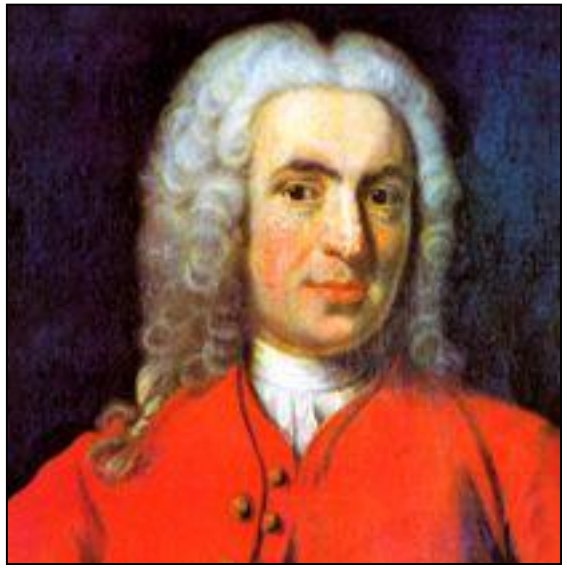
Introduction to Bioinformatics, 2024

Lecture by Bent Petersen
Slides by Anders Gorm Pedersen and Rasmus Wernersson

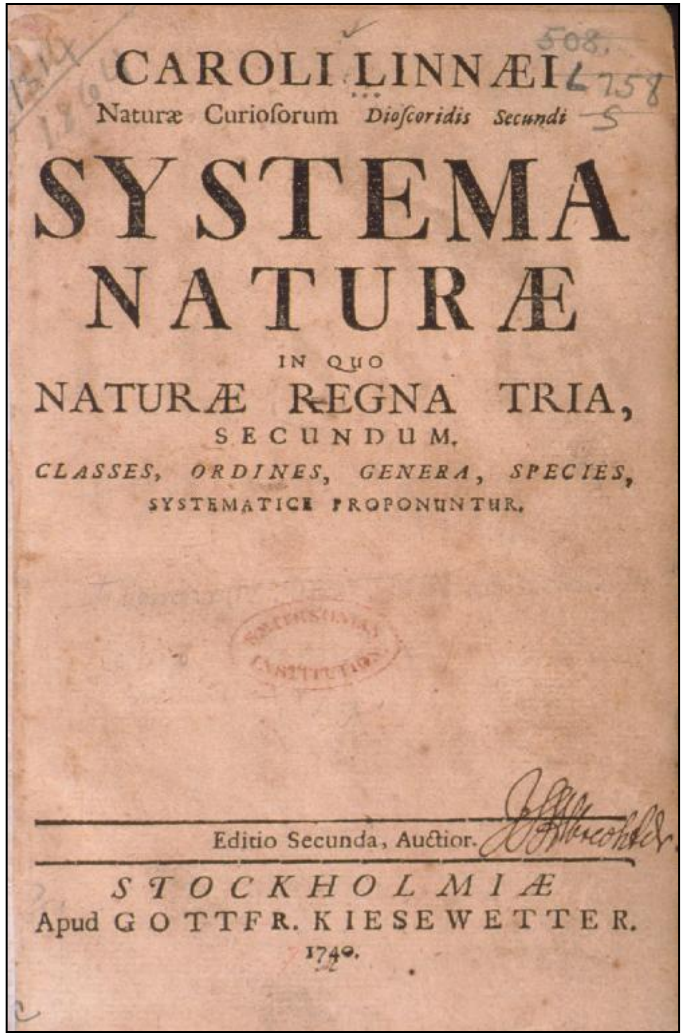
On evolution and sequences

Classification: Linnaeus

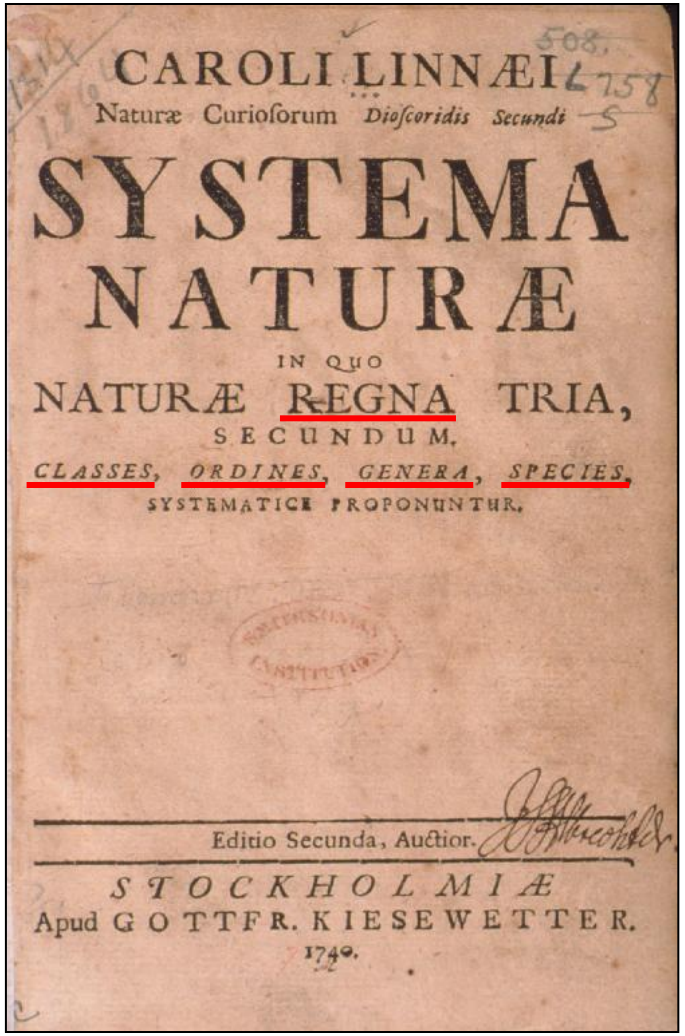


Carl Linnaeus

1707-1778



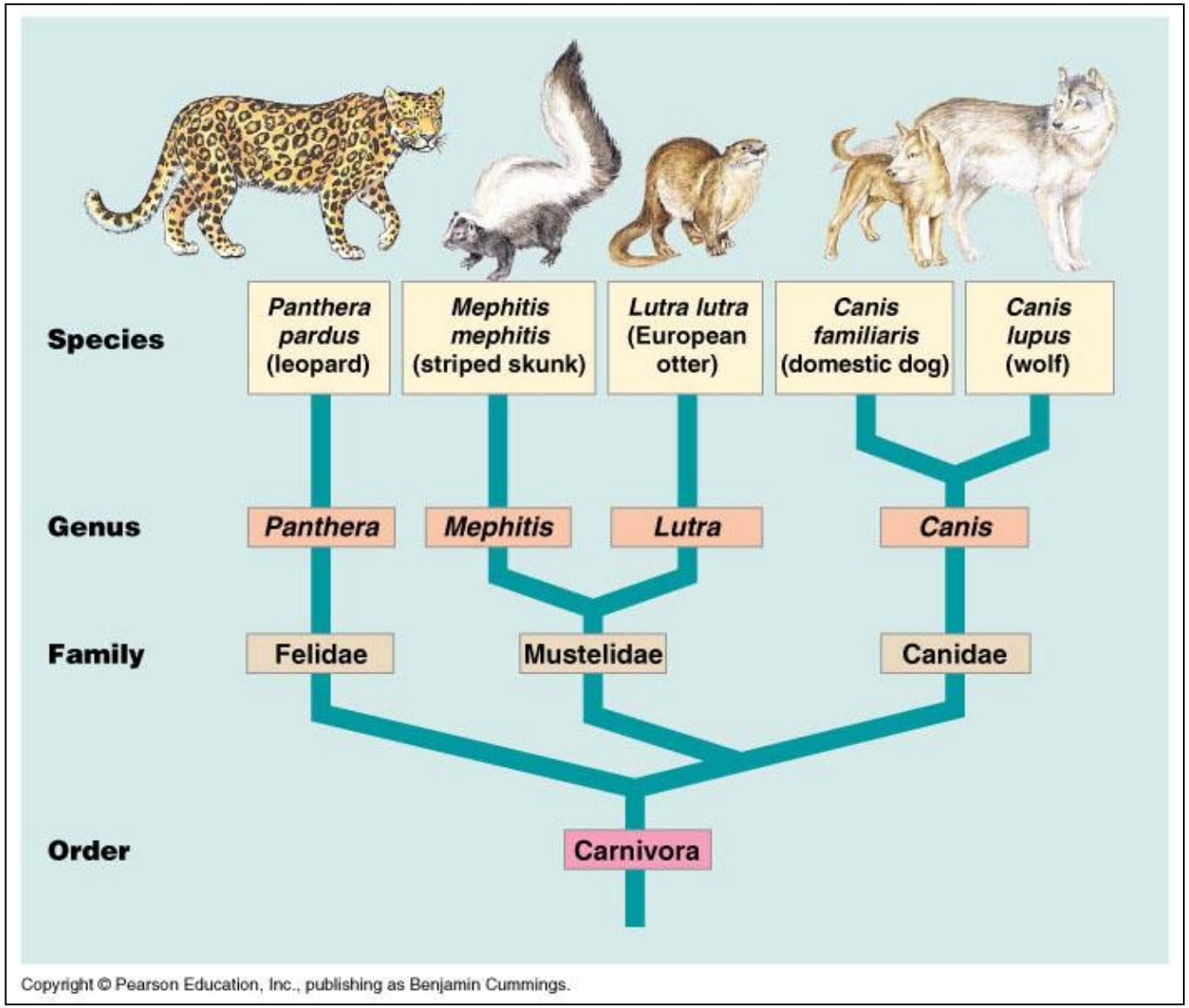
