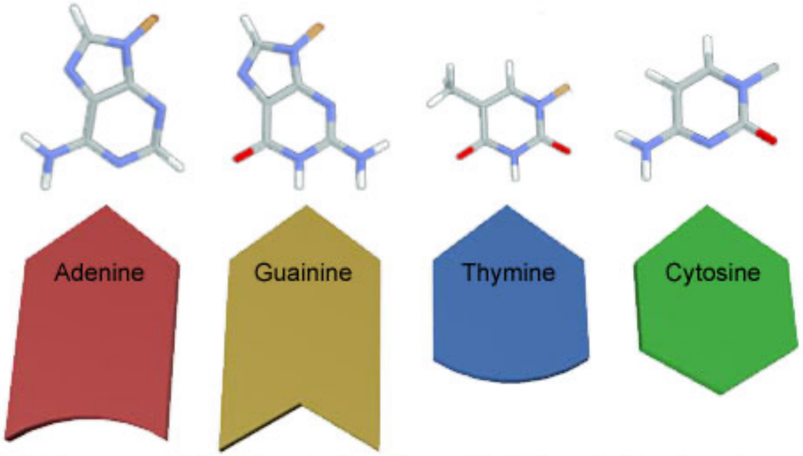
BIOR93 Module 5

Understanding trees and modelling of DNA substitution

DNA as a source of information

- **DNA has four characters**

Figure B-3: The Four Nitrogenous Bases



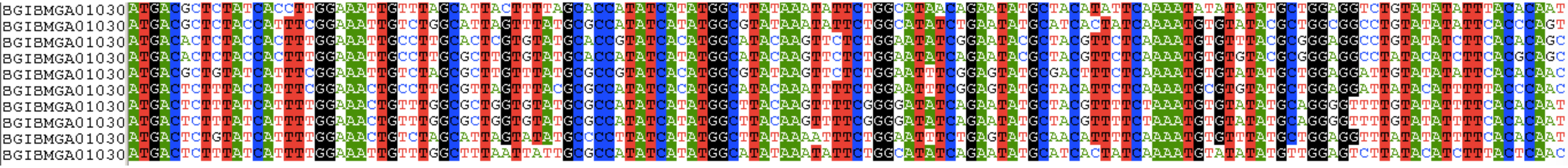
Purines

Pyrimidines

Each base has a distinct shape that can be used to distinguish it form the others. 3D representations of the four bases are shown, with the corresponding chemical structures drawn above.

Homology: Definition

- **Homology: similarity that is the result of inheritance from a common ancestor - identification and analysis of homologies is central to phylogenetic systematics**
- **An alignment is a hypothesis of positional homology between bases/amino acids**



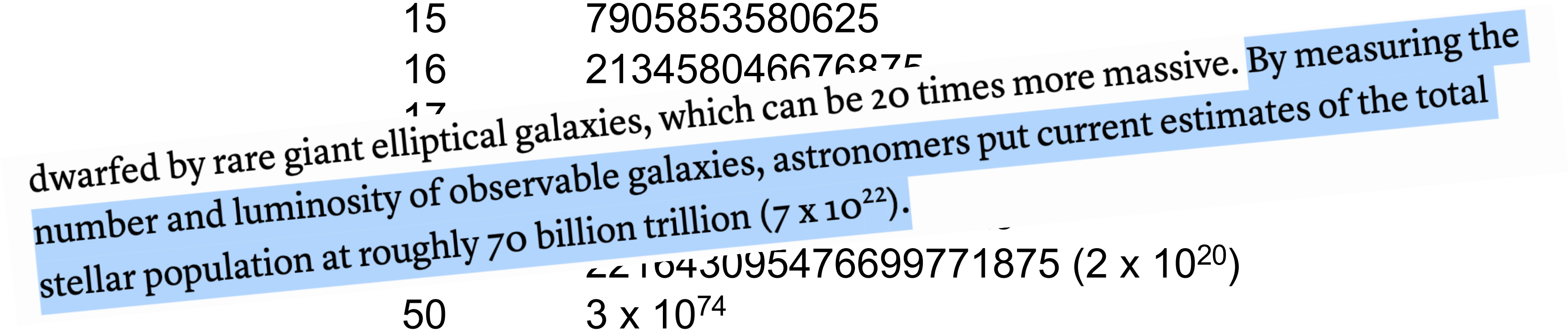
```
BGIBMGA01030 ATGACGCTCTATCACCTTGGAAATTGTTTAGCATTACTTTTAGCACCATATCATATGGCTTATAAATATTCTGGCATAACAGAATATGCTACATATTCAAAATATATATATGCTGGAGGTCTGTATATATTTACACAAT
BGIBMGA01030 ATGACGCTCTACCATTTCGGAAATTGTCTGGCATTAGTTTATGCGCCATATCATATGGCGTATAAATATTCTGGCATATCTGAATATGCATCACTATCAAAATGTGTATACGCTGGCGGCCTGTATATATTCACCCAGT
BGIBMGA01030 ATGACACTCTACCACTTTGGAAATTGCCTTGCACTCGTGTATGCACCGTATCACATGGCATACAAGTTCTCTGGAATATCGGAATACGCTACGTTCTCAAAATGTGTTTACGCGGGAGGCCTGTATATCTTCACACAGC
BGIBMGA01030 ATGACACTCTACCACTTTGGAAATTGCCTTGCGCTTGTGTATGCACCATATCATATGGCATACAAGTTCTCTGGAATATCAGAATACGCTACGTTCTCAAAATGTGTGTACGCGGGAGGCCTATACATCTTCACGCAGC
BGIBMGA01030 ATGACGCTGTATCATTTCGGAAATTGTCTAGCGCTTGTTTATGCGCCGTATCACATGGCGTATAAGTTCTCTGGAATTTCGGAGTATGCGACTTTCTCAAAATGTGTATATGCTGGAGGATTGTATATATTCACACAAC
BGIBMGA01030 ATGACTCTTTACCATTTCGGAAACTGCCTTGCGTTAGTTTACGCGCCATATCACATGGCATACAAATTTTCTGGAATTTCGGAGTATGCTACATTCTCAAAATGCGTGTATGCTGGAGGATTATACATTTTTACCCAAC
BGIBMGA01030 ATGACTCTTTATCATTTTGGAAACTGTTTGGCGCTGGTGTATGCGCCATATCATATGGCTTACAAGTTTTCGGGGATATCAGAATATGCTACGTTTTCTAAATGTGTATATGCAGGGGTTTTGTATATTTTCACACAAT
BGIBMGA01030 ATGACTCTTTATCATTTTGGAAACTGTTTGGCGCTGGTGTATGCGCCATATCATATGGCTTACAAGTTTTCGGGGATATCAGAATATGCTACGTTTTCTAAATGTGTATATGCAGGGGTTTTGTATATTTTCACACAAT
BGIBMGA01030 ATGACTCTGTATCATTTTGGAAACTGTCTAGCATTAGTATATGCCCCTTATCATATGGCTTATAAAATTTCTGGAATTTCTGAGTATGCAACATTTTCAAAATGTGTTTATGCTGGAGGTTTATATATTTTCACACAAT
BGIBMGA01030 ATGACTCTTTATCATTTTGGAAATTGTTTGGCTTTAATTATTGCGCCATATCATATGGCATATAAATATTCTGGCATATCAGAATATGCATCACTATCAAAATGTATATATGTTGGAGTCTTATACATCTTTACTCAAC
```

The Tree

Finding the optimal trees

Numbers of possible trees for N taxa

1	1
2	1
3	1
4	3
5	15
6	105
7	945
8	10395
9	135135
10	2027025
11	34459425
12	654729075
13	13749310575
14	316234143225
15	7905853580625
16	213458046676875
17	
	221643095476699771875 (2 x 10^{20})
50	3 x 10^{74}

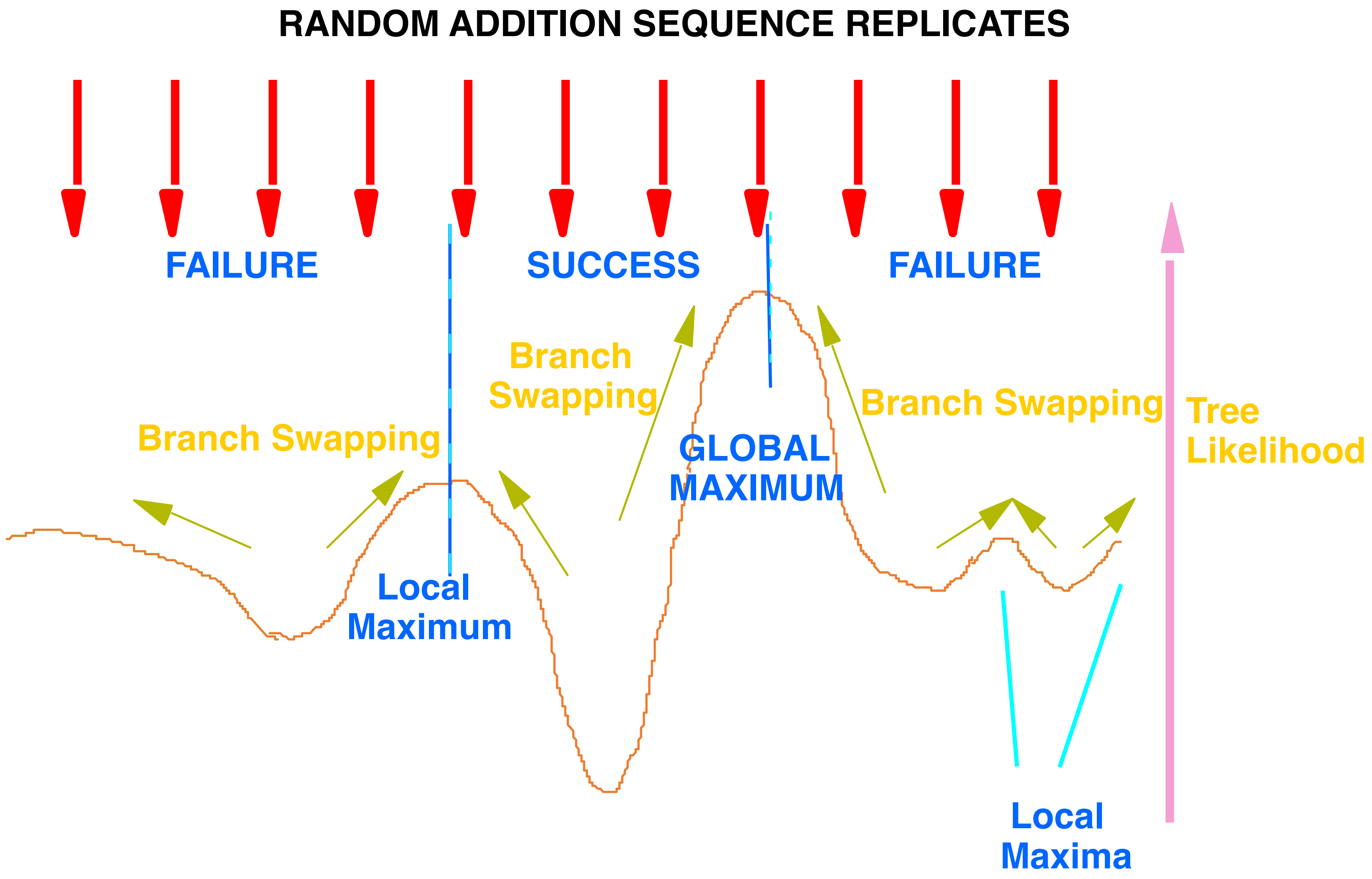


dwarfed by rare giant elliptical galaxies, which can be 20 times more massive. By measuring the number and luminosity of observable galaxies, astronomers put current estimates of the total stellar population at roughly 70 billion trillion (7 x 10^{22}).

