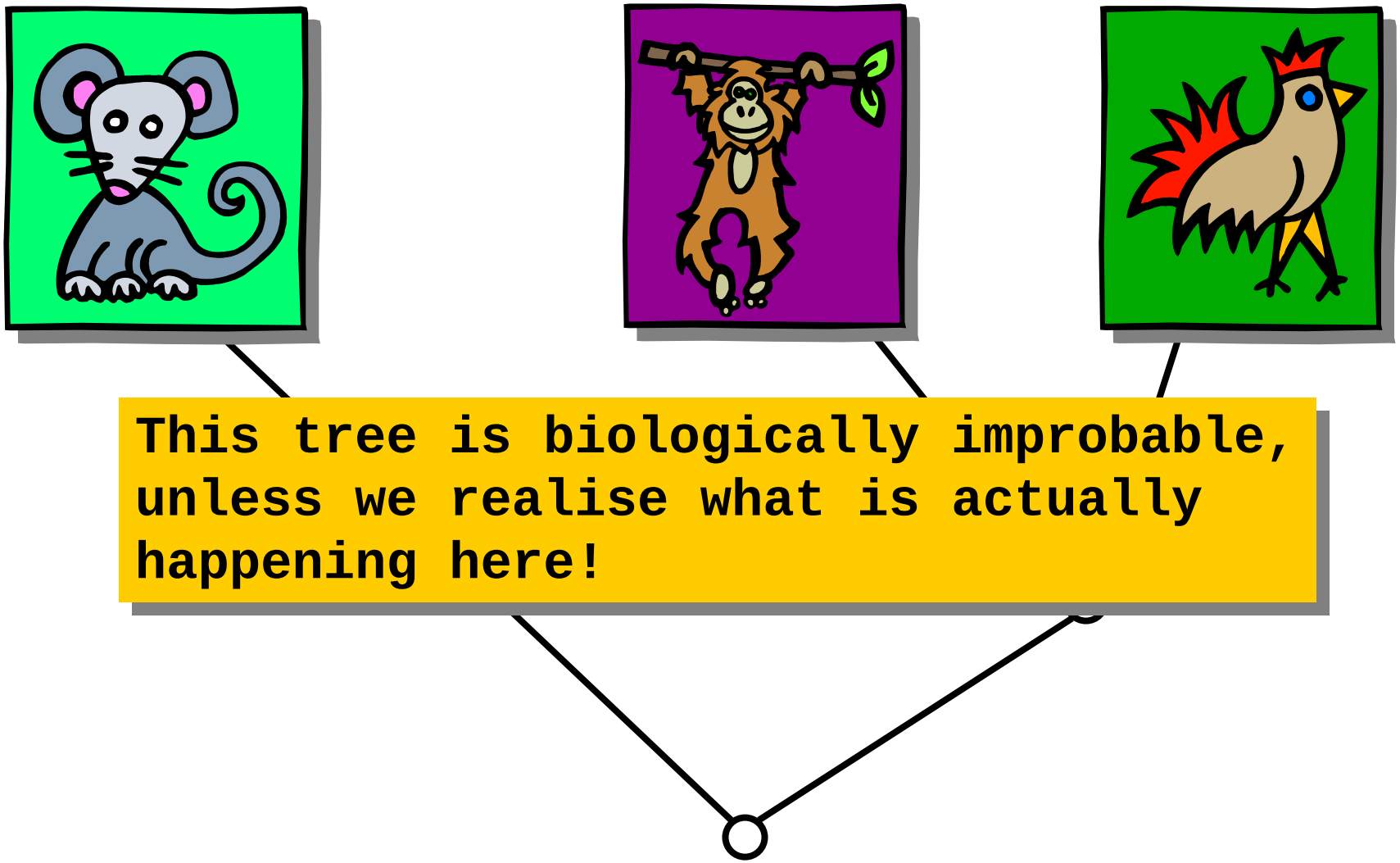


Biological terms

Orthology vs Paralogy

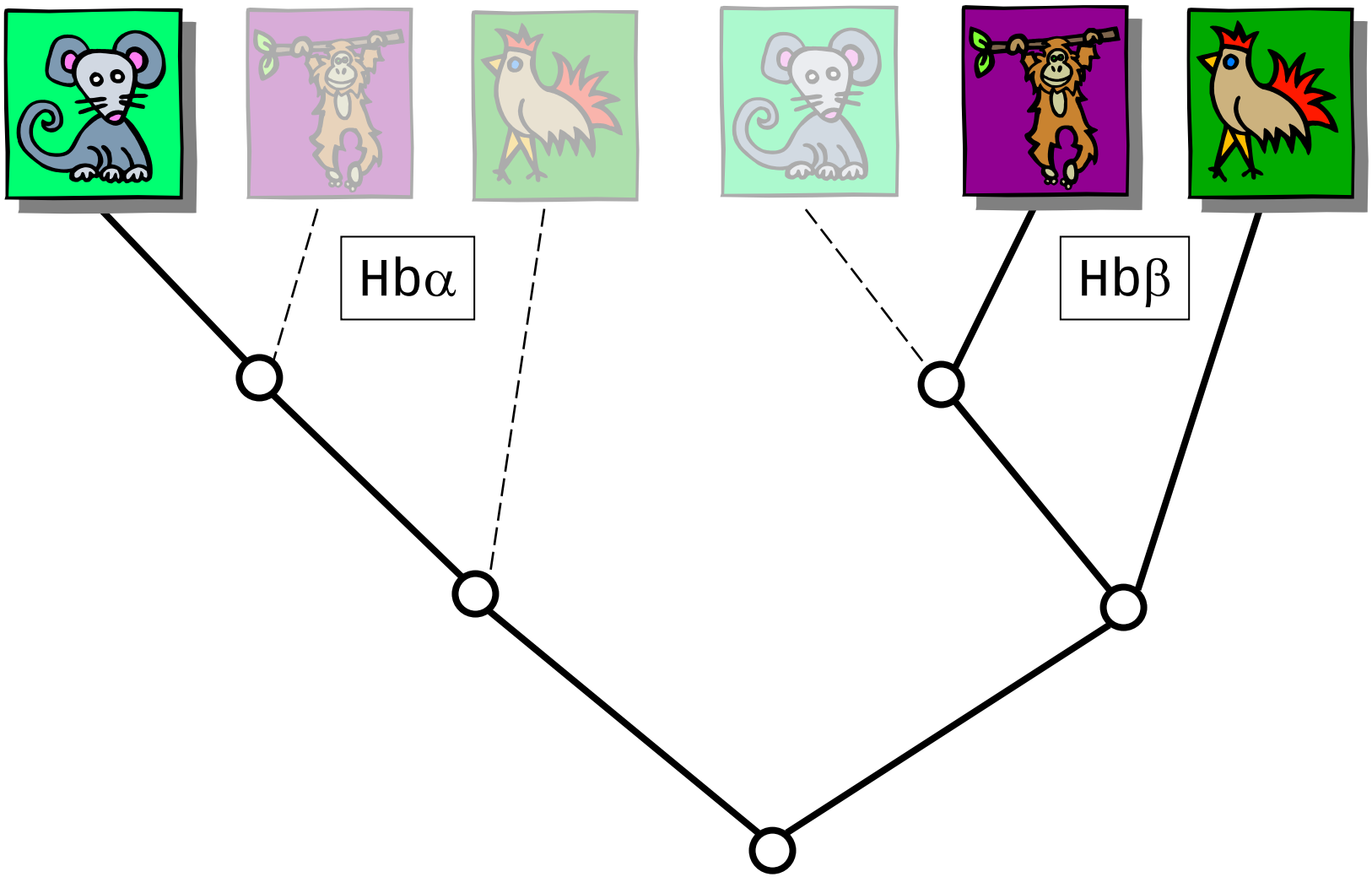
- Orthologs: the sequences have diverged by speciations
  - E.g. human, mouse and chicken α hemoglobin
- Paralogs: the sequences have diverged by gene duplication
  - E.g. the α and β hemoglobin genes
- Xenologs: genes acquired by horizontal gene transfer

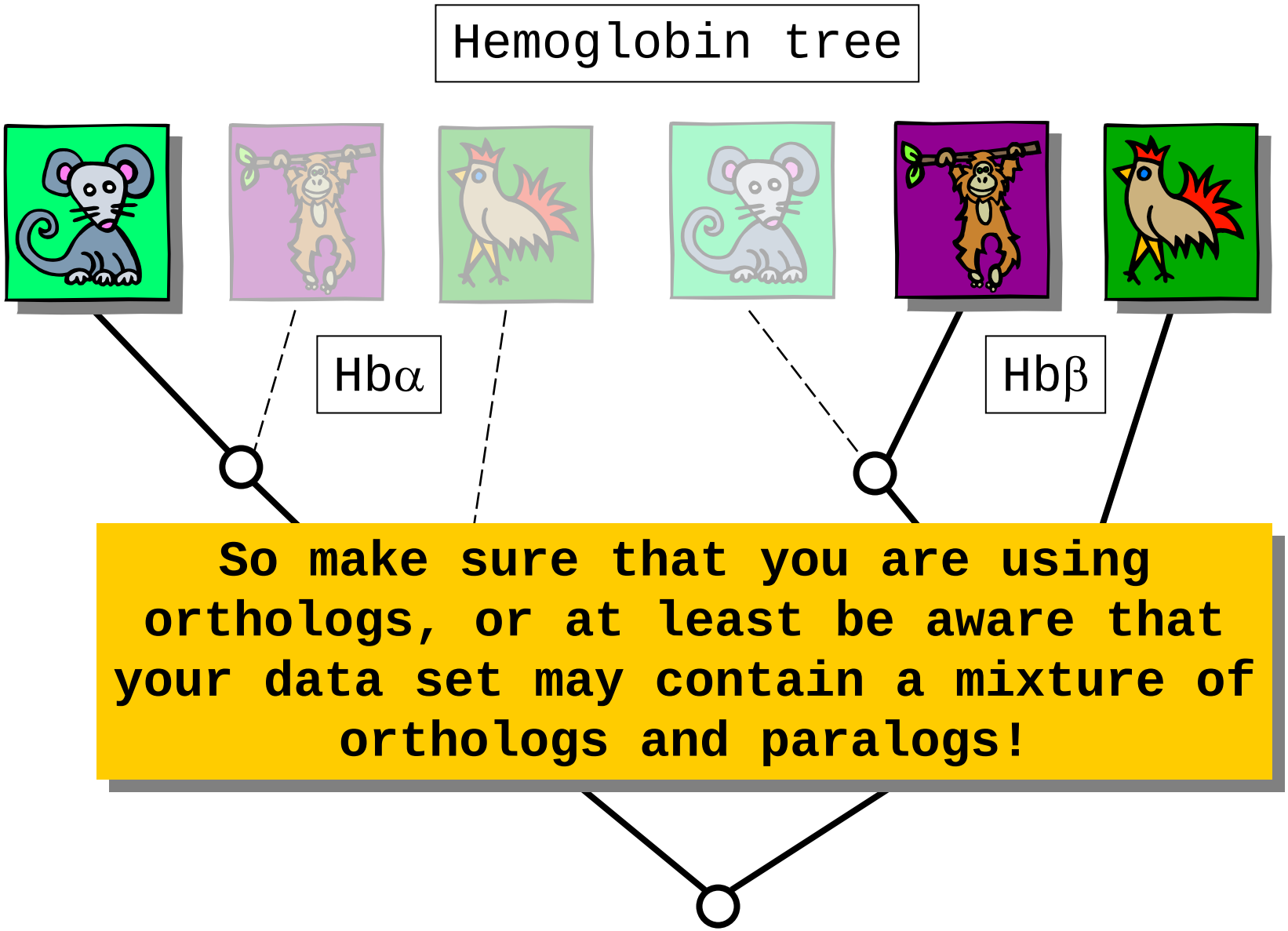
Hemoglobin tree

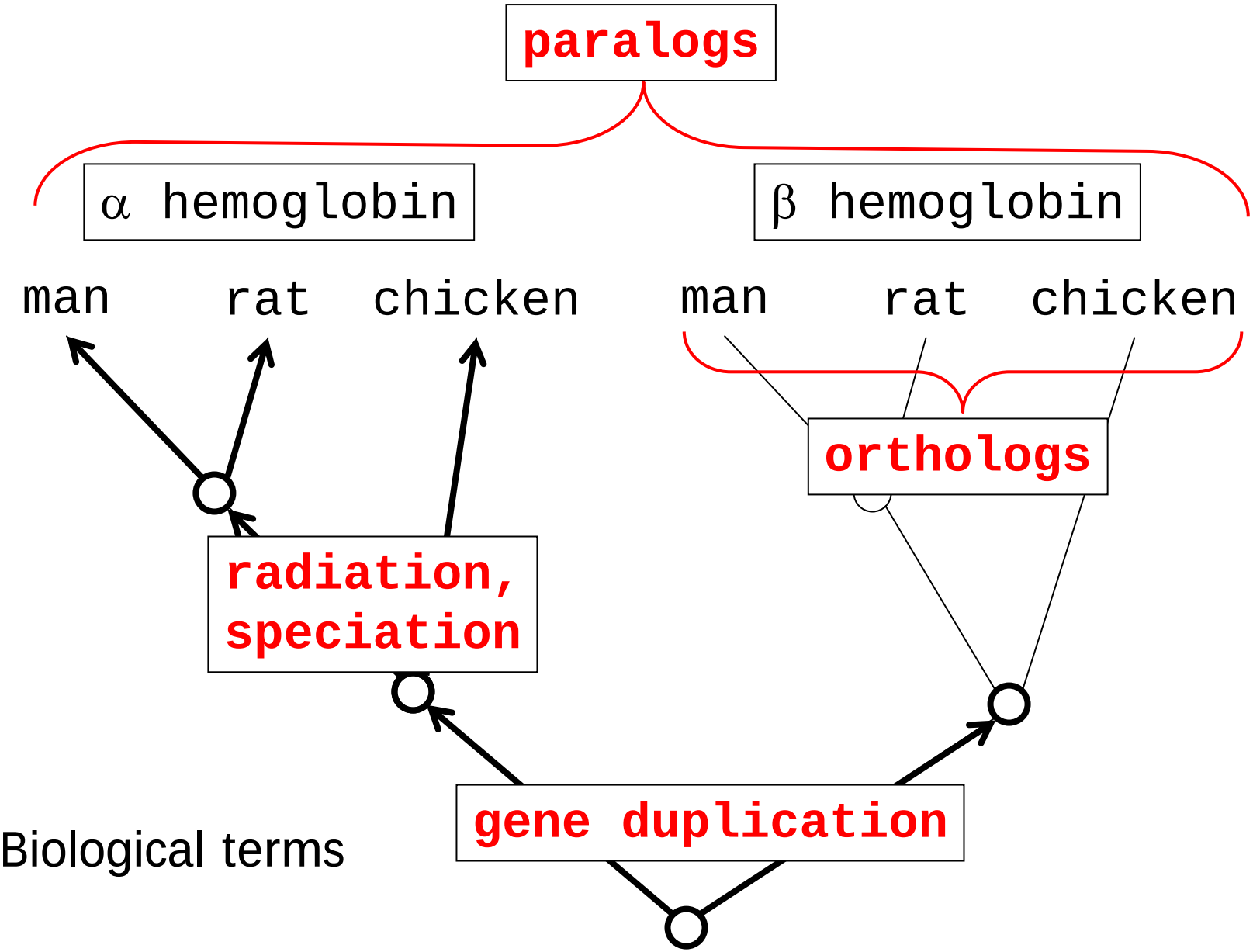


This tree is biologically improbable, unless we realise what is actually happening here!