Classification: Linnaeus



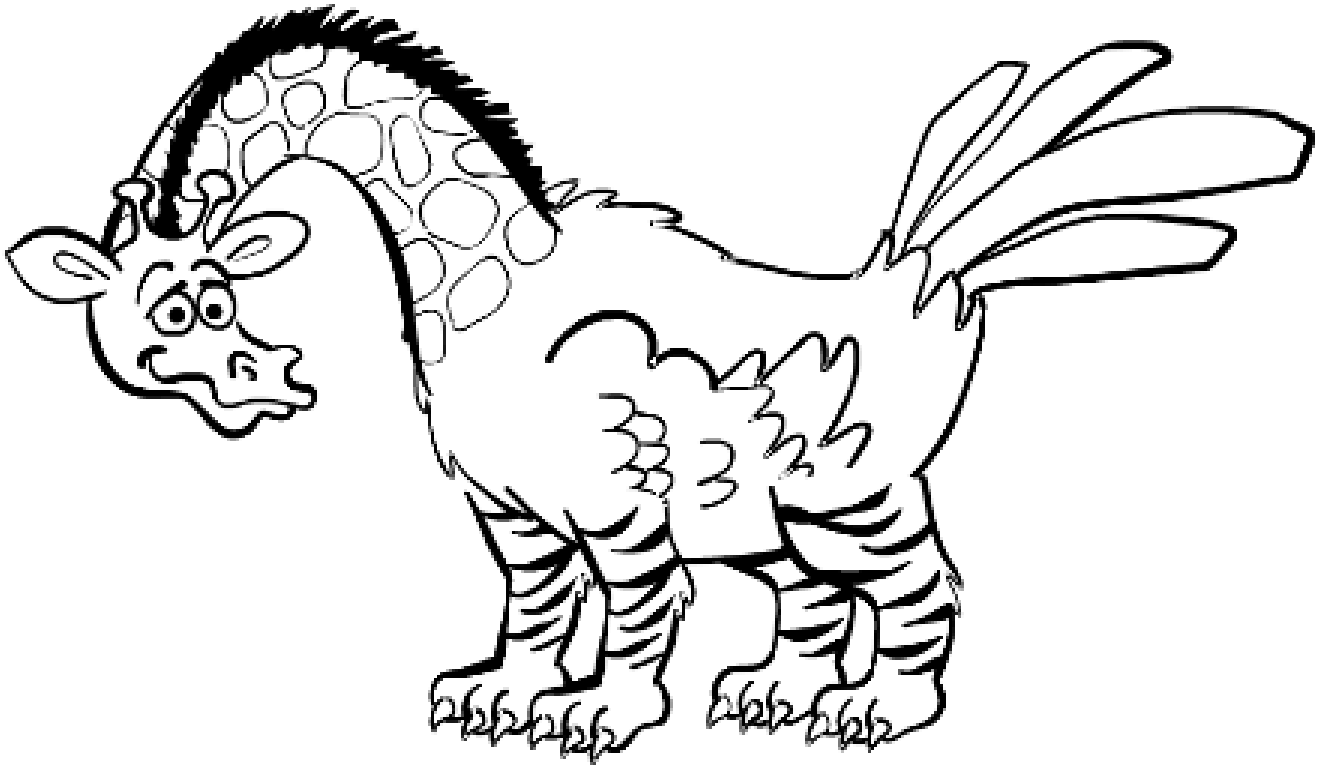
CAROLI LINNÆI
Naturæ Curioſorum Dioſcoridis Secundi
SYSTEMA
NATURÆ
IN QUO
NATURÆ REGNA TRIA,
SECUNDUM,
CLASSES, ORDINES, GENERA, SPECIES,
SYSTEMATICE PROPONUNTUR.
Editio Secunda, Auctior.
STOCKHOLMIÆ
Apud GOTTFR. KIESEWETTER.
1740.

- Hierarchical system
  - Kingdom
  - Phylum
  - Class
  - Order
  - Genus
  - Species

