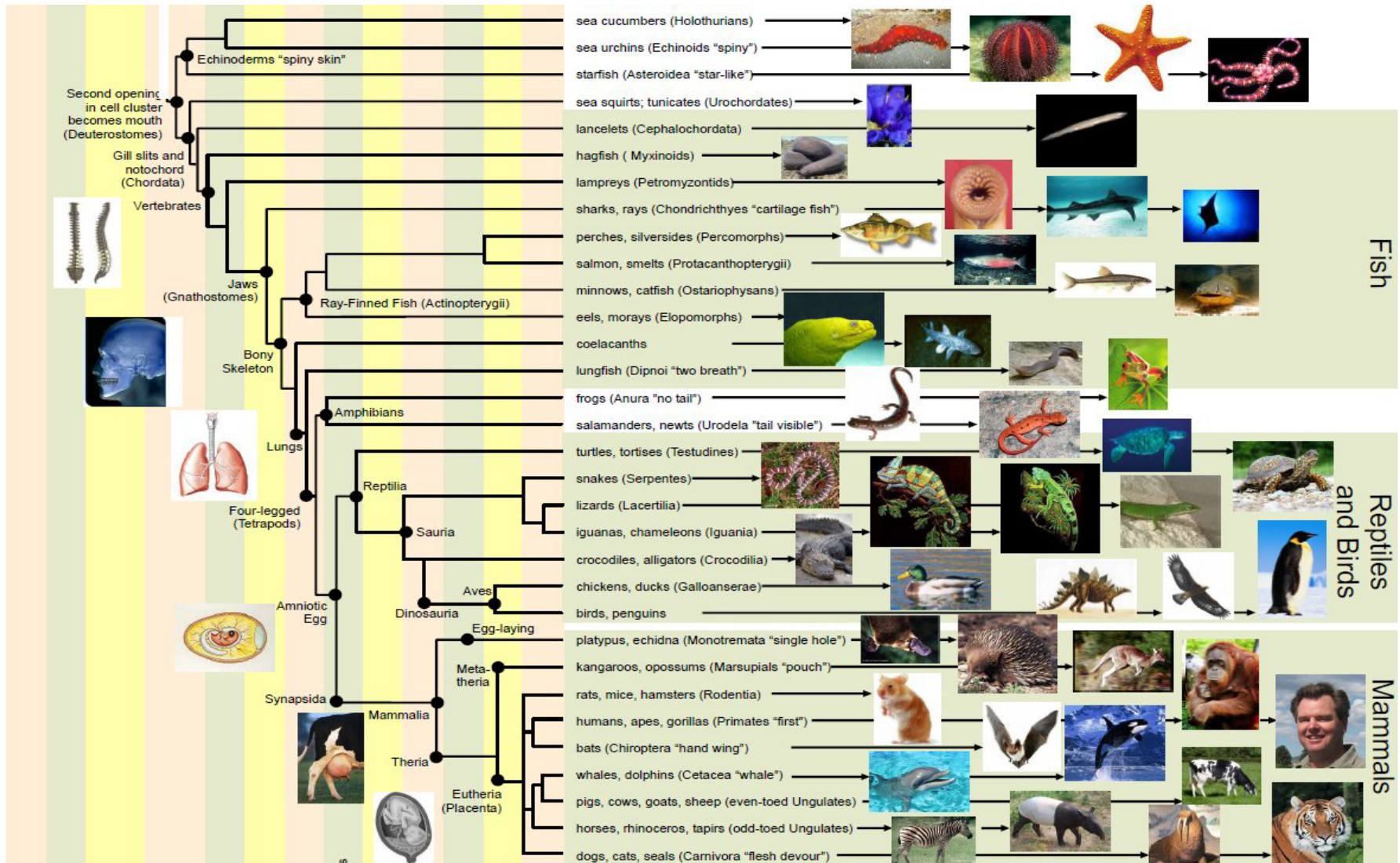
