Molecular Evolution
& Phylogenetics

Introduction: historical perspective; Mechanisms of molecular evolution and speciation

Jean-Baka Domelevo Entfellner
Bioinformatics Community of Practice,
BecA-ILRI Hub, September 2018

Learning Objectives

- understand the history of evolutionary thoughts, from 18^{th} to 20^{th} century
- know about the modern concepts of evolution, species and speciation
- understand the different mechanisms of evolution at molecular level
- understand the mechanisms of speciation at population level

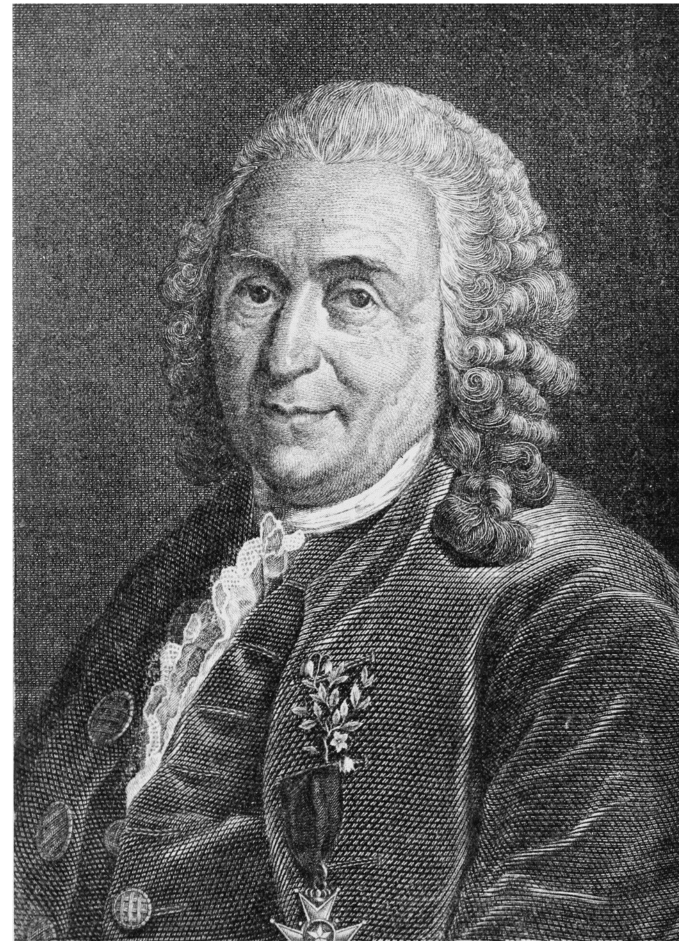
Learning Outcomes

- be able to name a few figures of early evolution theory and identify their contribution
- know what are the different mechanisms involved in evolution
- know what are the different mechanisms leading to speciation

Introduction:
historical perspective

Carl von Linné (Linnaeus), 1707-1778

- Swedish zoologist and botanist
- Father of modern taxonomy (nomenclature & classification of organisms): kingdoms, classes, orders, genera, species, subspecies/varieties.
- Proposed **classification according to physical characteristics**
- Formalised the *binomial nomenclature* used since then to name species (e.g. *Homo sapiens*, *Canis lupus*, *Oryza sativa*)
- ***Systema Naturæ*** (1st ed. 1735, 10th ed. 1758)
- pointed out the relatedness between humans and monkeys/apes
- paved the way for the advent of the theory of evolution



Binomial nomenclature

Gray wolf

Temporal range: Middle Pleistocene - present (810,000-0 years BP[1])

Eurasian wolf (*Canis lupus lupus*).