Classification depicted as a tree

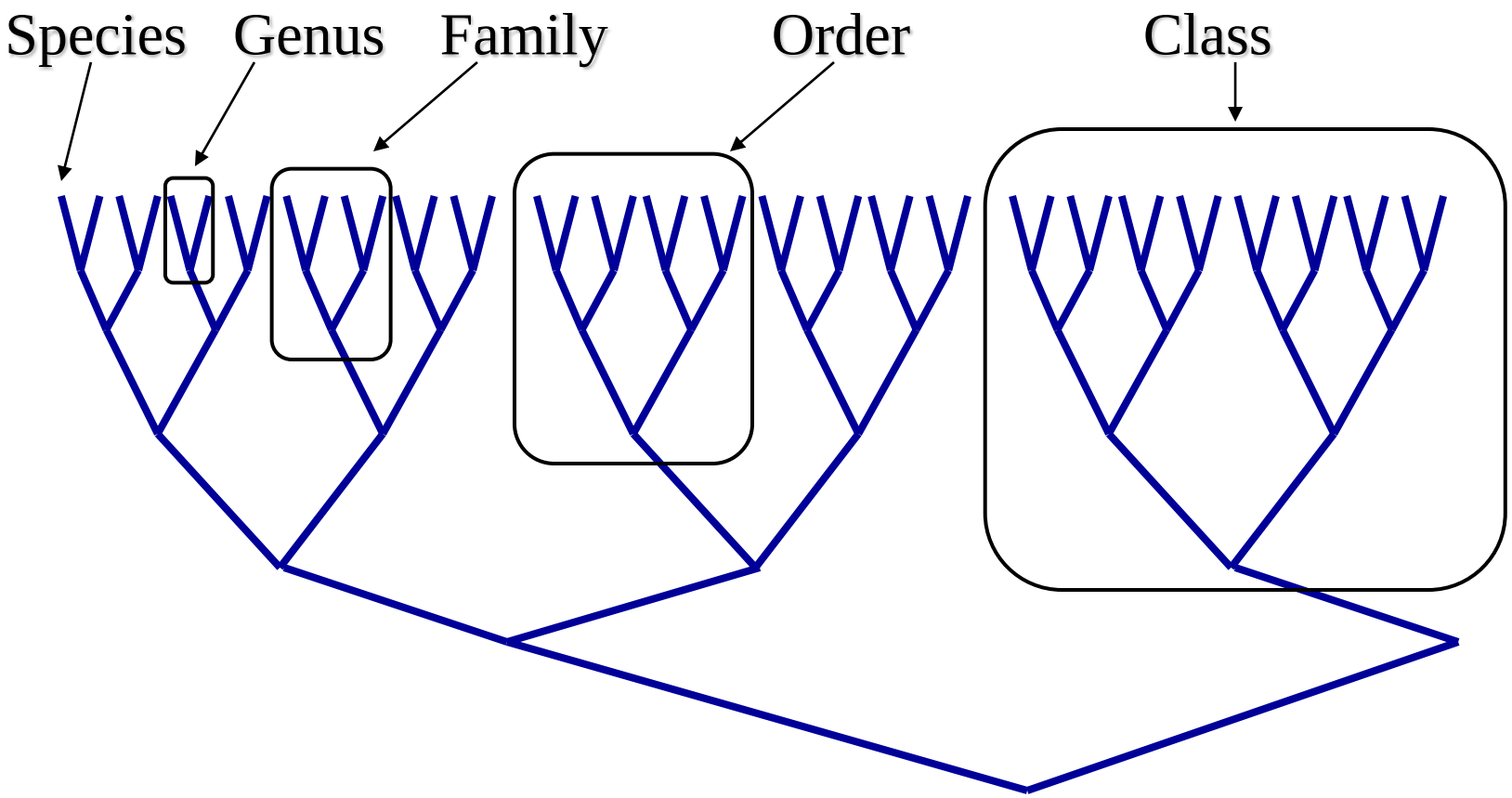


Observation: no “mixed animals”



Source: www.dr.dk/oline

Classification depicted as a tree



Comparison of limbs

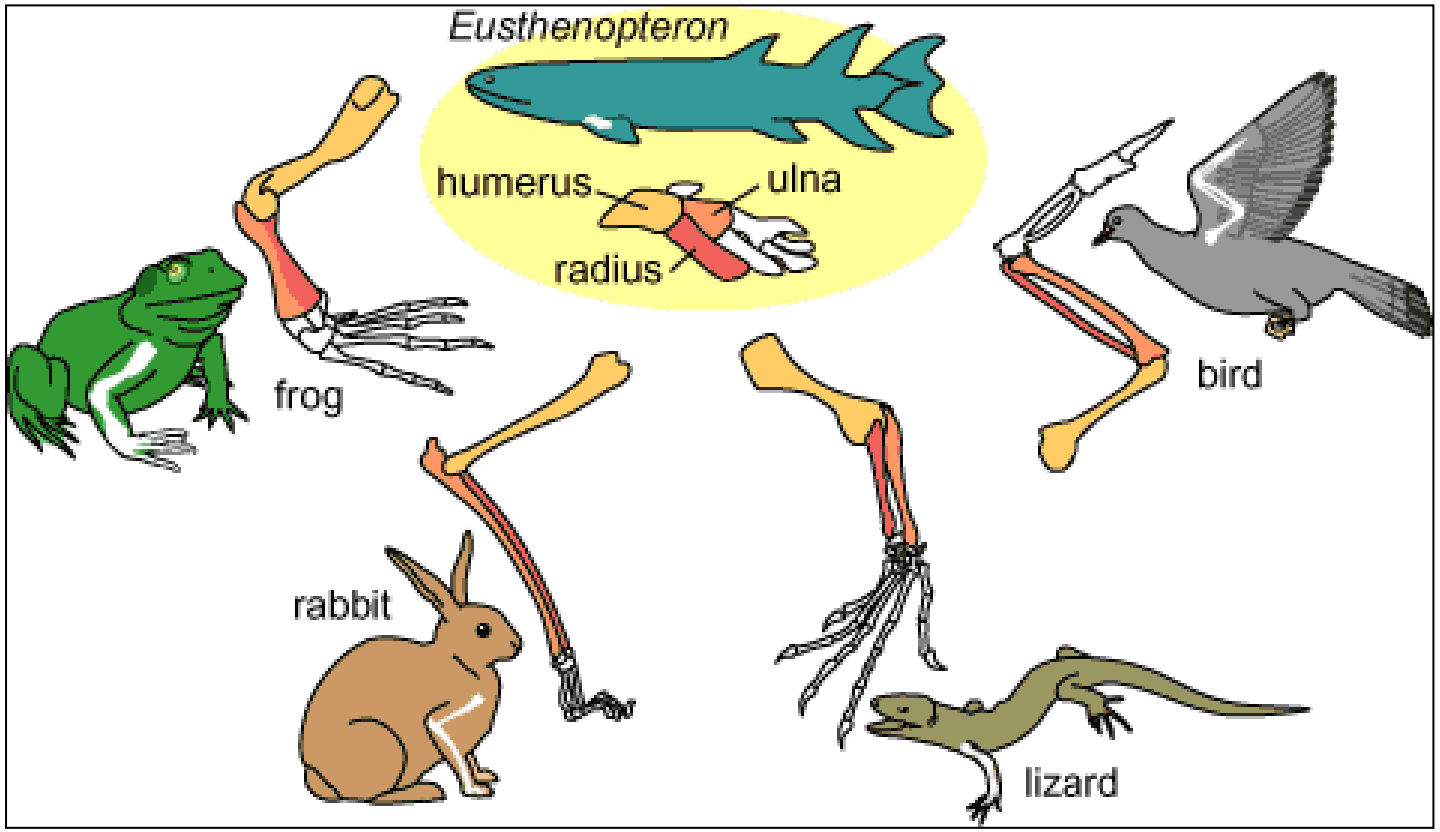
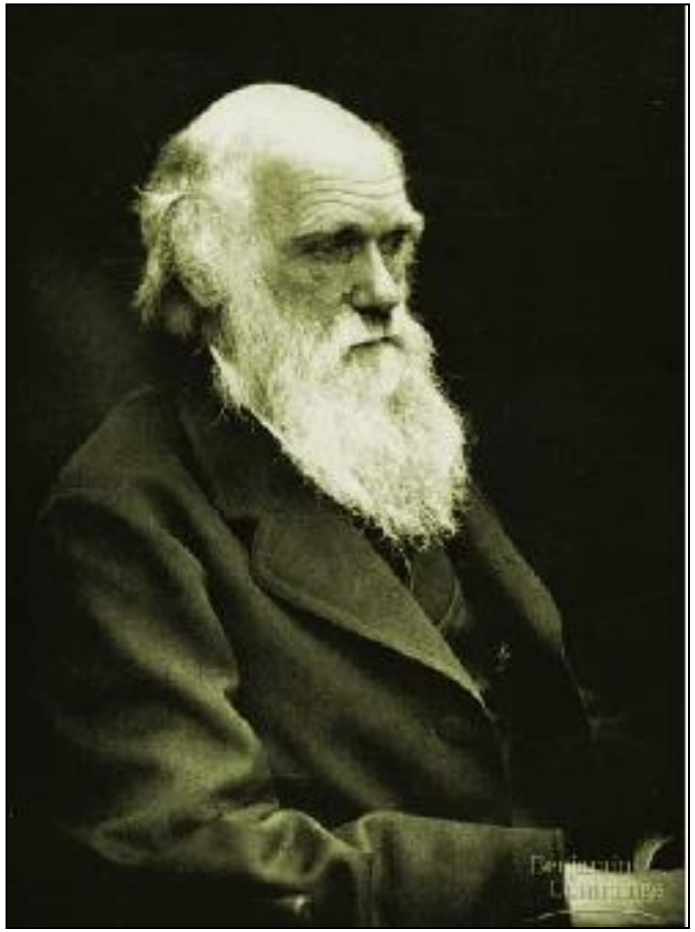