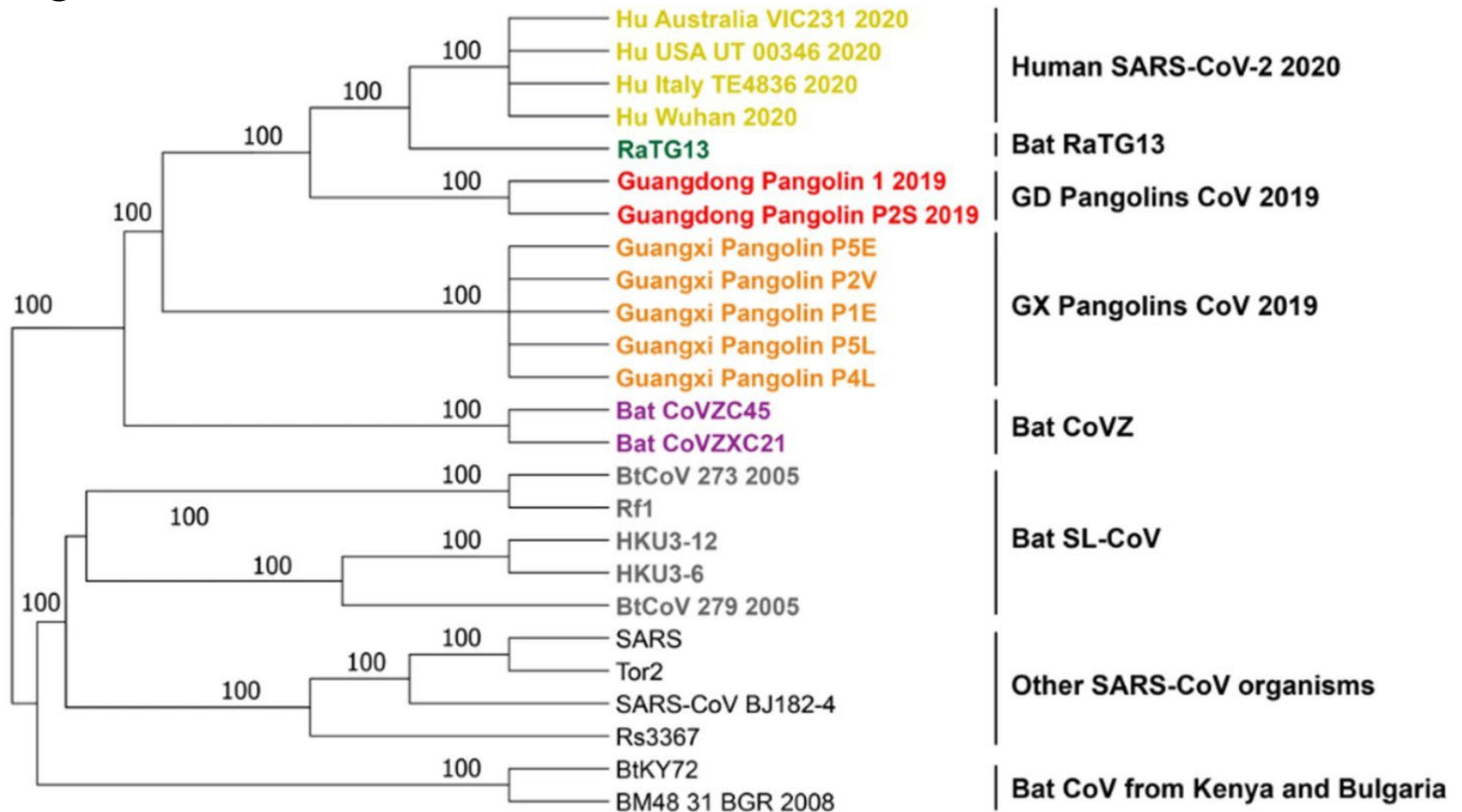


