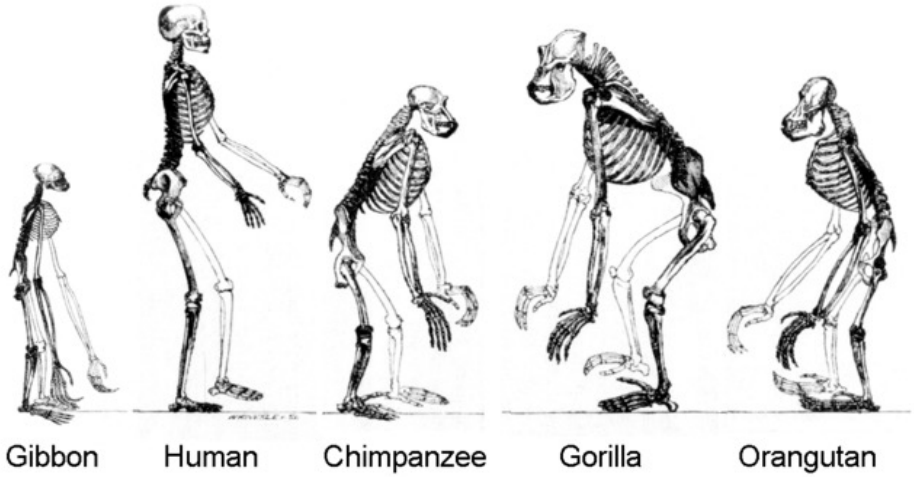
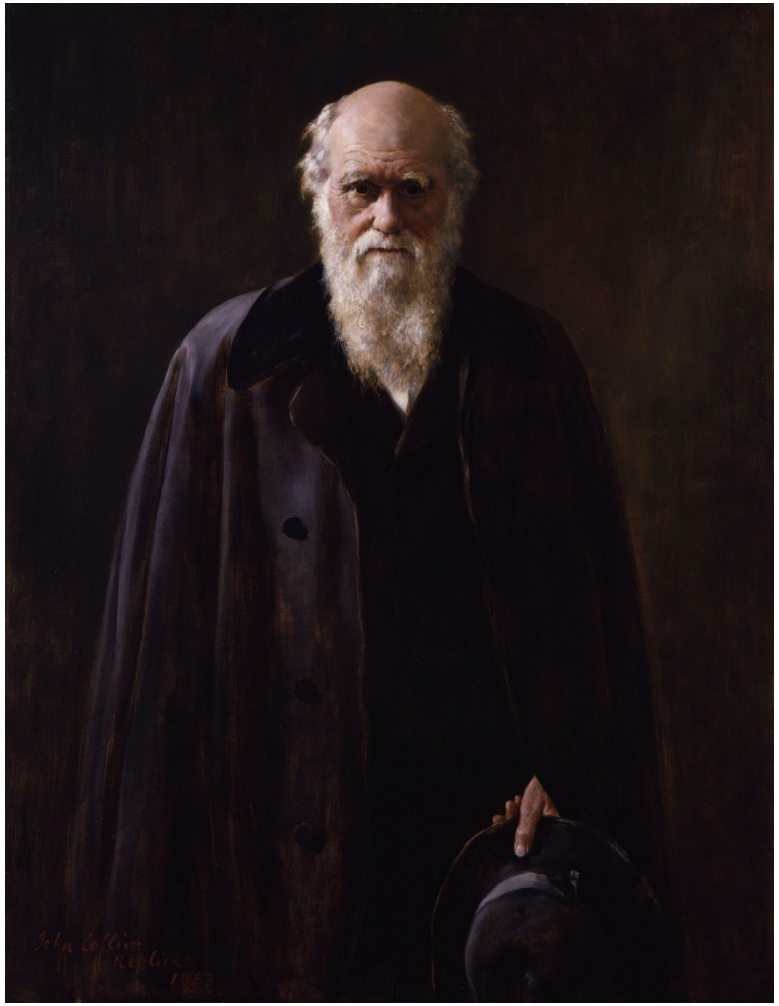
What makes us Human? Data Science and Genomics