Conservation status

Extinct | Threatened | Least Concern

EX EW CR EN VU NT **LC**

Least Concern (IUCN 3.1)[2]

Scientific classification

Kingdom:	Animalia
Phylum:	Chordata
Class:	Mammalia
Order:	Carnivora
Family:	Canidae
Genus:	*Canis*
Species:	***C. lupus***

Binomial name

Canis lupus

Linnaeus, 1758[3]

Subspecies

refer Subspecies of *Canis lupus*

Formica rufa

Formica rufa worker

Conservation status

Extinct | Threatened | Lower Risk

EX EW CR EN VU cd **nt** lc

Near Threatened (IUCN 2.3)[1]

Scientific classification

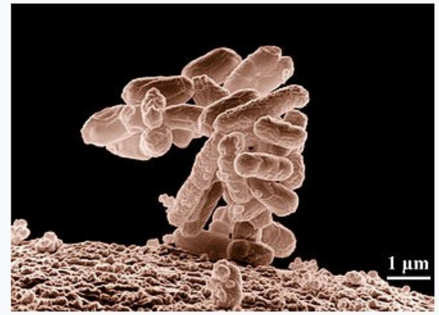
Kingdom:	Animalia
Phylum:	Arthropoda
Class:	Insecta
Order:	Hymenoptera
Family:	Formicidae
Subfamily:	Formicinae
Genus:	*Formica*
Species:	***F. rufa***

Binomial name

Formica rufa

Linnaeus, 1761

Escherichia coli



Scientific classification

Domain:	Bacteria
Phylum:	Proteobacteria
Class:	Gammaproteobacteria
Order:	Enterobacteriales
Family:	Enterobacteriaceae
Genus:	*Escherichia*
Species:	***E. coli***

Binomial name

Escherichia coli

(Migula 1895)

Castellani and Chalmers 1919

Synonyms

Bacillus coli communis Escherich 1885

Brown bear

Temporal range: 0.5–0 Ma

PreЄ Є O S D C P T J K PgN

Middle Pleistocene-Holocene

Kodiak bear (*U. a. middendorffi*) at the Kodiak National Wildlife Refuge in Alaska

Conservation status

Extinct | Threatened | Least Concern

EX EW CR EN VU NT **LC**

Least Concern (IUCN 3.1)[1]

Scientific classification

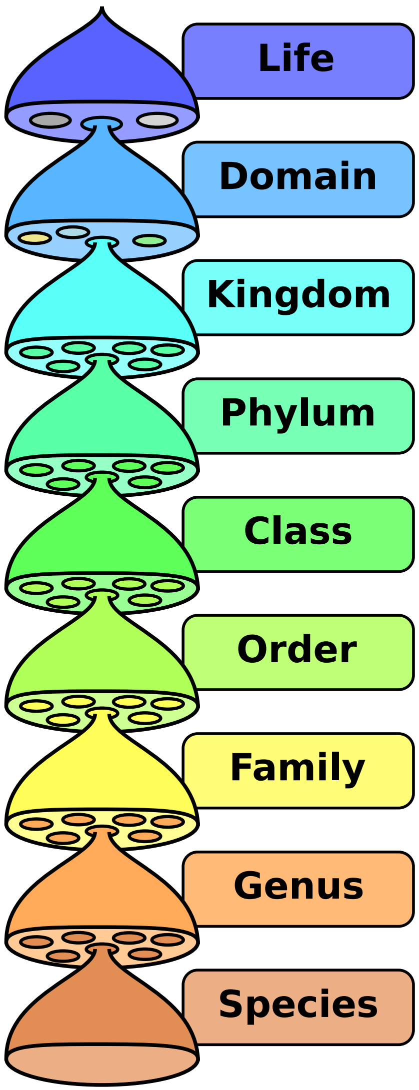
Kingdom:	Animalia
Phylum:	Chordata
Class:	Mammalia
Order:	Carnivora
Family:	Ursidae
Genus:	*Ursus*
Species:	***U. arctos***