Hemoglobin tree

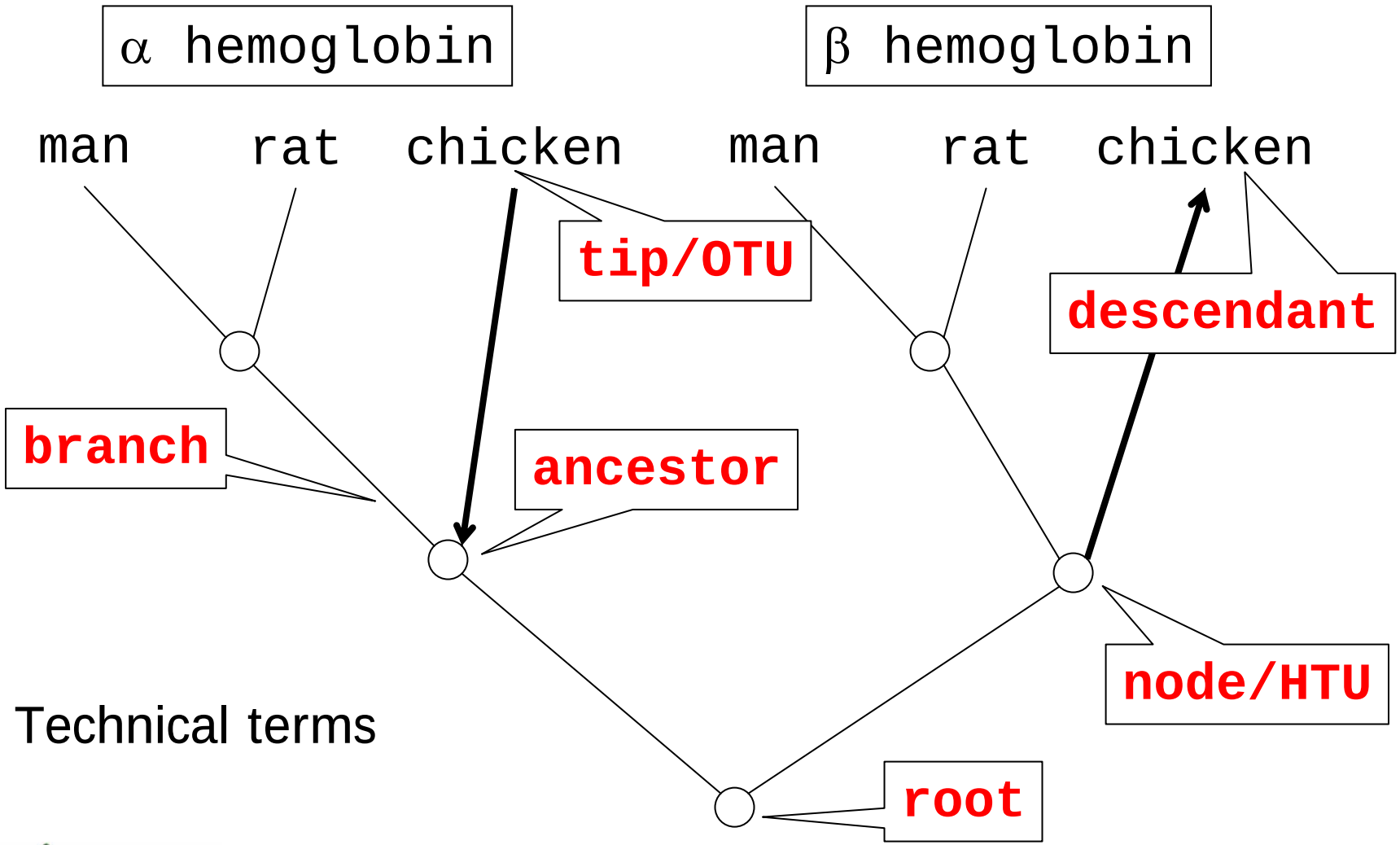



Hemoglobin tree
Hbα
Hbβ
So make sure that you are using orthologs, or at least be aware that your data set may contain a mixture of orthologs and paralogs!

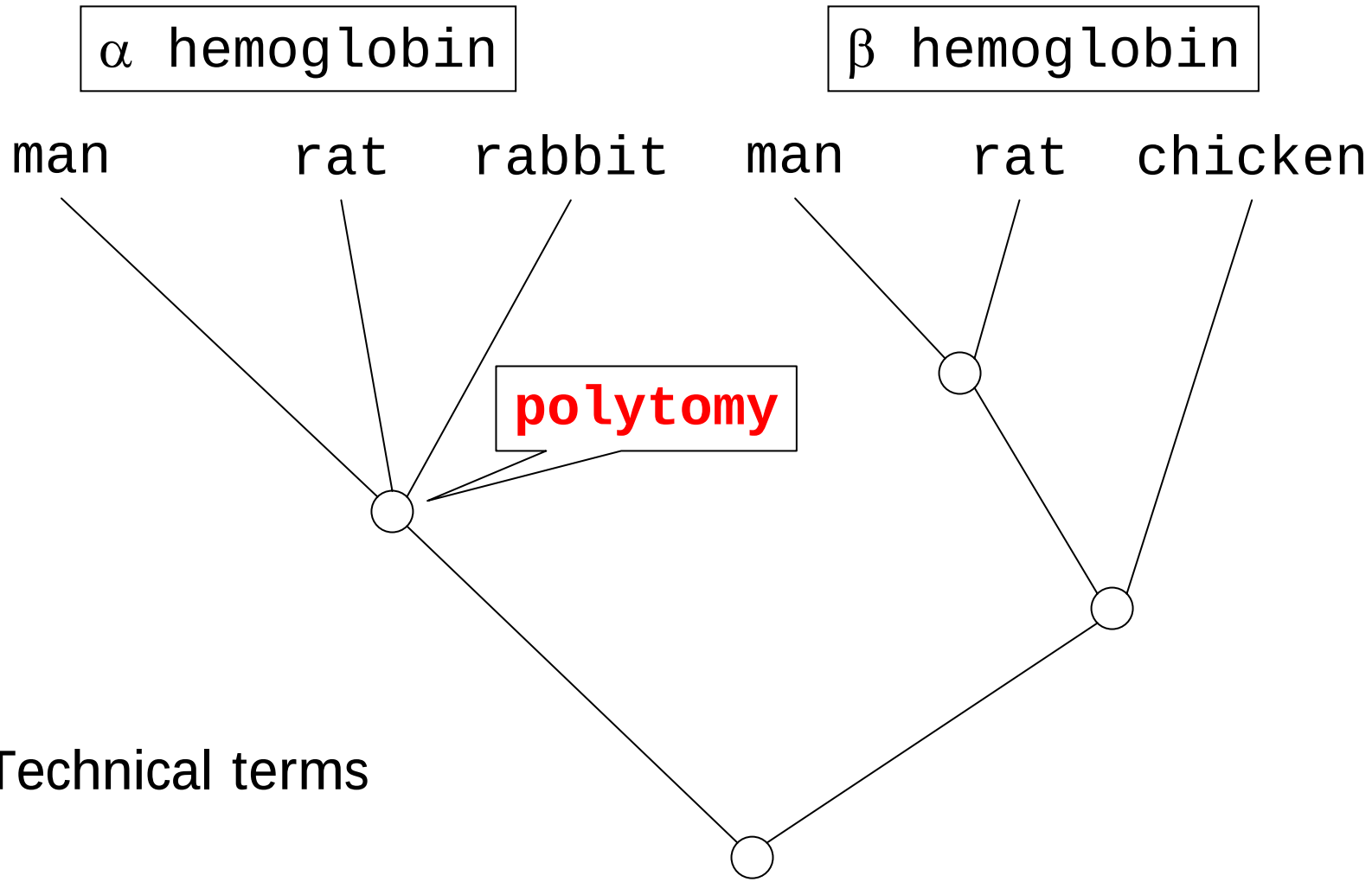


Biological terms

OTU = Operational Taxonomic Unit
HTU = Hypothetical Taxonomic Unit



Technical terms



Technical terms

Rooted versus Unrooted

- Rooted:
  - The tree reflects the branching pattern (*order* of evolutionary events) starting from the oldest common ancestor

- Unrooted:
  - The tree only shows the evolutionary *relationships* between the descendants

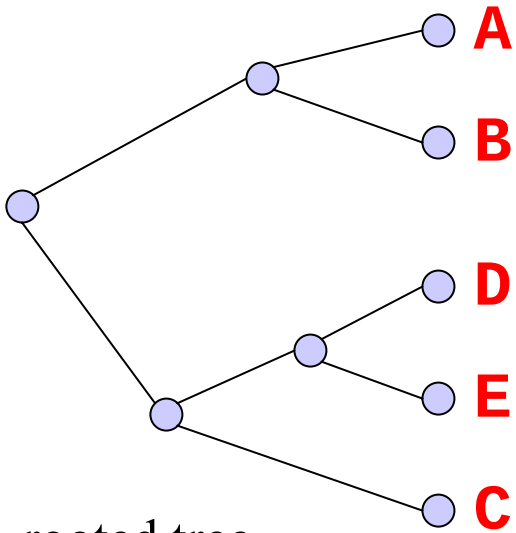
Rooted versus Unrooted

- You can only reconstruct a rooted tree if there is a constant evolutionary clock present, i.e. if there is little variation between the rates of change in the branches
- This may not necessarily be true

- Most algorithms construct unrooted trees

Tree Representations

- The Newick (PHYLIP) bracket notation:



rooted tree

Grouping information

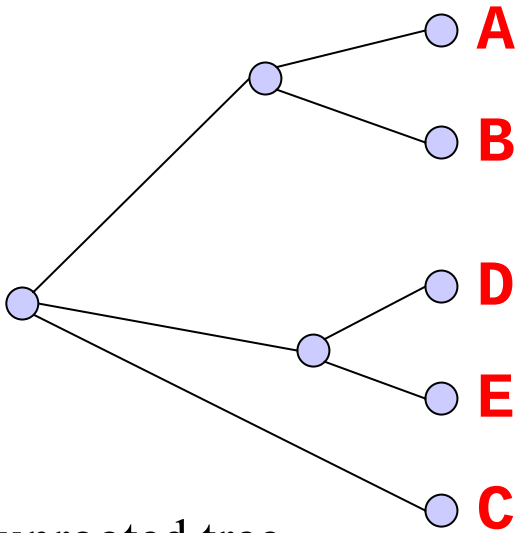
((A , B) , (C , (D , E))) ;

Grouping, including branch lengths

((A:1.1 , B:1.3):0.5 , C:1.4 , (D:0.8 , E:0.7):0.4) ;

Tree Representations

- The Newick (PHYLIP) bracket notation:



unrooted tree

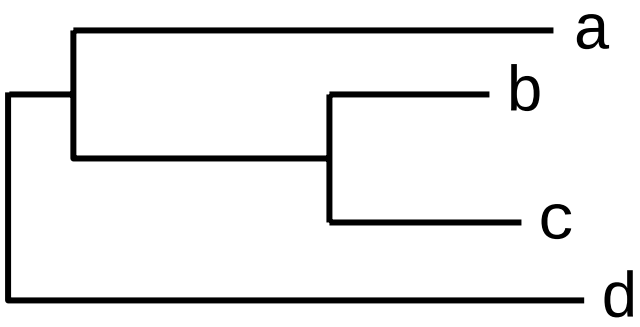
Grouping information

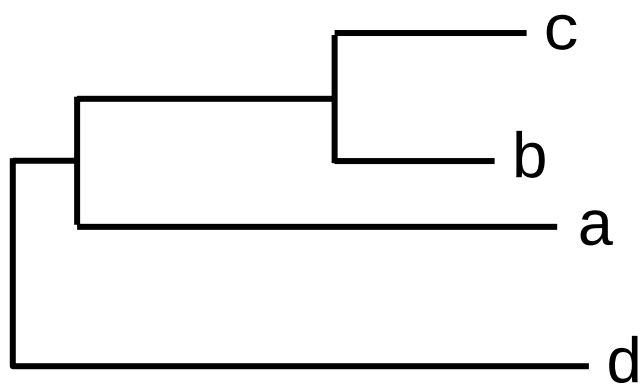
((A , B) , (C , (D , E))) ;

Grouping, including branch lengths

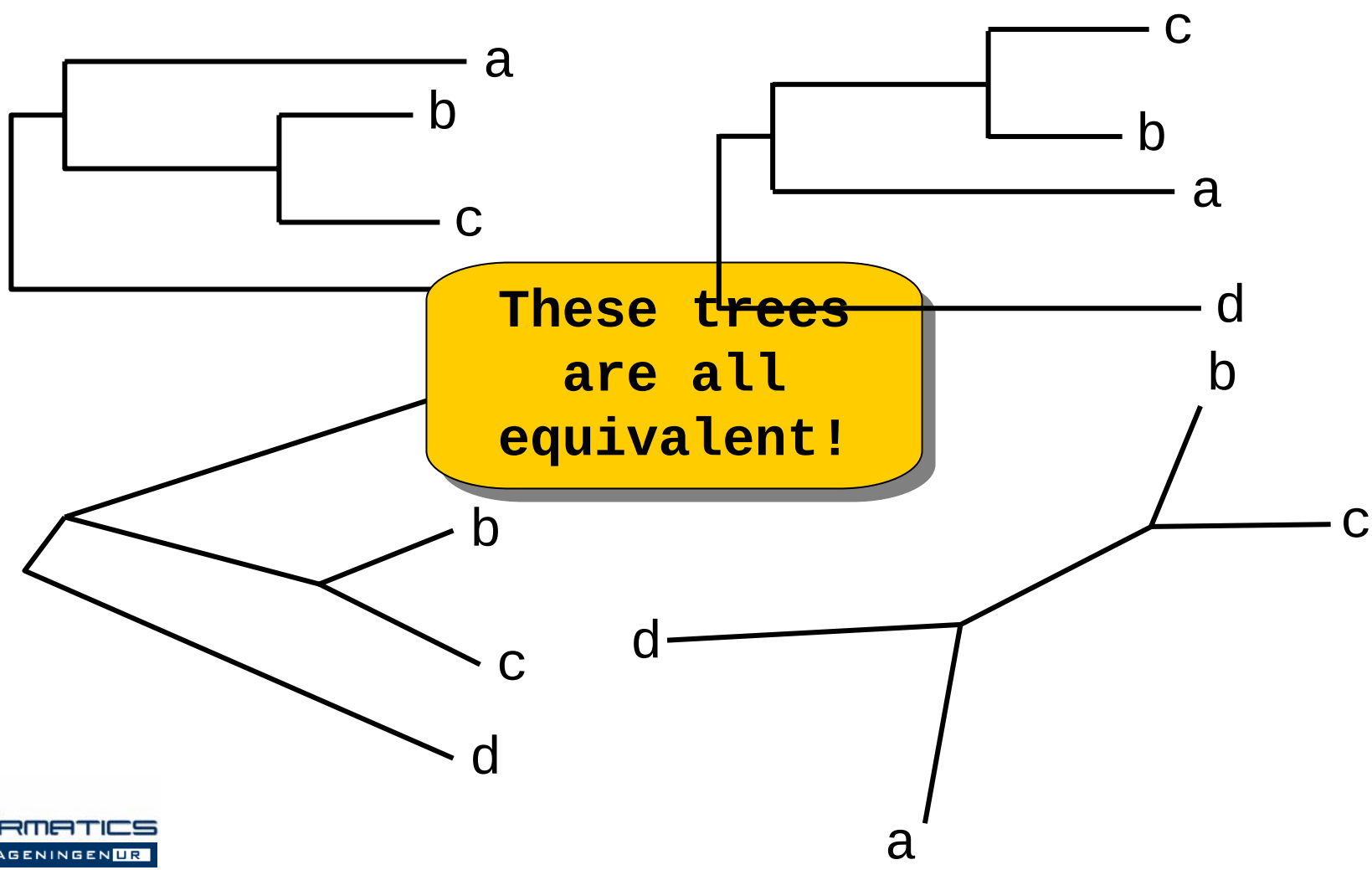
((A:1.1 , B:1.3):0.5 , C:1.4 , (D:0.8 , E:0.7):0.4) ;

Tree Representations





Tree Representations



Why Molecular Phylogeny

BIOINFORMATICS WAGENINGEN UR

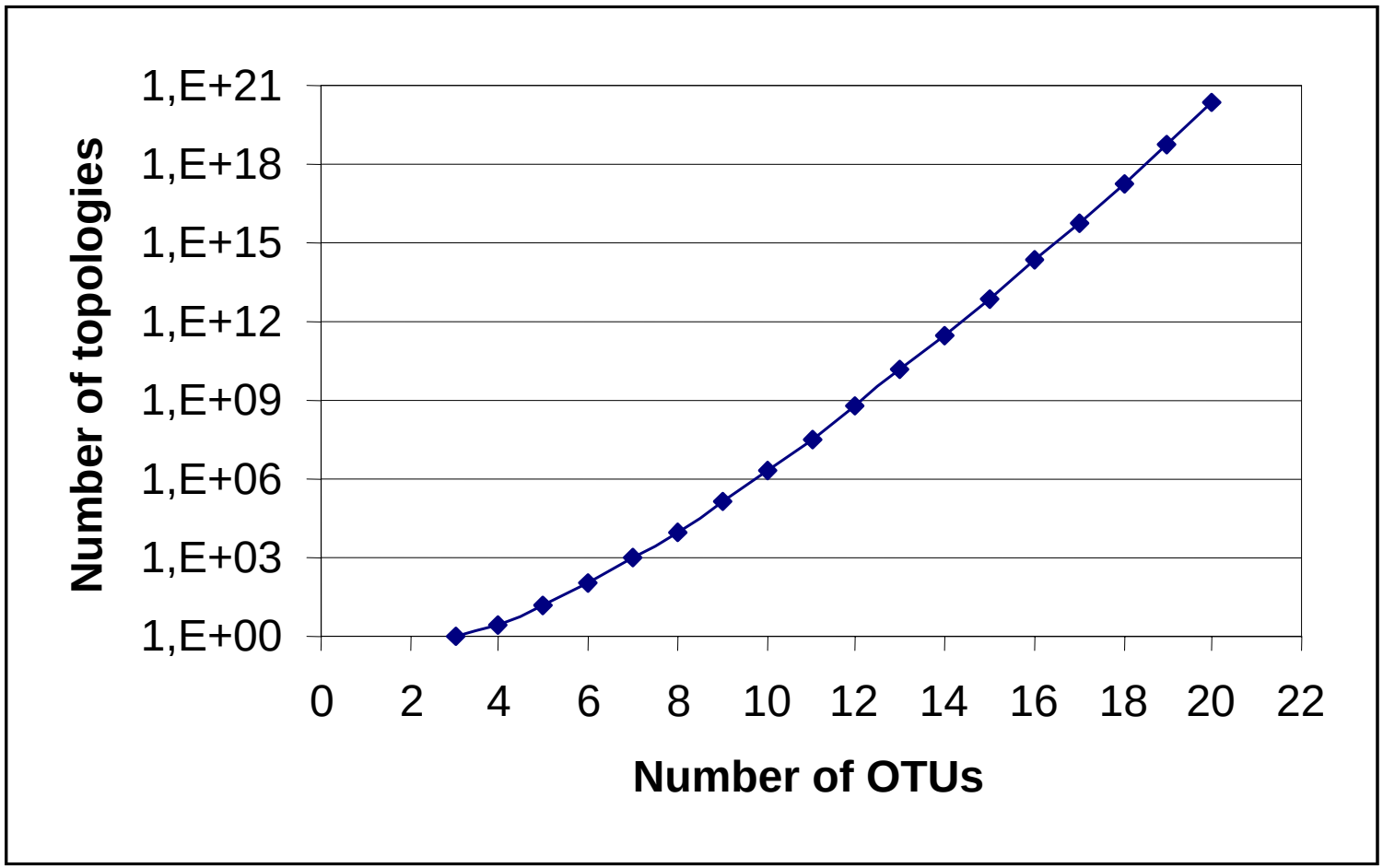
The Goal

- Find the true evolutionary tree of
  - species evolution
  - gene evolution
- This information can be used to
  - predict the function of genes and proteins
  - design or alter protein function
  - study the origin of life on earth