How can we find the most optimal tree?

https://skyandtelescope.org/astronomy-resources/how-many-stars-are-there/

Tree space may be populated by local optima and islands of optimal trees



Finding optimal trees - exact solutions

- **Exact solutions can only be used for small numbers of taxa**

- **Exhaustive search examines all possible trees**
- **Branch and bound does not examine all trees, but will find optimal tree(s)**

- **Typically used for problems with 10–20 taxa**

Finding optimal trees - heuristics

- **The number of possible trees increases faster than exponentially with the number of taxa making exhaustive searches impractical for many data sets (an NP-complete problem)**
- **Heuristic methods are used to search tree space for optimal trees by building or selecting an initial tree and swapping branches to search for better ones**
- **The trees found are not guaranteed to be optimal - they are best guesses**

Finding optimal trees - heuristics

- **Stepwise addition**

Asis - the order in the data matrix

Closest - starts with shortest 3-taxon tree, adds taxa in order that produces the least increase in tree length (greedy heuristic)

Simple - the first taxon in the matrix is taken as a reference - taxa are added to it in the order of their decreasing similarity to the reference

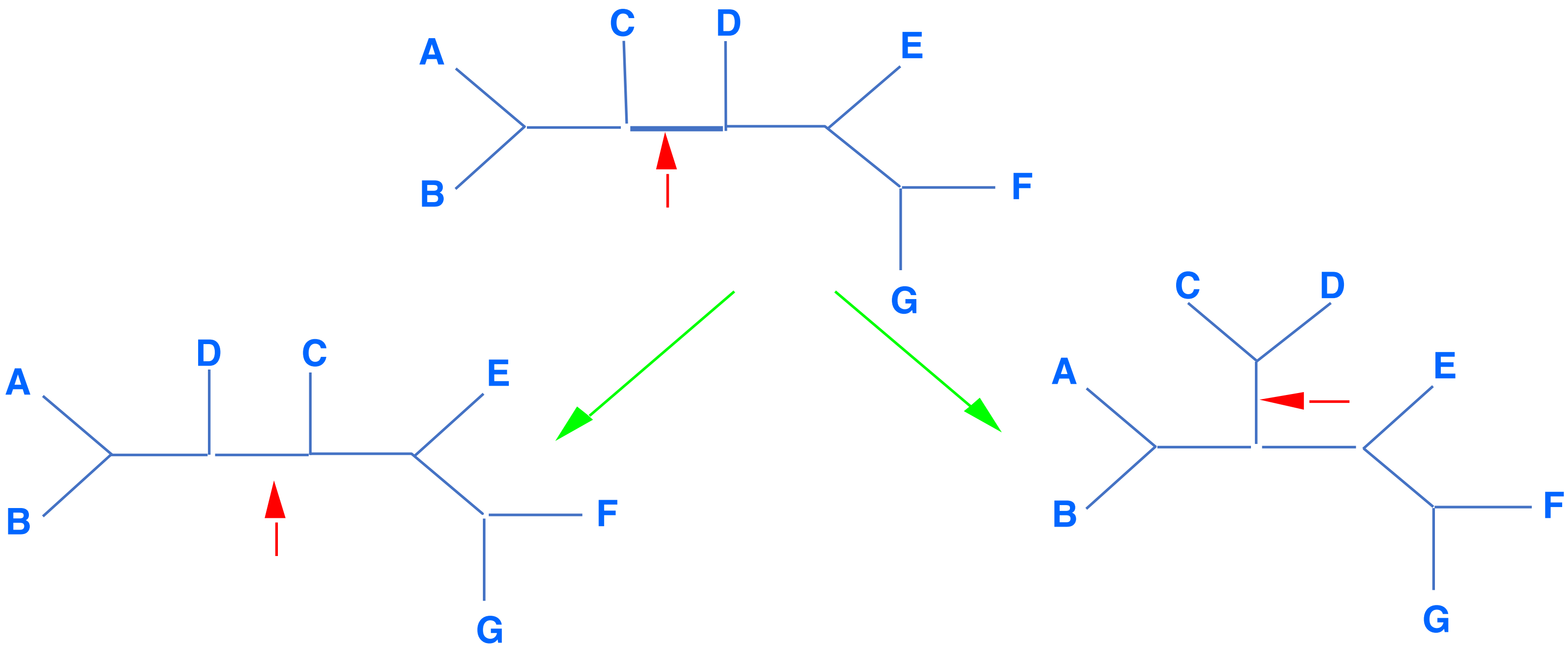
Random - taxa are added in a random sequence, many different sequences can be used

Finding optimal trees – branch swapping

- **Nearest neighbor interchange (NNI)**
- **Subtree pruning and regrafting (SPR)**
- **Tree bisection and reconnection (TBR)**

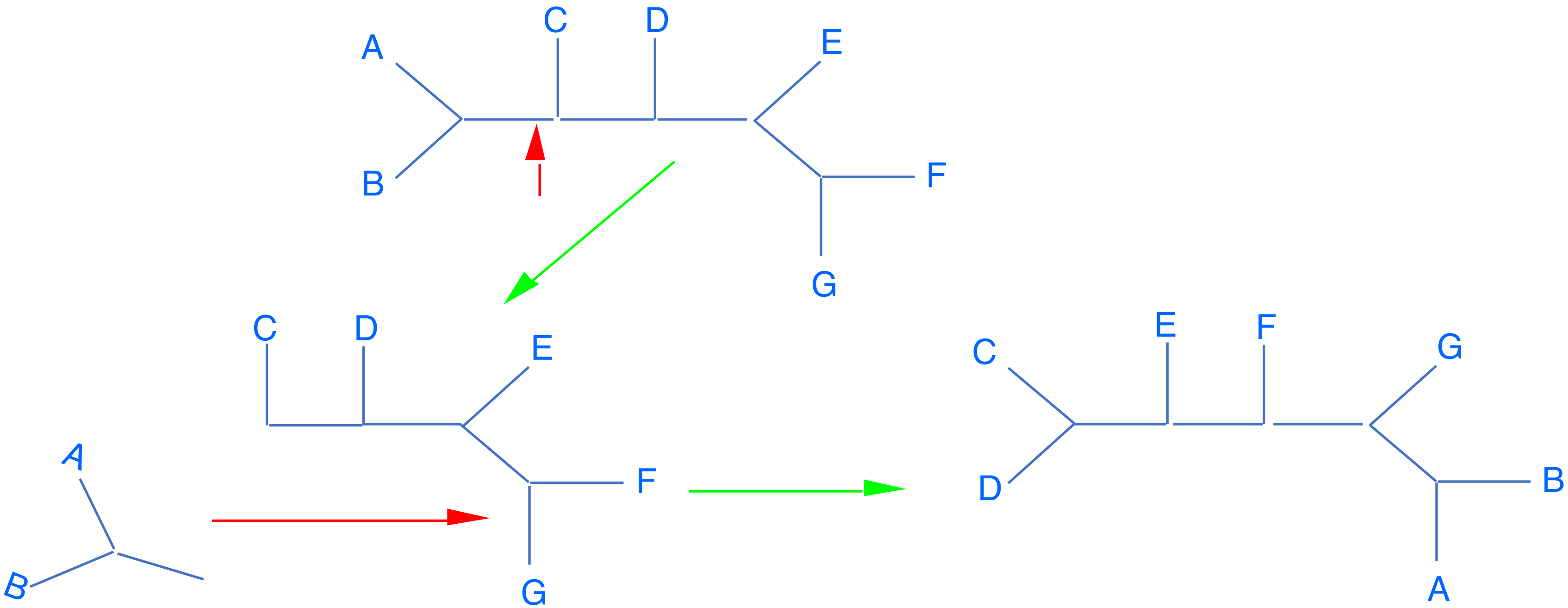
Moving through treespace

Nearest neighbor interchange (NNI)



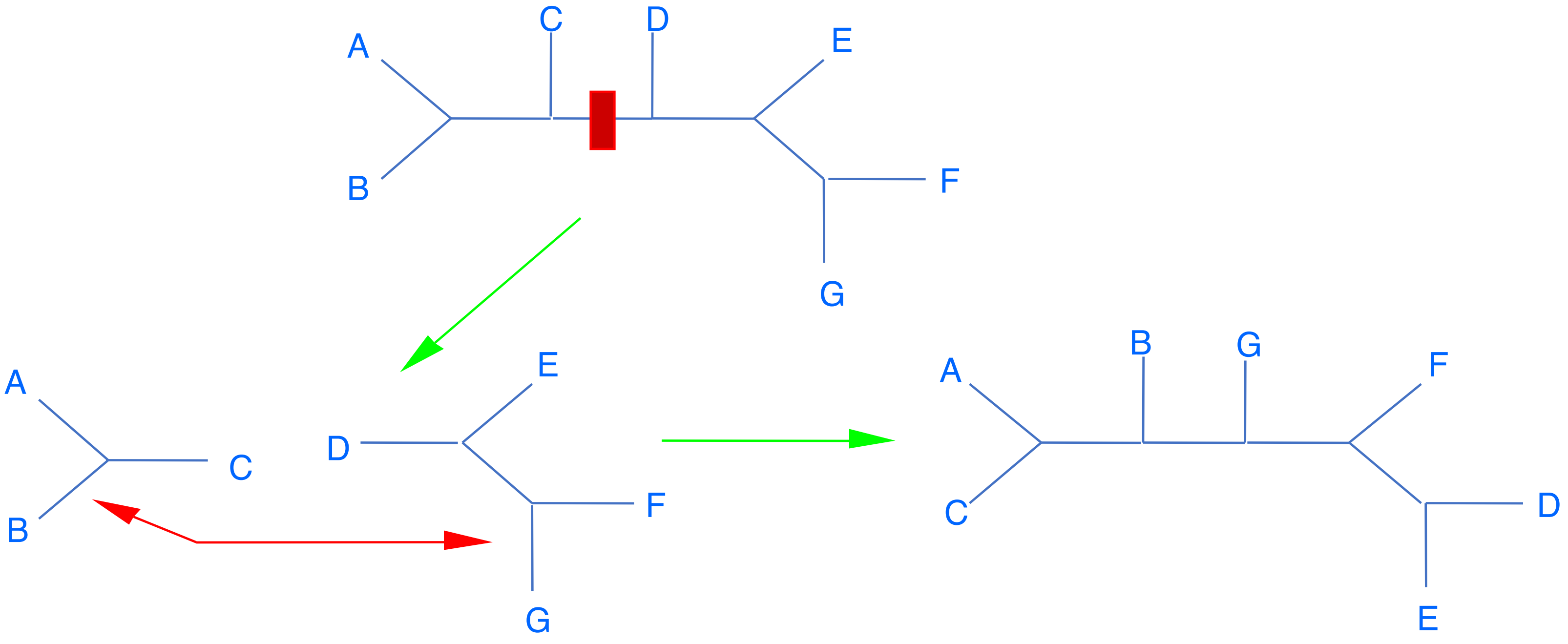
Moving through treespace

Subtree pruning and regrafting (SPR)