Special Topics in DS2000: Introduction to Programming with Data

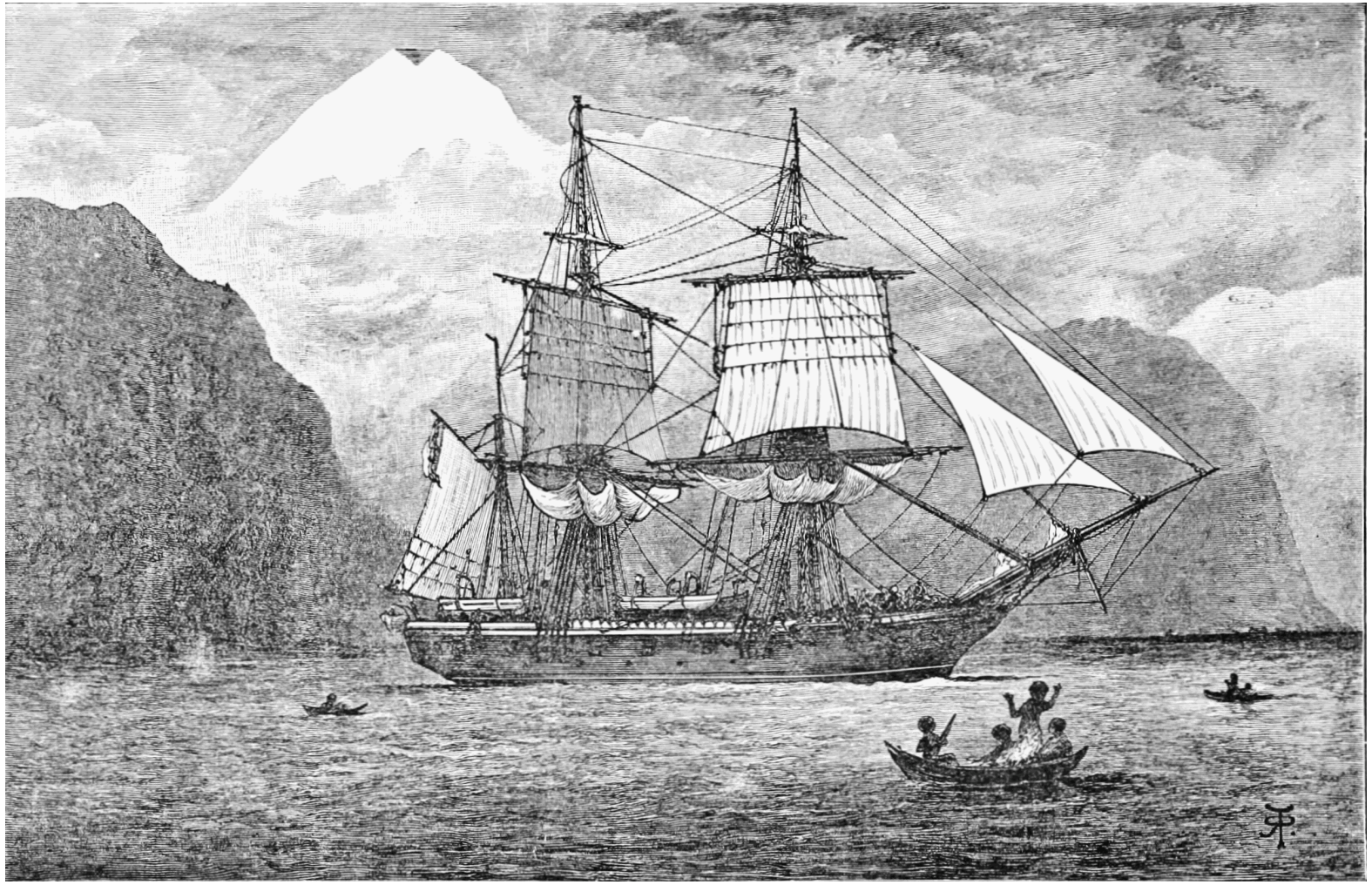


Charles Darwin



1809 - 1882

On the Origin of Species (1859)