The Problem(s)

- Trying to fit our data to every possible tree topology rapidly becomes infeasible
  - This type of problem is called intractable or NP-complete
- The more realistic the evolutionary model is, the more complex it is
  - For complex models it is difficult to generate realistic estimates of model parameters

Number of Possible Trees



BIOINFORMATICS WAGENINGEN UR

# Taxa	Factor	# Unrooted trees
3	1	1
4	3	3
5	5	15
6	7	105
7	9	945
8	11	10,395
9	13	135,135
10	15	2,027,025
11	17	34,459,425
12	19	654,729,075
13	21	13,749,310,575
14	23	316,234,143,225
15	25	7,905,853,580,625
16	27	213,458,046,676,875
17	29	6,190,283,353,629,375
18	31	191,898,783,962,510,625
19	33	6,332,659,870,762,850,625
20	35	221,643,095,476,699,771,875

$$B(T) = \prod_{i=3}^{T} (2i - 5)$$

The Solution(s)

- Only calculate every possible topology for a small number of OTUs
- Use approximations for larger data sets
  - Stepwise addition
  - Branch swapping (also: BNNI)
  - Quartet puzzling
  - Star decomposition
  - Branch-and-bound

Tree Building Methods

- Distances
- Parsimony
- Maximum likelihood (ML)
- Bayesian (MCMC)

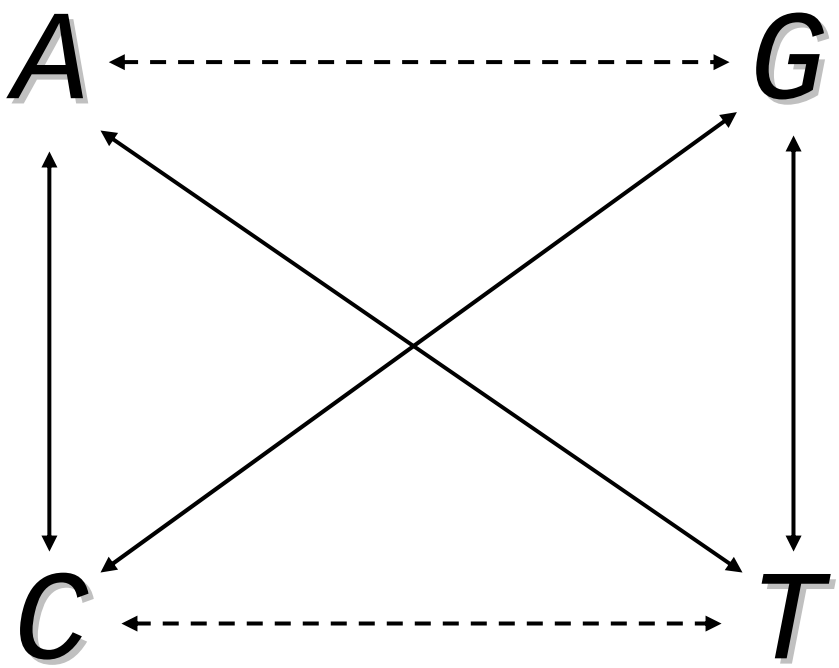
Distance Methods

- Find the topology that fits a set of pairwise distances best
- Distances can be derived from:
  - Protein sequences
  - Gene frequencies
  - DNA hybridisation
  - Codon usage
  - DNA sequences
  - Restriction patterns
  - Immunological data

Distance Methods

- Algorithm:
  - 1 - Calculate a matrix of pairwise dissimilarity values between sequences
  - 2 - Find the closest pair, and cluster them
  - 3 - Calculate distance of pair to all other entries in the matrix
  - 4 - Repeat steps 2 and 3 until done
- There may be an input-order dependency!

Nucleotide Substitution Models



Transitions

Transversions

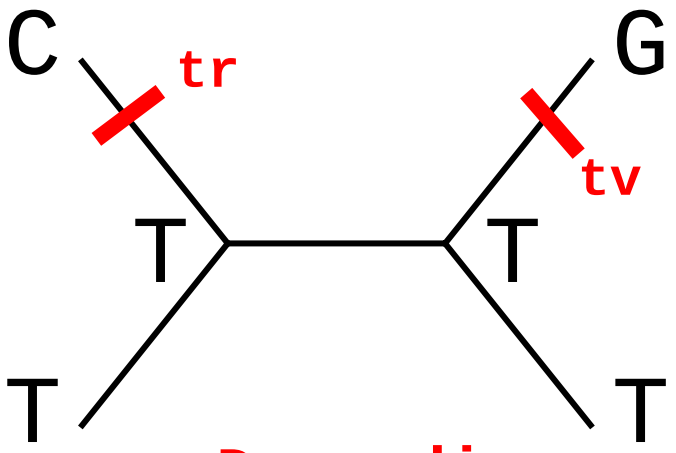
In many cases: Tr/Tv > 1

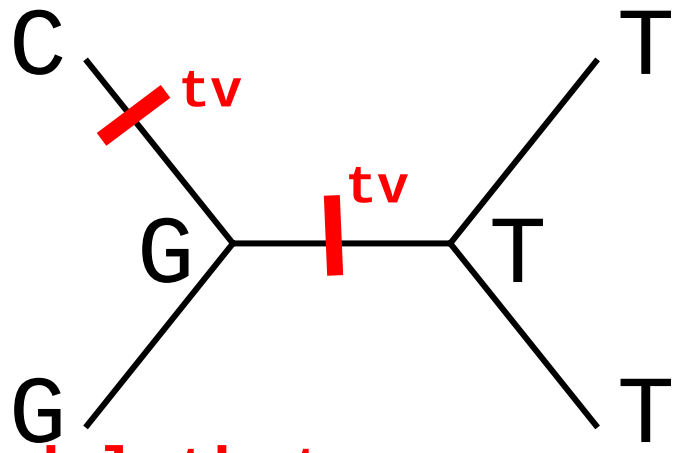
General Substitution Model

	A	**T**	**C**	**G**
A	$1-\alpha_{12}-\alpha_{13}-\alpha_{14}$	α_{12}	α_{13}	α_{14}
T	α_{21}	$1-\alpha_{21}-\alpha_{23}-\alpha_{24}$	α_{23}	α_{24}
C	α_{31}	α_{32}	$1-\alpha_{31}-\alpha_{32}-\alpha_{34}$	α_{34}
G	α_{41}	α_{42}	α_{43}	$1-\alpha_{41}-\alpha_{42}-\alpha_{43}$

Nucleotide Substitution Models

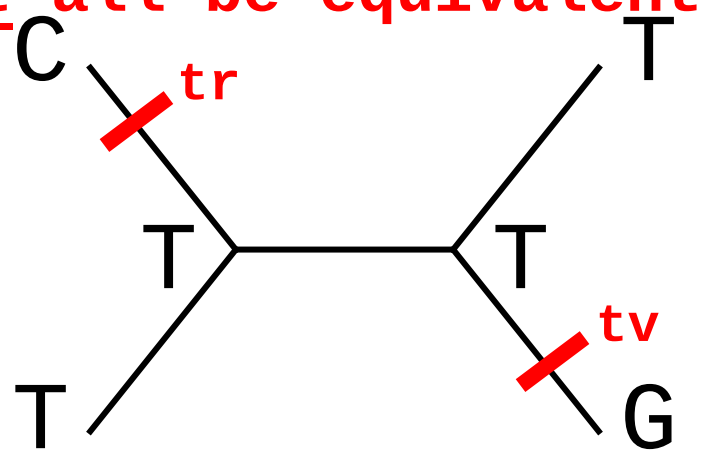
- One-parameter (Jukes &Cantor 1969)
  - no distinction between “tr” and “tv” events
- Two-parameter (Kimura 1980)
  - different values for “tr” and “tv” events
- Three-parameter (Kimura 1981)
- Four-parameter (Tajima & Nei 1984)
- Six-parameter (Hasegawa 1985)





Depending on the model that you choose, these topologies may or may not all be equivalent!

T
T
C
G



BIOINFORMATICS WAGENINGEN UR

PAM250

C	12																		
S	0	2																	
T	-2	1	3																
P	-3	1	0	6															
A	-2	1	1	1	2														
G	-3	1	0	-1	1	5													
N	-4	1	0	-1	0	0	2												
D	-5	0	0	-1	0	1	2	4											
E	-5	0	0	-1	0	0	1	3	4										
Q	-5	-1	-1	0	0	-1	1	2	2	4									
H	-3	-1	-1	0	-1	-2	2	1	1	3	6								
R	-4	0	-1	0	-2	-3	0	-1	-1	1	2	6							
K	-5	0	0	-1	-1	-2	1	0	0	1	0	3	5						
M	-5	-2	-1	-2	-1	-3	-2	-3	-2	-1	-2	0	0	6					
I	-2	-1	0	-2	-1	-3	-2	-2	-2	-2	-2	-2	-2	2	5				
L	-6	-3	-2	-3	-2	-4	-3	-4	-3	-2	-2	-3	-3	4	2	6			
V	-2	-1	0	-1	0	-1	-2	-2	-2	-2	-2	-2	-2	2	4	2	4		
F	-4	-3	-3	-5	-4	-5	-4	-6	-5	-5	-2	-4	-5	0	1	2	-1	9	
Y	0	-3	-3	-5	-3	-5	-2	-4	-4	-4	0	-4	-4	-2	-1	-1	-2	7	10
W	-8	-2	-5	-6	-6	-7	-4	-7	-7	-5	-3	2	-3	-4	-5	-2	-6	0	0	17
	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W

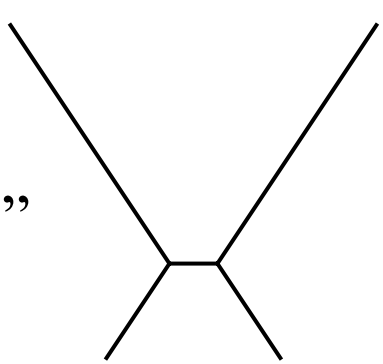
BIOINFORMATICS WAGENINGEN UR

Distance Methods

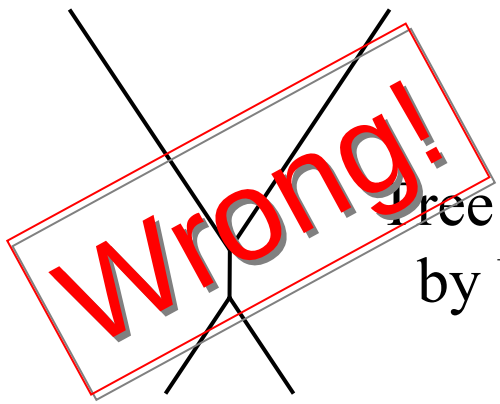
- Commonly used methods
  - UPGMA ⇐ **DO NOT USE THIS METHOD!!!**
  - Fitch (Fitch & Margoliash)
    - uses stepwise addition and branchswapping
  - Neighbor-Joining (Saitou & Nei)
    - uses star-decomposition
  - Principle Co-ordinates Analysis (Higgins)
    - does not give a binary tree, only spatial clustering

Fatal Attraction

- Some algorithms tend to cluster short branches together (most noticeably UPGMA)
- This is known as “long branch attraction”



“True” tree



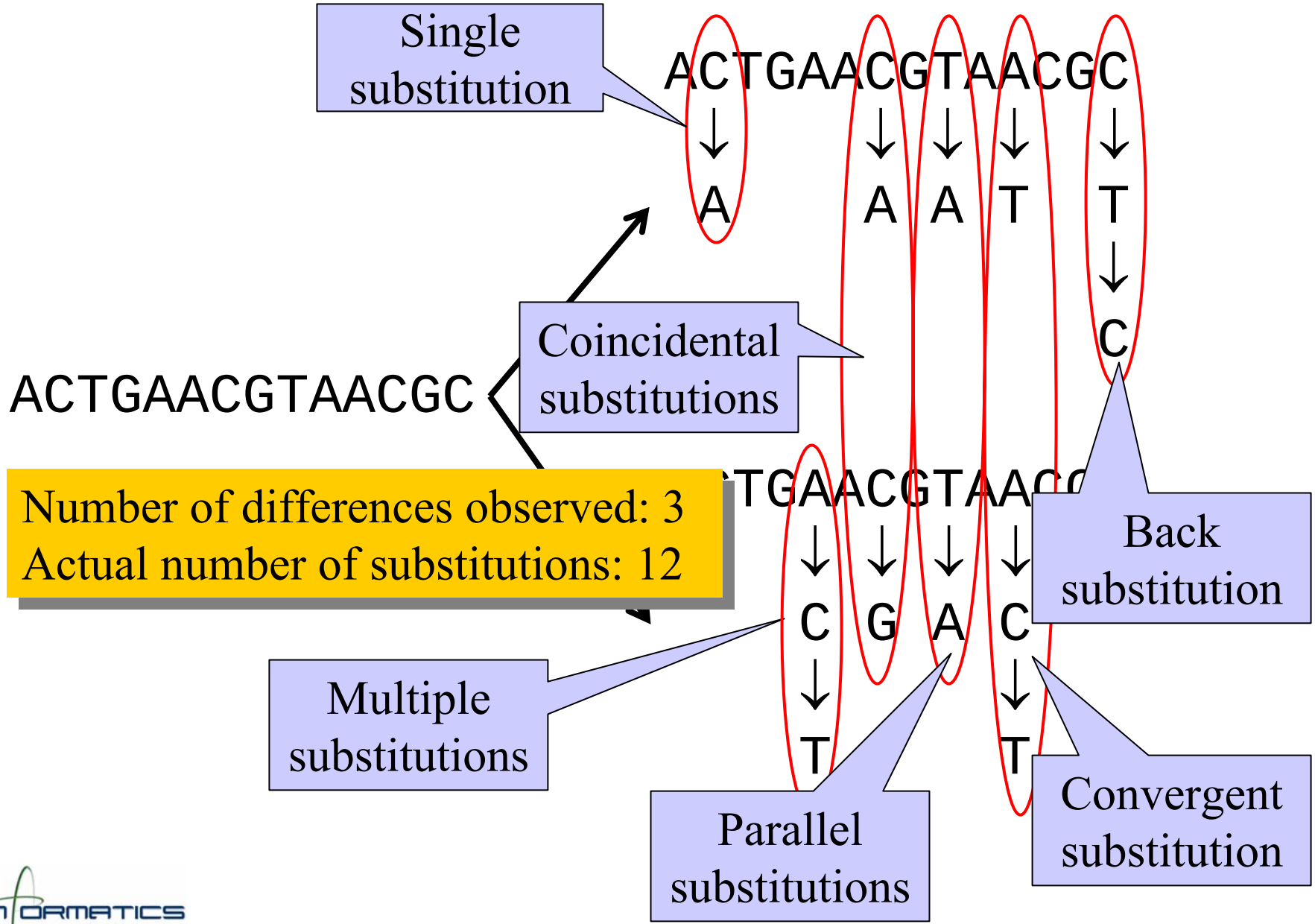
Tree calculated by UPGMA

Fatal Attraction Overcome

- The Neighbor-Joining method does not suffer from long branch attraction, by calculating the “closeness” of each pair of OTUs, corrected for its distance to all the other OTUs.