Moving through treespace

Tree bisection and reconnection (TBR)



Consensus methods

Multiple optimal trees

- **Many methods can yield multiple equally optimal trees**
- **We can further select among these trees with additional criteria, but**
- **Typically, relationships common to all the optimal trees are summarised with *consensus trees***

Consensus methods

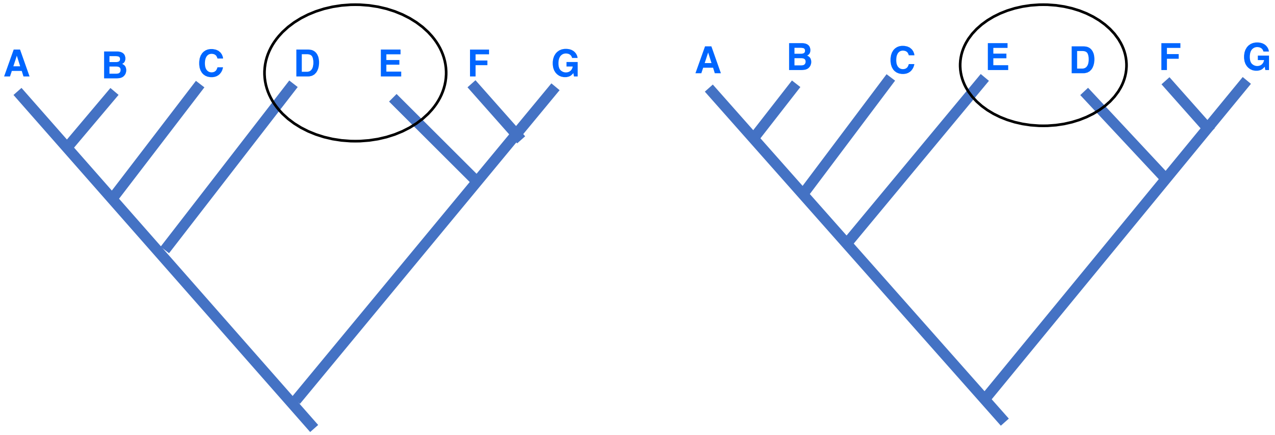
- **A consensus tree is a summary of the agreement among a set of fundamental trees**
- **There are many consensus methods that differ in:**
  1. **the kind of agreement**
  2. **the level of agreement**
- **Consensus methods can be used with multiple trees from a single analysis or from multiple analyses**

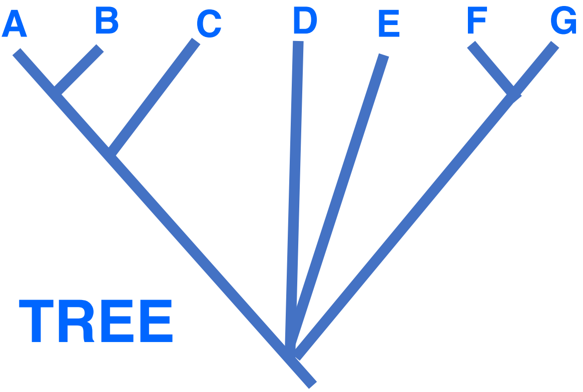
Strict consensus methods

- **Strict consensus methods require agreement across all the fundamental trees**
- **They show only those relationships that are unambiguously supported by the parsimonious interpretation of the data**
- **The commonest method (*strict component consensus*) focuses on clades/components/full splits**
- **This method produces a consensus tree that includes all and only those full splits found in all the fundamental trees**
- **Other relationships (those in which the fundamental trees disagree) are shown as unresolved polytomies**

Strict consensus methods

TWO FUNDAMENTAL TREES





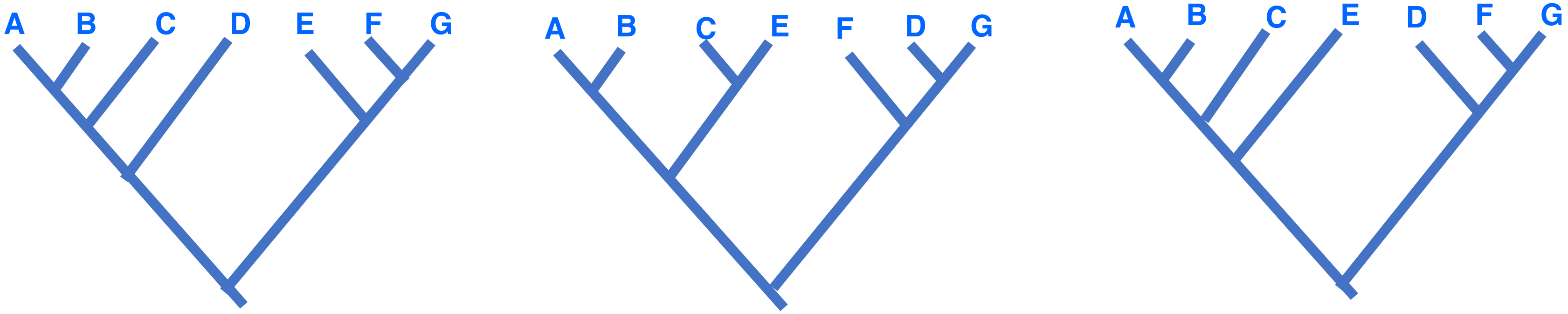
STRICT CONSENSUS TREE

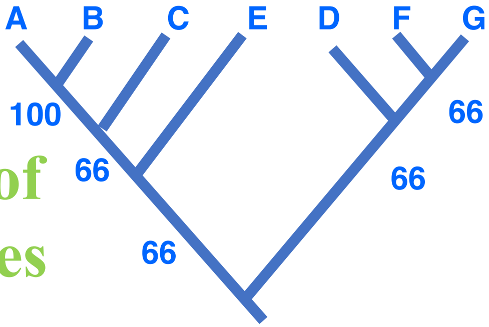
Majority-rule consensus methods

- **Majority-rule consensus methods require agreement across a majority of the fundamental trees**
- **May include relationships that are not supported by the most parsimonious interpretation of the data**
- **The commonest method focuses on clades/components/full splits**
- **This method produces a consensus tree that includes all and only those full splits found in a majority (>50%) of the fundamental trees**
- **Other relationships are shown as unresolved polytomies**
- **Of particular use in bootstrapping**

Majority rule consensus

THREE FUNDAMENTAL TREES



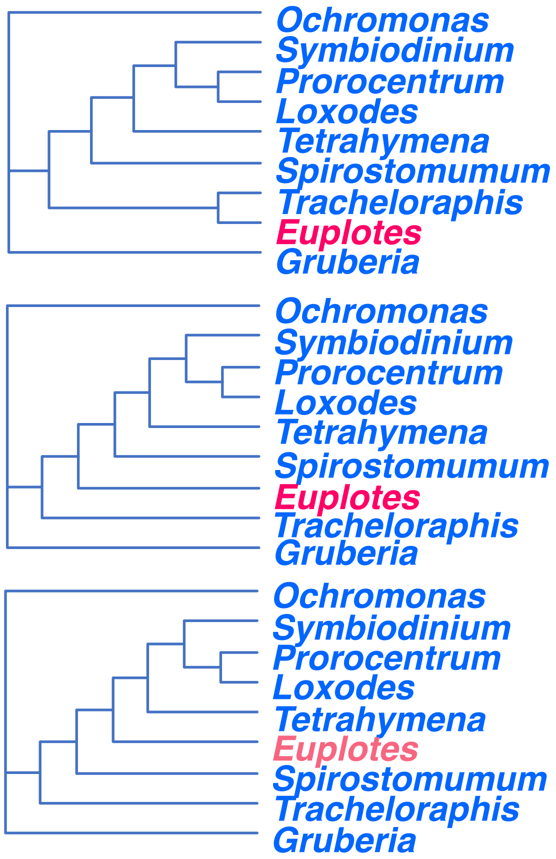


Numbers indicate frequency of clades in the fundamental trees

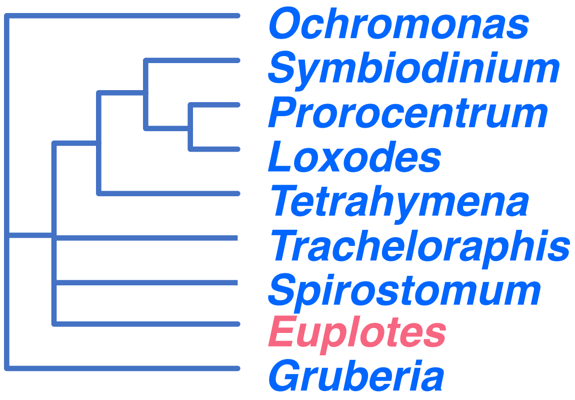
MAJORITY-RULE CONSENSUS TREE

Consensus methods

Three fundamental trees

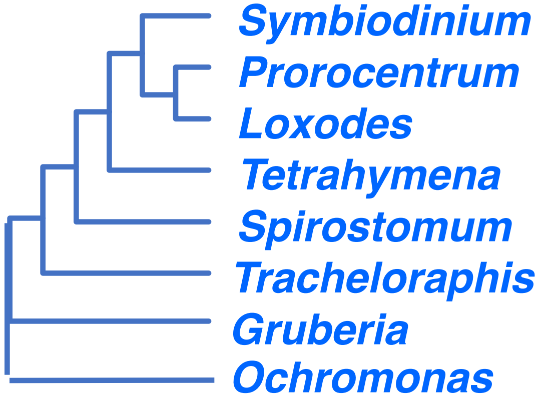


Strict (component)

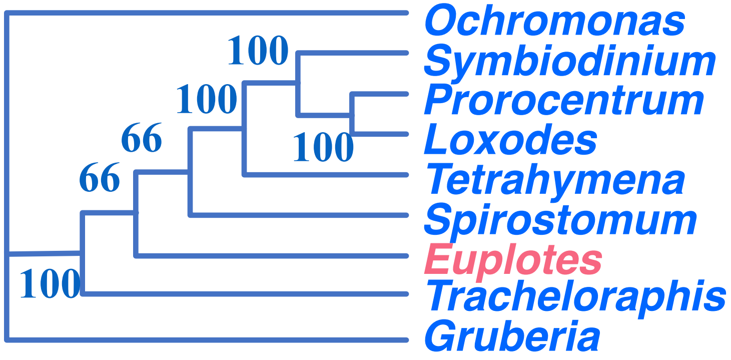


Strict reduced cladistic

Euplotes excluded



Majority-rule



Consensus methods – use

- **Currently majority-rule methods mainly used**
  - **bootstrapping**
  - **Bayesian methods**
- **Reduced methods can be useful to identify problem taxa**
  - **E.g. RogueNaRok**
- **Strict methods mainly used in parsimony analyses**
  - **rarely used with molecular data**