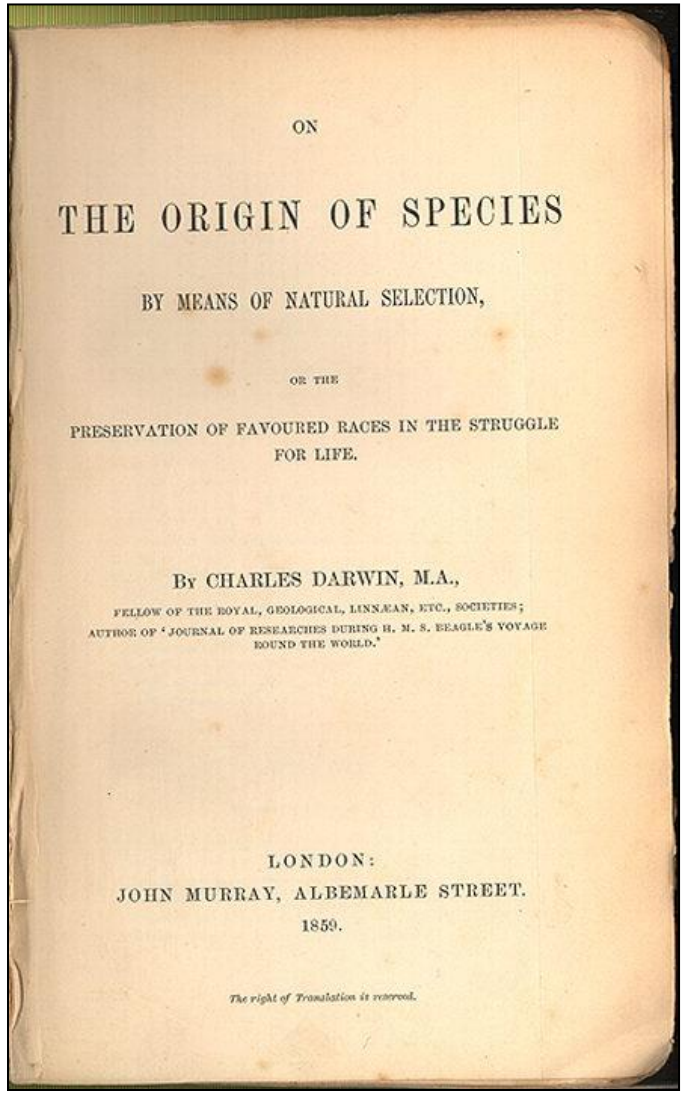
Image source: http://evolution.berkeley.edu

Theory of evolution



Charles Darwin

1809-1882



ON

THE ORIGIN OF SPECIES

BY MEANS OF NATURAL SELECTION,

OR THE

PRESERVATION OF FAVOURED RACES IN THE STRUGGLE FOR LIFE.

By CHARLES DARWIN, M.A.,

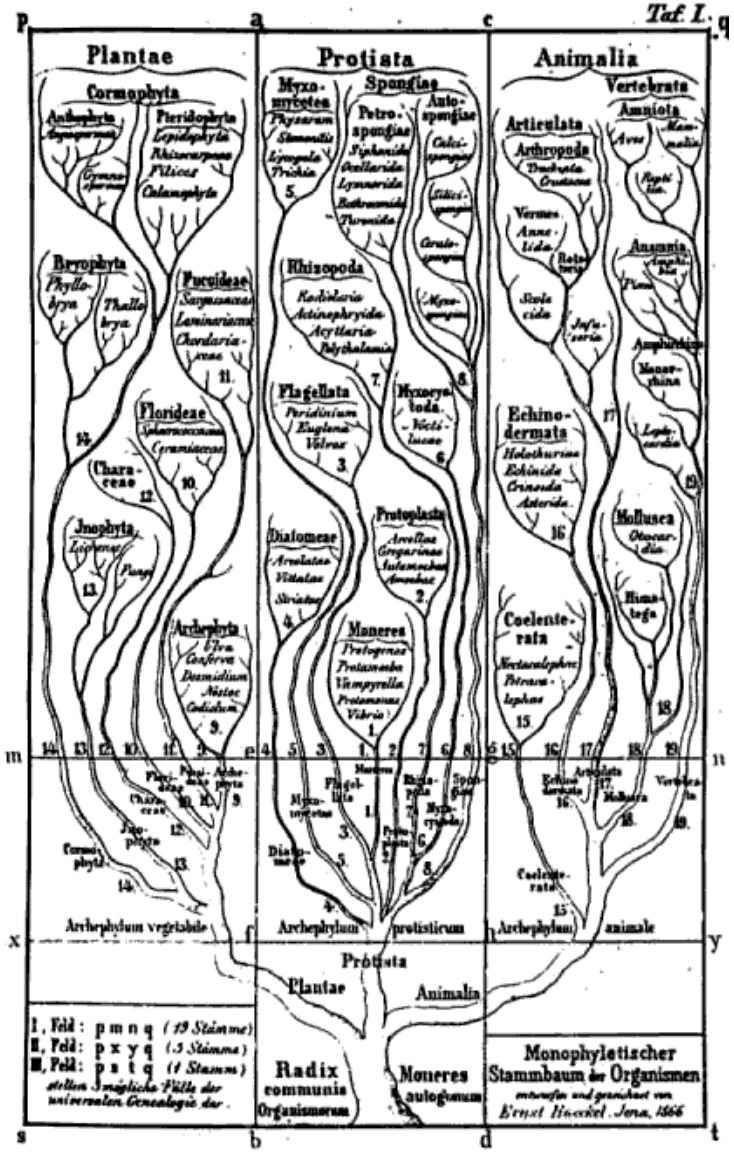
FELLOW OF THE ROYAL, GEOLOGICAL, LINNÆAN, ETC., SOCIETIES;
AUTHOR OF 'JOURNAL OF RESEARCHES DURING H. M. S. BEAGLE'S VOYAGE ROUND THE WORLD.'

LONDON:
JOHN MURRAY, ALBEMARLE STREET.
1859.

The right of Translation is reserved.

Phylogenetic basis of systematics

- Linnaeus:
  Ordering principle is God.
- Darwin:
  Ordering principle is shared descent from common ancestors.
- Today, systematics is explicitly based on phylogeny.