Phylogenetics



Phylogenetics is useful for:

- Design of biodiversity conservation policies
- Detecting the origin of pathogens (e.g. SARSCov-2)
- Studying molecular evolution of viruses
- Detecting origin and evolution of cancer cells
- Linguistics



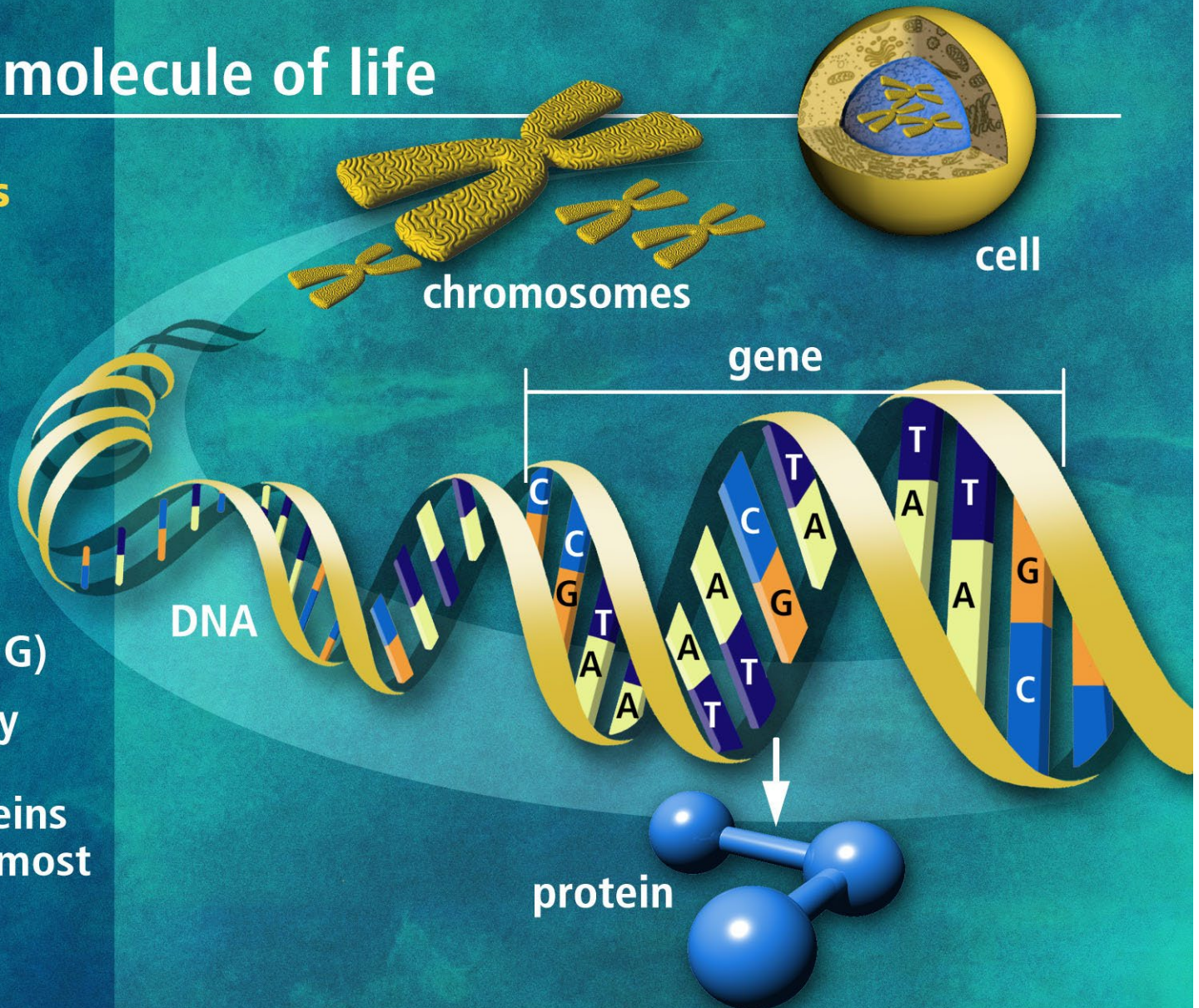
Genome

DNA the molecule of life

Trillions of cells

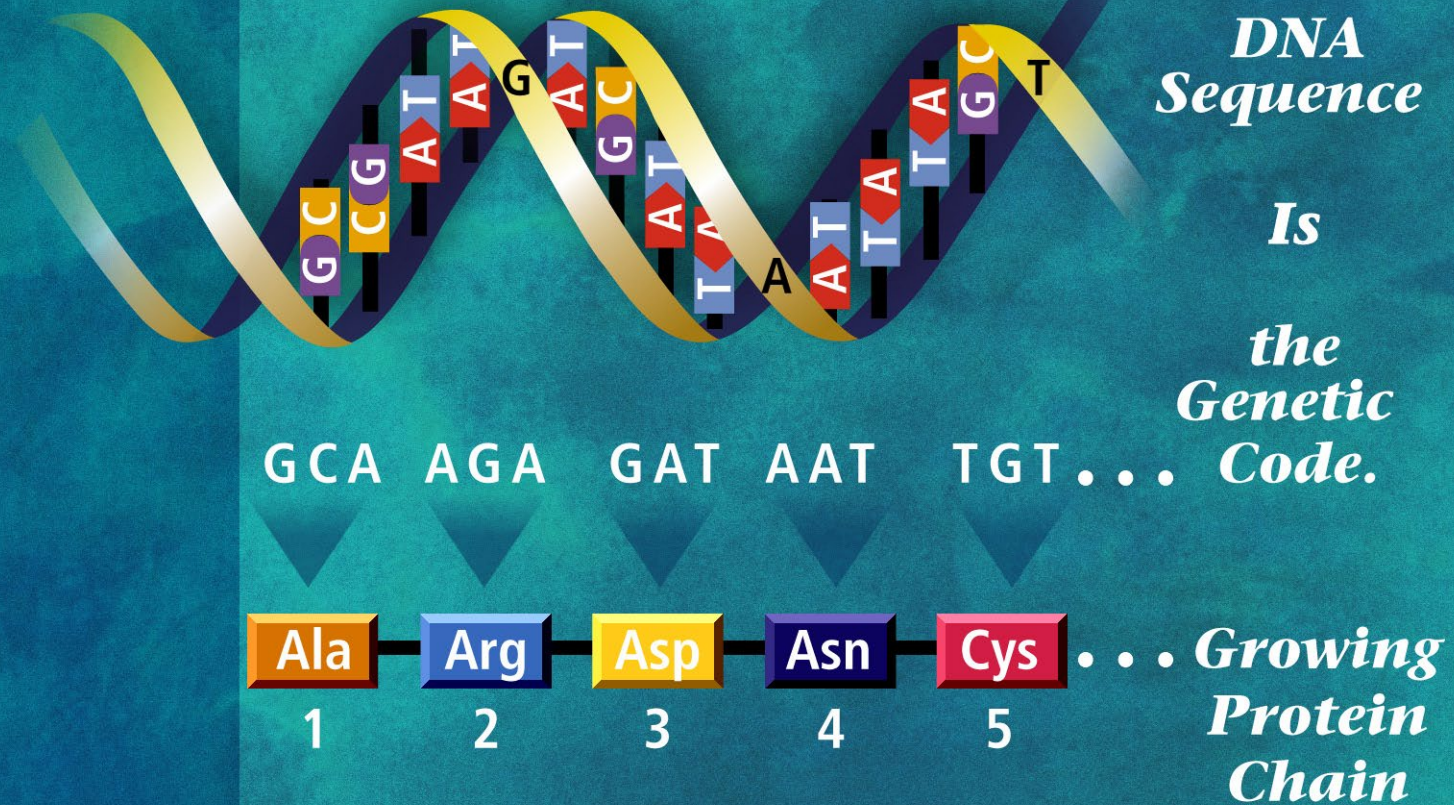
Each cell:

- 46 human chromosomes
- 2 meters of DNA
- 3 billion DNA subunits (the bases: A, T, C, G)
- Approximately 30,000 genes code for proteins that perform most life functions



Protein coding genes

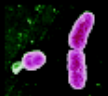






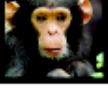
DNA Genetic Code Dictates Amino Acid Identity and Order



Translation: the genetic code

		SECOND BASE					
		U	C	A	G		
FIRST BASE	U	UUU } Phe	UCU } Ser	UAU } Tyr	UGU } Cys	THIRD BASE	
		UUC } Phe	UCC } Ser	UAC } Tyr	UGC } Cys		
		UUA } Leu	UCA } Ser	UAA } Stop	UGA } Stop		
		UUG } Leu	UCG } Ser	UAG } Stop	UGG } Trp		
	C	CUU } Leu	CCU } Pro	CAU } His	CGU } Arg		
		CUC } Leu	CCC } Pro	CAC } His	CGC } Arg		
		CUA } Leu	CCA } Pro	CAA } Gln	CGA } Arg		