Binomial name

Ursus arctos

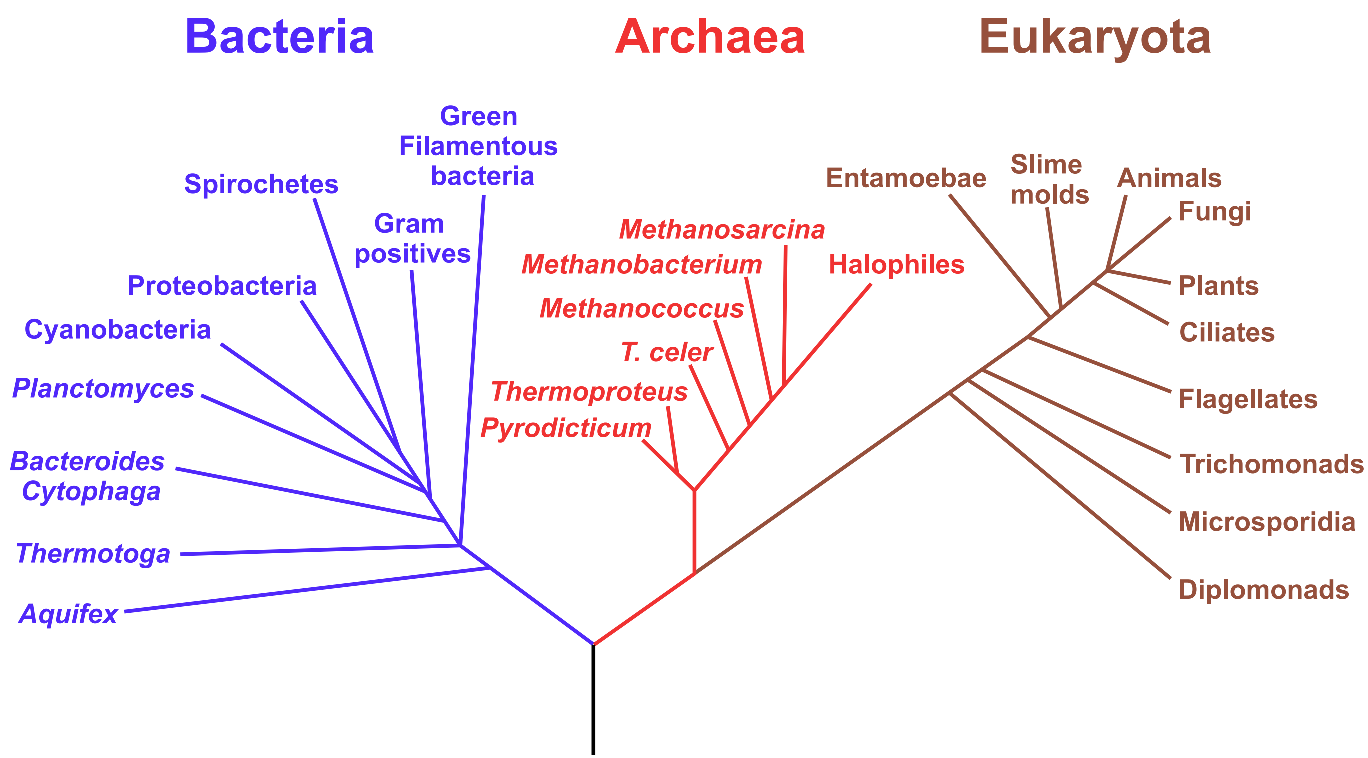
Linnaeus, 1758

Hierarchical classification of Life

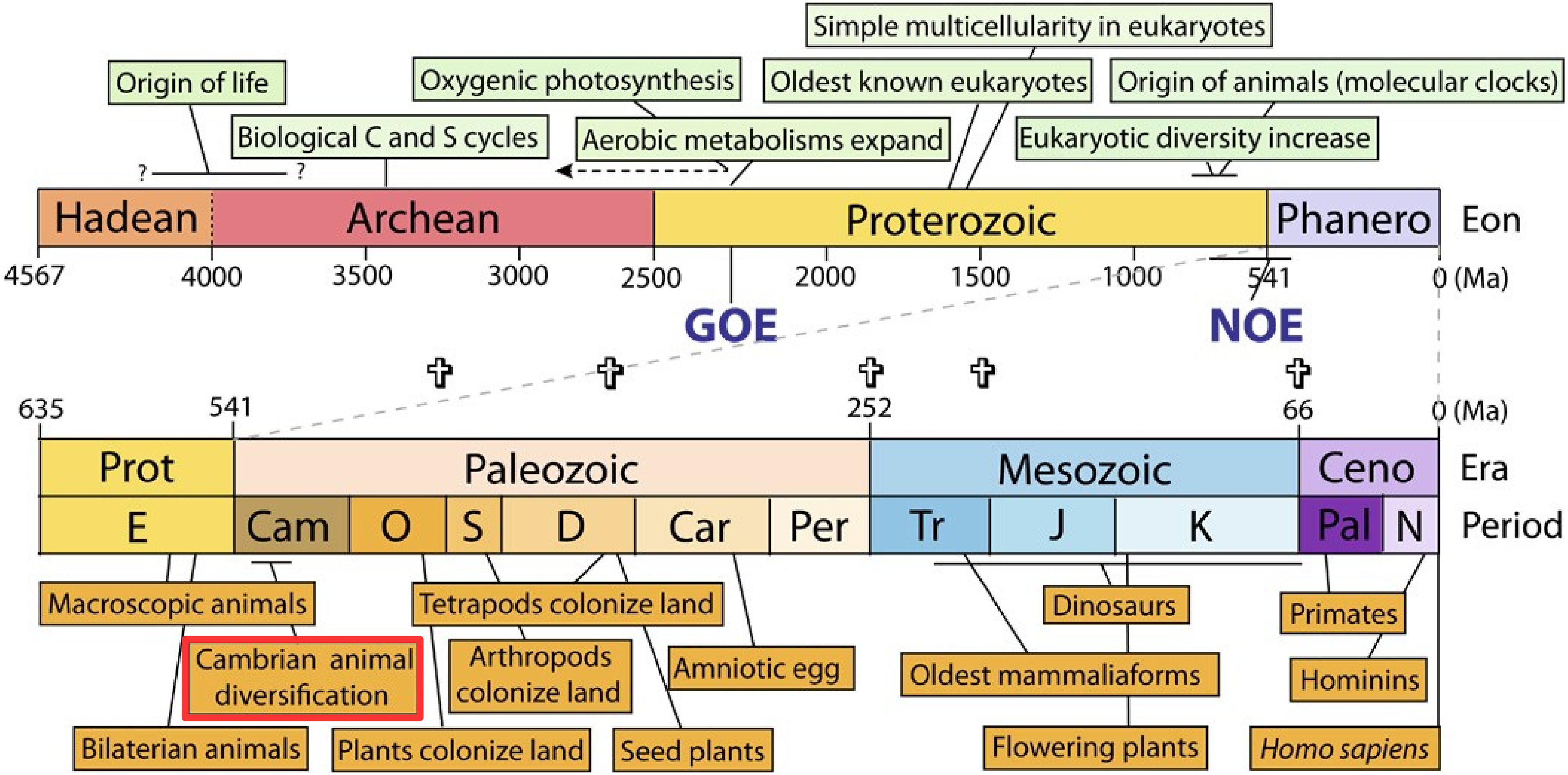


Left: the modern hierarchy of taxonomic ranks

Below: the three-domain classification of Carl Woese, in a phylogenetic tree inferred from the 16S ribosomal RNA



A brief "timetable" of evolution



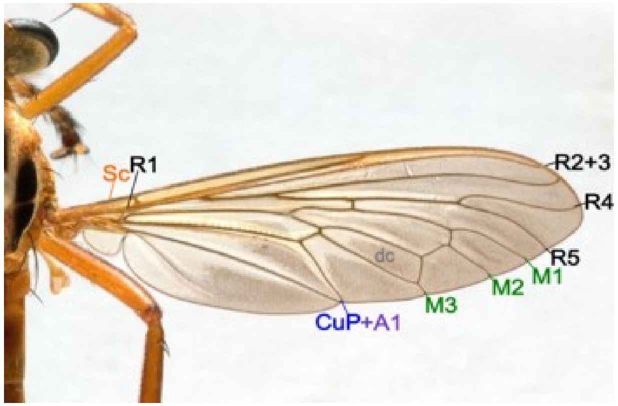
GOE: Great Oxygenation Event. NOE: Neoproterozoic Oxygenation Event