Natural Selection: Darwin's four postulates

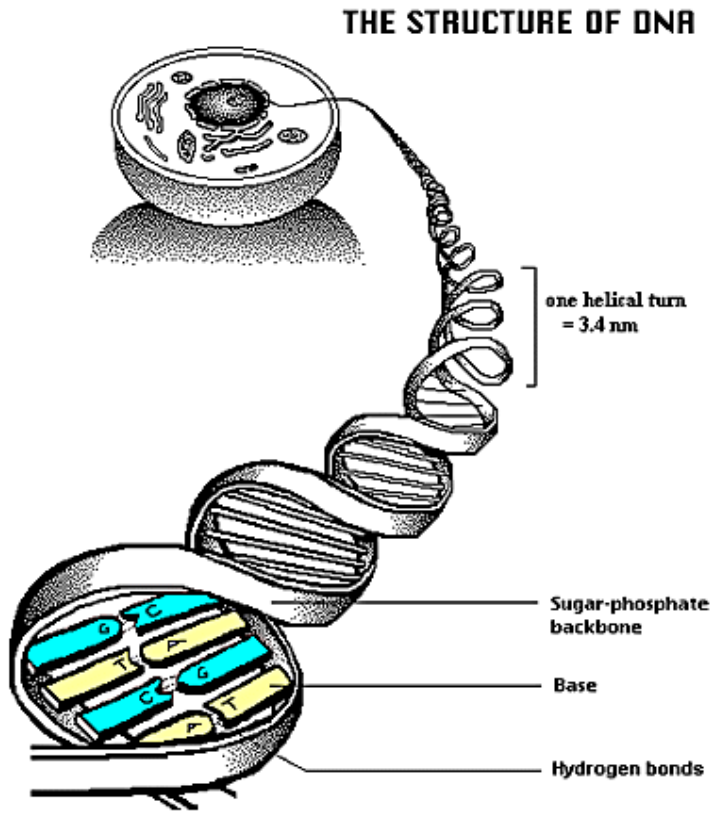
- More young are produced each generation than can survive to reproduce.
- Individuals in a population vary in their characteristics.
- Some differences among individuals are based on genetic differences.
- Individuals with favorable characteristics have higher rates of survival and reproduction.

- Evolution by means of natural selection
- Presence of "design-like" features in organisms:
- Quite often features are there "for a reason"

Evolution at the sequence level

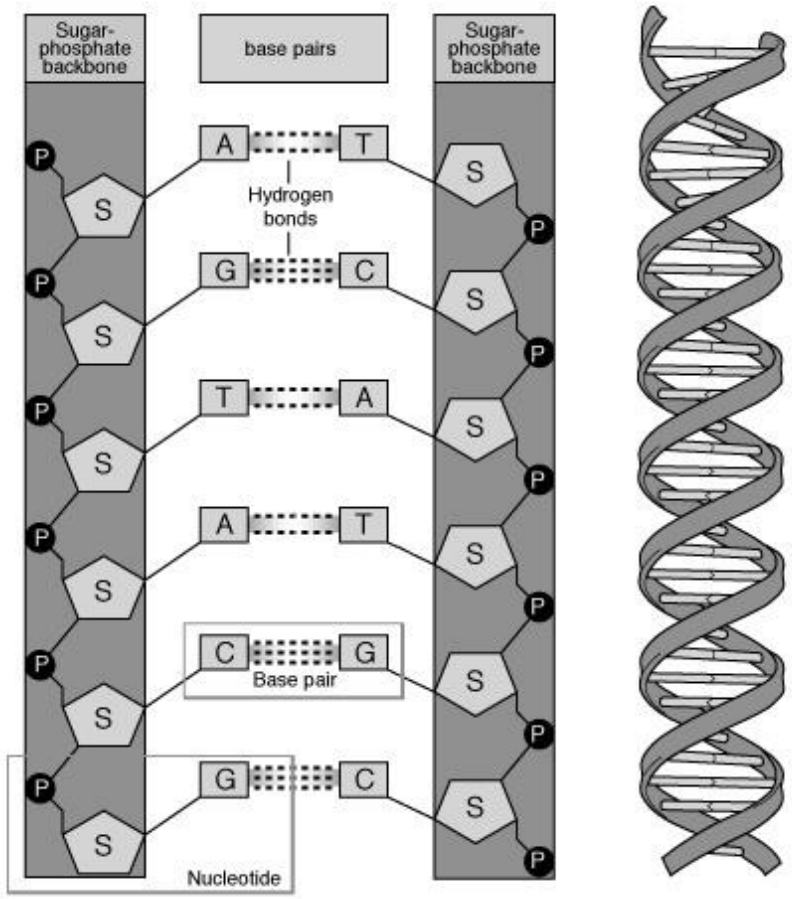
Recap about DNA

- DNA contains the recipes of how to make protein / enzymes.
- Every time a cells divides its DNA is duplicated, and each daughter cell gets a copy.

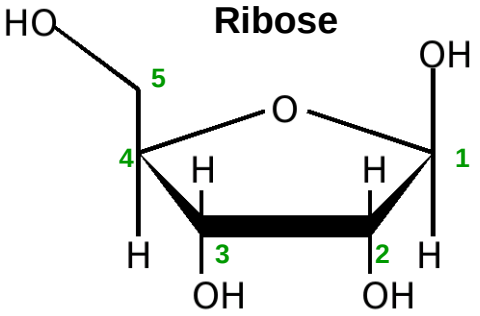


The DNA alphabet

- The *information* in the DNA is written in a four letter code: **A**, **T**, **G**, **C**.
- The DNA can be “sequenced” and the result stored in a computer file.
- ATGGCCCTGTGGAT



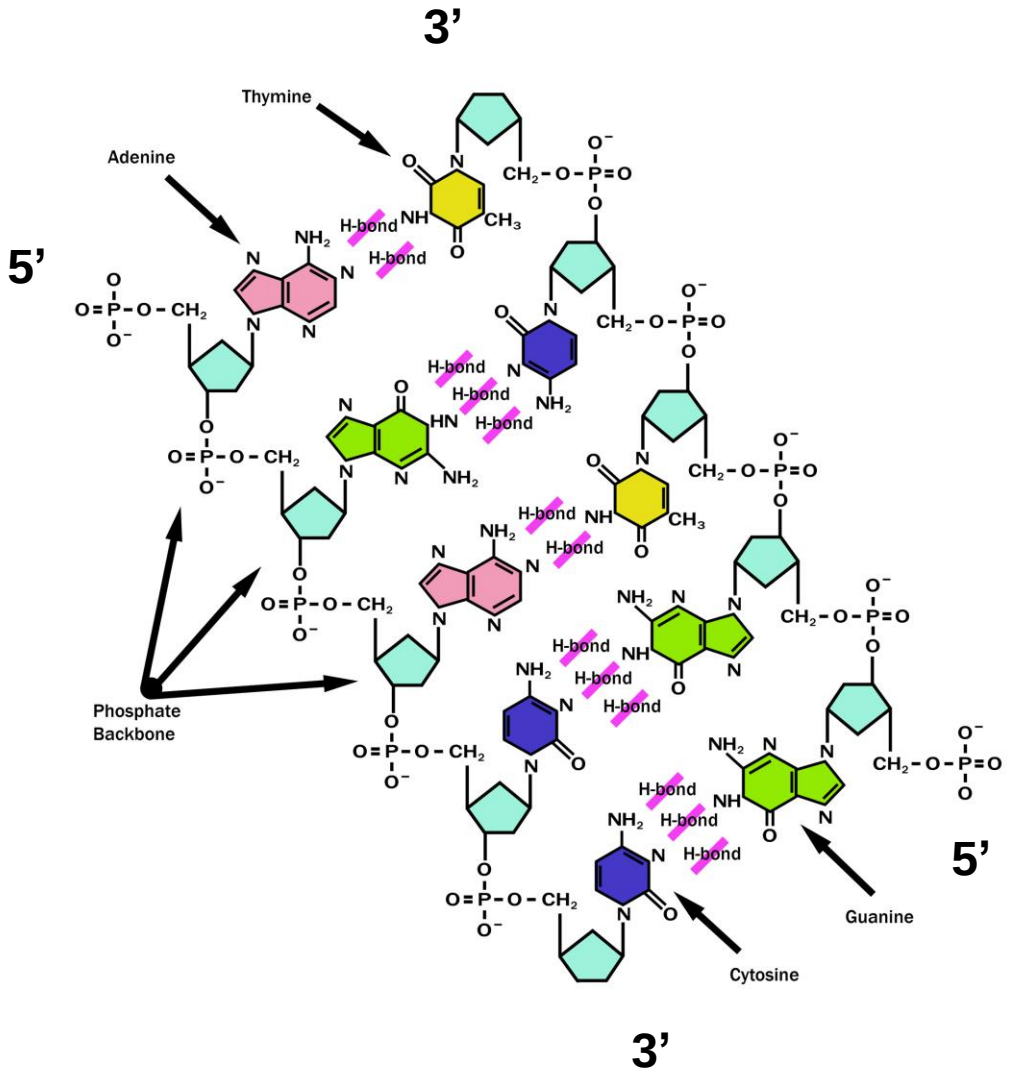
DNA is always written 5’ → 3’