Fatal Attraction Overcome 2

- From experiments by Hillis (1996) it turns out that long branch attraction is less prone to occur when analysing large phylogenies
  - In these large trees covariation is less probable when the long branches are well enough separated



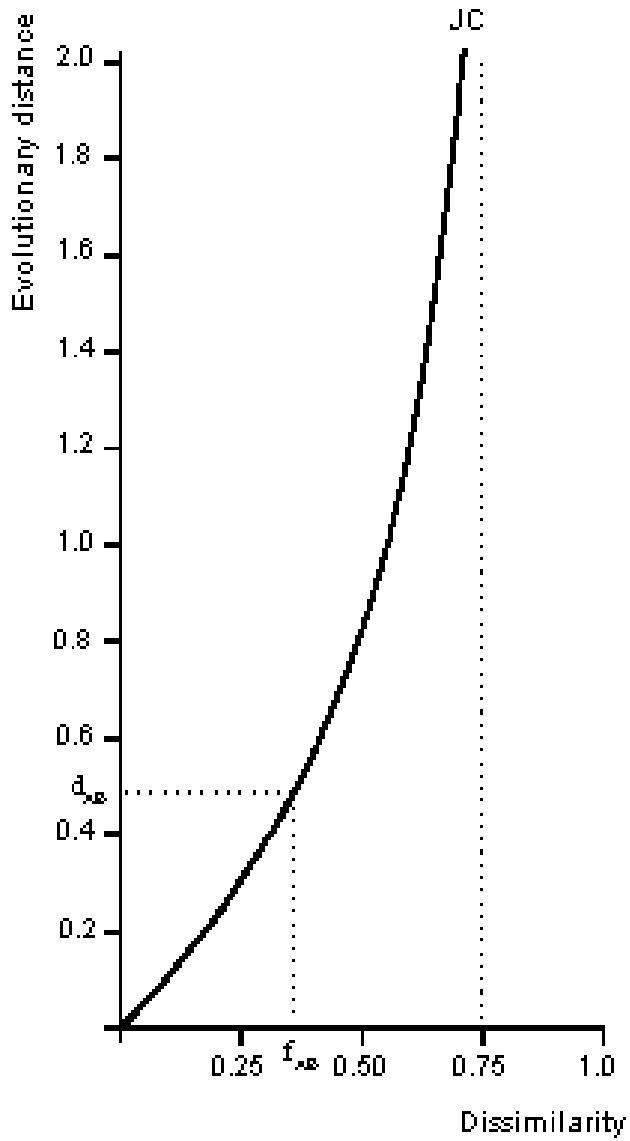
After Li (1997)

Multiple Hits

- Correction for multiple superimposed substitutions (Jukes-Cantor, 1969)

$$D' = -\frac{3}{4}\ln\left(1-\frac{4D}{3}\right)$$

- Method gets saturated at 75% dissimilarity!



Kimura's Method for Proteins

$$d = -\ln\left(1 - p - 0.2\,p^2\right)$$

where p is the proportion of amino acids differing between the two protein sequences

Distance Methods

- Speed: ☺
  - Can handle very large data sets
- Completeness: ☹
  - only tree found

However… Pearson's "Generalized NJ" investigates also sub-optimal trees!

Parsimony

- Uses the sequences directly to construct ancestral sequences
- It applies Ockham's Razor:

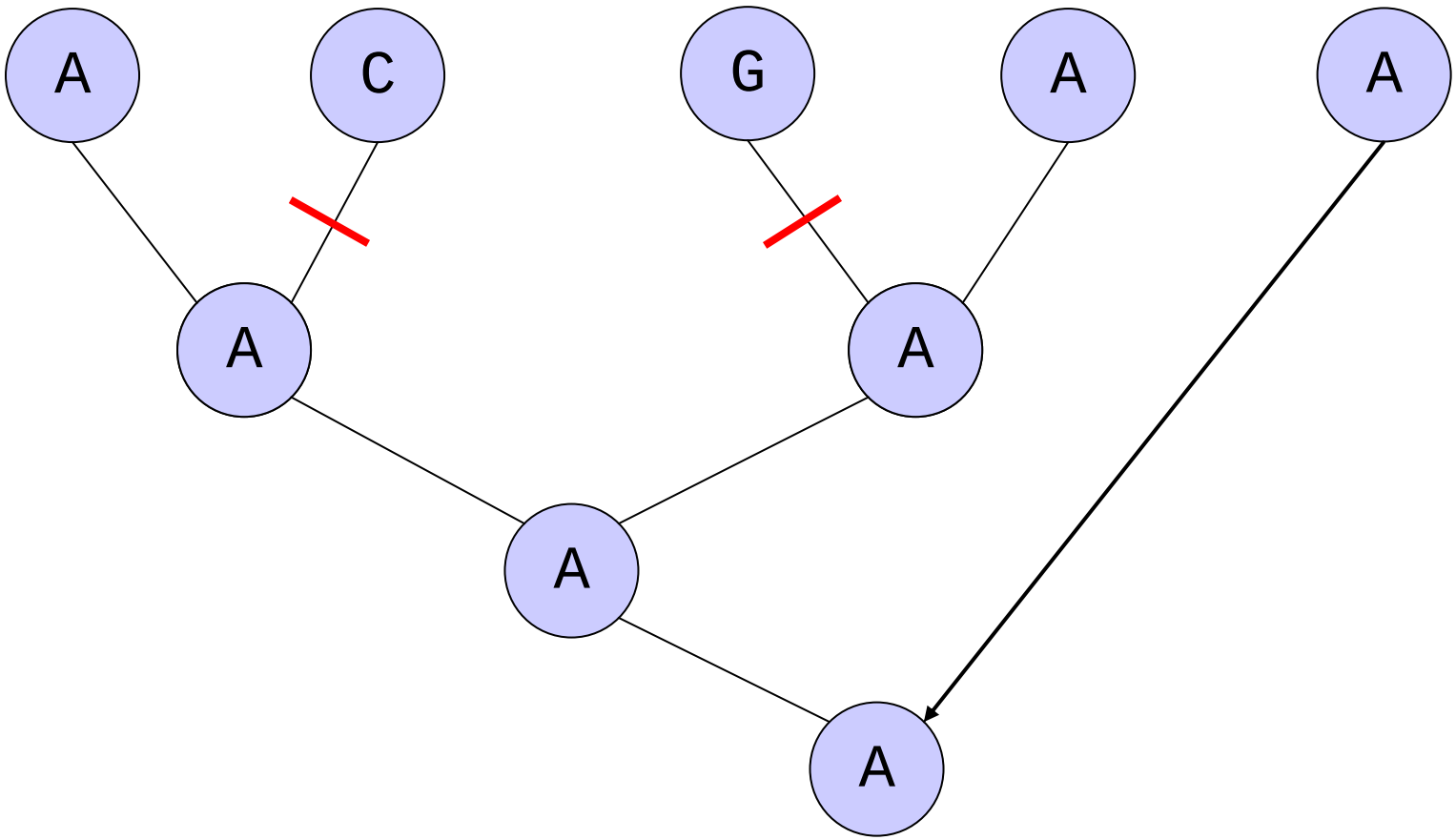
  Essentia non sunt multiplicanda praeter necessitatem

  Don’t use more concepts than necessary

  William of Ockham (1288-1348)

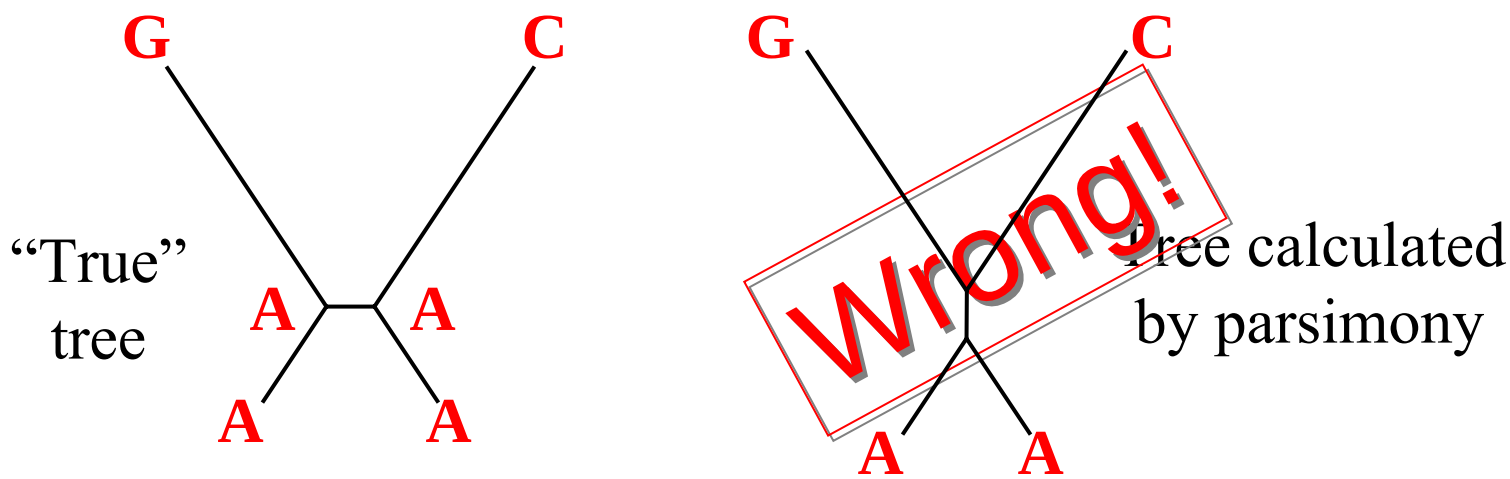
- Accept the tree(s) that has (have) the least number of substitions needed

Parsimony



Oh No, Not Again!

- Also maximum parsimony can suffer from long branch attraction!



Parsimony

- Speed: 😐
  - Can handle moderately large data sets
- Completeness: ☺
  - shows all trees that have equal length

Maximum Likelihood

- Find the tree(s) that has the highest likelihood of fitting a given evolutionary model
- This method allows sites to evolve independently, and thus allows for different rates of change in both lineages and between sites

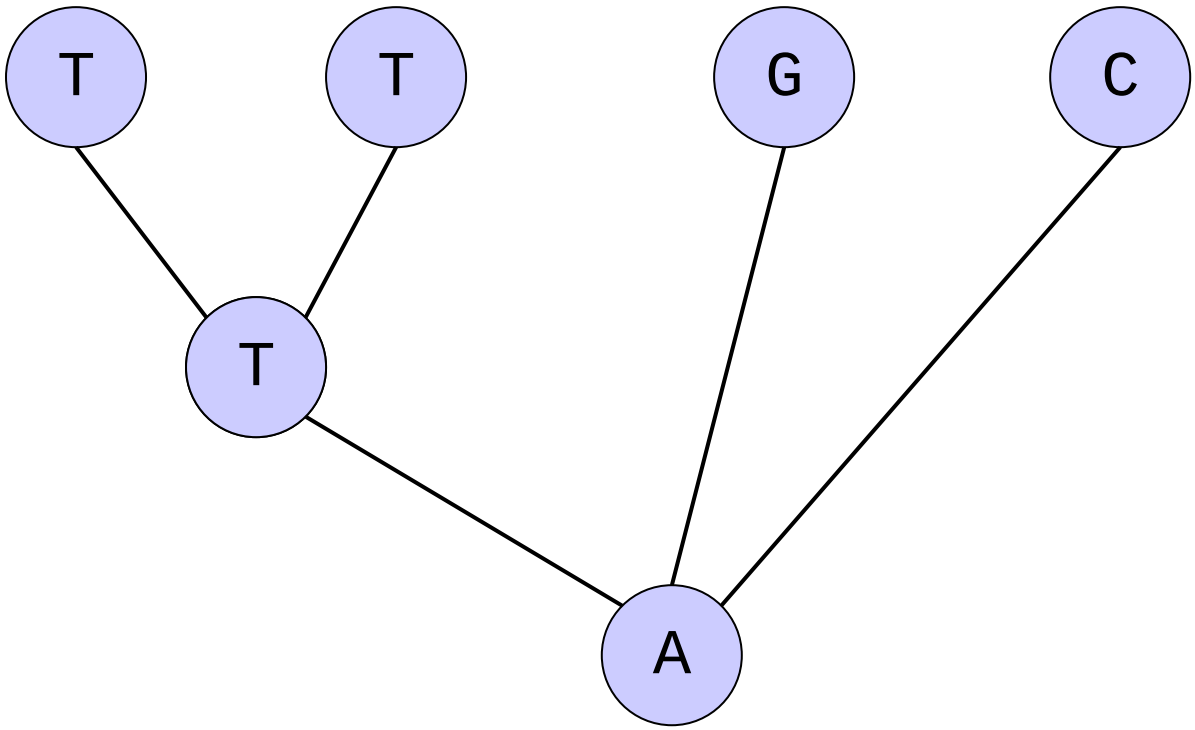
Maximum Likelihood

- The hypothesis that has the greatest likelihood is called the maximum likelihood estimate
- For simple systems a maximum likelihood estimator can be determined analytically
  - e.g. the arithmetic mean is the maximum likelihood estimate under a normal probability distribution

Sequence-based ML

- In ML we use a model in which we calculate the likelihood that the alignment supports a tree topology
- We have to provide the probabilities that one nucleotide (or amino acid) can be replaced by another

Sequence-based ML



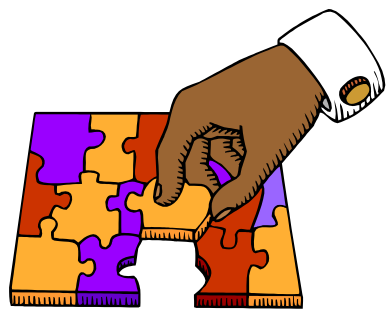
$$P_{tree} = P_A \times P_{AG} \times P_{AC} \times P_{AT} \times P_{TT} \times P_{TT}$$

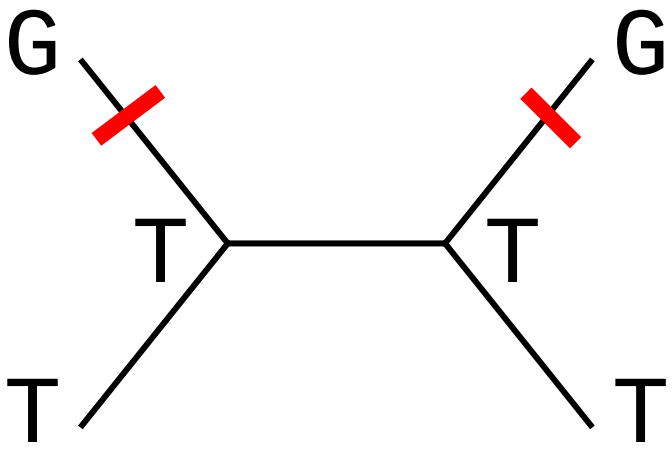
Maximum Likelihood Methods

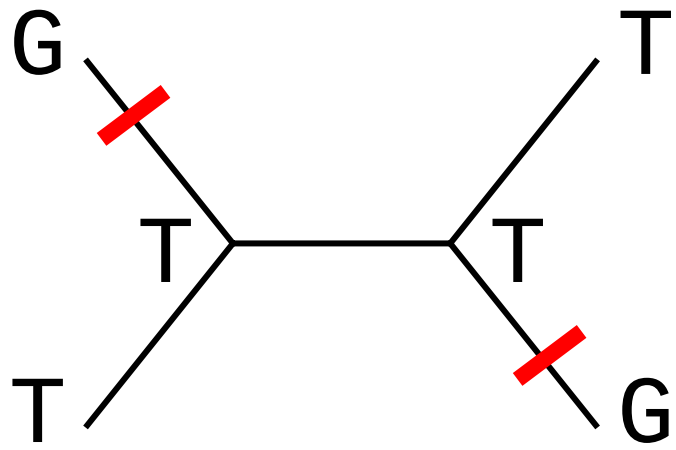
- Speed: ☹
  - Can handle only very small data sets
- Completeness: ☺
  - all trees found

Quartet Puzzling

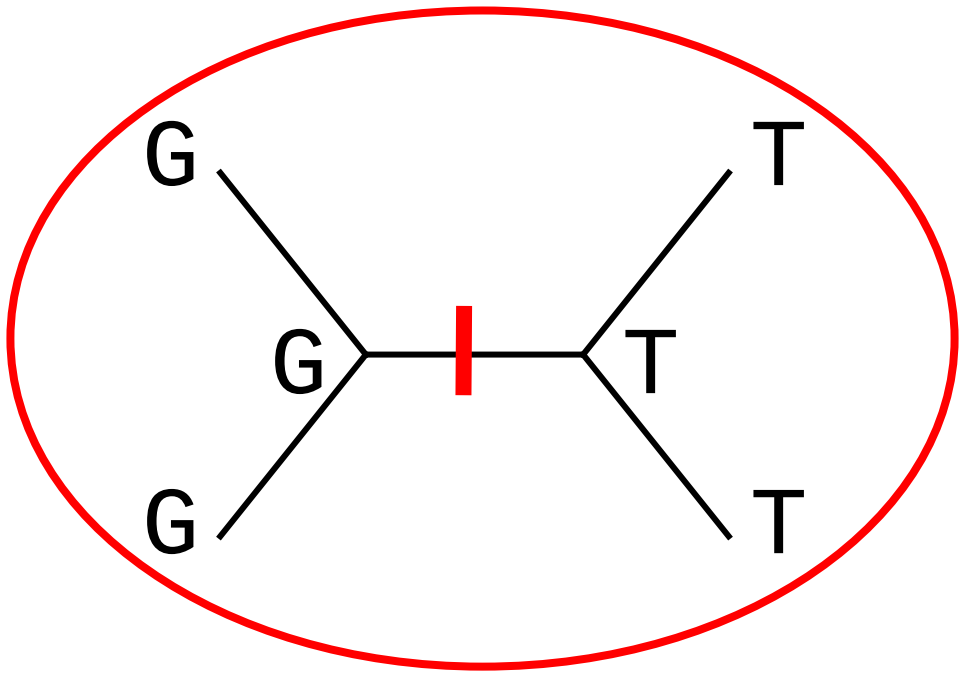
- Calculate the maximum likelihood tree for all quartet trees that can be constructed from a given set of sequences
- If there is a tie, than randomly choose one
- Reassemble all quartets found into the full species tree
  - Strimmer & von Häseler





