Introduction to Molecular Phylogeny

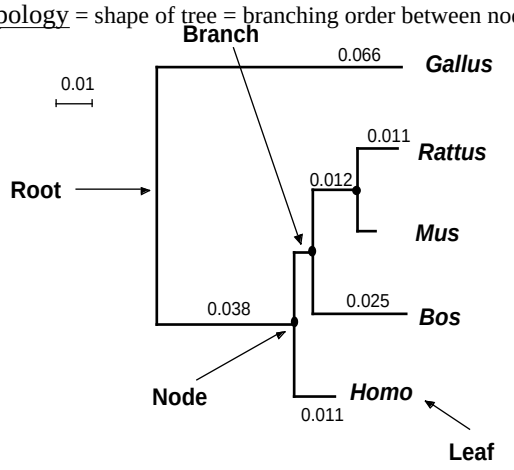
- Starting point: a set of homologous, aligned DNA or protein sequences
- Result of the process: a tree describing evolutionary relationships between studied sequences
  = a genealogy of sequences
  = a phylogenetic tree

```
CLUSTAL W (1.74) multiple sequence alignment

Xenopus         ATGCATGGGCCAACATGACCAGGAGTTGGTGTCGGTCCAAACAGCGTT---GGCTCTCTA
Gallus          ATGCATGGGCCAGCATGACCAGCAGGAGGTAGC---CAAAATAACACCAACATGCAAATG
Bos             ATGCATCCGCCACCATGACCAGCAGGAGGTAGCACCCAAAACAGCACCAACGTGCAAATG
Homo            ATGCATCCGCCACCATGACCAGCAGGAGGTAGCACTCAAAACAGCACCAACGTGCAAATG
Mus             ATGCATCCGCCACCATGACCAGCAGGAGGTAGCACTCAAAACAGCACCAACGTGCAAATG
Rattus          ATGCATCCGCCACCATGACCAGCGGGAGGTAGCTCTCAAAACAGCACCAACGTGCAAATG
                ******  **** *********  *  ***  *   * *** * *           *
```

Phylogenetic Tree

- Internal branch : between 2 nodes. External branch : between a node and a leaf
- Horizontal branch length is proportional to evolutionary distances between sequences and their ancestors (unit = substitution / site).
- Tree Topology = shape of tree = branching order between nodes



Alignment and Gaps

- The quality of the alignment is essential : each column of the alignment (site) is supposed to contain homologous residues (nucleotides, amino acids) that derive from a common ancestor.

  ==> Unreliable parts of the alignment must be omitted from further phylogenetic analysis.

- Most methods take into account only substitutions ; gaps (insertion/deletion events) are not used.

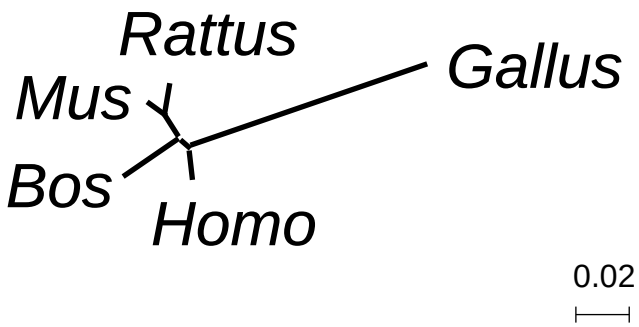
  ==> gaps-containing sites are ignored.

```
Xenopus     ATGCATGGGCCAACATGACCAGGAGTTGGTGTCggtCCAAACAGCGTT---GGCTCTCTA
Gallus      ATGCATGGGCCAGCATGACCAGCAGGAGGTAGC---CAAAATAACACCaacATGCAAATG
Bos         ATGCATCCGCCACCATGACCAGCAGGAGGTAGCagtCAAAACAGCACCaacGTGCAAATG
Homo        ATGCATCCGCCACCATGACCAGCAGGAGGTAGCagtCAAAACAGCACCaacGTGCAAATG
Mus         ATGCATCCGCCACCATGACCAGCAGGAGGTAGCactCAAAACAGCACCaacGTGCAAATG
Rattus      ATGCATCCGCCACCATGACCAGCGGGAGGTAGCtctCAAAACAGCACCaacGTGCAAATG
```