5’ AGCC 3’

3’ TCGG 5’

5’ ATGGCCAGGTAA 3’



DNA backbone: http://en.wikipedia.org/wiki/DNA
(Deoxy)ribose: http://en.wikipedia.org/

DNA can change via mutations

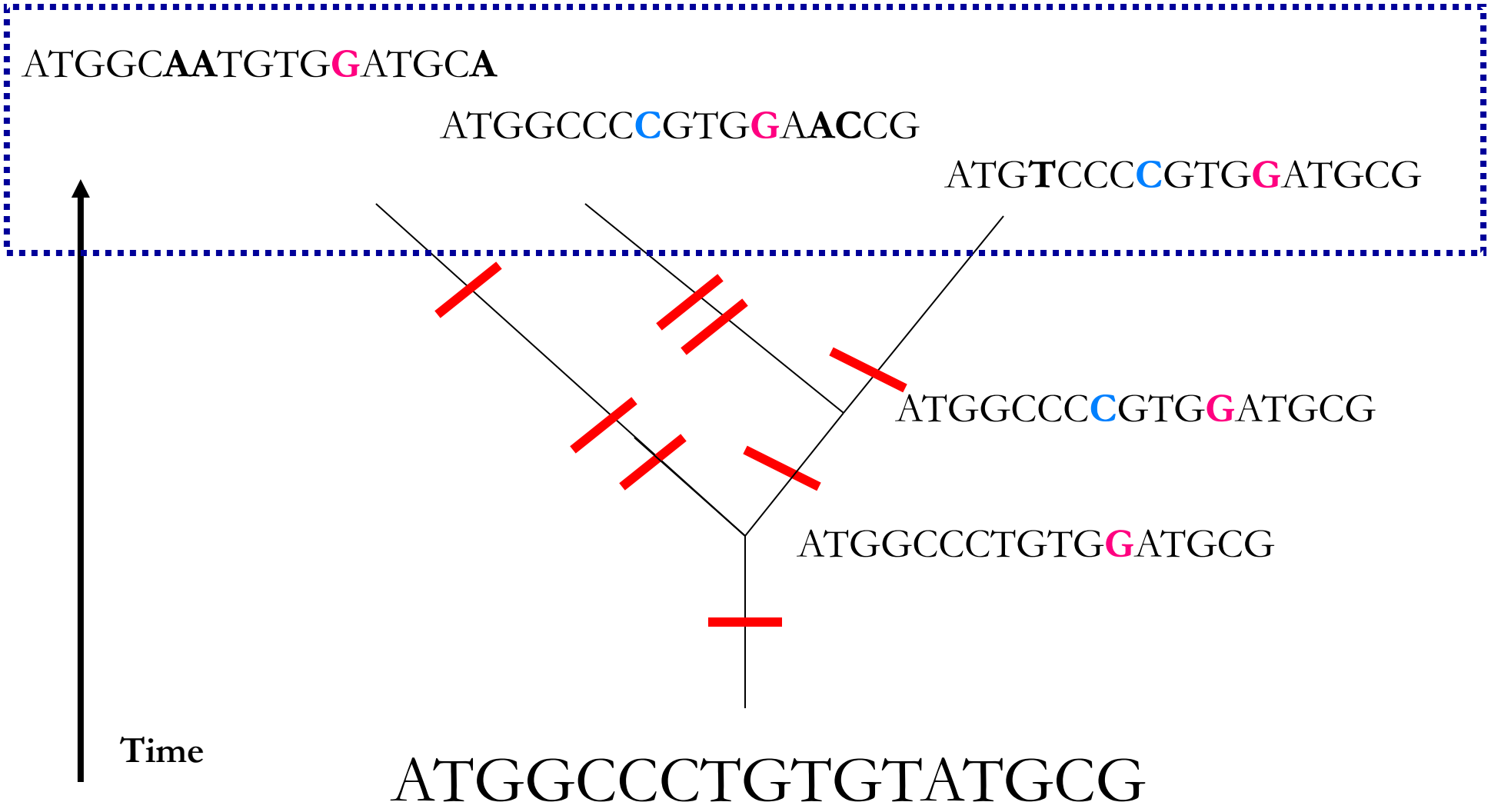
- ATGGCCCTGTGGATGCG

Focus for this example: point mutations (SNPs)