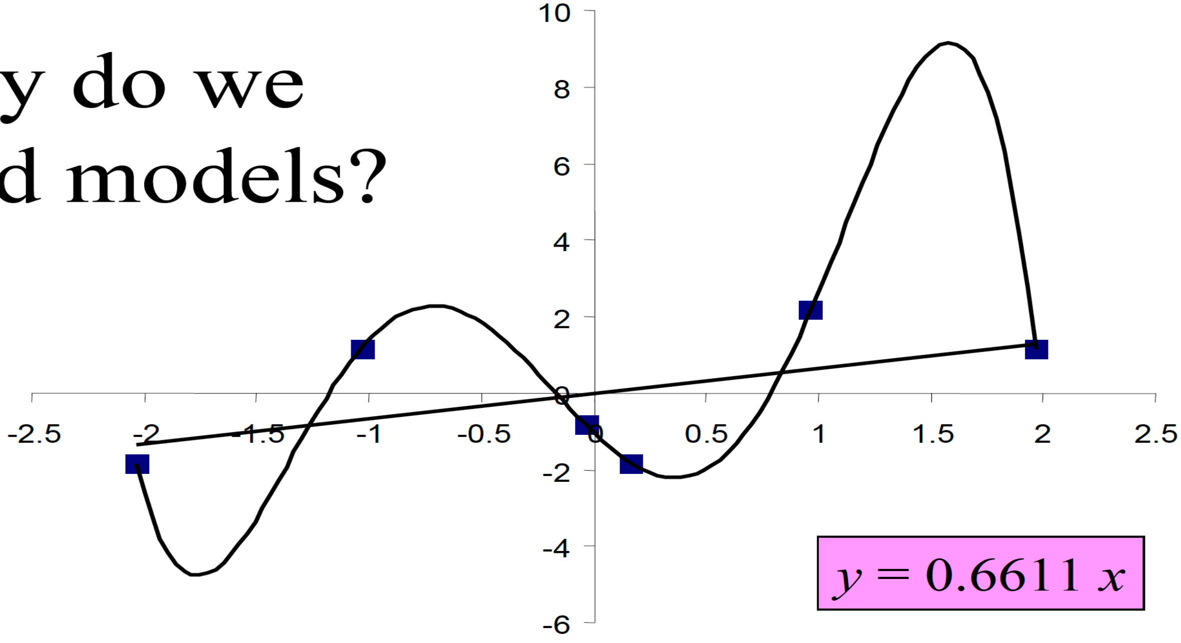
Take home messages

- **Statements of homology are the basis of phylogenetics**
- **Alignments of molecular sequences are very strong statements of positional homology**
- **Finding an optimal tree is not a trivial task**

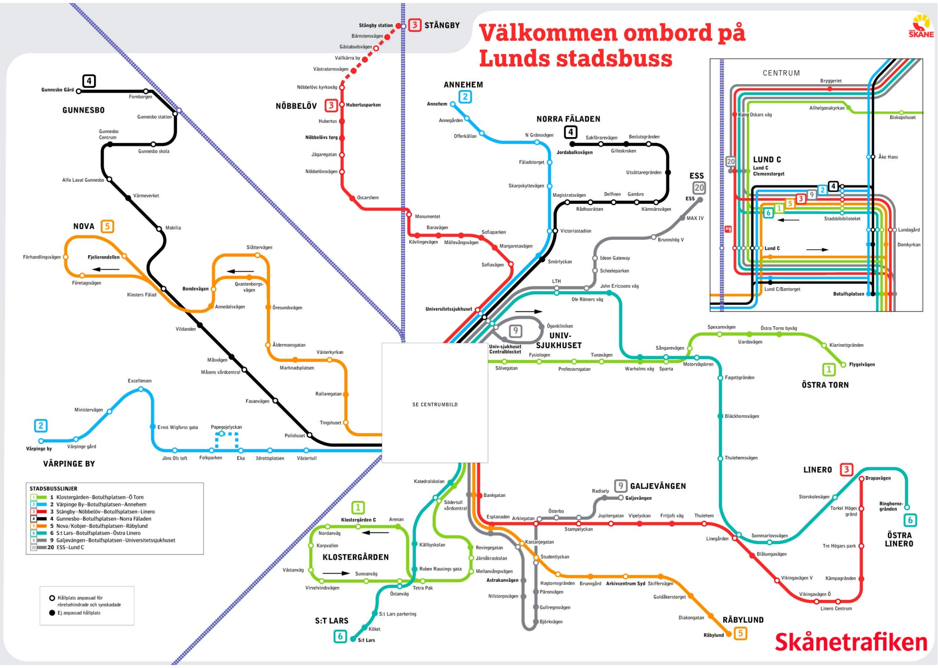
Modelling DNA Sequence Evolution

$$y = -1.5972\, x^5 + 23.167\, x^4 - 126.18\, x^3 + 319.17\, x^2 - 369.22\, x + 155.67$$

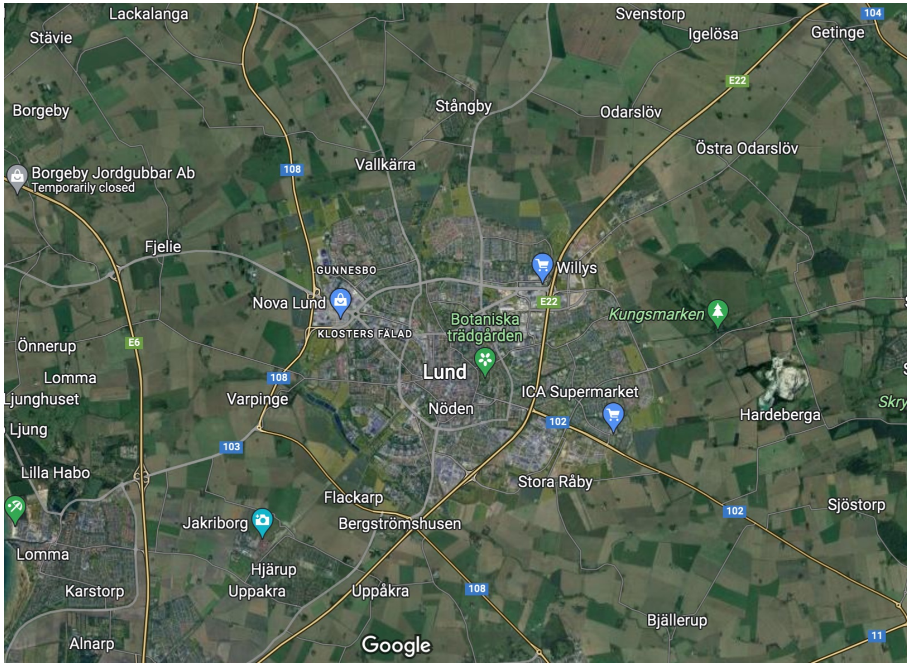
Why do we need models?



$$y = 0.6611\, x$$


Välkommen ombord på Lunds stadsbuss
Skånetrafiken
SKÅNE
STÅNGBY
NÖBBELÖV
ANNEHEM
GUNNESBO
NORRA FÄLADEN
ESS
NOVA
CENTRUM
LUND C
UNIV-SJUKHUSET
ÖSTRA TORN
SE CENTRUMBILD
VÄRPINGE BY
LINERO
GALJEVÄNGEN
ÖSTRA LINERO
KLOSTERGÅRDEN
S:T LARS
RÅBYLUND
STADSBUSSLINJER
1 Klostergården–Botulfsplatsen–Ö Torn
2 Värpinge By–Botulfsplatsen–Annehem
3 Stångby–Nöbbelöv–Botulfsplatsen–Linero
4 Gunnesbo–Botulfsplatsen–Norra Fäladen
5 Nova/Kobjer–Botulfsplatsen–Råbylund
6 S:t Lars–Botulfsplatsen–Östra Linero
9 Galjevången–Botulfsplatsen–Universitetssjukhuset
20 ESS–Lund C
Hållplats anpassad för rörelsehindrade och synskadade
Ej anpassad hållplats

A simplified map of bus routes in Lund



A realistic map of Lund

- **Which one would you use to get around Lund by bus?**

Models: an overview

- **In general, models help us predict the future based on our observations**
- **With more parameters, models have a better fit to the data (observations)**
- **Underparamaterized models: poor fit to the observed data**
- **Overparameterized models: poor prediction of future observations**
- **Choosing best models based on different criteria**
  - **Likelihood ratio tests, AIC, BIC, Bayes factors**

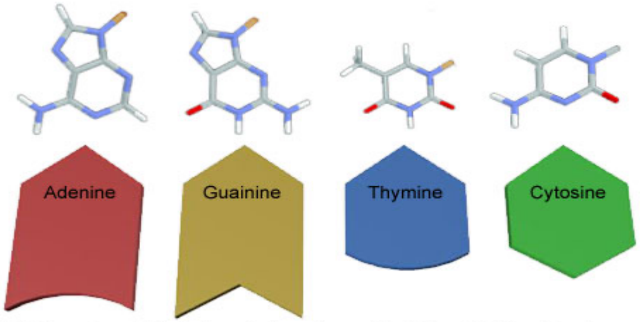
What do we model in DNA sequence evolution?

- **Nucleotide substitutions**
  - **The rate at which each nucleotide is replaced by each alternative nucleotide**

What is the challenge?

- **DNA has only four characters**

Figure B-3: The Four Nitrogenous Bases



Each base has a distinct shape that can be used to distinguish it form the others. 3D representations of the four bases are shown, with the corresponding chemical structures drawn above.

Saturation in sequence data

- **Saturation is due to multiple changes at the same site subsequent to lineage splitting**
- **Models of evolution attempt to infer the missing information through correcting for “multiple hits”**
- **Most data will contain some fast evolving sites which are potentially saturated (e.g. in proteins often codon position 3)**
- **In severe cases the data become essentially random and all information about relationships can be lost**

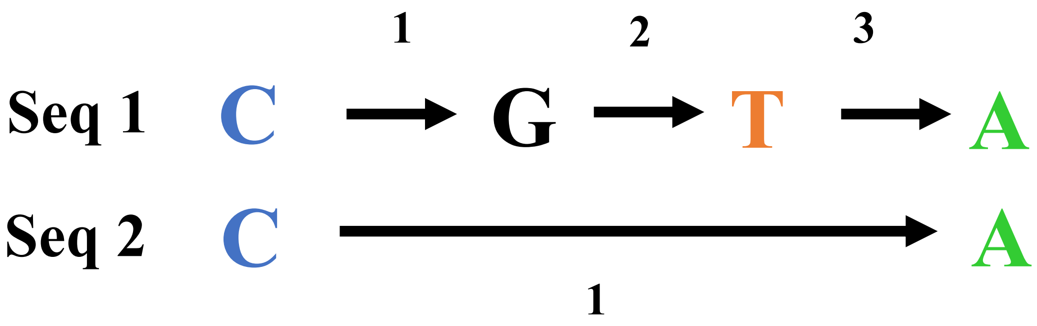
Multiple changes at a single site
- hidden changes

```
Seq 1   AGCGAG
Seq 2   GCGGAC
```