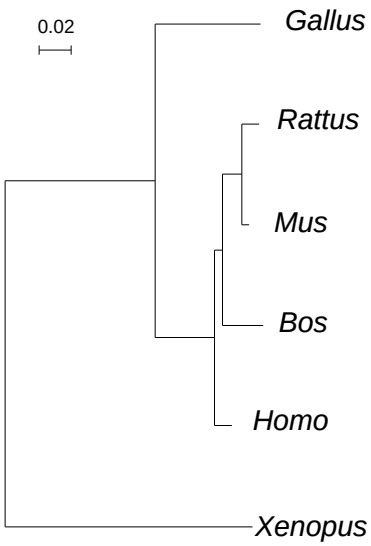
Rooted and Unrooted Trees

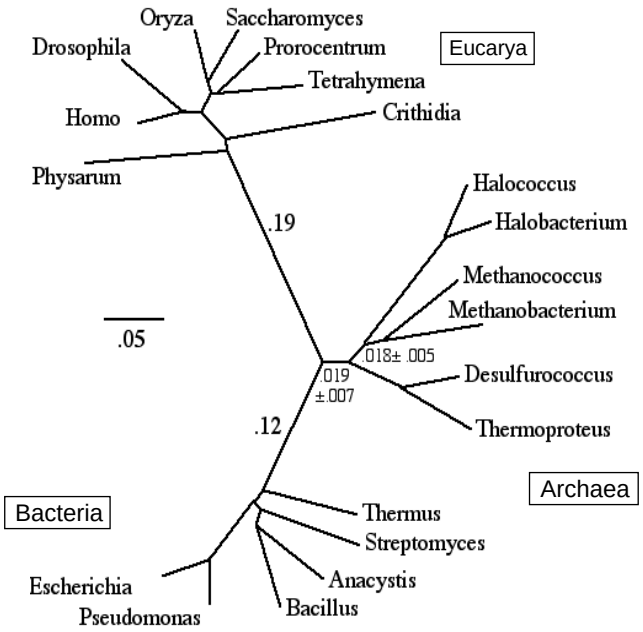
- Most phylogenetic methods produce unrooted trees. This is because they detect differences between sequences, but have no means to orient residue changes relatively to time.
- Two means to root an unrooted tree :
  - The outgroup method : include in the analysis a group of sequences known *a priori* to be external to the group under study; the root is by necessity on the branch joining the outgroup to other sequences.
  - Make the molecular clock hypothesis : all lineages are supposed to have evolved with the same speed since divergence from their common ancestor. The root is at the equidistant point from all tree leaves.

Unrooted Tree



Rooted Tree





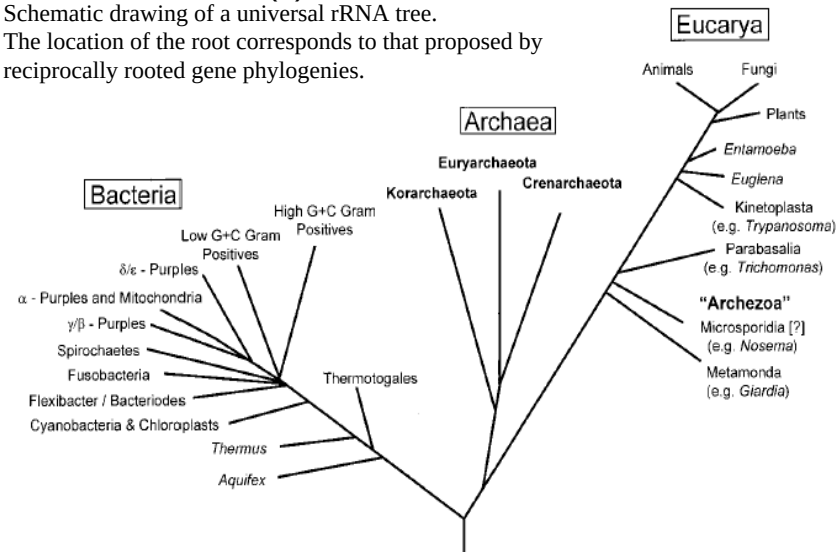
Universal phylogeny (1)

deduced from comparison of SSU and LSU rRNA sequences (2508 homologous sites) using Kimura's 2-parameter distance and the NJ method.

The absence of root in this tree is expressed using a circular design.

Universal phylogeny (2)

Schematic drawing of a universal rRNA tree.
The location of the root corresponds to that proposed by reciprocally rooted gene phylogenies.



Brown & Doolittle (1997) Microbiol.Mol.Biol.Rev. 61:456-502

Number of possible tree topologies for n taxa

$$N_{trees} = 3.5.7...(2n-5) = \frac{(2n-5)!}{2^{n-3}(n-3)!}$$

n	N_{trees}
4	3
5	15
6	105
7	945
...	...
10	2,027,025
...	...
20	~ 2×10^{20}

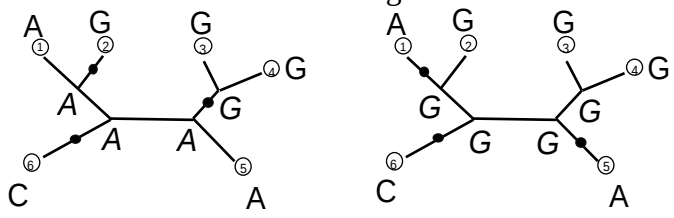
Methods for Phylogenetic reconstruction

Three main families of methods :

- Parsimony
- Distance methods
- Maximum likelihood methods

Parsimony (1)

- Step 1: for a given tree topology (shape), and for a given alignment site, determine what ancestral residues (at tree nodes) require the smallest total number of changes in the whole tree.
  Let *d* be this total number of changes.