Andrew H. Knoll, and Martin A. Nowak Sci Adv 2017;3:e1603076

ScienceAdvances AAAS

Classification caveat: analogy ≠ homology

Analogy: same function

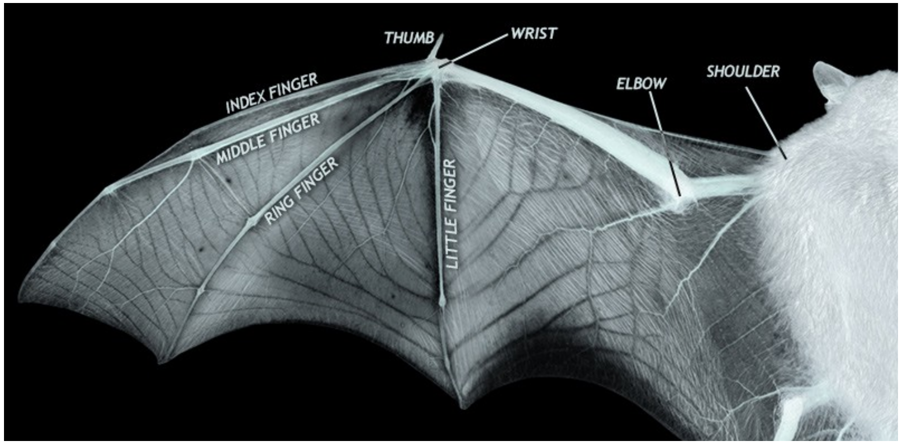


Diptera (fly)

Baird’s sandpiper (shorebird)

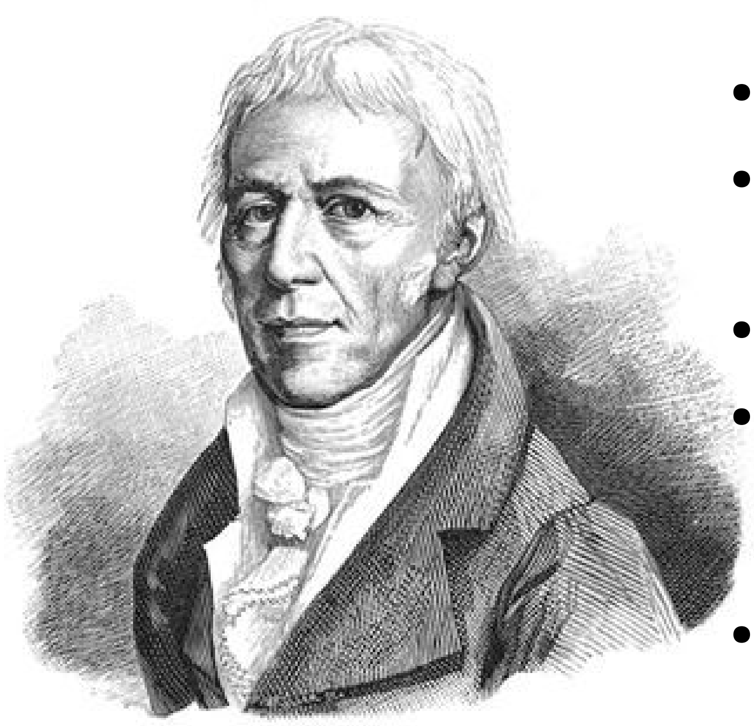
Homology: same evolutionary origin

Human forelimb



Bat’s wing (Chiroptera)

Jean-Baptiste de Lamarck (1744-1829)