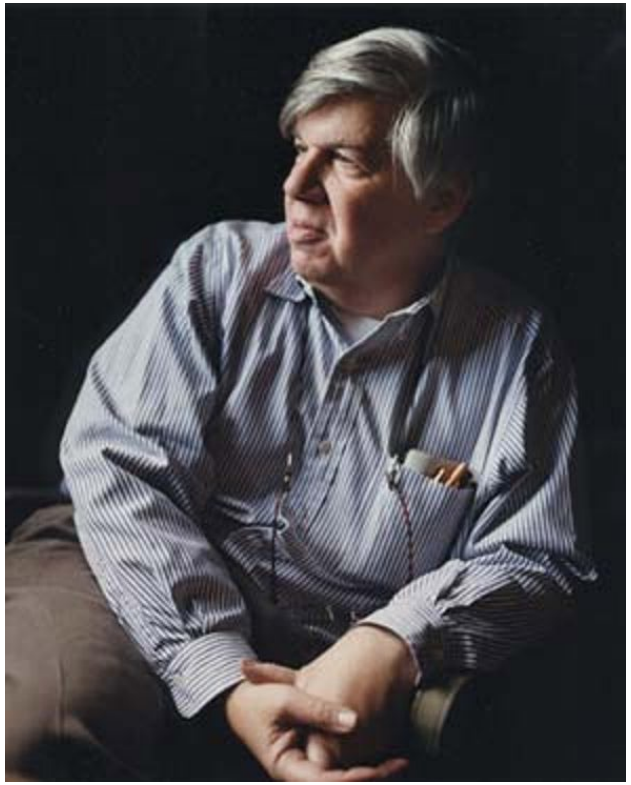
HMS Beagle

Voyage of the Beagle: 1831 – 1836

Natural Selection

[The] basis of natural selection is simplicity itself – two undeniable facts and an inescapable conclusion.

-- Ever Since Darwin (1977)