Example: At this site and for this tree shape, at least 3 substitution events are needed to explain the nucleotide pattern at tree leaves. Several distinct scenarios with 3 changes are possible.

Parsimony (2)

- Step 2:
  - Compute *d* (step 1) for each alignment site.
  - Add *d* values for all alignment sites.
  - This gives the length *L* of tree.

- Step 3:
  - Compute *L* value (step 2) for each possible tree shape.
  - Retain the shortest tree(s)
    = the tree(s) that require the smallest number of changes
    = the most parsimonious tree(s).

Some properties of Parsimony

- Several trees can be equally parsimonious (same length, the shortest of all possible lengths).
- The position of changes on each branch is not uniquely defined
  => parsimony does not allow to define tree branch lengths in a unique way.
- The number of trees to evaluate grows extremely fast with the number of processed sequences :
  - ⇨ Parsimony can be very computation - intensive.
  - ⇨ The search for the shortest tree must often be restricted to a fraction of the set of all possible tree shapes (heuristic search)
    => there is no mathematical certainty of finding the shortest (most parsimonious) tree.

Building phylogenetic trees by distance methods

General principle :

Sequence alignment

⬇ (1)

Matrix of evolutionary distances between sequence pairs

⬇ (2)

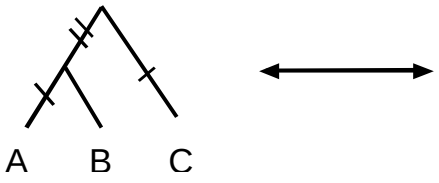
(unrooted) tree

- (1) Measuring evolutionary distances.
- (2) Tree computation from a matrix of distance values.

Correspondence between trees and distance matrices

•Any phylogenetic tree induces a matrix of distances between sequence pairs

• “Perfect” distance matrices correspond to a single phylogenetic tree



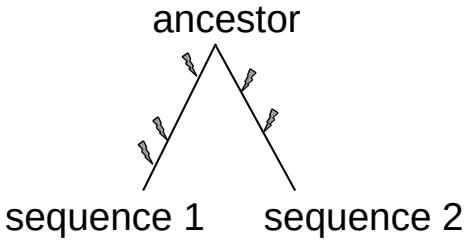
	A	B	C
A	0		
B	1	0	
C	4	3	0

tree

Distance matrix

Evolutionary Distances

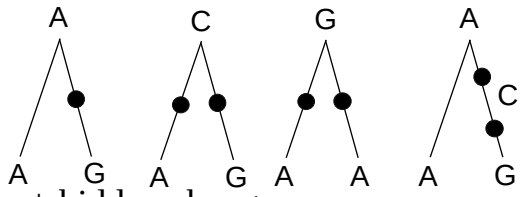
- They measure the total number of substitutions that occurred on <u>both</u> lineages since divergence from last common ancestor.
- Divided by sequence length.
- Expressed in substitutions / site



Quantification of evolutionary distances (1):

The problem of hidden or multiple changes

- *D* (true evolutionary distance) ≥ fraction of observed differences (*p*)



- *D* = *p* + hidden changes
- Through hypotheses about the nature of the residue substitution process, it becomes possible to estimate *D* from observed differences between sequences.
- Estimated *D* : *d*

Quantification of evolutionary distances(2):

Jukes and Cantor's distance (DNA)

- Hypotheses of the model (Jukes & Cantor, 1969) :
  (a) All sites evolve independently and following the same process.
  (b) All substitutions have the same probability.
  (c) The base substitution process is constant in time.
- Quantification of evolutionary distance (*d*) as a function of the fraction of observed differences (*p*):

$$d = -\frac{3}{4}\ln\left(1 - \frac{4}{3}p\right)$$

$$V(d) = \frac{9p(1-p)}{(3-4p)^2 N}$$

N = number of compared sites

p	*d*
0,10	0,11
0,20	0,23
0,40	0,57
0,60	1,21
0,75	+ ∞

Quantification of evolutionary distances (3):

Poisson distances (proteins)

- Hypotheses of the model :
  (a) All sites evolve independently and following the same process.
  (b) All substitutions have the same probability.
  (c) The amino acid substitution process is constant in time.
- Quantification of evolutionary distance (d) as a function of the fraction of observed differences (p) :

$$d = - \ln(1 - p)$$

- !! The hypotheses of the Jukes-Cantor and the Poisson models are very simplistic !!

Quantification of evolutionary distances (3bis):

PAM and Kimura's distances (proteins)

- Hypotheses of the model (Dayhoff, 1979) :
  (a) All sites evolve independently and following the same process.
  (b) Each type of amino acid replacement has a given, empirical probability :
  Large numbers of highly similar protein sequences have been collected; probabilities of replacement of any a.a. by any other have been tabulated.
  (c) The amino acid substitution process is constant in time.
- Quantification of evolutionary distance (d) :
  the number of replacements most compatible with the observed pattern of amino acid changes and individual replacement probabilities.
- Kimura's empirical approximation : $d = - \ln(1 - p - 0.2\, p^2)$
  (Kimura, 1983) where p = fraction of observed differences

Quantification of evolutionary distances (4):

Kimura's two parameter distance (DNA)

- Hypotheses of the model :
  (a) All sites evolve independently and following the same process.
  (b) Substitutions occur according to two probabilities :
  One for transitions, one for transversions.
  Transitions : G <–>A or C <–>T Transversions : other changes
  (c) The base substitution process is constant in time.