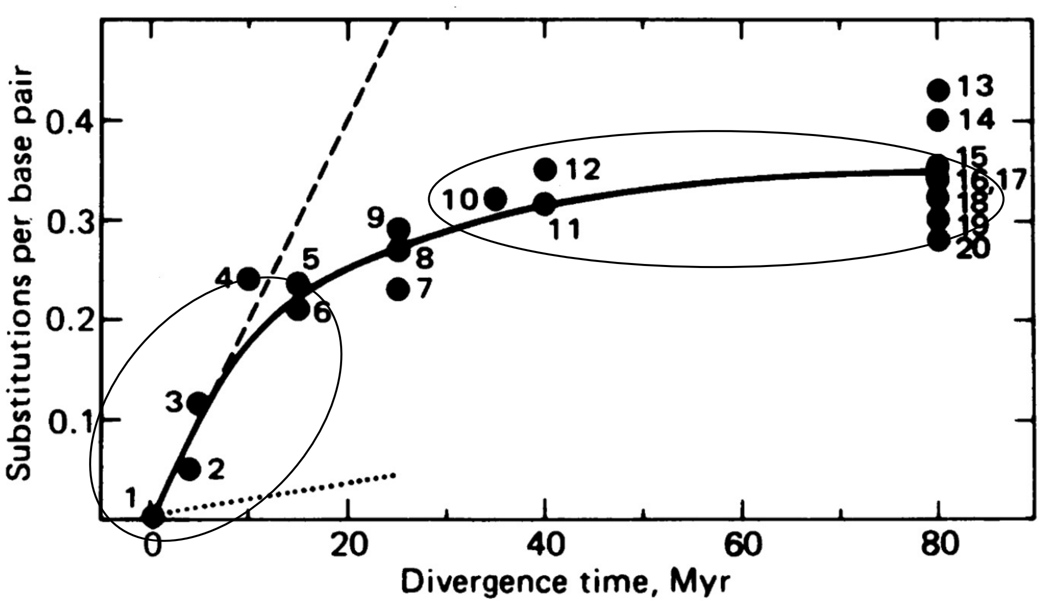
Multiple changes at a single site
- hidden changes

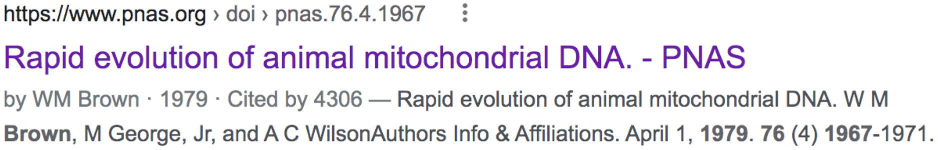
```
Ancest GGCGCG
Seq 1  AGCGAG
Seq 2  GCGGAC
```

Number of changes



“Multiple hits” or saturation




https://www.pnas.org › doi › pnas.76.4.1967

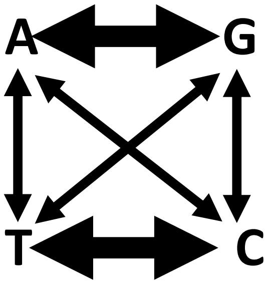
Rapid evolution of animal mitochondrial DNA. - PNAS

by WM Brown · 1979 · Cited by 4306 — Rapid evolution of animal mitochondrial DNA. W M **Brown**, M George, Jr, and A C WilsonAuthors Info & Affiliations. April 1, **1979**. **76** (4) **1967**-1971.

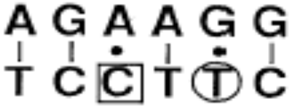
Brown et al. 1979. PNAS 76:1967

Substitution types

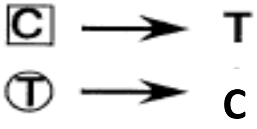
- **Purines: A, G**
- **Pyrimidines: C, T**
- **Transversions**
  - **Pu --> Pyr**
  - **Pyr --> Pu**
- **Transitions – more common**
  - **Pu --> Pu**
  - **Pyr --> Pyr**



Pur - Pyr mispairs lead to transitions

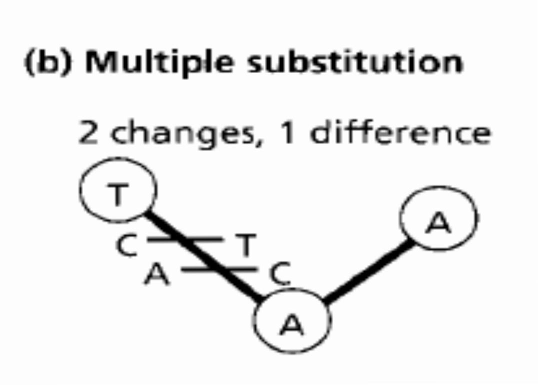


In next round of replication



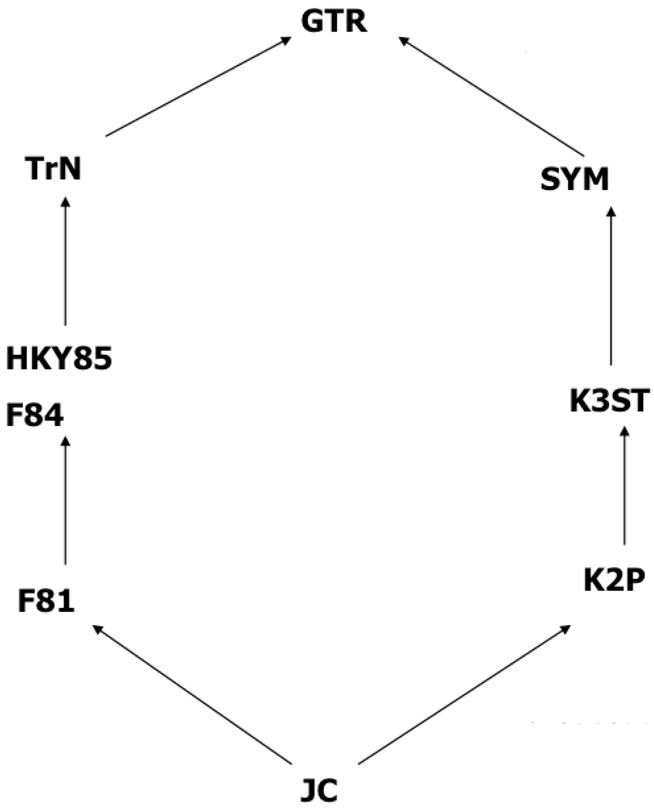
Saturation in sequence data:

- **Saturation is due to multiple substitutions at the same site subsequent to lineage splitting**
- **Models of evolution attempt to infer the missing information through correcting for "multiple hits"**
- **Most data will contain some fast evolving sites which are potentially saturated**
  - **e.g. in protein-coding genes codon position 3**



Saturation in sequence data (cont.)

- **In severe cases the data become essentially random and all information about relationships can be lost**
- **Probabilistic models of sequence evolution are used to calculate expected distances**



GTR

TrN

SYM

HKY85

F84

K3ST

F81

K2P

JC

Modelling nucleotide substitutions

- **These dynamics can be modelled over a tree and they are incoporated into distance methods, maximum likelihood, and Bayesian inference**
- **Models incorporate information about the rates at which each nucleotide is replaced by each alternative nucleotide**
    - **For DNA this can be expressed as a 4 x 4 rate matrix (known as the Q matrix)**
- **Other model parameters may include:**
    - **Site by site rate variation (aka among-site rate variation – ASRV)**

Corrections for multiple substitutions:

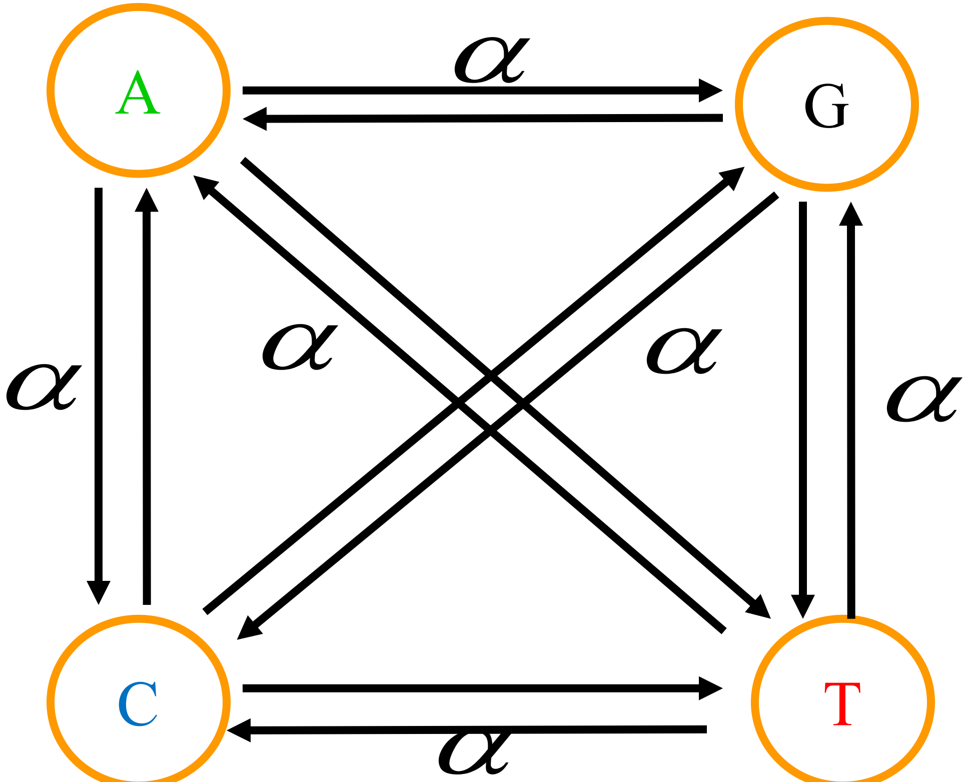
First DNA subtitution model

Jukes & Cantor (1969) assumptions:

1. **A = T = G = C No nucleotide bias**
2. **Every base changes to every other base with equal probability (no TS/TV bias)**
3. **All sites change with the same probability (no ASRV - among-site rate variation)**

Also: probability of substitution & base composition remains constant over time/across lineages

Jukes-Cantor model



t = time

- α = the rate of substitution (α changes from A to G every t)
- The rate of substitution for each nucleotide is 3α
- In t steps there will be 3αt changes

The Q matrix

To

From

$$\begin{array}{ccccc} & A & C & G & T \\ A & -3\alpha & \alpha & \alpha & \alpha \\ C & \alpha & -3\alpha & \alpha & \alpha \\ G & \alpha & \alpha & -3\alpha & \alpha \\ T & \alpha & \alpha & \alpha & -3\alpha \end{array}$$

The Jukes-Cantor model: the simplest model

$$\begin{array}{ccccc} & A & C & G & T \\ A & -3\alpha & \alpha & \alpha & \alpha \\ C & \alpha & -3\alpha & \alpha & \alpha \\ G & \alpha & \alpha & -3\alpha & \alpha \\ T & \alpha & \alpha & \alpha & -3\alpha \end{array}$$

JC model: one parameter model

1) It assumes that all bases are equally frequent (p=0.25)
2) It assumes that all sites can change and they do so at the same rate of α

The Jukes-Cantor model: the simplest model

	A	C	G	T
A	—	α	α	α
C	α	—	α	α
G	α	α	—	α
T	α	α	α	—

JC model: one parameter model

1) It assumes that all bases are equally frequent (p=0.25)
2) It assumes that all sites can change and they do so at the same rate of α