- ATGGCCCTGTGGATGCG

- ATGGCCCT**A**TGGATGCG

A history of mutations

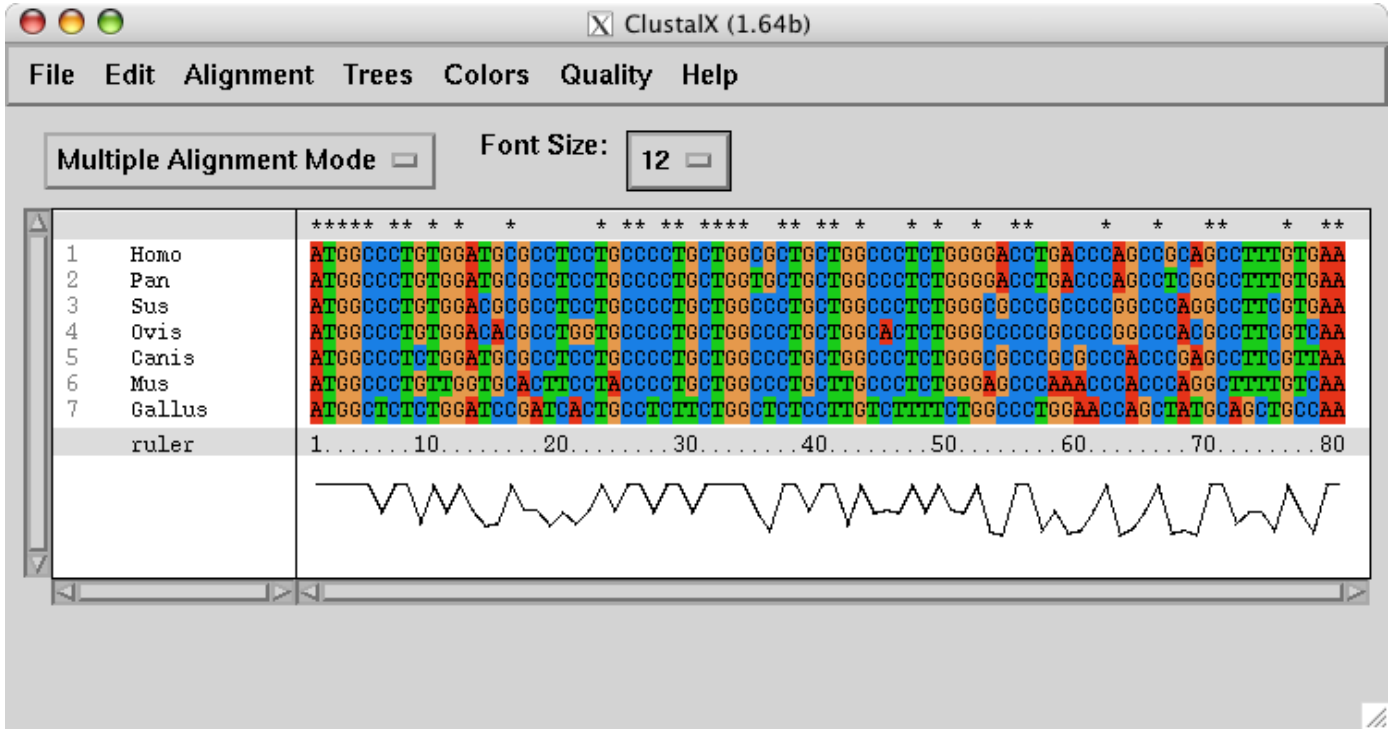


Alignment of related DNA sequences

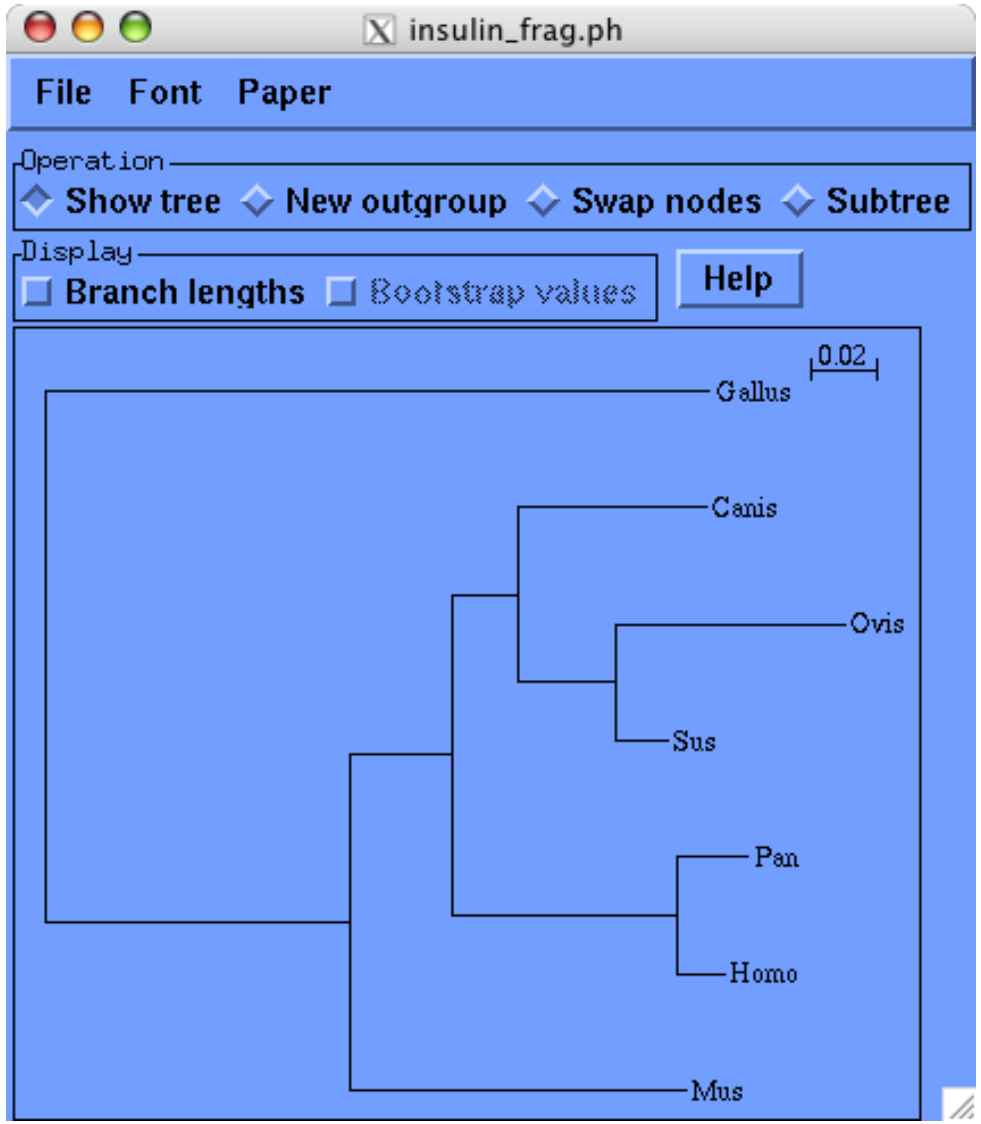
- Species1: ATGGC**AA**TGTG**G**ATGC**A**
- Species2: ATGGCCC**C**GTG**G**A**AC**CG
- Species3: ATG**T**CCC**C**GTG**G**ATGCG

Real life example:
Alignment of Insulin from 7 vertebrates

```
Homo:    ATGGCCCTGTGGATGCGCCTCCTGCCCCTGCTGGCGCTGCTGGCCCTCTGGGGACCTGACCCAGCCGCAGCCTTTGTGAA
Pan:     ATGGCCCTGTGGATGCGCCTCCTGCCCCTGCTGGTGCTGCTGGCCCTCTGGGGACCTGACCCAGCCTCGGCCTTTGTGAA
Sus:     ATGGCCCTGTGGACGCGCCTCCTGCCCCTGCTGGCCCTGCTGGCCCTCTGGGCGCCCGCCCCGGCCCAGGCCTTCGTGAA
Ovis:    ATGGCCCTGTGGACACGCCTGGTGCCCCTGCTGGCCCTGCTGGCACTCTGGGCCCCCGCCCCGGCCCACGCCTTCGTCAA
Canis:   ATGGCCCTCTGGATGCGCCTCCTGCCCCTGCTGGCCCTGCTGGCCCTCTGGGCGCCCGCGCCCACCCGAGCCTTCGTTAA
Mus:     ATGGCCCTGTTGGTGCACTTCCTACCCCTGCTGGCCCTGCTTGCCCTCTGGGAGCCCAAACCCACCCAGGCTTTTGTCAA
Gallus:  ATGGCTCTCTGGATCCGATCACTGCCTCTTCTGGCTCTCCTTGTCTTTTCTGGCCCTGGAACCAGCTATGCAGCTGCCAA
```