- Quantification of evolutionary distance (d) as a function of the fraction of observed differences (p: transitions, q: transversions):

$$d = -\frac{1}{2}\ln[(1-2p-q)\sqrt{1-2q}\,]$$

Kimura (1980) J. Mol. Evol. 16:111

Quantification of evolutionary distances (5):

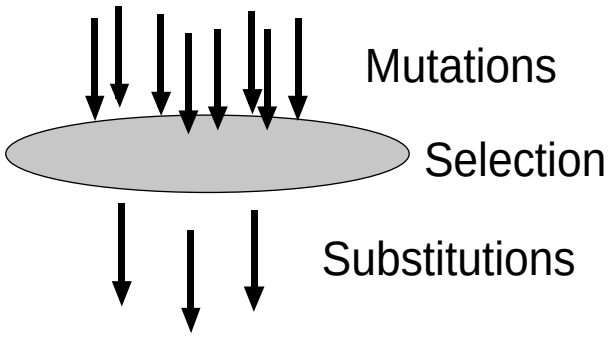
Synonymous and non-synonymous distances (coding DNA): Ka, Ks

- Hypothesis of previous models :
  (a) All sites evolve independently and following the same process.
- Problem: in protein-coding genes, there are two classes of sites with very different evolutionary rates.
  - non-synonymous substitutions (change the a.a.): slow
  - synonymous substitutions (do not change the a.a.): fast
- Solution: compute <u>two</u> evolutionary distances
  - Ka = non-synonymous distance
  - Ka = nbr. non-synonymous substitutions / nbr. non-synonymous sites
  - Ks = synonymous distance
  - Ks = nbr. synonymous substitutions / nbr. synonymous sites

The genetic code

TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys
TTC	Phe	TCC	Ser	TAC	Tyr	TGC	Cys
TTA	Leu	TCA	Ser	TAA	stop	TGA	stop
TTG	Leu	TCG	Ser	TAG	stop	TGG	Trp
CTT	Leu	CCT	Pro	CAT	His	CGT	Arg
CTC	Leu	CCC	Pro	CAC	His	CGC	Arg
CTA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CTG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser
ATC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
ATA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
ATG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly
GTC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GTA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GTG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Substitution rate = f (mutation, selection)



NB: the vast majority of mutations are either neutral (i.e. have no phenotypic effect), or deleterious. Advantageous mutations are very rare.

Quantification of evolutionary distances (6):

Calculation of Ka and Ks

- The details of the method are quite complex. Roughly :
  - Split all sites of the 2 compared genes in 3 categories :
    I: non degenerate, II: partially degenerate, III: totally degenerate
  - Compute the number of non-synonymous sites = I + 2/3 II
  - Compute the number of synonymous sites = III + 1/3 II
  - Compute the numbers of synonymous and non-synonymous changes
  - Compute, with Kimura's 2-parameter method, Ka and Ks

- Frequently, one of these two situations occur :
  - Evolutionarily close sequences : Ks is informative, Ka is not.
  - Evolutionarily distant sequences : Ks is saturated , Ka is informative.

Li, Wu & Luo (1985) Mol.Biol.Evol. 2:150

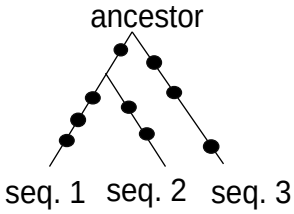
Ka and Ks : example

# sites	observed diffs.	J & C	K2P	K_A	K_S
10254	0.077	0.082	0.082	0.035	0.228

Urotrophin gene of rat (AJ002967) and mouse (Y12229)

Saturation: loss of phylogenetic signal

- When compared homologous sequences have experienced too many residue substitutions since divergence, it is impossible to determine the phylogenetic tree, whatever the tree-building method used.



- NB: with distance methods, the saturation phenomenon may express itself through mathematical impossibility to compute *d*. Example: Jukes-Cantor: $p > 0.75 \Rightarrow d \rightarrow \infty$ and $V(d) \rightarrow \infty$
- NB: often saturation may not be detectable

Quantification of evolutionary distances (7):

Other distance measures

- Several other, more realistic models of the evolutionary process at the molecular level have been used :
  - Accounting for biased base compositions (Tajima & Nei).
  - Accounting for variation of the evolutionary rate across sequence sites.
  - etc ...