		CUG } Leu	CCG } Pro	CAG } Gln	CGG } Arg		
	A	AUU } Ile	ACU } Thr	AAU } Asn	AGU } Ser		
		AUC } Ile	ACC } Thr	AAC } Asn	AGC } Ser		
		AUA } Met	ACA } Thr	AAA } Lys	AGA } Arg		
		AUG } Met	ACG } Thr	AAG } Lys	AGG } Arg		
	G	GUU } Val	GCU } Ala	GAU } Asp	GGU } Gly		
		GUC } Val	GCC } Ala	GAC } Asp	GGC } Gly		
		GUA } Val	GCA } Ala	GAA } Glu	GGA } Gly		
		GUG } Val	GCG } Ala	GAG } Glu	GGG } Gly		

Evolutionary process: changes, insertions and deletions of nucleotides

	Year	Common Name	Species	# of Chromosomes	Size (base pairs)
	1995	Bacterium	<i>Haemophilus influenzae</i>	1	1.8×10^6
	1996	Yeast	<i>Saccharomyces cerevisiae</i>	16	1.2×10^7
	1998	Worm	<i>Caenorhabditis elegans</i>	6	1.0×10^8
	1999	Fruit Fly	<i>Drosophila melanogaster</i>	4	1.3×10^8
	2000	Human	<i>Homo sapiens</i>	23	3.4×10^9
	2002	Mouse	<i>Mus musculus</i>	20	3.5×10^9
	2004	Rat	<i>Rattus norvegicus</i>	21	2.9×10^9
	2005	Chimpanzee	<i>Pan troglodytes</i>	24	3.6×10^9

Check the information at www.ensembl.org

Multiple sequence alignment

Which nucleotides in each genome have evolved from the same nucleotide at the common ancestor?