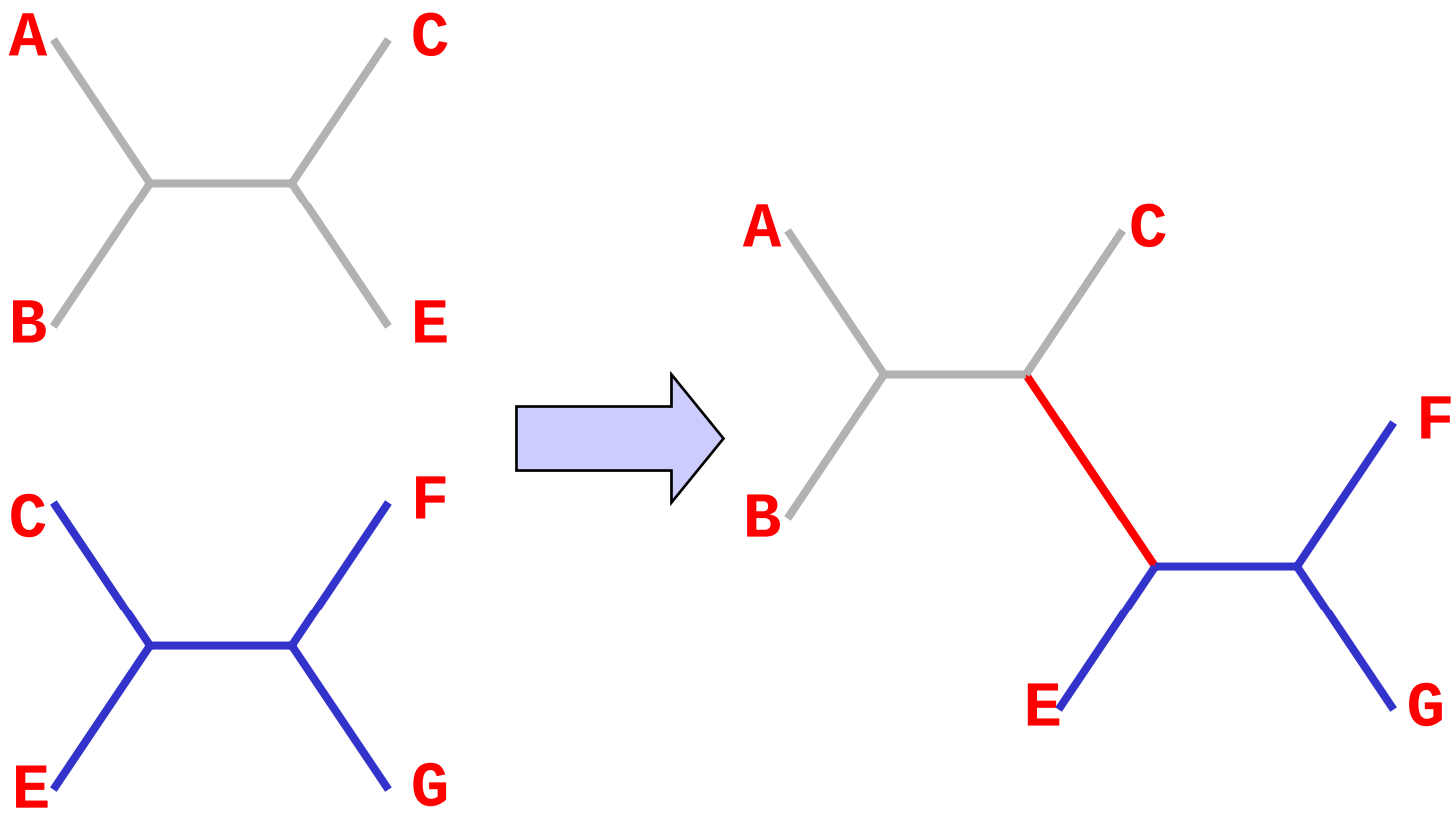


T
T
G
G



BIOINFORMATICS WAGENINGEN UR

Quartet Puzzling

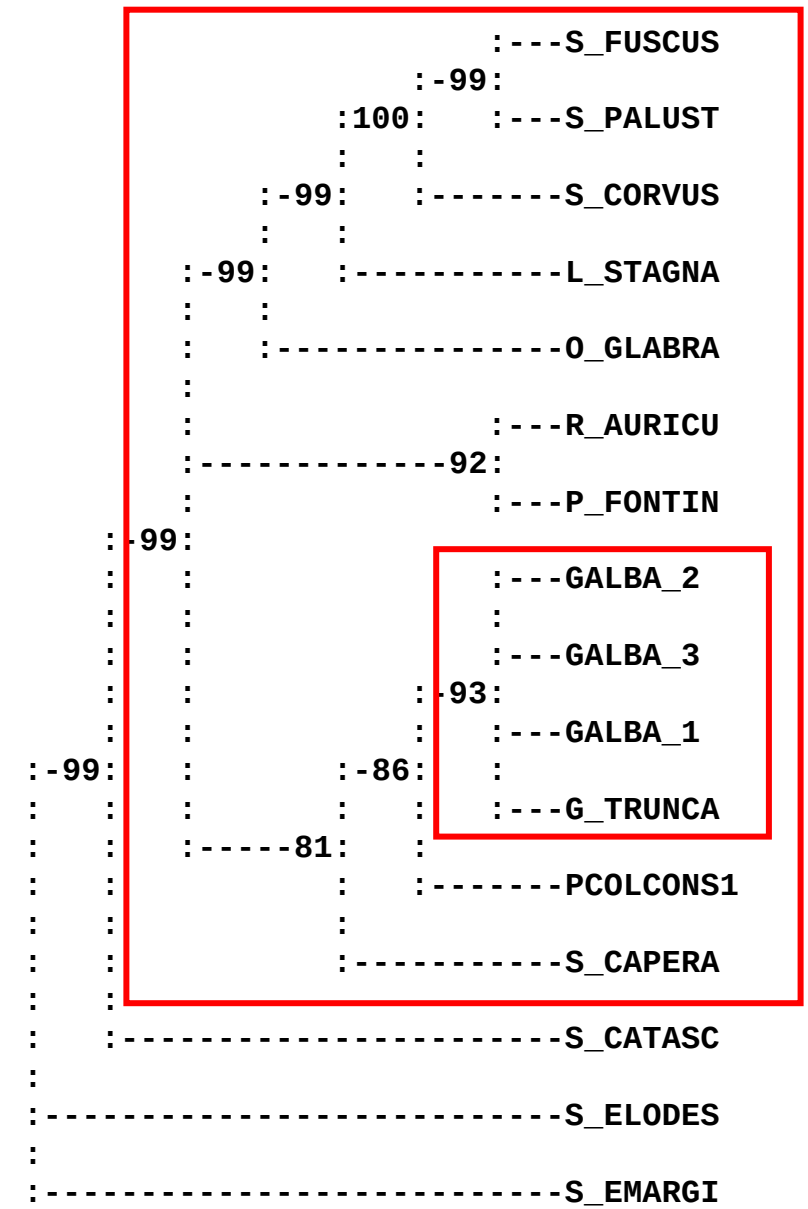


These (sub)trees are consistent within each other

Quartet Puzzling

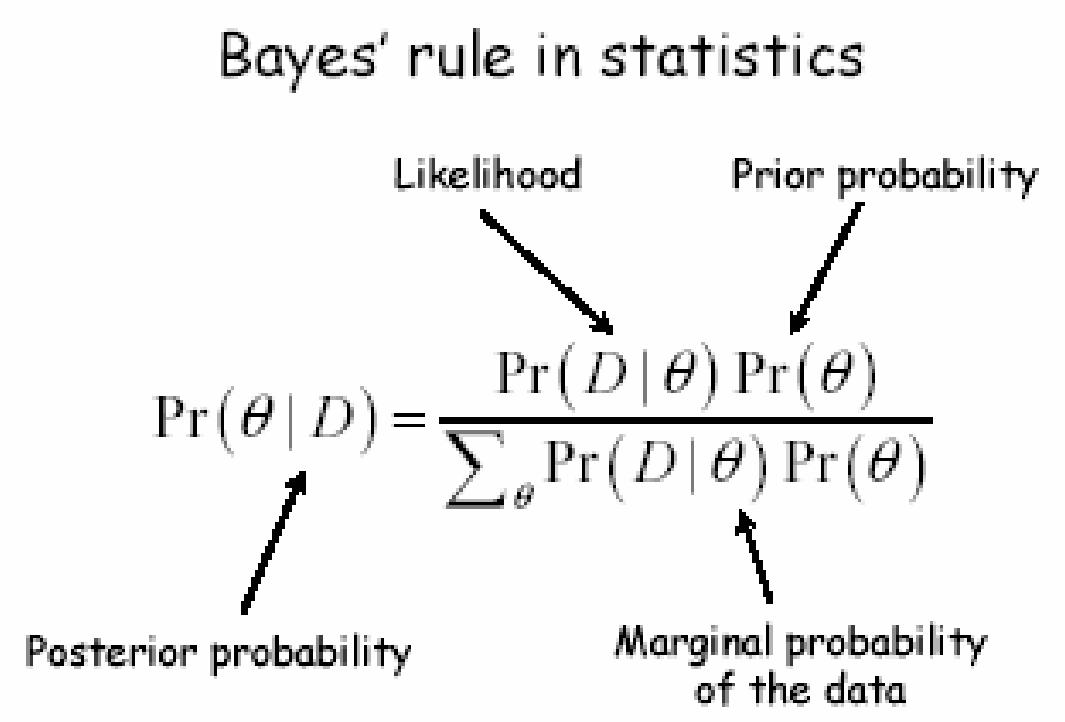
- Speed: ☹
  - Can handle only small data sets
- Completeness: ☺
  - If no complete separation of OTUs is possible, tree will be not completely resolved

This quartet puzzling tree is not completely resolved!



BIOINFORMATICS WAGENINGEN UR

Bayesian Methods



Bayesian Methods

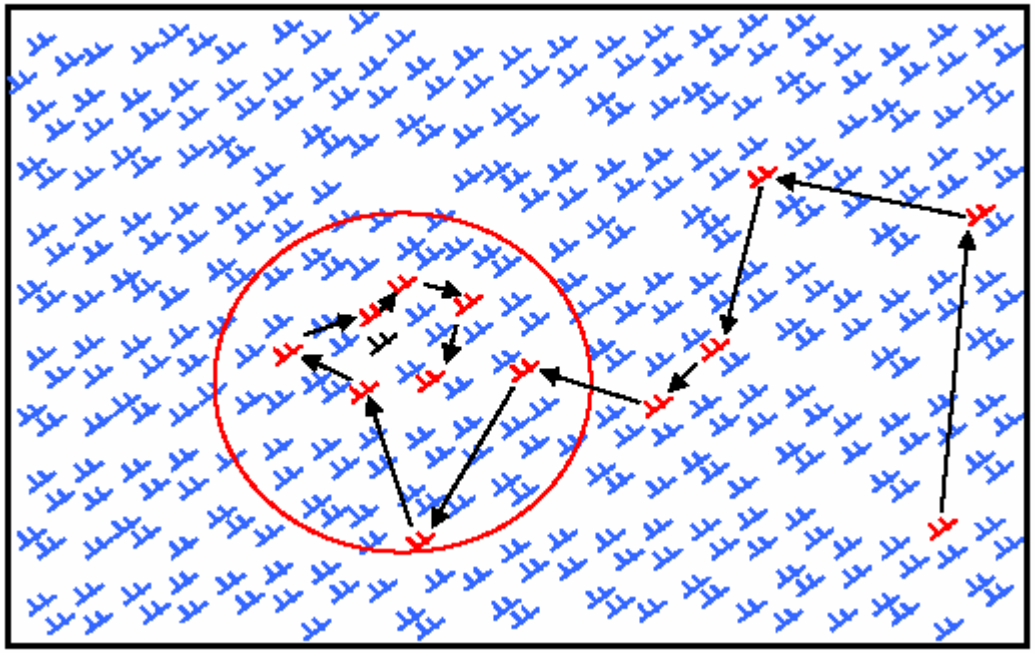
- **Start** with random tree and arbitrary initial values for branch lengths and model parameters
- **Each generation** consists of one of these (chosen at random):
  - Propose a **new tree** and either accept or reject the move
  - Propose (and either accept or reject) a **new model parameter**
- Every k generations, save tree, branch lengths and all model parameters (i.e. **sample the chain**)
- After n generations, **summarise sample** using histograms, means, credibility intervals, etc.

Markov-Chain Monte Carlo

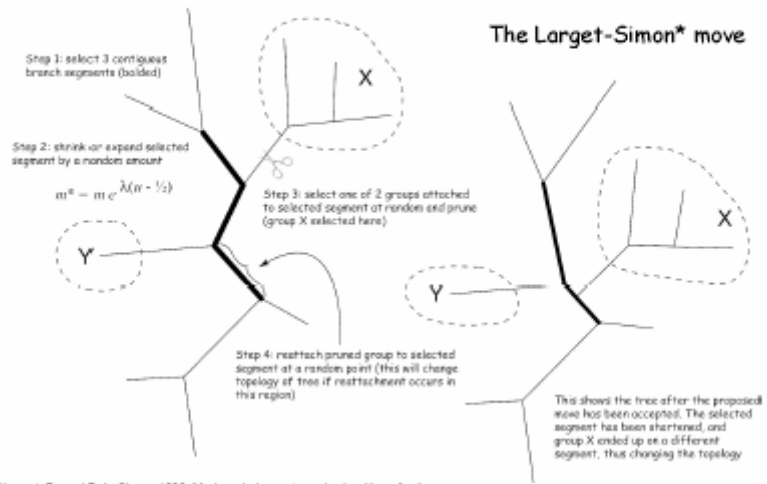
- MCMC methods
  - generate a long chain of phylogenetic trees (tree proposal and acceptance mechanisms)
  - randomly sample from the converged chain
  - calculate event or evolutionary process in each
- Tree acceptance mechanism: Metropolis-Hastings Algorithm
  - Accept new tree with p=1.0 if $L(Tn+1) > L(Tn)$

    otherwise…
  - accept with probability $\propto L(Tn+1)/ L(Tn)$

MCMC Sampling



Moving through treespace



*Larget, B., and D. L. Simon. 1999. Markov chain monte carlo algorithms for the Bayesian analysis of phylogenetic trees. Molecular Biology and Evolution 16: 750-759.

BIOINFORMATICS WAGENINGEN UR

Bayesian Methods

- Speed: ☹
  - Can handle only relatively small data sets
- Completeness: ☺
  - depending on number of generations, the best tree(s) can be found

Among Site Rate Variation

- Not all sites in a sequence will evolve at the same rate. For instance
  - third base position in a codon will show faster rate changes than the first and second position
  - active site residues in a protein will be more slowly evolving than other residues

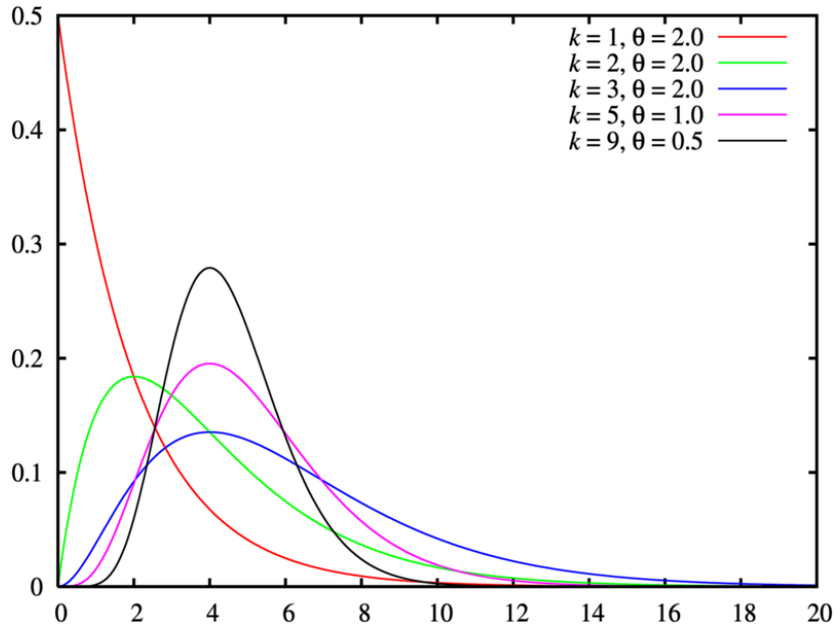
Among Site Rate Variation

- Some methods allow the use of categories or a gamma distribution to mimic ASRV

- These typically are distance, maximum-likelihood, or bayesian methods

Gamma Distribution

- A shape variable (usually called α) varies the character of the distribution
- With α=∞ all sites change at the same rate; an extreme where only a few sites vary and the majority of sites is invariant or change very slowly is obtained with α approaching to 0
- In between are cases that resemble exponential (α=1), Poisson (α≈2) and normal distributions (α>10)



Method	Speed	Completeness
Distances		
Parsimony		
ML		
Puzzle		
Bayes		

Please note that completeness is only valid within the boundaries set by the model and its parameters!