Improvements on Jukes-Cantor

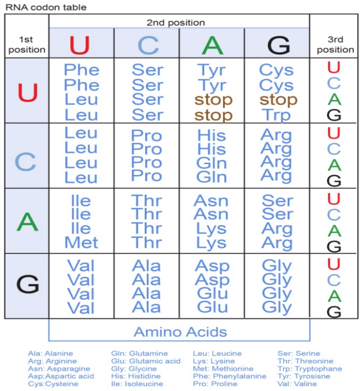
- **Allow base frequencies to be unequal to accommodate e.g. sequences such as these**

AAACCTGGATTTACCGAGATTTAAGCGATATATTGCAATGC

34% A 17% C

29% T 20% G

- **Allow transitions to be more common than transversions, in fact, allow separate estimates of the probability of change of all six possible nucleotide substitutions**
- **Allow the probability of substitution to change along the molecule - ASRV**

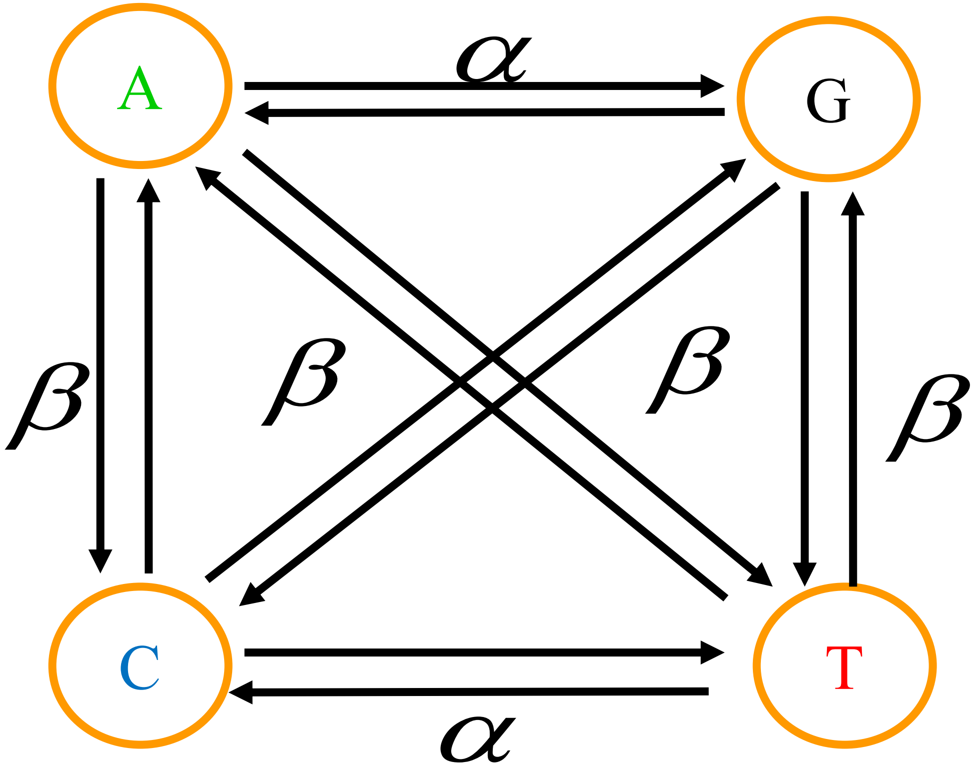


RNA codon table

1st position	2nd position U	 C	 A	 G	3rd position
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr stop stop	Cys Cys stop Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G
	Amino Acids				

Ala: Alanine
Arg: Arginine
Asn: Asparagine
Asp:Aspartic acid
Cys:Cysteine
Gln: Glutamine
Glu: Glutamic acid
Gly: Glycine
His: Histidine
Ile: Isoleucine
Leu: Leucine
Lys: Lysine
Met: Methionine
Phe: Phenylalanine
Pro: Proline
Ser: Serine
Thr: Threonine
Trp: Tryptophane
Tyr: Tyrosisne
Val: Valine

Kimura (1980) model: K2P



α = transitions β = transversions

The Kimura model has 2 parameters

	A	C	G	T
A	—	β	α	β
C	β	—	β	α
G	α	β	—	β
T	β	α	β	—

K2P model is more realistic, but still

1) It assumes that all bases are equally frequent (p=0.25)
2) There are two substitution types (transitions – α and transversions - β

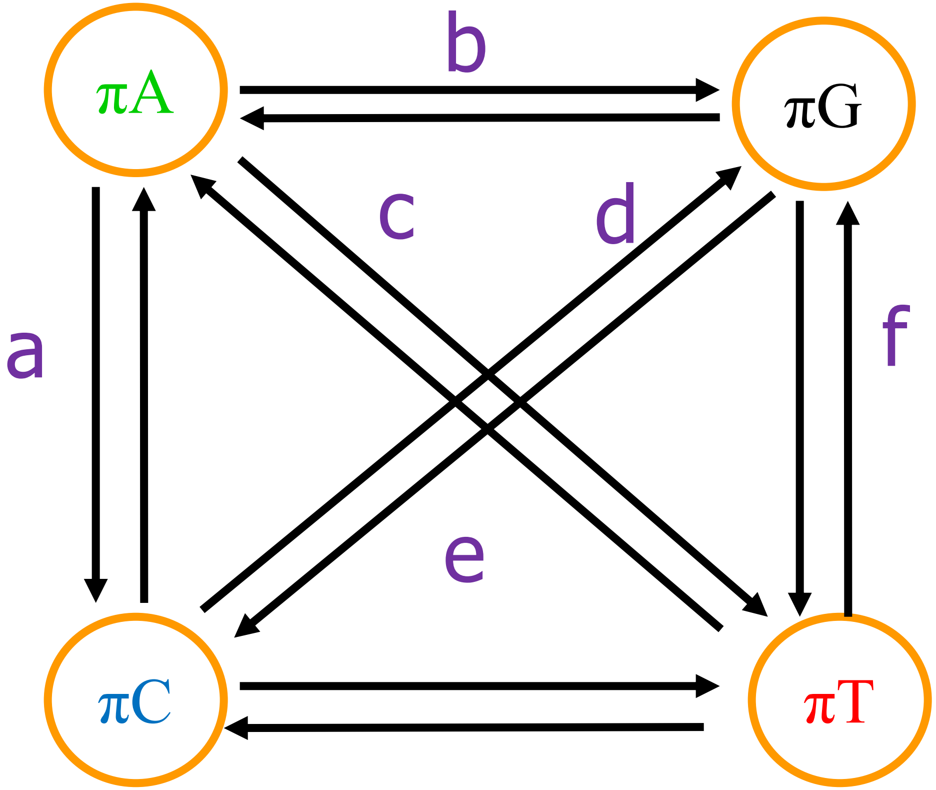
The Hasegawa-Kishino-Yano model

$$\begin{array}{ccccc}
 & A & C & G & T \\
A & - & \pi_C\beta & \pi_G\alpha & \pi_T\beta \\
C & \pi_A\beta & - & \pi_G\beta & \pi_T\alpha \\
G & \pi_A\alpha & \pi_C\beta & - & \pi_T\beta \\
T & \pi_A\beta & \pi_C\alpha & \pi_G\beta & -
\end{array}$$

HKY model:

1) Base frequencies are allowed to vary: πA, πC, πG, πT
2) There are two substitution types (transitions – α and transversions – β)

The General Time-Reversible model



The General Time-Reversible model (GTR)

$$\begin{array}{c|cccc} & \mathrm{A} & \mathrm{C} & \mathrm{G} & \mathrm{T} \\ \hline \mathrm{A} & - & \pi_C a & \pi_G b & \pi_T c \\ \mathrm{C} & \pi_A a & - & \pi_G d & \pi_T e \\ \mathrm{G} & \pi_A b & \pi_C d & - & \pi_T f \\ \mathrm{T} & \pi_A c & \pi_C e & \pi_G f & - \end{array}$$

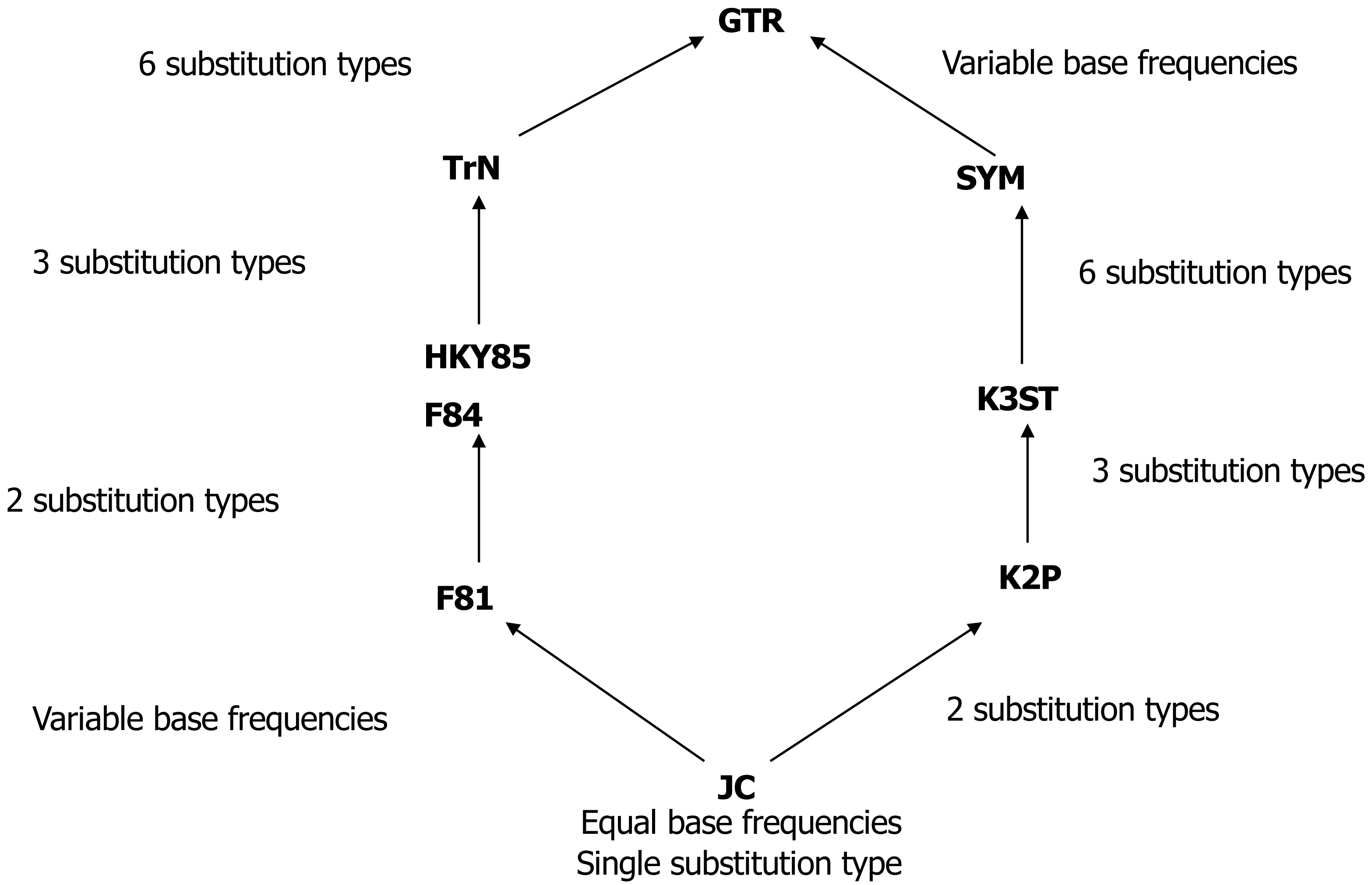
GTR model:

1) Base frequencies are allowed to vary: πA, πC, πG, πT
2) There are six substitution types: a, b, c, d, e, f

The most commonly used models

- **Almost all models used are special cases of one model:**
    - **The general time reversible model - GTR**

ACAGGTGAGGCTCAGCCAATTTGAGCTTTGTCGATAGGT



GTR

6 substitution types

Variable base frequencies

TrN

SYM

3 substitution types

6 substitution types

HKY85

F84

K3ST

3 substitution types

2 substitution types

F81

K2P

Variable base frequencies

2 substitution types

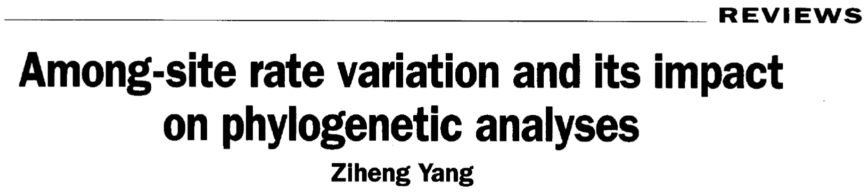
JC

Equal base frequencies

Single substitution type

Modelling among-site rate variation (ASRV)

- **All of the models so far assume that the rate of change is the same for every position in the alignment**
- **Variable vs. invariable sites**
- **Two classes of invariable sites**
  - **Highly restricted "not free to vary"**
  - **not observed to vary but in fact variable**
    - **due to convergence or reversal**
    - **% invariable sites can't be calculated by simple sequence comparison**



REVIEWS

Among-site rate variation and its impact on phylogenetic analyses

Ziheng Yang

https://www.sciencedirect.com › science › article › pii ⋮

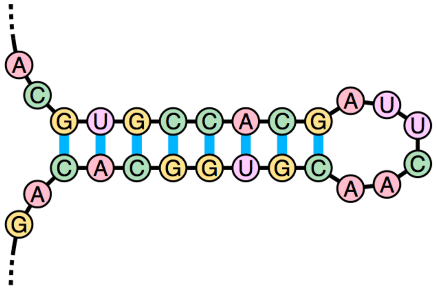
Among-site rate variation and its impact on phylogenetic ...

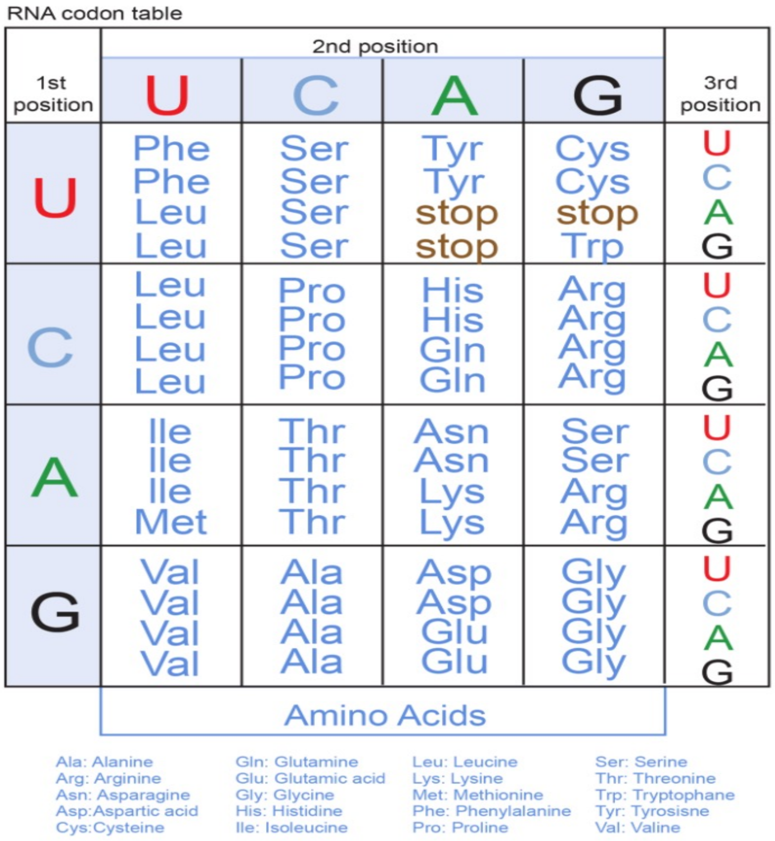
by Z Yang · 1996 · Cited by 1342 — Recent **analyses** show that failure to account for **rate variation** can have drastic **effects**, leading to biased dating of speciation events, biased...

Yang (1996) TREE 11(9): 367–372

Why is modelling ASRV important?

- **Protein-coding genes – 1st, 2nd, 3rd codon positions evolve differently from each other**
- **RNA molecules – stems and loops**
- **Introns vs. exons**




RNA codon table

1st position	2nd position				3rd position
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	stop	stop	A
	Leu	Ser	stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Amino Acids

Ala: Alanine
Arg: Arginine
Asn: Asparagine
Asp:Aspartic acid
Cys:Cysteine

Gln: Glutamine
Glu: Glutamic acid
Gly: Glycine
His: Histidine
Ile: Isoleucine

Leu: Leucine
Lys: Lysine
Met: Methionine
Phe: Phenylalanine
Pro: Proline

Ser: Serine
Thr: Threonine
Trp: Tryptophane
Tyr: Tyrosisne
Val: Valine

Modelling among-site rate variation (ASRV)

- **The most common additional parameters are:**
  - **A correction for the proportion of sites which are invariable (parameter *I*)**
  - **A correction for variable site rates at those sites which can change (parameter gamma, *G*)**
- **All models can be supplemented with these parameters (e.g. GTR+*I*+*G*, HKY+*I*+*G*)**

Modelling among-site rate variation with Gamma distribution

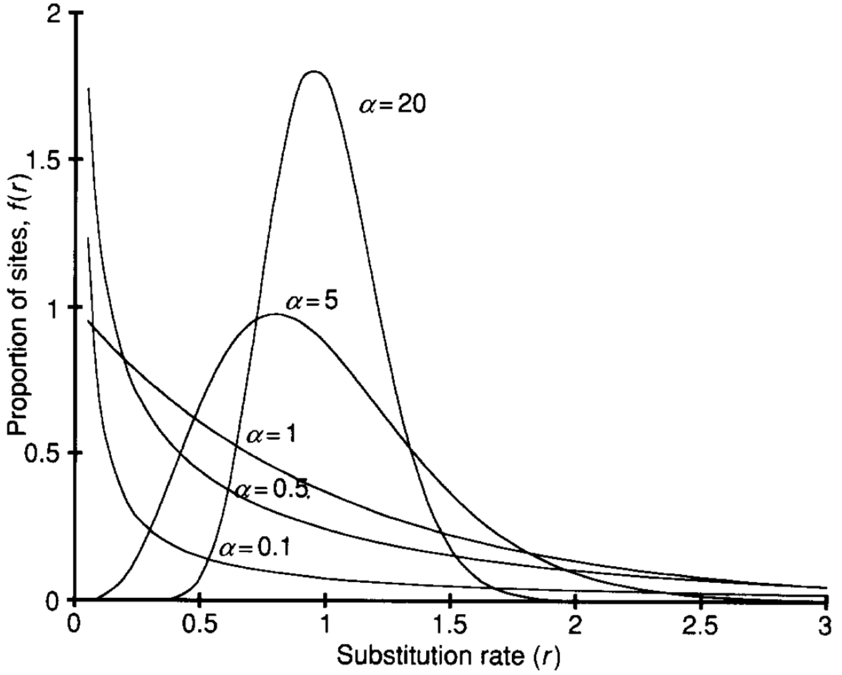


Fig. 1. The density function, $f(r)$, of the gamma distribution of substitution rates at sites (r). The gamma distribution has a shape parameter α and a scale parameter β, with mean α/β and variance α/β^2. Since the rate is a proportional factor,

Gamma distribution:

Relative substitution rates for different α values

Fig. 1 from Yang 1996:

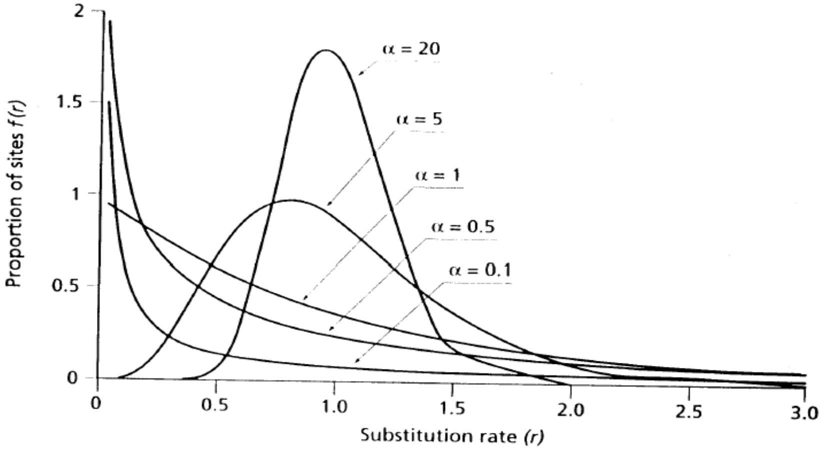
Alpha – the shape parameter of the gamma distribution

Smaller alpha = higher ASRV

Yang (1996) TREE 11(9): 367–372

Another method for modelling ASRV

- **Gamma distribution is always unimodal**
  - **Not necessarily the case in our dataset!**
- **Flexible rate heterogeneity across sites model**
  - **Probability distribution free model so that you can find the distribution that fits your data (FreeRate Model)**
  - **Implemented in IQ-TREE**