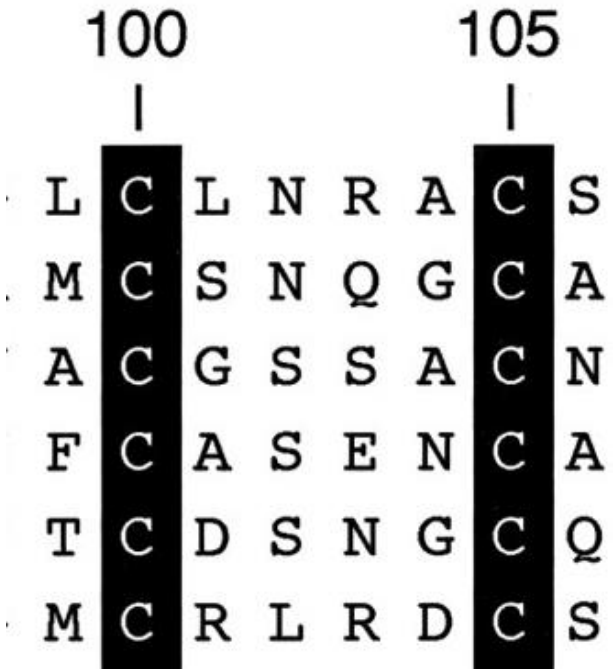
… shown as a phylogenetic tree

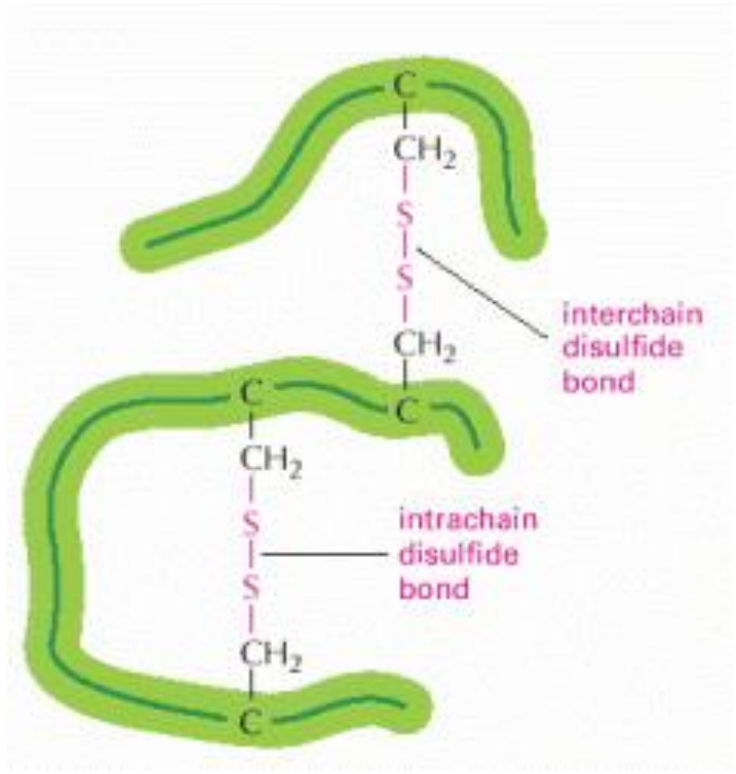


Interpretation of Multiple Alignments

Conserved features assumed to be important for functionality

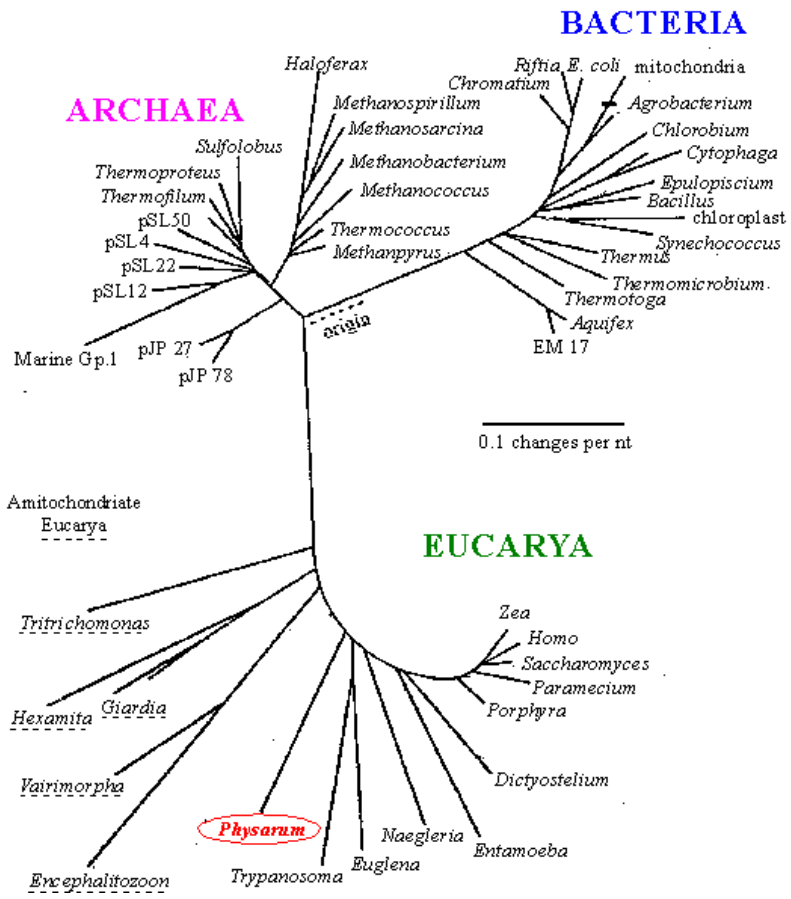
For instance: conserved pairs of cysteines indicate a possible disulphide bridge





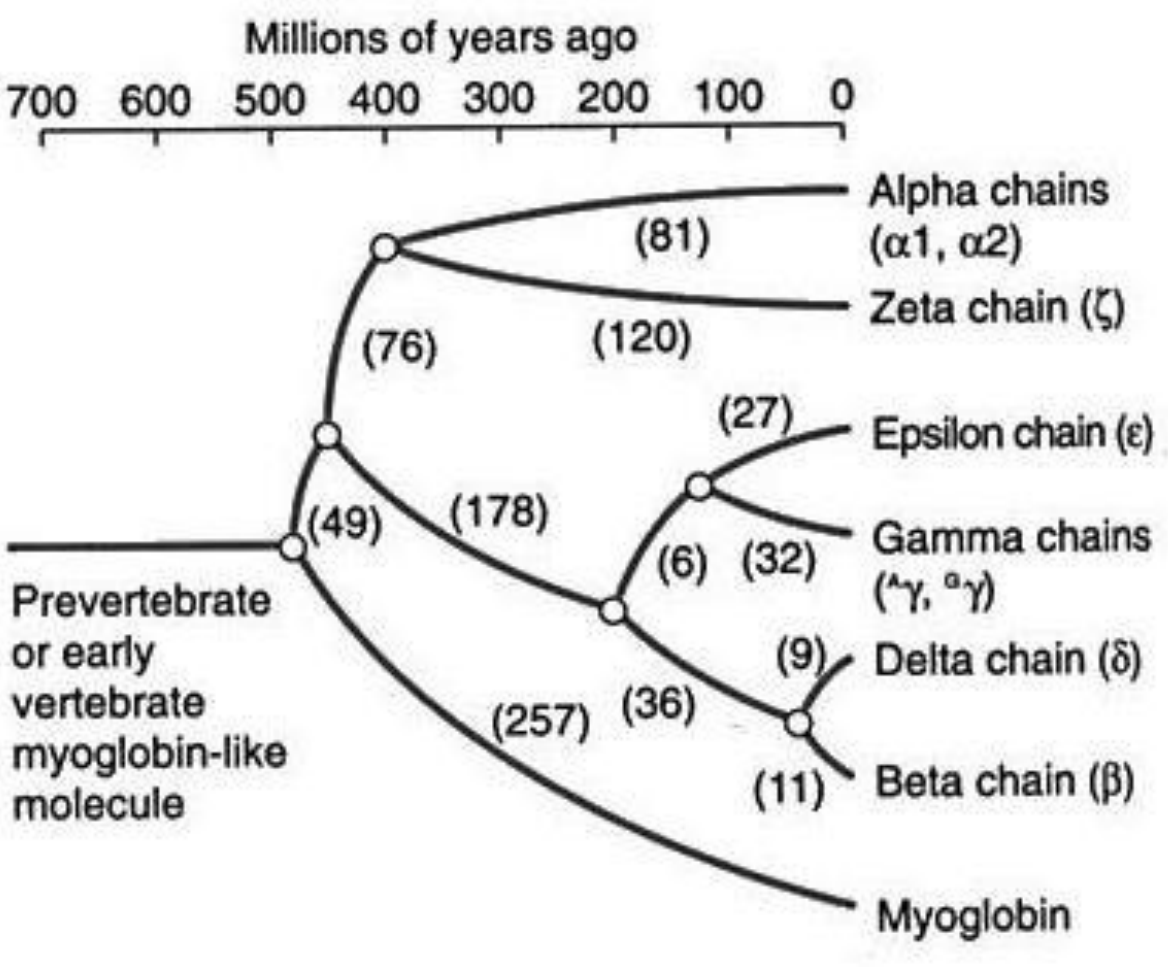
Sequences are related

- Darwin: all organisms are related through descent with modification
- Prediction: similar molecules have similar functions in different organisms



Protein synthesis carried out by very similar RNA-containing molecular complexes (ribosomes) that are present in all known organisms

Sequences are related – part II



Related oxygen-binding proteins in humans