Building phylogenetic trees by distance methods

General principle :

Sequence alignment

⬇ (1)

Matrix of evolutionary distances between sequence pairs

⬇ (2)

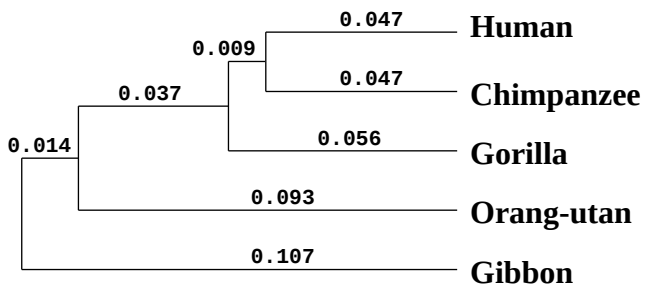
(unrooted) tree

- (1) Measuring evolutionary distances.
- (2) Tree computation from a matrix of distance values.

A (bad) method : UPGMA

	Human	Chimpanzee	Gorilla	Orang-utan	Gibbon
Human	-	0.088	0.103	0.160	0.181
Chimpanzee	0.094	-	0.106	0.170	0.189
Gorilla	0.111	0.115	-	0.166	0.189
Orang-utan	0.180	0.194	0.188	-	0.188
Gibbon	0.207	0.218	0.218	0.216	-

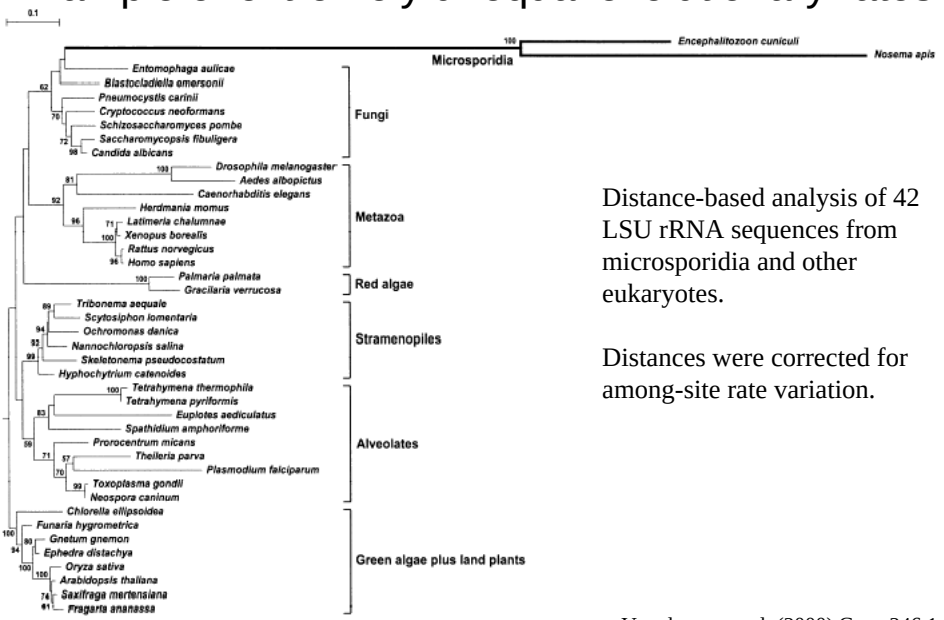
Proportion of differences (*p*) (above diagonal) and Kimura's 2-parameter distances (*d*) (below) for mitochondrial DNA sequences (895 bp).



Resulting UPGMA tree

d(Gibbon,[Human+Chimp]) = 1/2 [d(Gibbon,Human) + d(Gibbon,Chimp)]

Example of extremely unequal evolutionary rates



Distance-based analysis of 42 LSU rRNA sequences from microsporidia and other eukaryotes.

Distances were corrected for among-site rate variation.

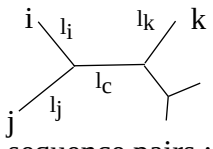
Van de peer et al. (2000) Gene 246:1

UPGMA : properties

- UPGMA produces a rooted tree with branch length.
- It is a very fast method.
- But UPGMA fails if evolutionary rate varies among lineages.
- UPGMA would not have recovered the fungal evolutionary origin of microsporidia.

==> need methods insensitive to rate variations.

Distance matrix -> tree (1): preliminary



- Let us consider the following tree :
- Let us consider two sets of distances between sequence pairs :
  - *d* = distance as measured on sequences
  - δ = distance induced by the above tree :

$$\delta_{i,j} = l_i + l_j \qquad \delta_{i,k} = l_i + l_c + l_k$$

- It is possible (with a computer) to compute branch lengths (l_i, l_j, l_c, etc.) so that distances δ correspond "best" to distances *d*.
  "Best" means that the divergence Δ between *d* and δ values is minimal :

$$\Delta = \sum_{1 \le x < y \le n} (d_{x,y} - \delta_{x,y})^2$$

- It is then possible to compute the total tree length, *S* :

$$S = l_i + l_j + l_c + \ldots + l_k + \ldots$$

Distance matrix -> tree (2):

The Minimum Evolution Method

- Step 1: for a given tree topology (shape), compute branch lengths that minimise Δ; compute tree length *S*.
- Step 2: repeat step 1 for all possible topologies.
  Keep the tree with smallest *S* value.