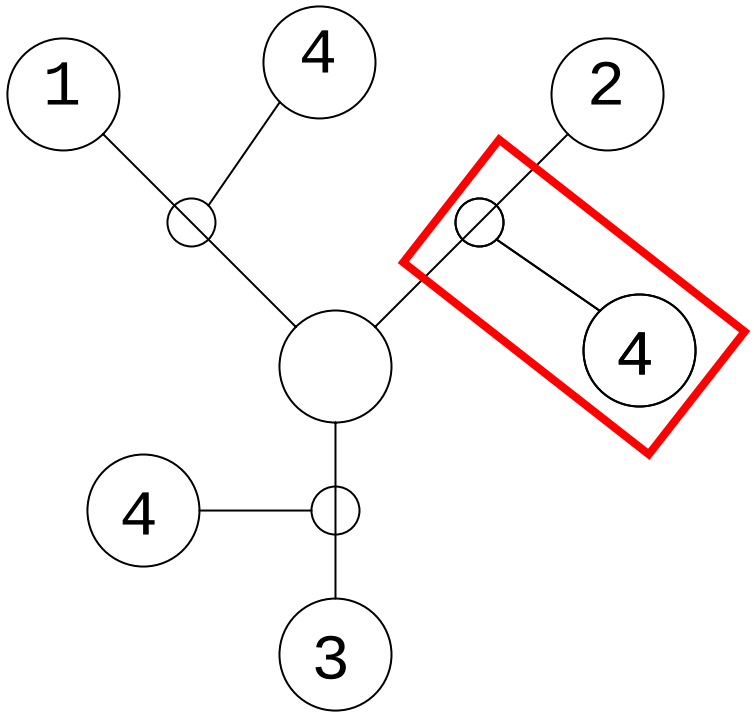
BIOINFORMATICS WAGENINGEN UR

The Mechanics

Stepwise Addition

- Algorithm:
  - Start off with a three-OTU tree
  - Try the next OTU at all possible branches, and retain the most favourable topology
  - Repeats this step until all OTUs have been processed
- There may be an input-order dependency!

Stepwise Addition

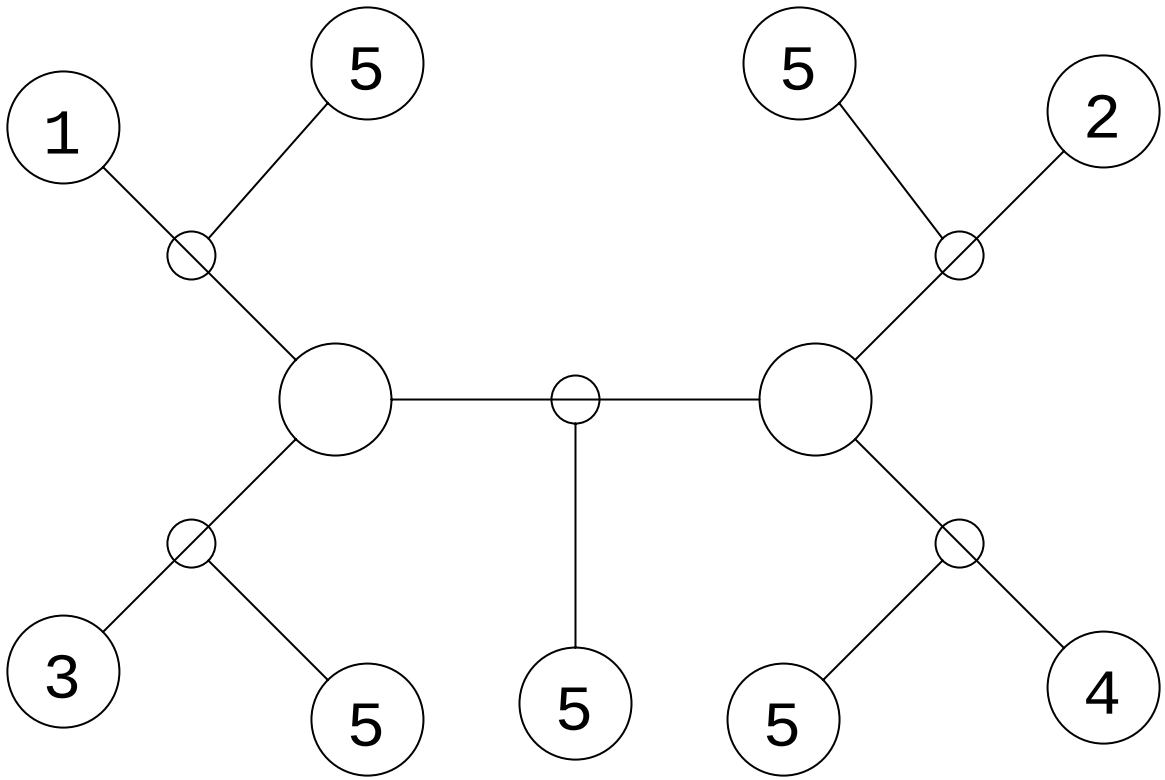


Steps=19

Steps=14

Steps=18

Stepwise Addition

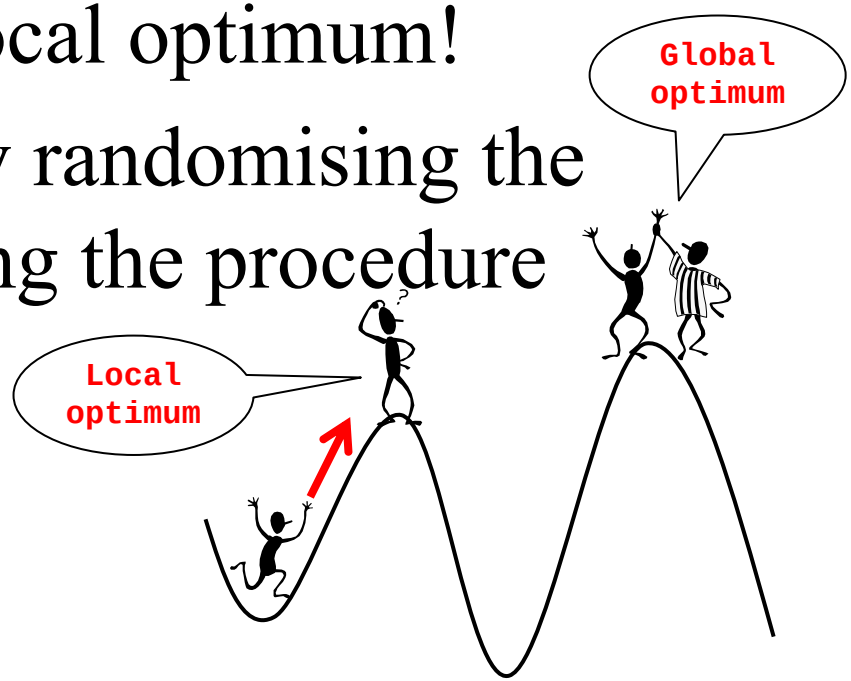


Time Saved in S-A

- Number of trees calculated for 10 OTUs
  - Exhaustive: `1*3*5*7*9*11*13*15 = 2,027,025`
  - Stepwise: `1+3+5+7+9+11+13+15 = 64`

Dependency On Input Order

- By adding OTU *i* at the most optimal position in a previously found (*i*-1) OTU tree, while aiming for the global optimum, we may get stuck in a local optimum!
- This may be detected by randomising the input order, and repeating the procedure several times

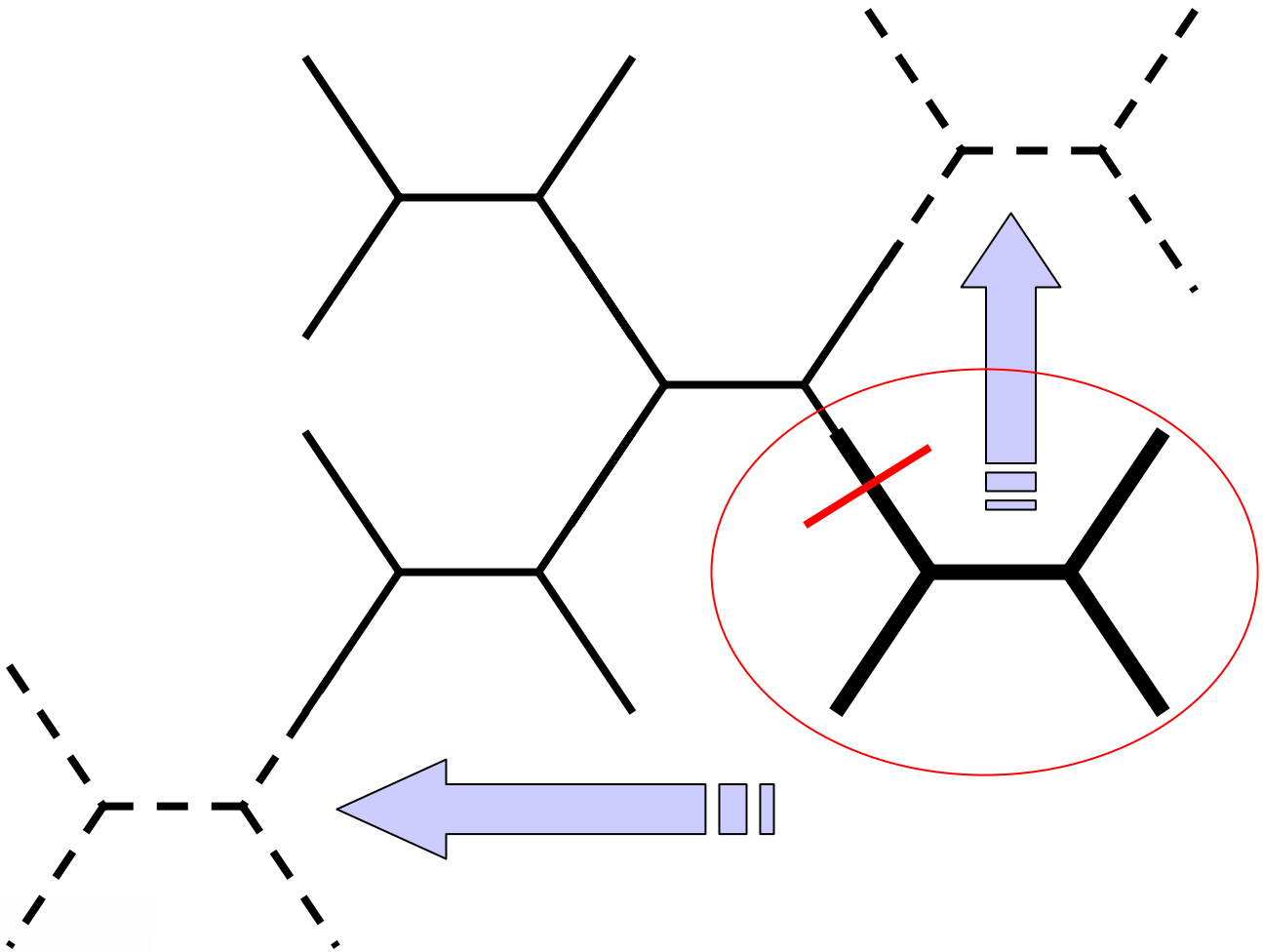


BIOINFORMATICS WAGENINGEN UR

Branch Swapping

- (Randomly) prunes a subtree from the main tree, and tries to add it at various locations; if this yields a better topology, this one is retained
- Is often used in conjunction with the stepwise addition algorithm
- Also used as a post-processing step: Balanced Nearest Neighbour Interchange (BNNI, Vinh & von Haeseler, 2005)

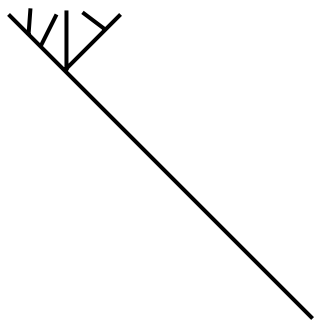
Branch Swapping



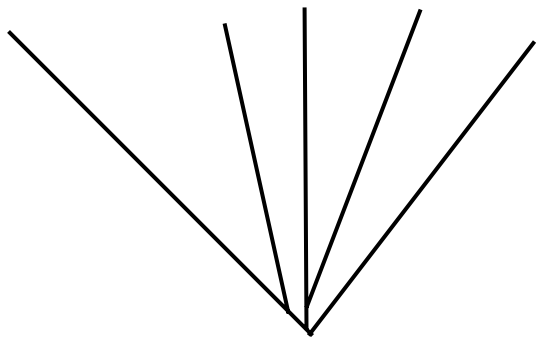
Informative Sites

Are All Residues Useable?

- Too little variation:
  - no resolution at the tips



- Too much variation:
  - no resolution at the root



- Small changes in the model may change the tree topology ⇒ hence tree is unreliable!

Non-informative Sites are ...

- sites that carry no information about any preferential tree topology, such as residue positions with:
  - no variation at all
  - only one difference
  - variations that allow more than one topology with equal “cost” (this may however depend upon the evolutionary model in use!)

The chef's recommendations …

- Remove identical sequences from the data
- Delete all non-informative residue positions from the alignment – **but only if** …

```
Alpha     A A C G T G G C C A A A T
Beta      A A G G T C G C C A A A C
Gamma     C A T T T C G T C A C A A
Delta     G G T A T T T C G G C C T
Epsilon   G G G A T C T C G G C C C
Zeta      G G G A T C T C G G C C C
```

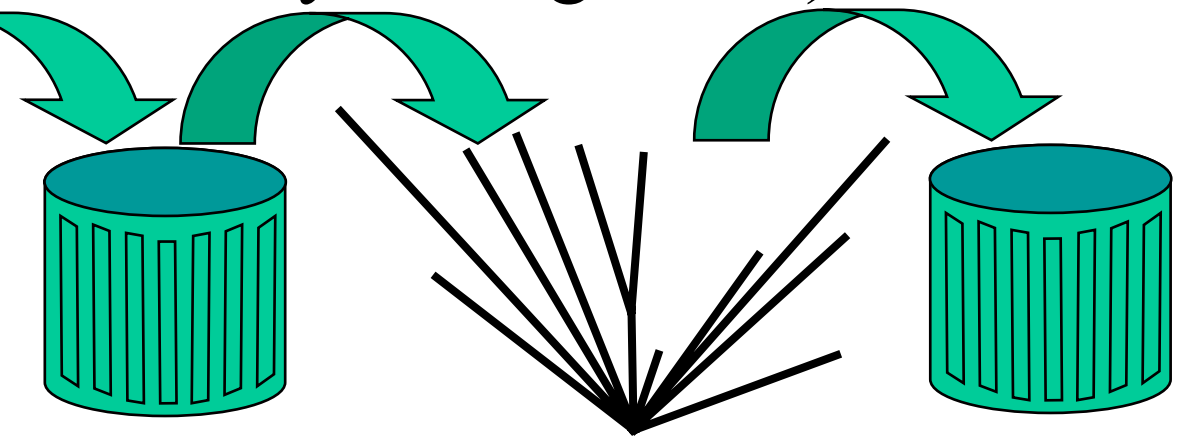
… but only if ...

- … this does **not** compromise the model:
  - Parsimony: unaffected
  - Distances: distances will change!
  - Maximum Likelihood: base frequencies and distances will change!

The Alignment

- The quality of an evolutionary tree can only be as good as the quality of its underlying data (in our case mostly an alignment)

```
 sauraSP21  GVD.GDNVRA K........ ..     .SS
baphidHSPA  RIS...VMSA K........ ./     .N..
avinelIBPB  HIE.VK..GA AP....... ..     AAGR
 ecoliIBPA  .IH.VR..GA KP........ ..RRIEIN..
 ecoliIBPB  NME.VS..GA AA........ ..QRIAISER
 lpneuGSPA  HMK.VT..GA KPRIIPIKSI SDQESNKKSK
  bjapHSPA  .VR.VE..GA KP........ ..RRI.IDNL
  bjapHSPB  YVQ.VK..GA KP........ ..RRIS.SGS
  bjapHSPD  YVQ.VK..NA KP........ ..RRIAINGA
  bjapHSPH  YVQ.VQ..SA KP........ ..RR.AI.AA
  bjapHSPE  YVQ.VK..DA KP........ ..RQ.AIDVA
   bjapHSPC  GID.EDKVSA N........ .VRR...RN
  bjapHSPF  DVD.QDKVAA K........ .VKRIAINGK
mint18kDa1  NLD.TDKIQA R........ ...KIAVGRG
 mav18kDa1  NLD.TDKIEA R........ ...KIAVG..
mlepHSP167  NLD.TERILA R........ ...KISVDRG
```



The “garbage in, garbage out” principle!