Stephen Jay Gould
Harvard Paleontologist
1941 - 2002

Variation and Inheritance

1. Organisms vary, and these variations are inherited (at least in part) by their offspring.

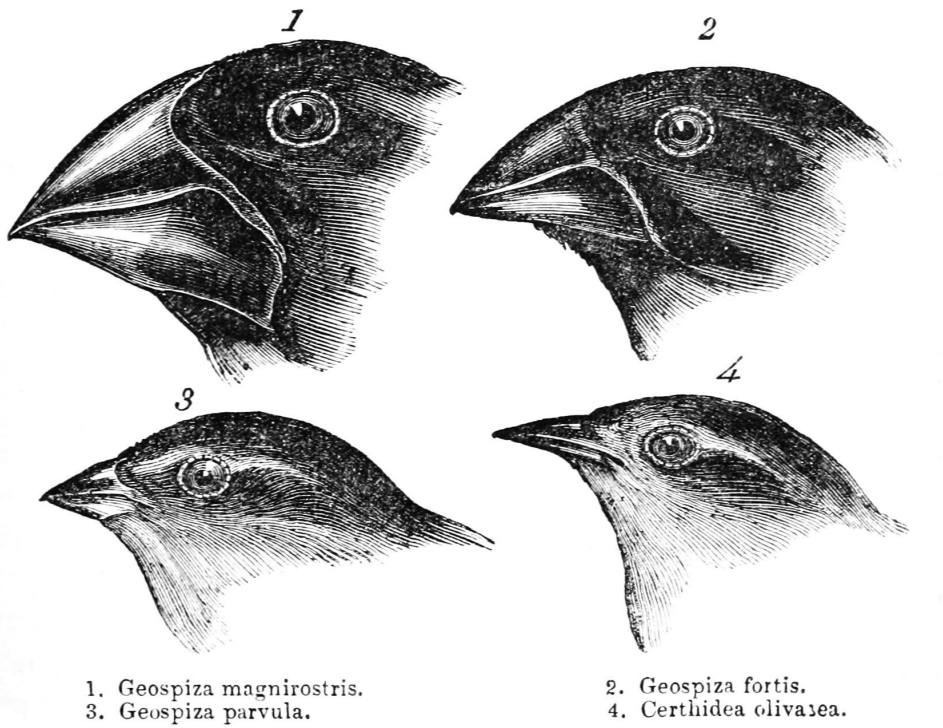
Population explosion

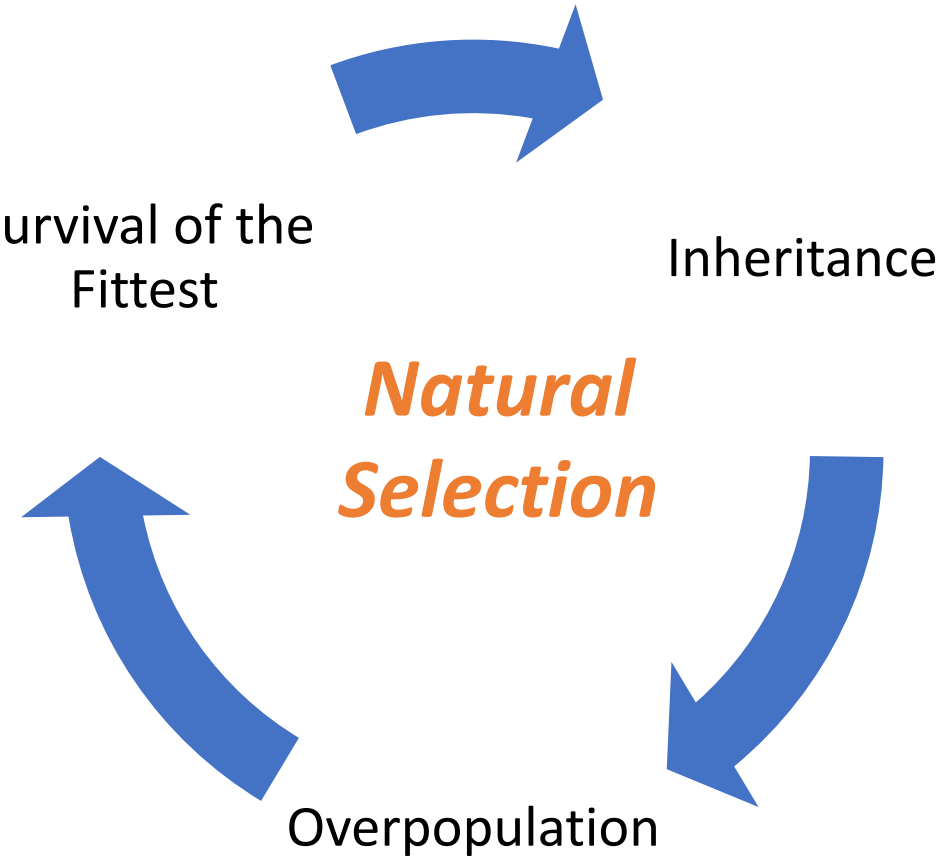
2. Organisms produce more offspring than can possibly survive.