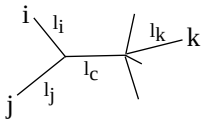
- Problem: this method is very computation intensive. It is practically not usable with more than ≈ 25 sequences.
  => approximate (heuristic) methods are used.

  Example: Neighbor-Joining.

Distance matrix -> tree (3):

The Neighbor-Joining Method: algorithm

- Start from a star - topology and progressively construct a tree as :
  - Step 1: Use *d* distances measured between the *N* sequences
  - Step 2: For <u>all</u> pairs *i* et *j*: consider the following tree topology, and compute $S_{i,j}$, the sum of all "best" branch lengths. (Saitou and Nei have found a simple way to compute $S_{i,j}$).



  - Step 3: Retain the pair (*i*,*j*) with smallest $S_{i,j\ value}$. Group *i* and *j* in the tree.

Saitou & Nei (1987) Mol.Biol.Evol. 4:406

Distance matrix -> tree (4):

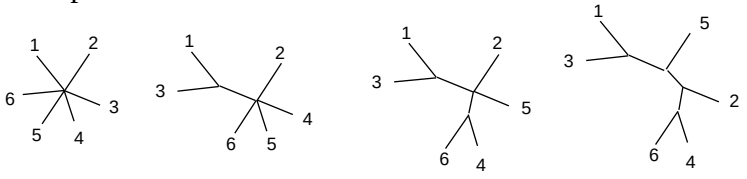
The Neighbor-Joining Method: algorithm (2)

- Step 4: Compute new distances *d* between *N*-1 objects: pair (*i*,*j*) and the *N*-2 remaining sequences.

$$d_{(i,j),k} = (d_{i,k} + d_{j,k}) / 2$$

- Step 5: Return to step 1 as long as $N \geq 4$.
  When $N = 3$, an (unrooted) tree is obtained

- Example



Distance matrix -> tree (5):

The Neighbor-Joining Method (NJ): properties

- NJ is a fast method, even for hundreds of sequences.
- The NJ tree is an approximation of the minimum evolution tree (that whose total branch length is minimum).
- In that sense, the NJ method is very similar to parsimony because branch lengths represent substitutions.
- NJ produces always unrooted trees, that need to be rooted by the outgroup method.
- NJ always finds the correct tree if distances are tree-like.
- NJ performs well when substitution rates vary among lineages. Thus NJ should find the correct tree if distances are well estimated.

Maximum likelihood methods

(program fastDNAml, Olsen & Felsenstein)

- Hypotheses
  - The substitution process follows a probabilistic model whose mathematical expression, but not parameter values, is known *a priori*.
  - Sites evolve independently from each other.
  - All sites follow the same substitution process (some methods use a more realistic hypothesis).
  - Substitution probabilities do not change with time on any tree branch. They may vary between branches.

Maximum likelihood methods (1)

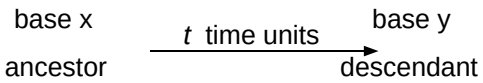
Simple example : one - parameter substitution model :

v = probability that a base changes per unit time

(fastDNAml uses a more elaborate model)

Maximum likelihood methods (2)

- Let us consider evolution along a tree branch :



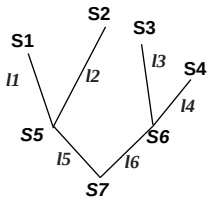
- Our probabilistic model allows to compute the probability of substitution x → y along this branch :

$$P_l(x,y) = \begin{cases} \frac{1}{4} - \frac{1}{4}e^{-\frac{4}{3}l} & \text{if } x = y \\ \frac{1}{4} - \frac{1}{4}e^{-\frac{4}{3}l} & \text{if } x \neq y \end{cases} \qquad \text{with } l = 3vt$$

- Quantity $l = 3vt$ is the average number of substitutions / site along this branch, *i.e.* the branch length.

Maximum likelihood algorithm (1)

- Step 1: Let us consider a given rooted tree, a given site, and a given set of branch lengths. Let us compute the probability that the observed pattern of nucleotides at that site has evolved along this tree.



S1, S2, S3, S4: observed bases at site in seq. 1, 2, 3, 4
S5, S6, S7: unknown and variable ancestral bases
l1, l2, …, l6: given branch lengths

$P(S1, S2, S3, S4) =$
$\Sigma_{S7}\Sigma_{S5}\Sigma_{S6} P(S7)\ P_{l5}(S7,S5)\ P_{l6}(S7,S6)\ P_{l1}(S5,S1)\ P_{l2}(S5,S2)\ P_{l3}(S6,S3)\ P_{l4}(S6,S4)$

where P(S7) is estimated by the average base frequencies in studied sequences.