BIOINFORMATICS WAGENINGEN UR

Further recommendations…

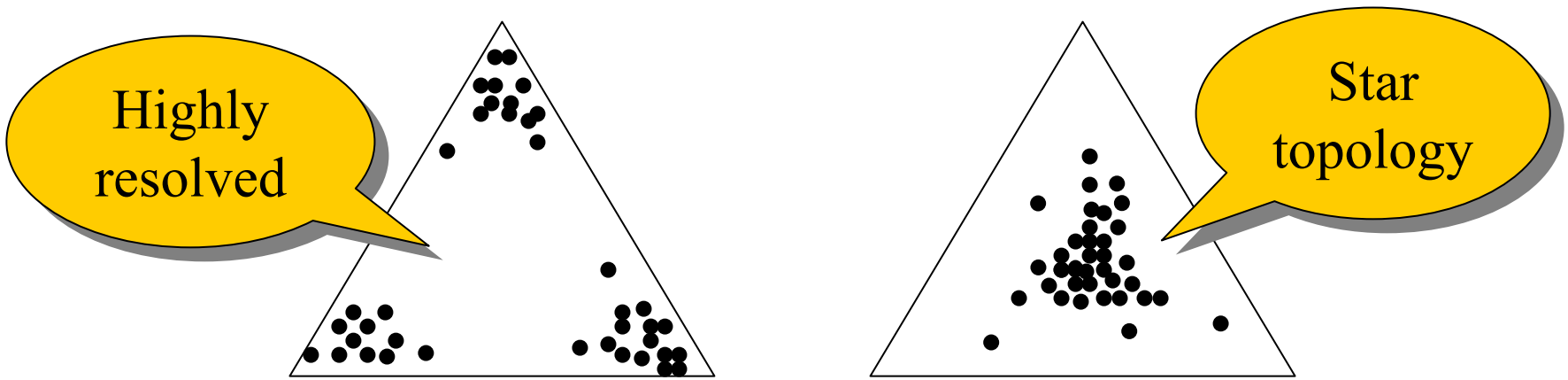
- The quality of an evolutionary tree can only be as good as the quality of its underlying alignment
- Make sure that your data sets contain only orthologous sequences

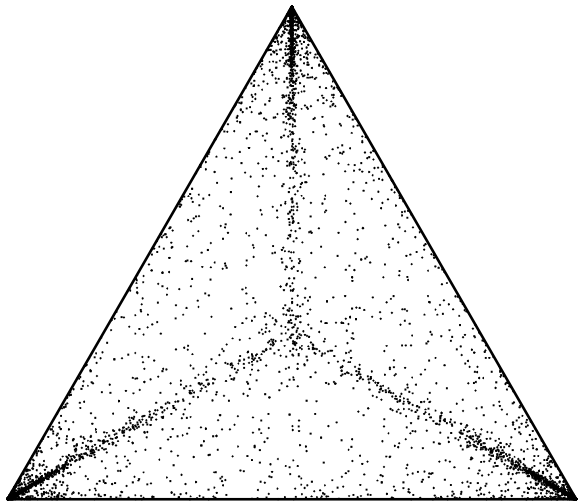
Statistical Evaluation

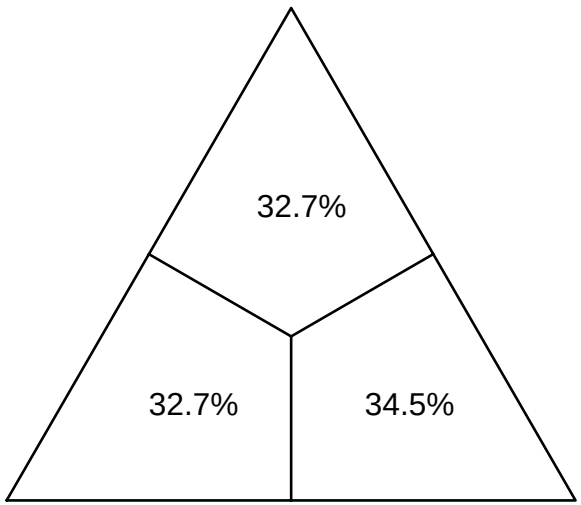
BIOINFORMATICS WAGENINGEN UR

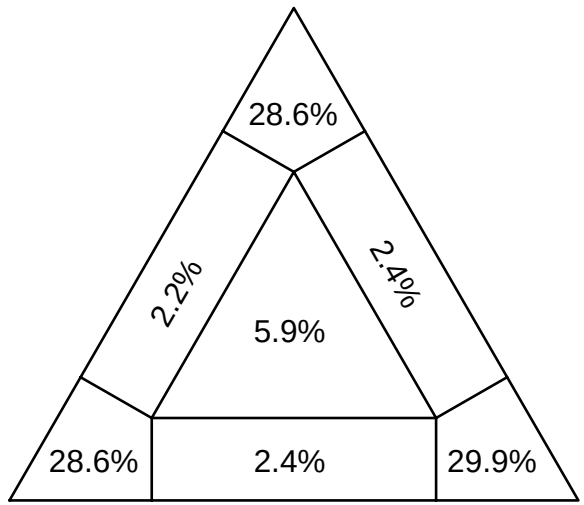
Maximum Likelihood Mapping

- For every quartet that can be constructed for a given set of sequences, calculate the ML values for all three possible topologies
- Plot the centre of gravity in a triangle







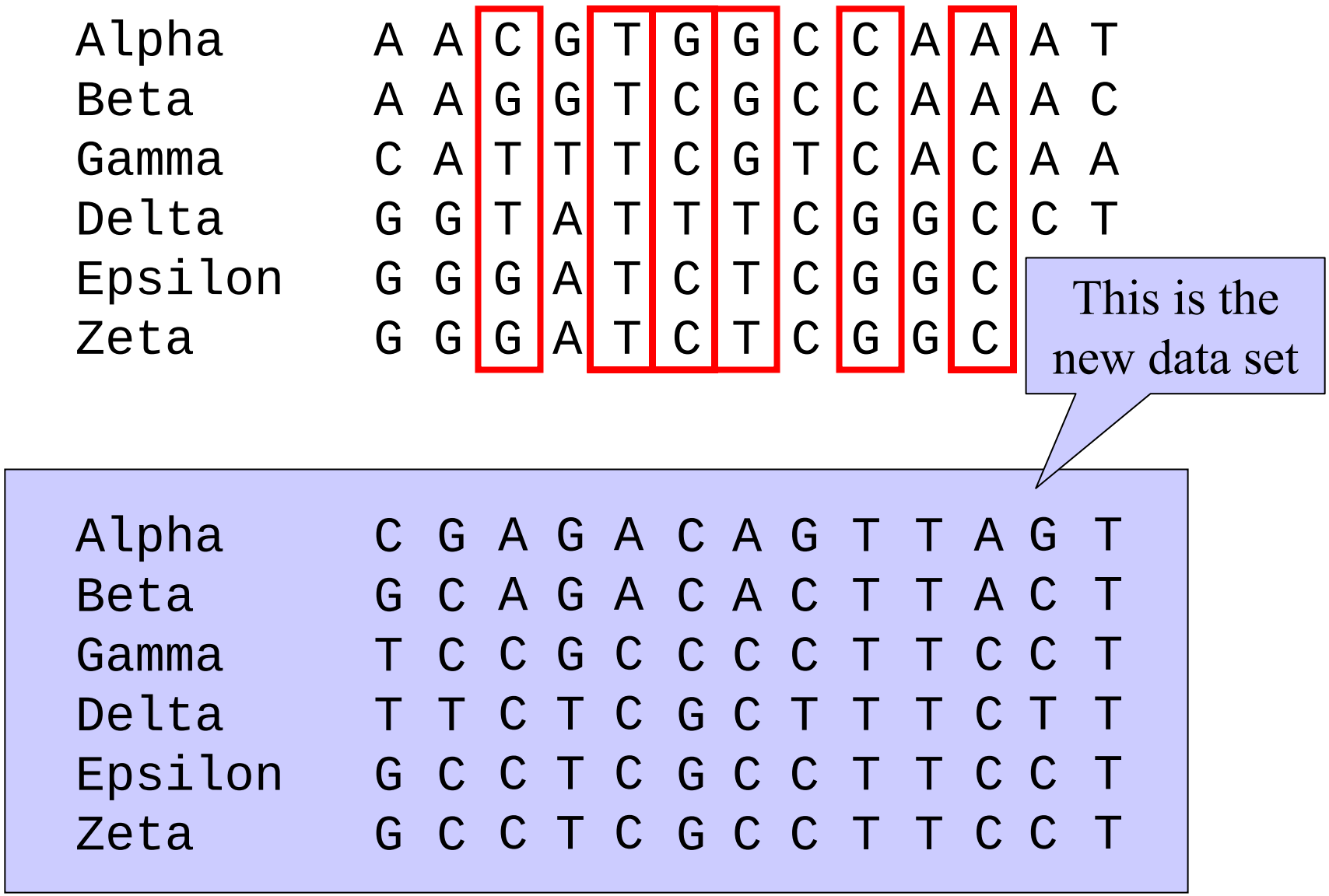


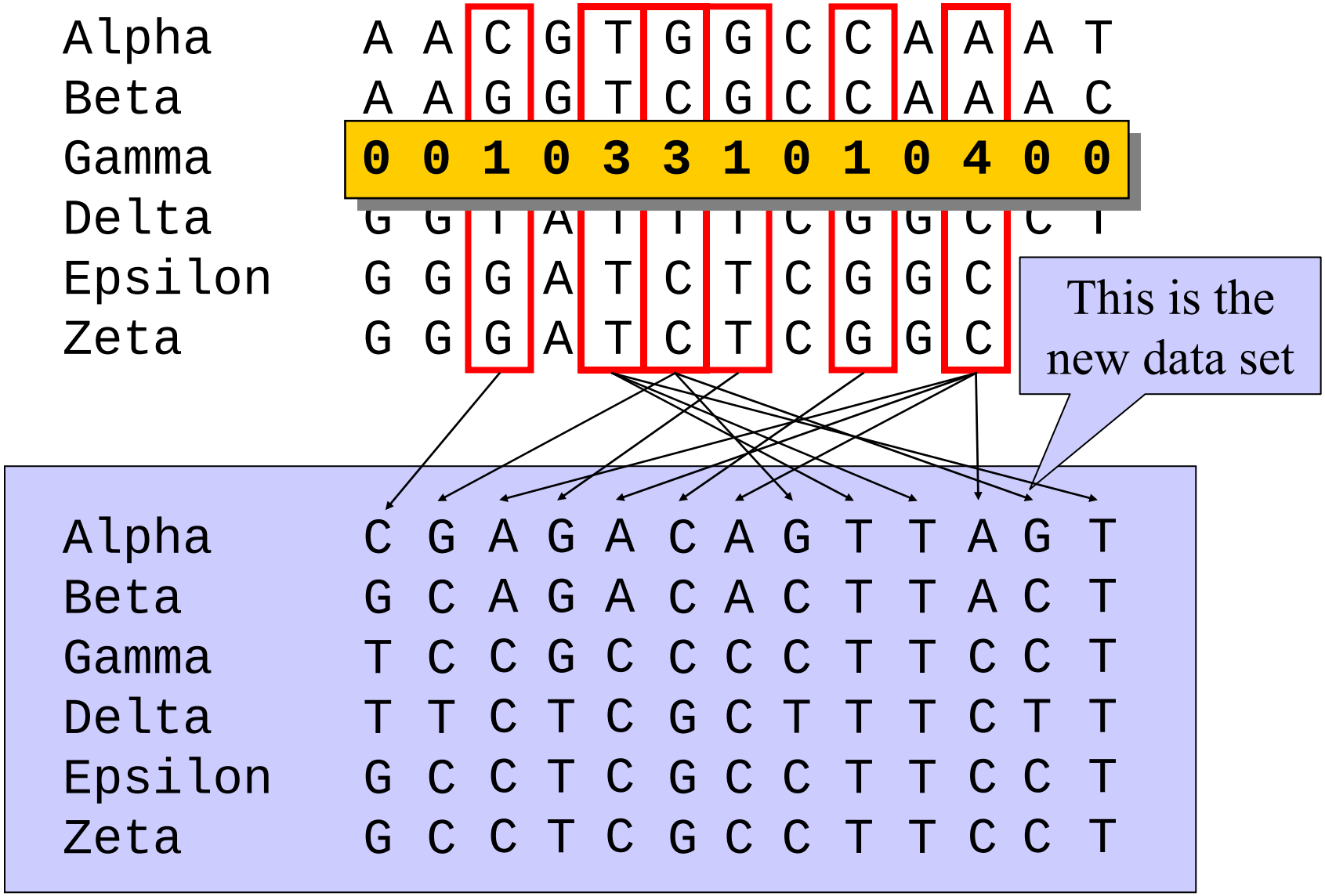
Bootstrapping

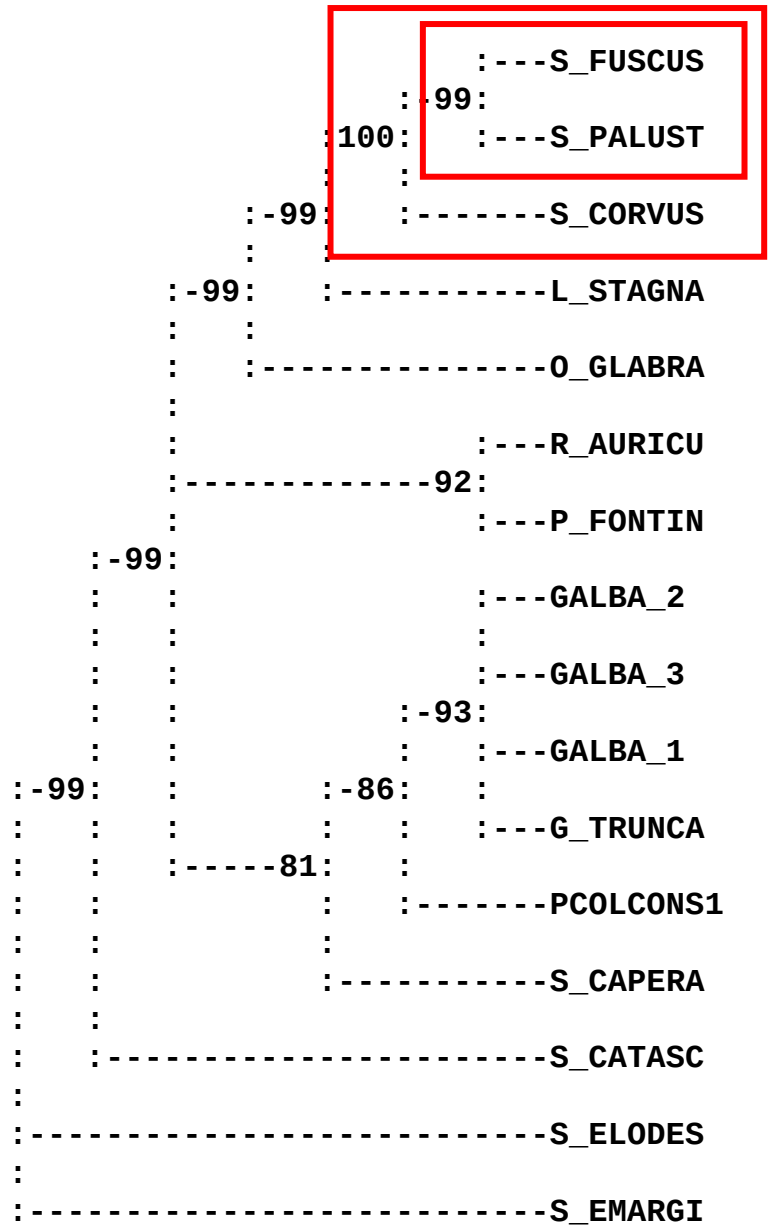
- With bootstrapping we try to find out to what degree the residues in the original multiple alignment support the tree, or parts of the tree, that we have found.
- In other words: it puts confidence limits to the degree of OTUs forming a monophyletic group.