Survival of the Fittest

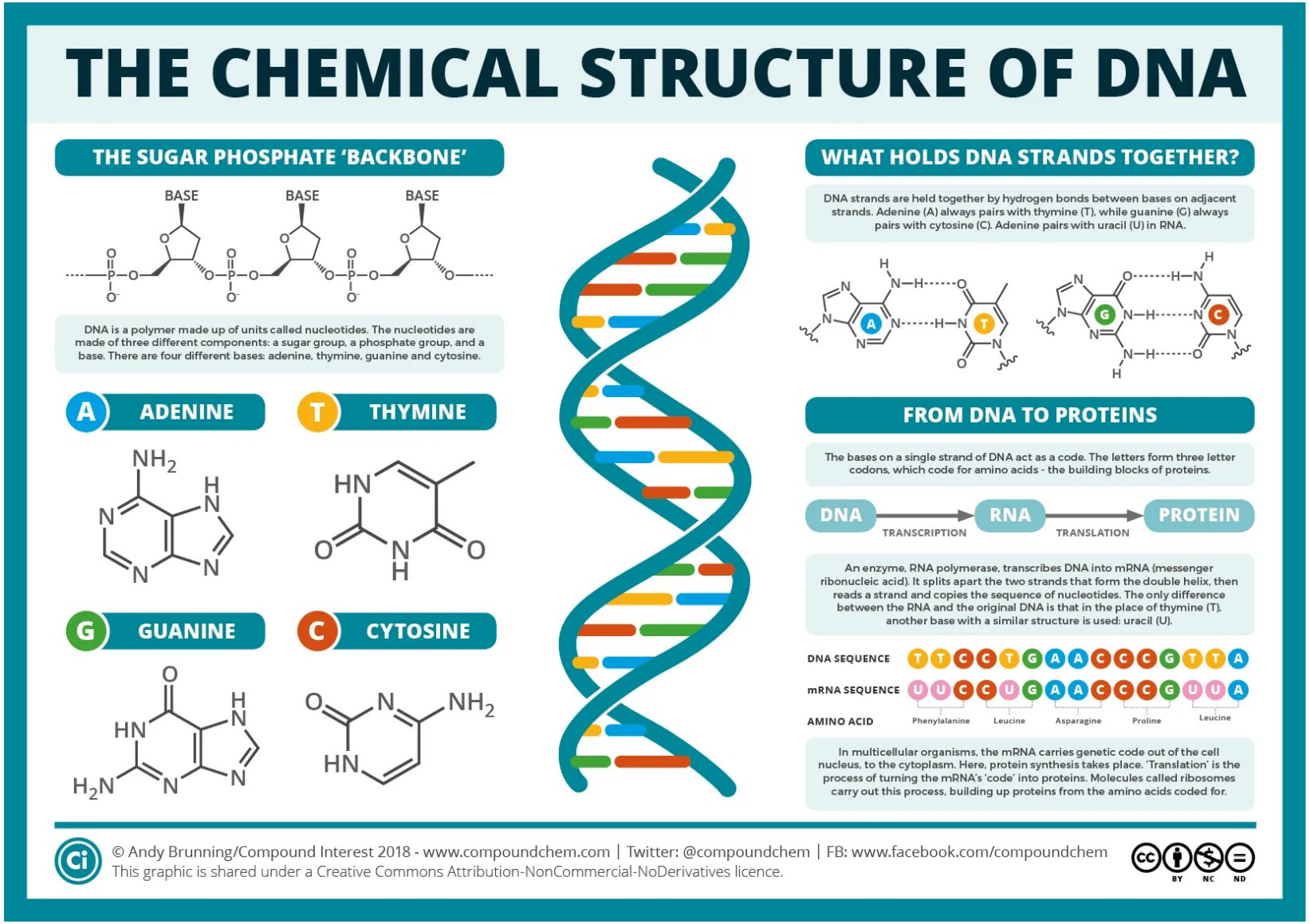
3. On average, offspring that vary most strongly in directions favored by the environment will survive and propagate.