Maximum likelihood algorithm (2)

- Step 2: Let us compute the probability that entire sequences have evolved :

  $P(Sq1, Sq2, Sq3, Sq4) = \Pi_{\text{all sites}} P(S1, S2, S3, S4)$

- Step 2: Let us compute branch lengths *l1, l2, …, l6* that give the highest $P(Sq1, Sq2, Sq3, Sq4)$ value. This is the *likelihood* of the tree.

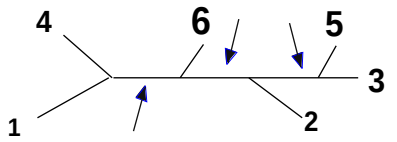
- Step 3: Let us compute the likelihood of all possible trees. The tree predicted by the method is that having the highest likelihood.

Maximum likelihood : properties

- This is the best justified method from a theoretical viewpoint.
- Sequence simulation experiments have shown that this method works better than all others in most cases.
- But it is a very computer-intensive method.
- It is nearly always impossible to evaluate all possible trees because there are too many. A partial exploration of the space of possible trees is done. The mathematical certainty of obtaining the most likely tree is lost.

Reliability of phylogenetic trees: the bootstrap

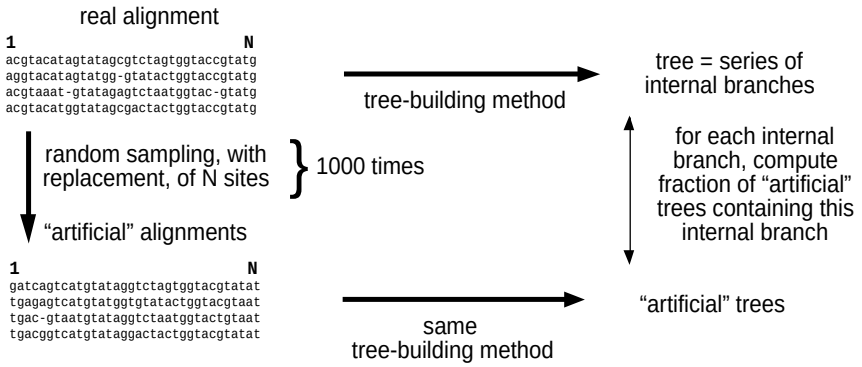
- The phylogenetic information expressed by an unrooted tree resides entirely in its internal branches.



internal branch separating group 1+4
from group 6+2+5+3

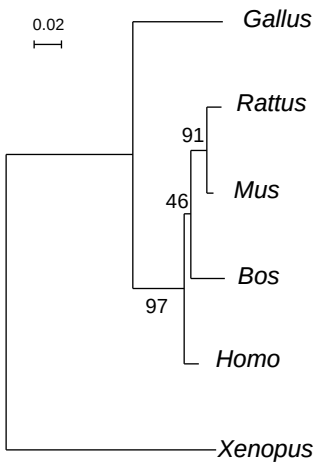
- The tree shape can be deduced from the list of its internal branches.
- Testing the reliability of a tree = testing the reliability of each internal branch.

Bootstrap procedure



The support of each internal branch is expressed as percent of replicates.

"bootstrapped” tree



Bootstrap procedure : properties

- Internal branches supported by ≥ 90% of replicates are considered as statistically significant.
- The bootstrap procedure only detects if sequence length is enough to support a particular node.
- The bootstrap procedure does not help determining if the tree-building method is good. A wrong tree can have 100 % bootstrap support for all its branches!

Gene tree ***vs.*** Species tree

- The evolutionary history of genes reflects that of species that carry them, except if :
  - horizontal transfer = gene transfer between species (*e.g.* bacteria, mitochondria)
  - Gene duplication : orthology/ paralogy

Orthology / Paralogy

speciation

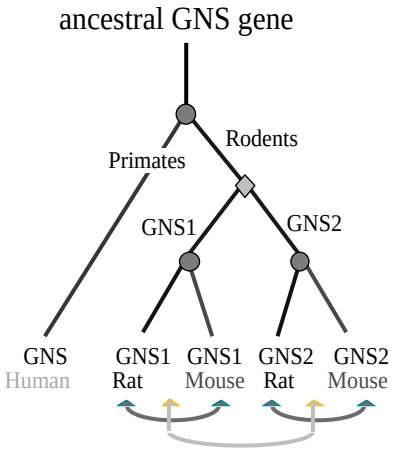
duplication

Homology : two genes are homologous iff they have a common ancestor.

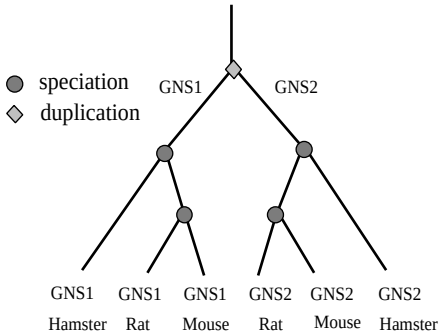
Orthology : two genes are orthologous iff they diverged following a speciation event.