Bootstrapping

- Resample (with replacement!) the data set randomly a large number of times (e.g. 1000 times)
- Generate evolutionary trees for all of these samples
- Obtain confidence limits from majority-rule consensus tree


Alpha A A C G T G G C C A A A T
Beta A A G G T C G C C A A A C
Gamma C A T T T C G T C A C A A
Delta G G T A T T T C G G C C T
Epsilon G G G A T C T C G G C
Zeta G G G A T C T C G G C
This is the new data set
Alpha C G A G A C A G T T A G T
Beta G C A G A C A C T T A C T
Gamma T C C G C C C C T T C C T
Delta T T C T C G C T T T C T T
Epsilon G C C T C G C C T T C C T
Zeta G C C T C G C C T T C C T


Alpha A A C G T G G C C A A A T
Beta A A G G T C G C C A A A C
Gamma 0 0 1 0 3 3 1 0 1 0 4 0 0
Delta G G T A T T T C G G C C T
Epsilon G G G A T C T C G G C
Zeta G G G A T C T C G G C
This is the new data set
Alpha C G A G A C A G T T A G T
Beta G C A G A C A C T T A C T
Gamma T C C G C C C C T T C C T
Delta T T C T C G C T T T C T T
Epsilon G C C T C G C C T T C C T
Zeta G C C T C G C C T T C C T



Random Trees

- Generate 1000 random tree topologies, and calculate their “length”.
- Plot the distribution of tree “lengths” versus the number of trees at this “length”.
- Analyse their frequency distribution.

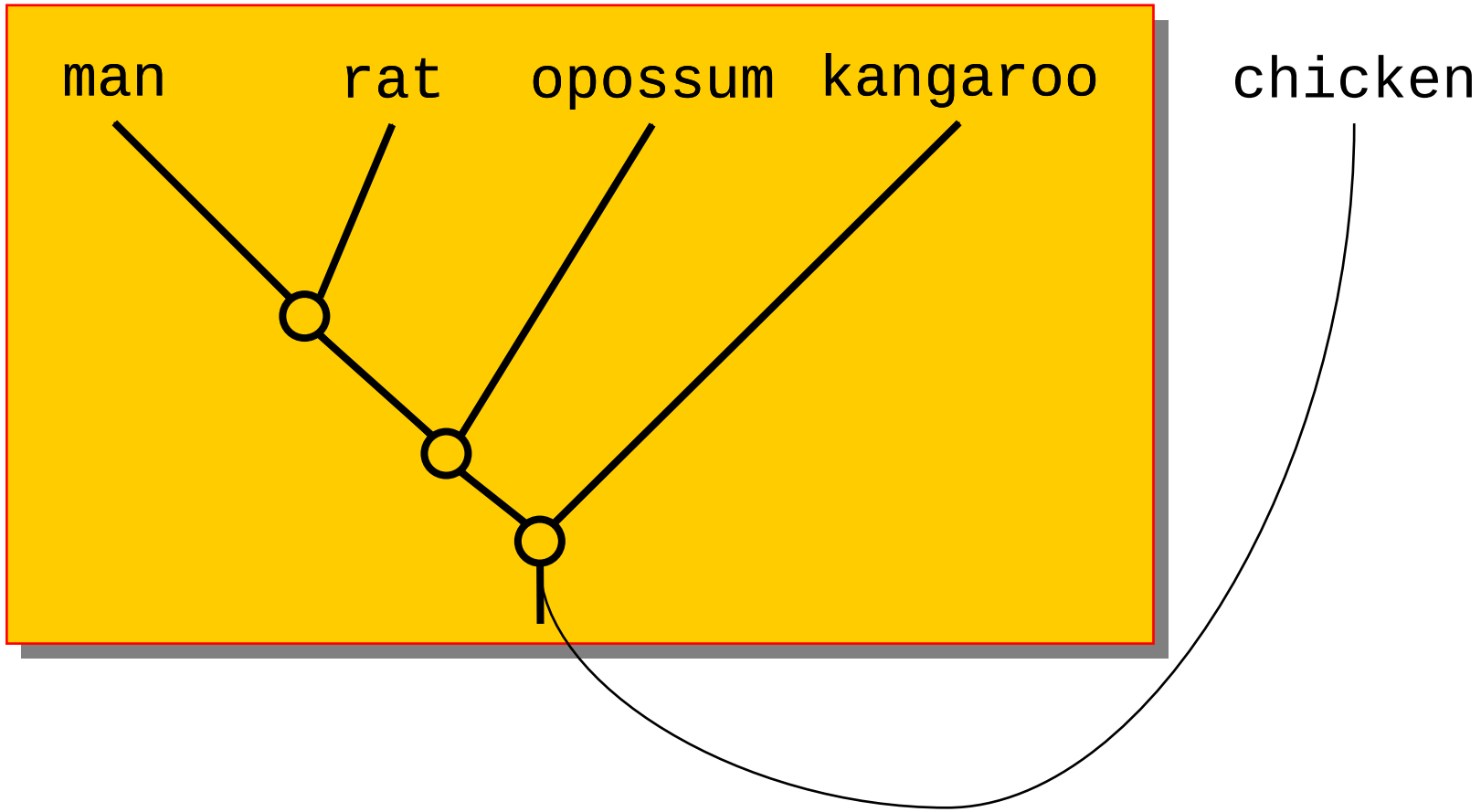
```
105 / 135135 nucml 2.3b3 "A/B:5.49 F" 9 OTUs 540 sites.
# <= 105 trees (top ranking for approx. ln L) in the top 30.0% range of TBL
# range    TBL      trees
#   <      9.87         0
#   5%     9.92        18
#  10%     9.98       104
#  15%    10.03       230
#  20%    10.08       588 *
#  25%    10.14      1004 **
#  30%    10.19      1904 ****
#  35%    10.25      3436 ********
#  40%    10.30      4789 ************
#  45%    10.35      6553 ****************
#  50%    10.41      8606 **********************
#  55%    10.46     10770 ****************************
#  60%    10.52     13814 ************************************
#  65%    10.57     17058 *********************************************
#  70%    10.62     19181 *************************************************
#  75%    10.68     17595 *********************************************
#  80%    10.73     13798 ************************************
#  85%    10.79      9189 ************************
#  90%    10.84      4636 ************
#  95%    10.89      1565 ****
# 100%    10.95       294
#          over         3
# approx. ln L -906.5 ... -909.2 diff 2.8, TBL 10.0 ... 9.9 diff -0.0
((((Peking_Chi,GG_Kenya),Jeddah),LV_Ethiopi),MSA,((Strainx,Tangil_Ben),(LRC-L_Keny,dd_India))); 0.0
((((Peking_Chi,GG_Kenya),Jeddah),LV_Ethiopi),(MSA,(LRC-L_Keny,dd_India)),(Strainx,Tangil_Ben)); 0.0
(((((Peking_Chi,GG_Kenya),Jeddah),LV_Ethiopi),(LRC-L_Keny,dd_India)),MSA,(Strainx,Tangil_Ben)); 0.0
(((Peking_Chi,GG_Kenya),(LV_Ethiopi,Jeddah)),MSA,((Strainx,Tangil_Ben),(LRC-L_Keny,dd_India))); 0.0
(((Peking_Chi,GG_Kenya),(LV_Ethiopi,Jeddah)),(MSA,(LRC-L_Keny,dd_India)),(Strainx,Tangil_Ben)); 0.1
((((Peking_Chi,GG_Kenya),Jeddah),LV_Ethiopi),MSA,((Strainx,(LRC-L_Keny,dd_India)),Tangil_Ben)); 0.1
```

Rooting the tree

Outgroup Rooting

- If we know on which branch the root must be, we can locate the root in a subtree

Outgroup Rooting

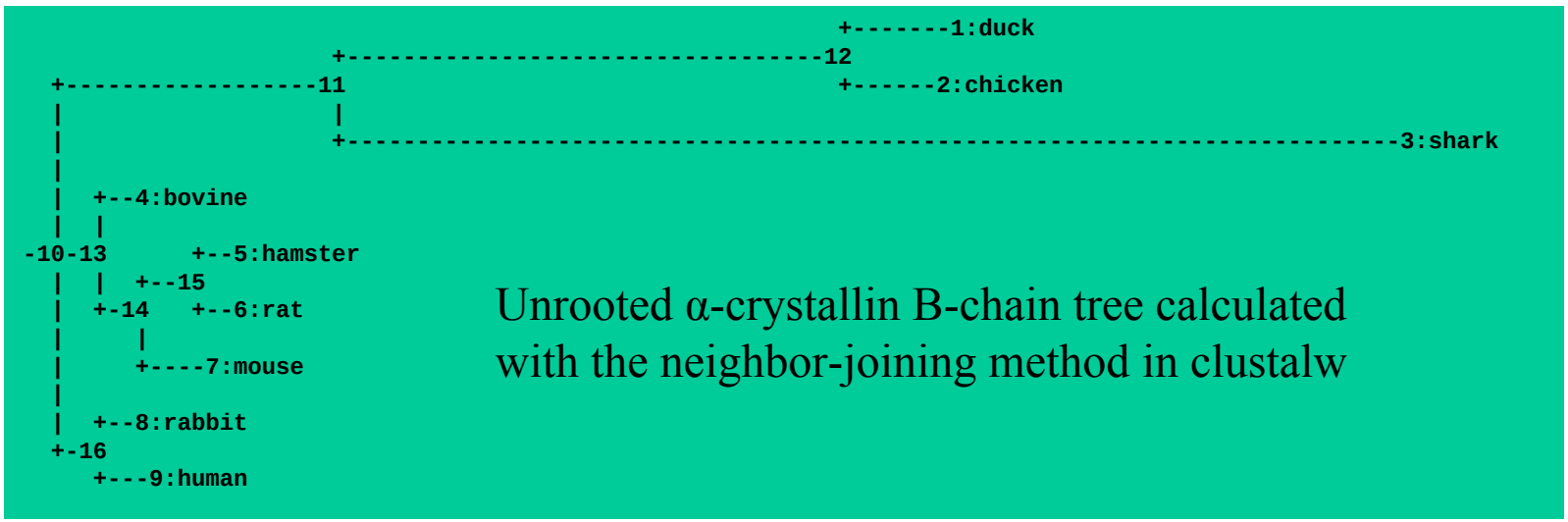


Outgroup Rooting

- If we know on which branch the root must be, we can locate the root in a subtree
- By including an OTU to the data set that is clearly located *outside* the data of interest (i.e. has diverged *before* a subset has diverged) we know where the root must be located in our subtree

Midpoint Rooting

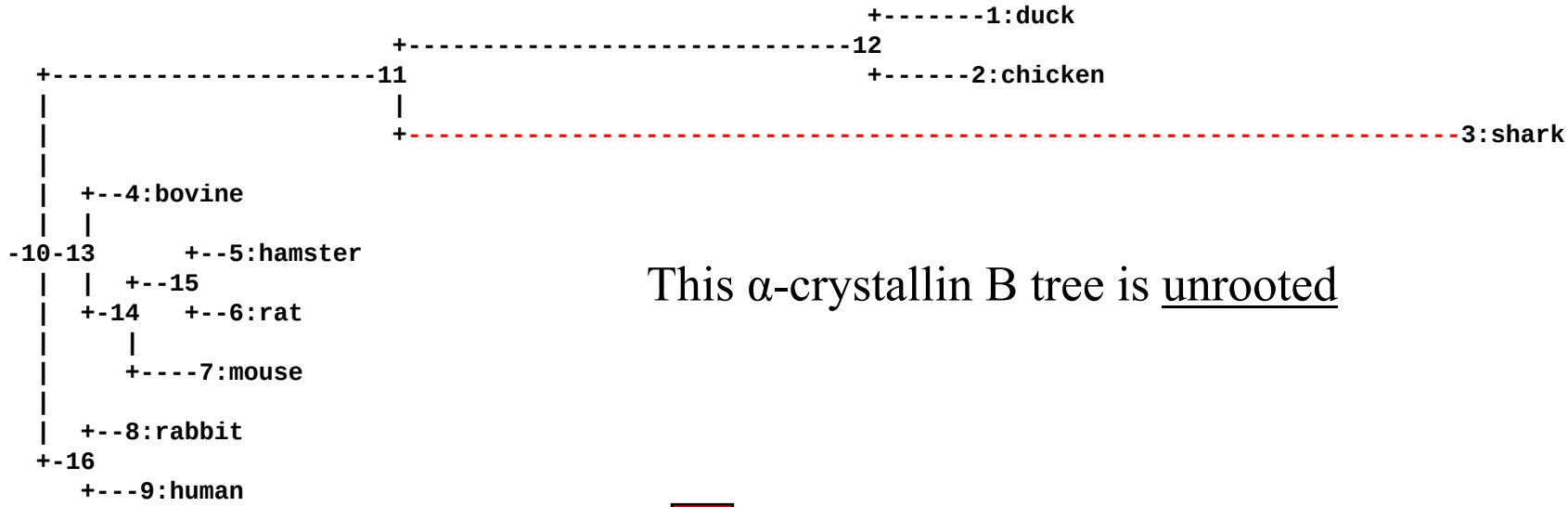
- Find a point on one branch from which all distances to all tips are (more or less) of equal length



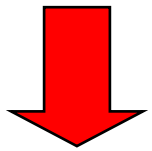
Unrooted α-crystallin B-chain tree calculated with the neighbor-joining method in clustalw

Midpoint Rooting

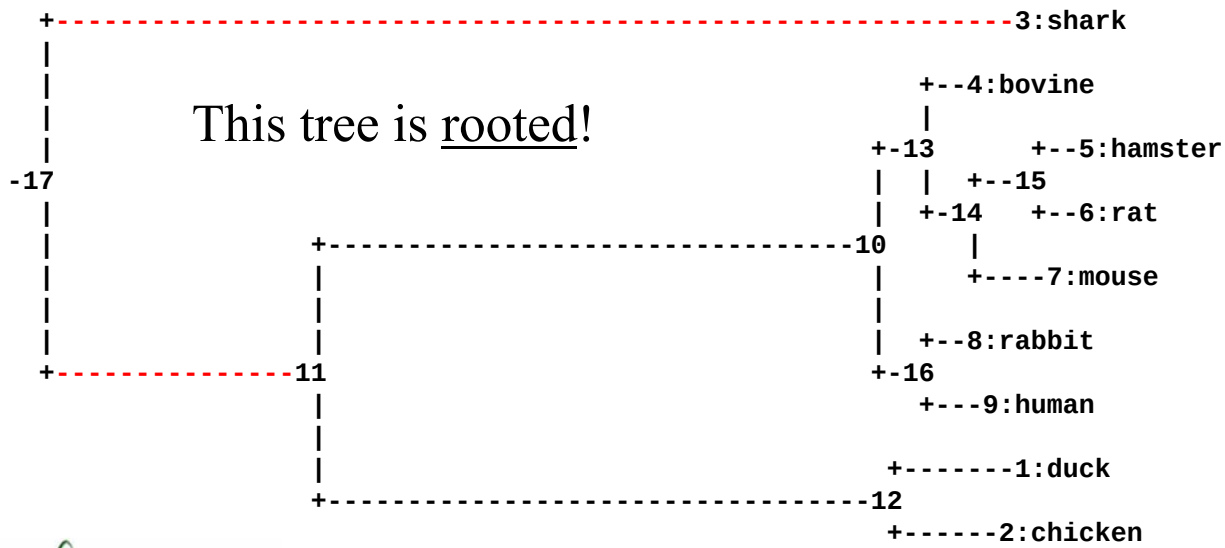
- Find a point on one branch from which all distances to all tips are (more or less) of equal length
  - You may imagine this as if the tree was made out of pieces of string, and you try to find a point from which all ends dangle at more or less the same height



This α-crystallin B tree is unrooted



This tree is rooted!



BIOINFORMATICS WAGENINGEN UR

Midpoint Rooting

- Find a point on one branch from which all distances to all tips are (more or less) of equal length

- Please note: this only works correctly when assuming that there is a constant molecular clock (equal evolutionary rates acting in all branches)!

Concluding Remarks

The Tree Is Influenced By ...

- Quality of the alignment
  - A bad alignment will generate a bad tree
- Choice of the construction algorithm
  - “Cutting corners” may not give you the optimal tree topology
  - Most algorithms will generate only the “best” tree, while disregarding “close” alternatives

The Tree Is Influenced By ...

- Choice of the evolutionary model. This is influenced by the selected
  - Tr/Tv ratio, gamma distribution, score matrix ...
  - The more realistic the model is, the more difficult it is to have realistic and reliable estimates of the parameters involved!

A Small Selection of Available Packages

- PHYLIP — Felsenstein
- PAUP* — Swofford
- MOLPHY — Adachi & Hasegawa
- Tree-Puzzle — Strimmer & von Häseler
- Mr.Bayes — Huelsenbeck & Ronquist
- MEGA3 — Kumar, Tamura & Nei
- PAML — Yang
  - and many, many more

http://evolution.genetics.washington.edu/phylip.html

Recommended Literature

- Molecular evolution (1997)
  - W.-H. Li
- Molecular evolution: a phylogenetic approach (1998)
  - R. Page and E. Holmes
- Inferring phylogenies (2003)
  - J. Felsenstein
- Phylogenetic trees made easy (2nd edition, 2004)
  - B.G. Hall

On-Line Documentation

- PAUP
  - `http://www.lms.si.edu/PAUP/`
- PHYLIP
  - `http://evolution.gs.washington.edu/phylip/phylip.html`