Evolution by Natural Selection



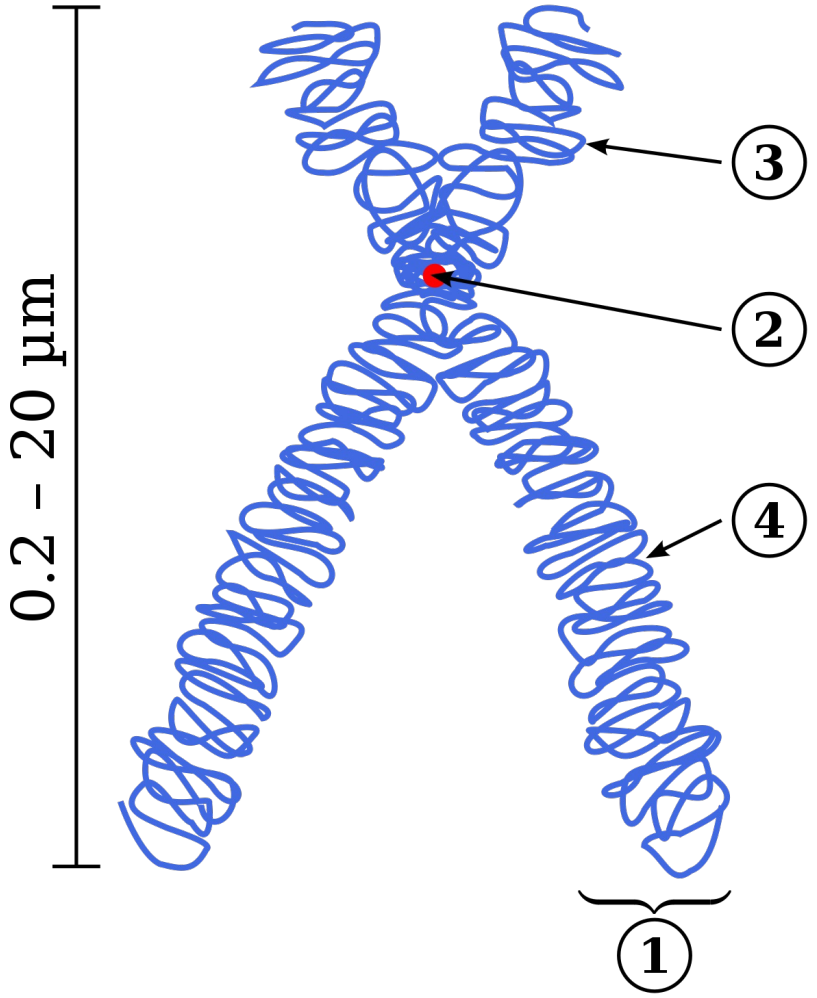


DNA Structure



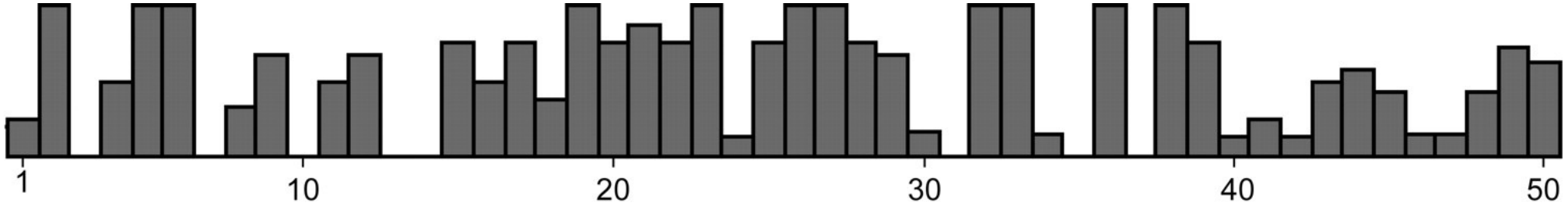
24 Chromosomes, 3 Billion Base Pairs

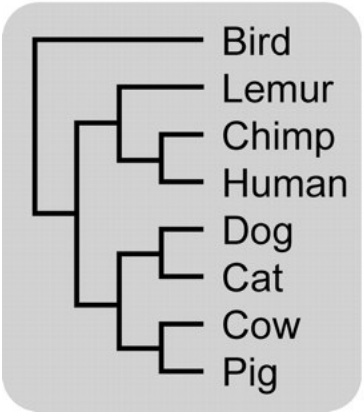
Chromosome	Length (bp)	Length (cm)	Weight (pg)	Weight (fg)	GC%
1	*248,956,422*	*8.14* ± 0.08	*0.25*	*254.57*	41.72
2	242,193,529	7.92 ± 0.08	*0.25*	247.65	40.23
3	198,295,559	6.48 ± 0.06	0.20	202.76	39.67
4	190,214,555	6.22 ± 0.06	0.19	194.49	*38.24*
5	181,538,259	5.93 ± 0.06	0.19	185.63	39.51
6	170,805,979	5.58 ± 0.05	0.17	174.65	39.61
7	159,345,973	5.21 ± 0.05	0.16	162.94	40.70
8	145,138,636	4.74 ± 0.05	0.15	148.41	40.16
9	138,394,717	4.52 ± 0.04	0.14	141.51	41.28
10	133,797,422	4.37 ± 0.04	0.14	136.81	41.54
11	135,086,622	4.42 ± 0.04	0.14	138.13	41.54
12	133,275,309	4.36 ± 0.04	0.14	136.28	40.77
13	114,364,328	3.74 ± 0.04	0.12	116.94	38.55
14	107,043,718	3.50 ± 0.03	0.11	109.46	40.83
15	101,991,189	3.33 ± 0.03	0.10	104.29	42.03
16	90,338,345	2.95 ± 0.03	0.09	92.38	44.58