Paralogy : two genes are paralogous iff they diverged following a duplication event.

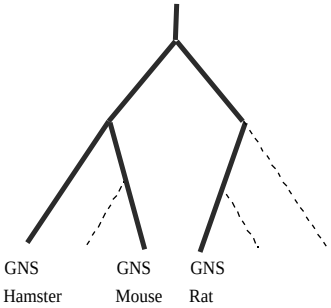
Orthology ≠ functional equivalence



Reconstruction of species phylogeny: artefacts due to paralogy



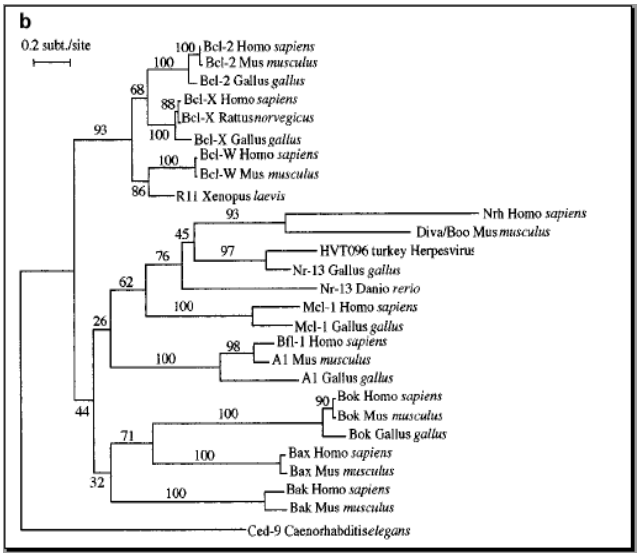
true tree



tree obtained with a partial sampling of homologous genes

!! Gene loss can occur during evolution : even with complete genome sequences it may be difficult to detect paralogy !!

Exploring the Bcl-2 family of inhibitors of apoptosis



Phylogenetic tree of the Bcl-2 family derived from the NJ method applied to PAM evolutionary distances (94 homologous sites).

The tree suggests human NRH, mouse Diva, chicken Nr-13, and *Danio* Nr-13 to be orthologous genes.

The tree also suggests the 2 mammalian genes have evolved much faster than other family members.

Aouacheria et al. (20001) Oncogene 20:5846

WWW resources for molecular phylogeny (1)

- Compilations
  - A list of sites and resources:
    `http://www.ucmp.berkeley.edu/subway/phylogen.html`
  - An extensive list of phylogeny programs
    `http://evolution.genetics.washington.edu/phylip/software.html`
- Databases of rRNA sequences and associated software
  - The rRNA WWW Server - Antwerp, Belgium.
    `http://rrna.uia.ac.be`
  - The Ribosomal Database Project – Michigan State University
    `http://rdp.cme.msu.edu/html/`

WWW resources for molecular phylogeny (2)

- Database similarity searches (Blast) :

`http://www.ncbi.nlm.nih.gov/BLAST/`
`http://www.infobiogen.fr/services/menuserv.html`
`http://bioweb.pasteur.fr/seqanal/blast/intro-fr.html`
`http://pbil.univ-lyon1.fr/BLAST/blast.html`

- Multiple sequence alignment

⇨ ClustalX : multiple sequence alignment with a graphical interface (for all types of computers).
`http://www.ebi.ac.uk/FTP/index.html` and go to 'software'

⇨ Web interface to ClustalW algorithm for proteins:
`http://pbil.univ-lyon1.fr/ and press "clustal"`

WWW resources for molecular phylogeny (3)

- Sequence alignment editor

⇨ SEAVIEW : for windows and unix
`http://pbil.univ-lyon1.fr/software/seaview.html`

- Programs for molecular phylogeny

⇨ PHYLIP : an extensive package of programs for all platforms
`http://evolution.genetics.washington.edu/phylip.html`
⇨ CLUSTALX : beyond alignment, it also performs NJ
⇨ PAUP* : a very performing commercial package
`http://paup.csit.fsu.edu/index.html`
⇨ PHYLO_WIN : a graphical interface, for unix only
`http://pbil.univ-lyon1.fr/software/phylowin.html`
⇨ WWW-interface at Institut Pasteur, Paris
`http://bioweb.pasteur.fr/seqanal/phylogeny`

WWW resources for molecular phylogeny (4)

- Tree drawing
  NJPLOT (for all platforms)
  `http://pbil.univ-lyon1.fr/software/njplot.html`
- Lecture notes of molecular systematics
  `http://www.bioinf.org/molsys/lectures.html`

WWW resources for molecular phylogeny (5)

- Books
  - Laboratory techniques
    `Molecular Systematics (2nd edition), Hillis, Moritz & Mable eds.; Sinauer, 1996.`
  - Molecular evolution
    `Fundamentals of molecular evolution (2nd edition); Graur & Li; Sinauer, 2000.`
  - Evolution in general
    `Evolution (2nd edition); M. Ridley; Blackwell, 1996.`