Kalyaanamoorthy et al. 2017 (Nature Methods) doi:10.1038/nmeth.4285

Modelling ASRV leads to greater improvement in fit than other parameters

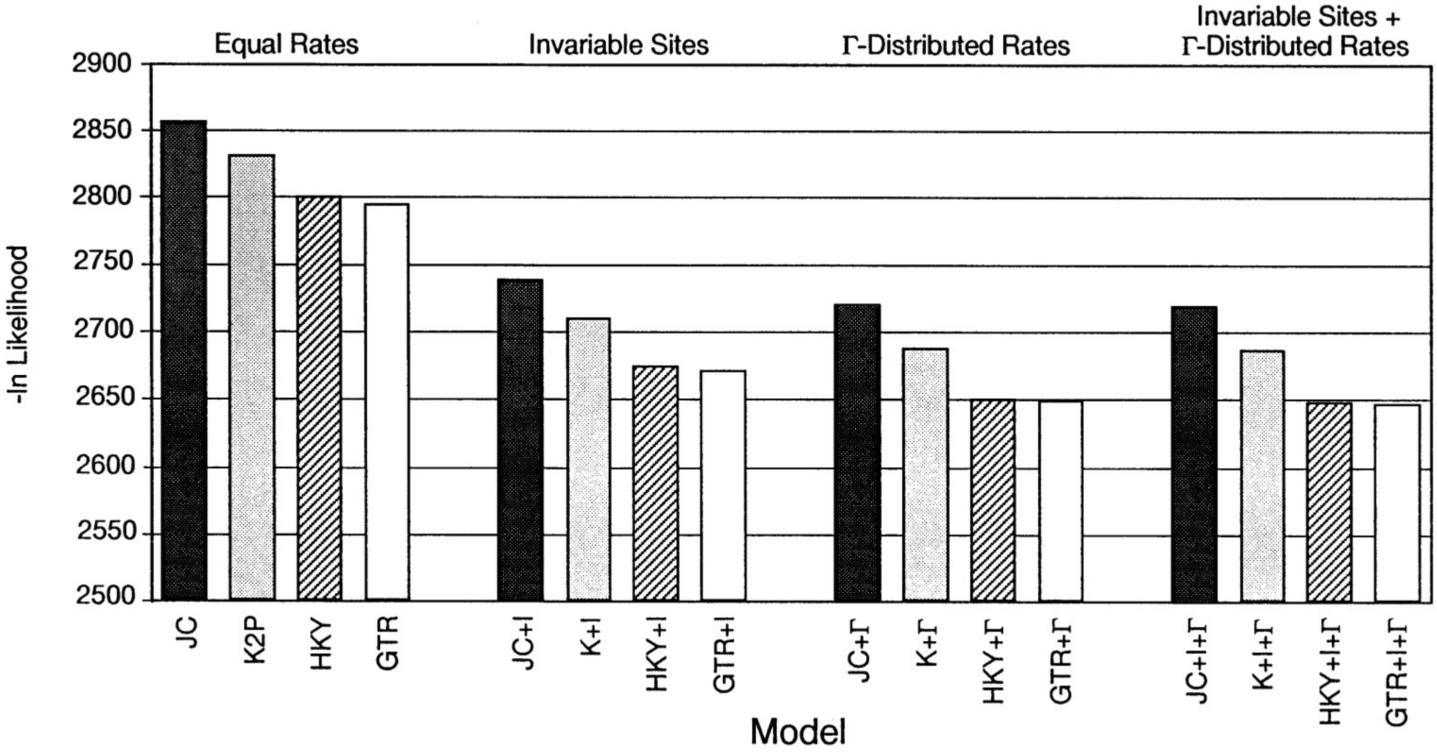


Fig. 4 from Frati et al. 1997. J. Mol. Evol. 44:145-158

Modelling ASRV leads to greater improvement in fit than other parameters

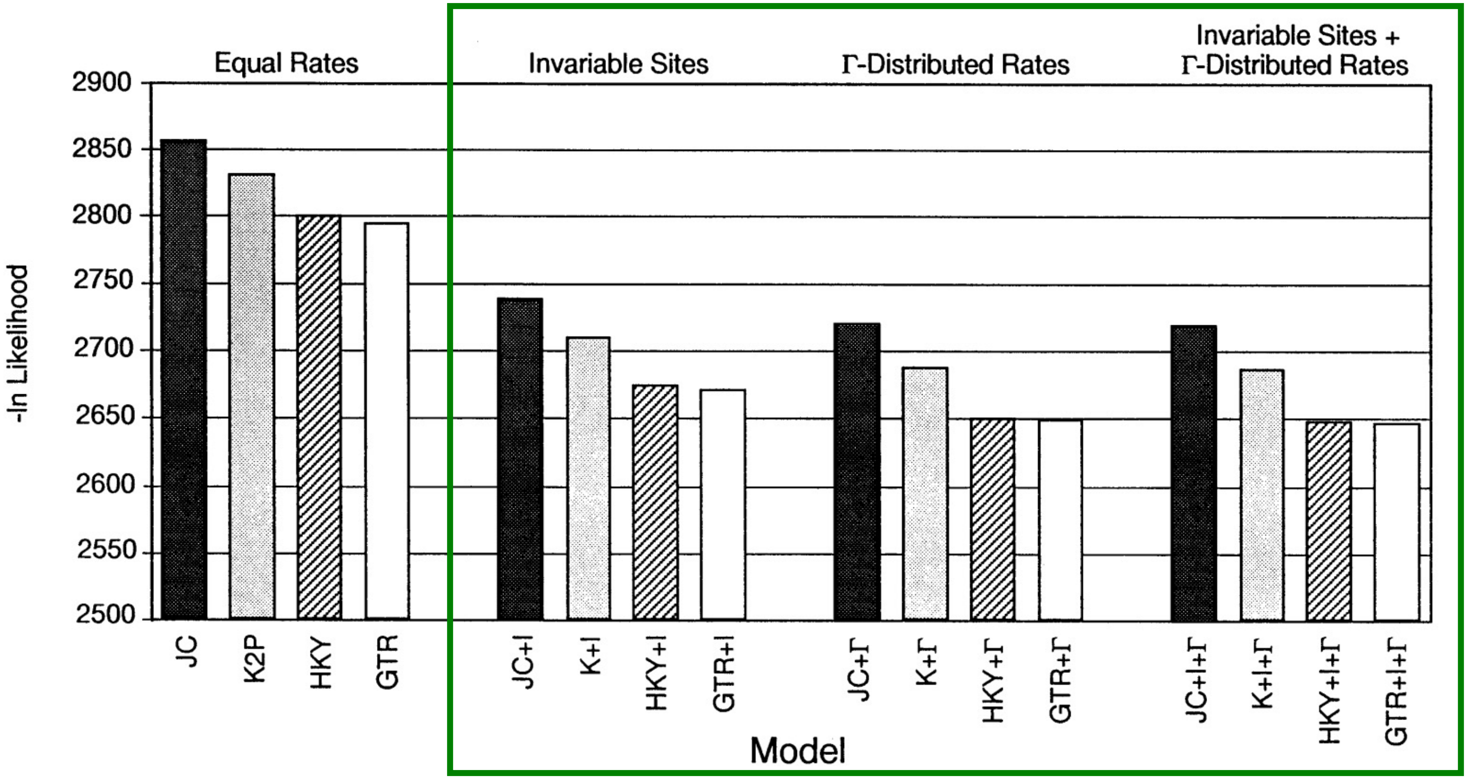
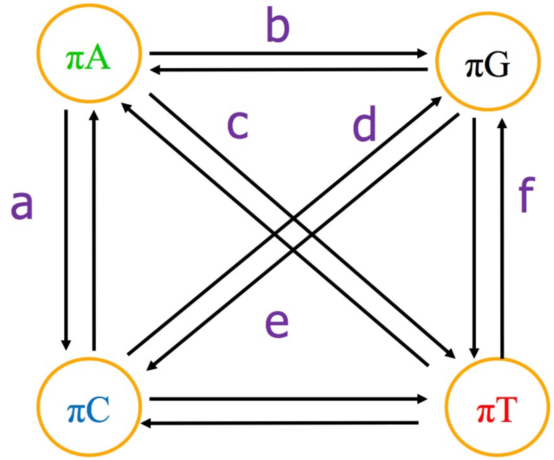
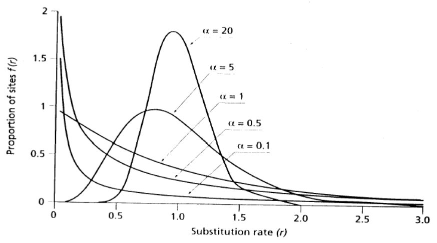


Fig. 4 from Frati et al. 1997. J. Mol. Evol. 44:145-158

Parameters in models of DNA evolution

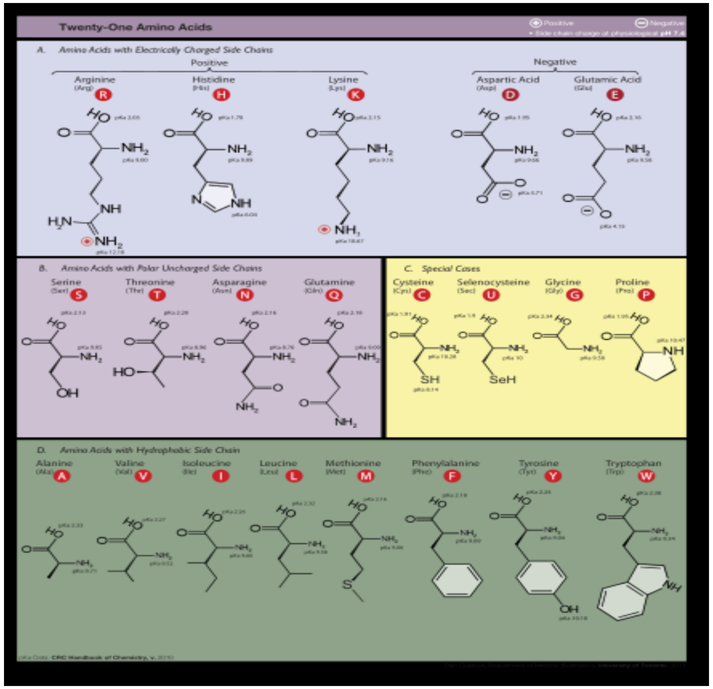
- **Numbers of parameters estimated:**
  - **Base composition**
    - **1 fixed, 3 estimated**
  - **Substitutions**
    - **up to 5; 1 fixed, 5 estimated**
  - **Among-site-rate variation**
    - **Gamma shape parameter = 1 parameter**
    - **Invariant sites = 1 parameter**
    - **Gamma + I = 2 parameters**





Models of amino acid substitution

- **Empirical and mechanistic models**
- **Empirical models: based on empirical AA replacement with matrices from different taxa**
  - **20 amino acids – 20x20 matrix too big for estimation**
  - **Examples: JTT, WAG, LG, MtREV (for mitochondria), Blosum62**
- **Mechanistic models:**
  - **e.g. codon models (61x61 matrix)**
  - **Tend to outperform empirical models BUT**
  - **Computationally very intensive**



Inferring phylogenies: methodological overview

- **Distance methods**
  - **A clustering method using pairwise distances between sequences (e.g. neighbour joining)**
  - **Covered in the assigned reading (chapter from Evolutionary Genetics)**
- **Discrete characters**
  - **Using an optimality criterion to choose the best tree**
    - **Maximum parsimony (Occam’s razor)**
      - **Best explanation is the simplest one (the one that minimizes the number of substitutions)**
      - **Doesn’t perform as well as model-based methods on molecular data**
      - **Still used for morphological characters**
    - **Maximum likelihood**
    - **Bayesian inference**

Distance – disadvantages

- **Prone to systematic errors**
- **Problems with missing data**
- **Generally outperformed by Maximum Likelihood and Bayesian methods in choosing the correct tree in computer simulations**
  - **See e.g. Ogden & Rosenberg (2006) Multiple Sequence Alignment Accuracy and Phylogenetic Inference. Syst. Biol. 55(2): 314–328 (DOI: 10.1080/10635150500541730)**