17	83,257,441	2.72 ± 0.03	0.09	85.14	45.32
18	80,373,285	2.63 ± 0.03	0.08	82.18	39.78
19	58,617,616	1.92 ± 0.02	0.06	59.95	*47.94*
20	64,444,167	2.11 ± 0.02	0.07	65.90	43.80
21	*46,709,983*	*1.53* ± 0.01	*0.05*	*47.76*	40.94
22	50,818,468	1.66 ± 0.02	*0.05*	51.97	47.00
X	156,040,895	5.10 ± 0.05	0.16	159.55	39.53
Y	57,227,415	1.87 ± 0.02	0.06	58.52	40.03
Total (1–22, X, Y)[a]	3,088,269,832	100.96 ± 0.97	3.16	3157.87	40.87



1. Chromatid
2. Centromere
3. Short arm
4. Long arm

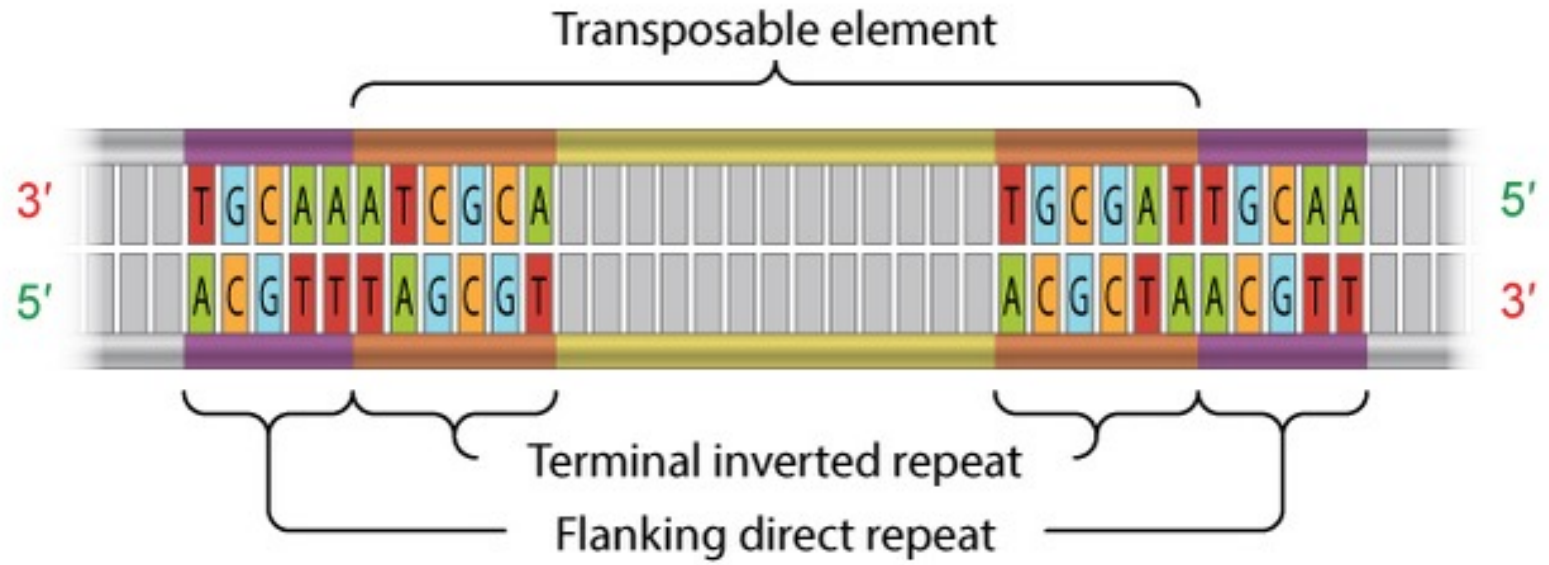
Sequence Alignment and Phylogenetic Trees






```
GGATGCAACTGGTAGTCCCGCGGACGGCTATGCTAGTCTAATCTCTGGCG
AGATGCAACTAGTTGTCTCGCGGACGGC--TGCTAGTCCATCT------A
AGAGGCAGCTGGTTGTCCCACAGACGGCCATGCTAGACCGGTTTCTACAA
AGAGGCACCTGGTTGTCCCGCAGACGGCCATGCTAGACCAGTTTCTACAA
----------TAACATGCGGCACGCGCATGCTAGTCCAATCGAAATCG
----------TAACATGCGGCACGCGCATGCTAGTCCAATTGAAATCG
------TAATATAAGGCACTAGCATGCTTGACGGAGTCCAATGGAGTTCC
------TAATATAAGGCACGCGCCTGCT------AGTCTAATGGAATTCG
```