Give this information in an **alignment**:

```
TGAAACTCTCCCAAAGATTGAGGCATTCTGCATTTGAATTTTATACAGAAAGTCATGAAT
TGAAACT-----GCTGAAG-ACTCTGTATCCCA--CCCAGTGGTGAGGGCATCGAG
*****          *** * * ***** ** * * * * * * * * * *
```

We assume alignments without gaps:

Human AACTTCGAGGCTTACCGCTG...

chimpanzee AAGGTCGATGCTCACCGATG...

gorilla AACGTCATGCTCACCGATG...

Phylogenetic reconstruction

Input: alignment of n DNA (genomic) sequences



AACTTCGAGGCTTACCGCTG...

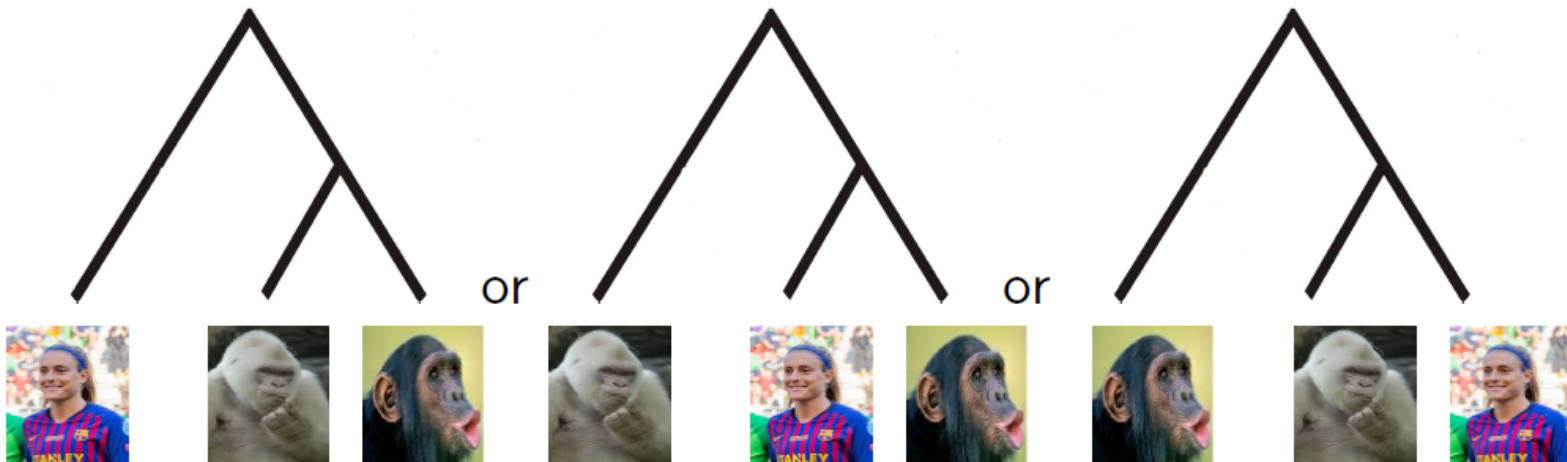


AA GGTCGATGCTCACCGATG...



AACGTCTATGCTCACCGATG...

Output: the most suitable tree



- **Branch lengths** are the expected number of mutations per site that have occurred along that branch.

Number of Taxa	Number of unrooted trees	Number of rooted trees
3	1	3
4	3	15
5	15	105
6	105	945
7	945	10395
8	10395	135135
9	135135	2027025
10	2027025	34459425
20	2.22E+020	8.20E+021
30	8.69E+036	4.95E+038
40	1.31E+055	1.01E+057
50	2.84E+074	2.75E+076
60	5.01E+094	5.86E+096
70	5.00E+115	6.85E+117
80	2.18E+137	3.43E+139

Bibliography

- Allman, E.S.; Rhodes, J.A., *The mathematics of phylogenetics*, lecture notes,
<https://jarhodesuaf.github.io/PhyloBook.pdf>