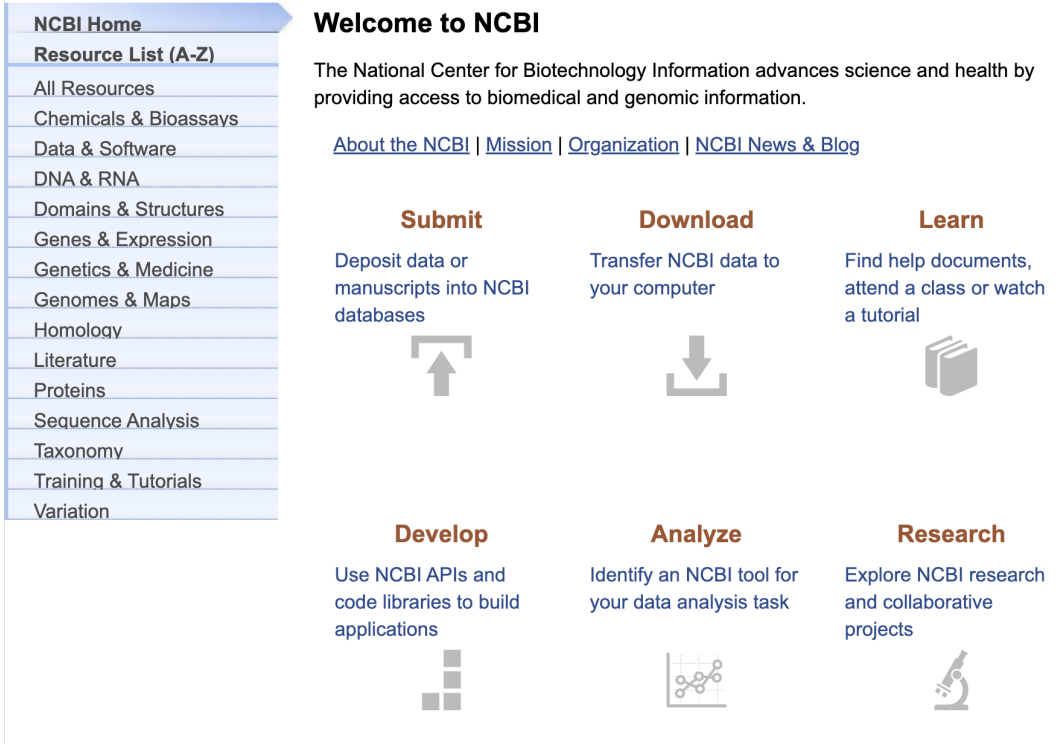
Barbara McClintock and Jumping Genes



1902 – 1992
Winner of the 1983 Nobel Prize in Physiology or Medicine
“for her discovery of mobile genetic elements”

Genomics Resources (and many more)

https://www.ncbi.nlm.nih.gov



https://uswest.ensembl.org/Homo_sapiens/Info/Index