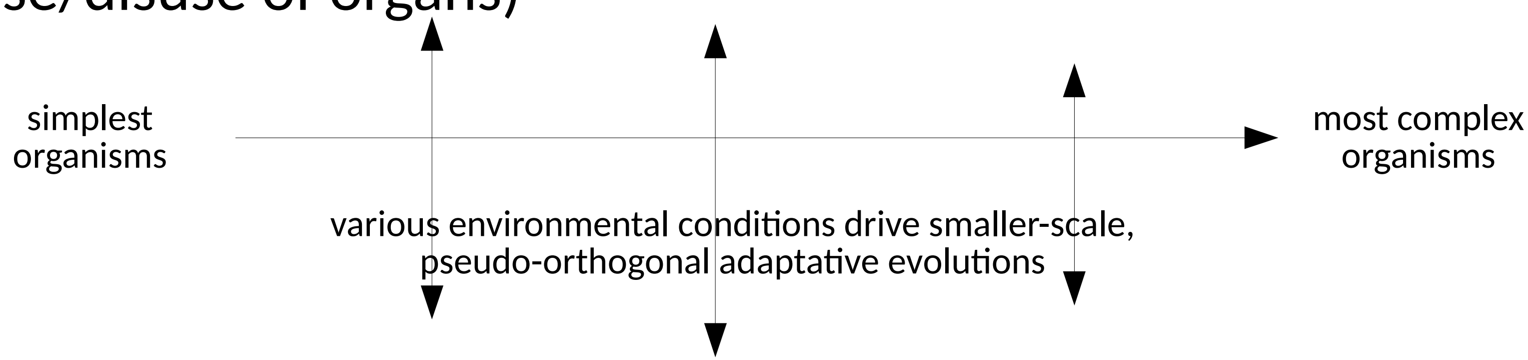
- French soldier, then botanist, academic and naturalist
- *Flore française* (1779)
- *Histoire des animaux sans vertèbres* (1815)
- Coined the term “biology”
- Strong interest in what tells apart living systems from mineral/inorganic stuff (“force vitale”)
- influenced by 17th- and 18th-century concepts of mechanical/alchemical biology (vital force, fluids…)
- influencial (?) to Darwin’s future work

Lamarckian evolution: a theory

Lamarck viewed evolution as a process driven by 2 main forces:

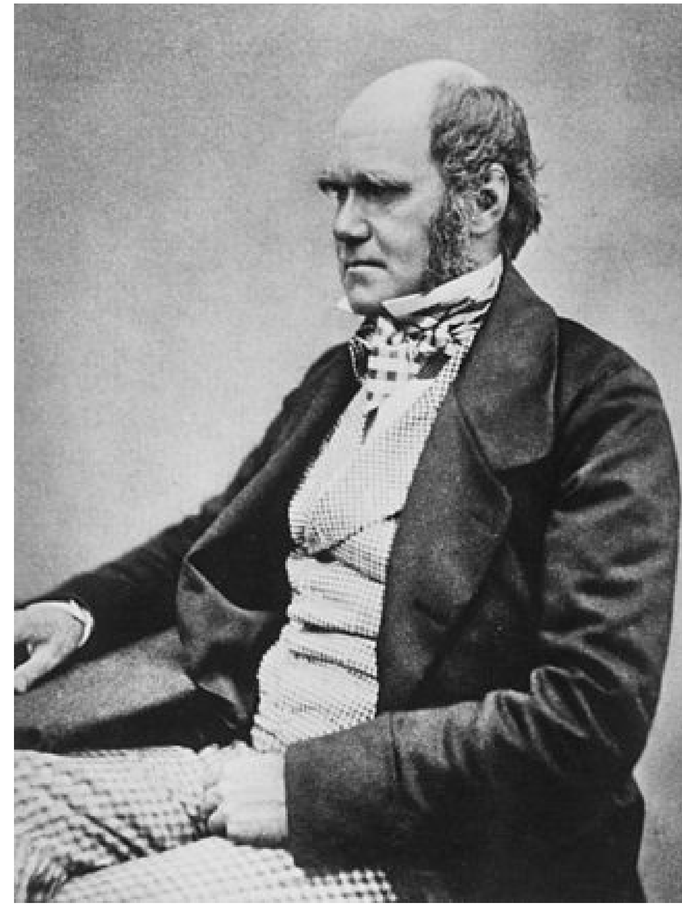
1) organisms evolve **towards greater complexity** from most basic organisms produced by **spontaneous generation** ("infusoires")

2) a second driving force make them evolve and diversify **to adapt to their various environments** (e.g. through use/disuse of organs)

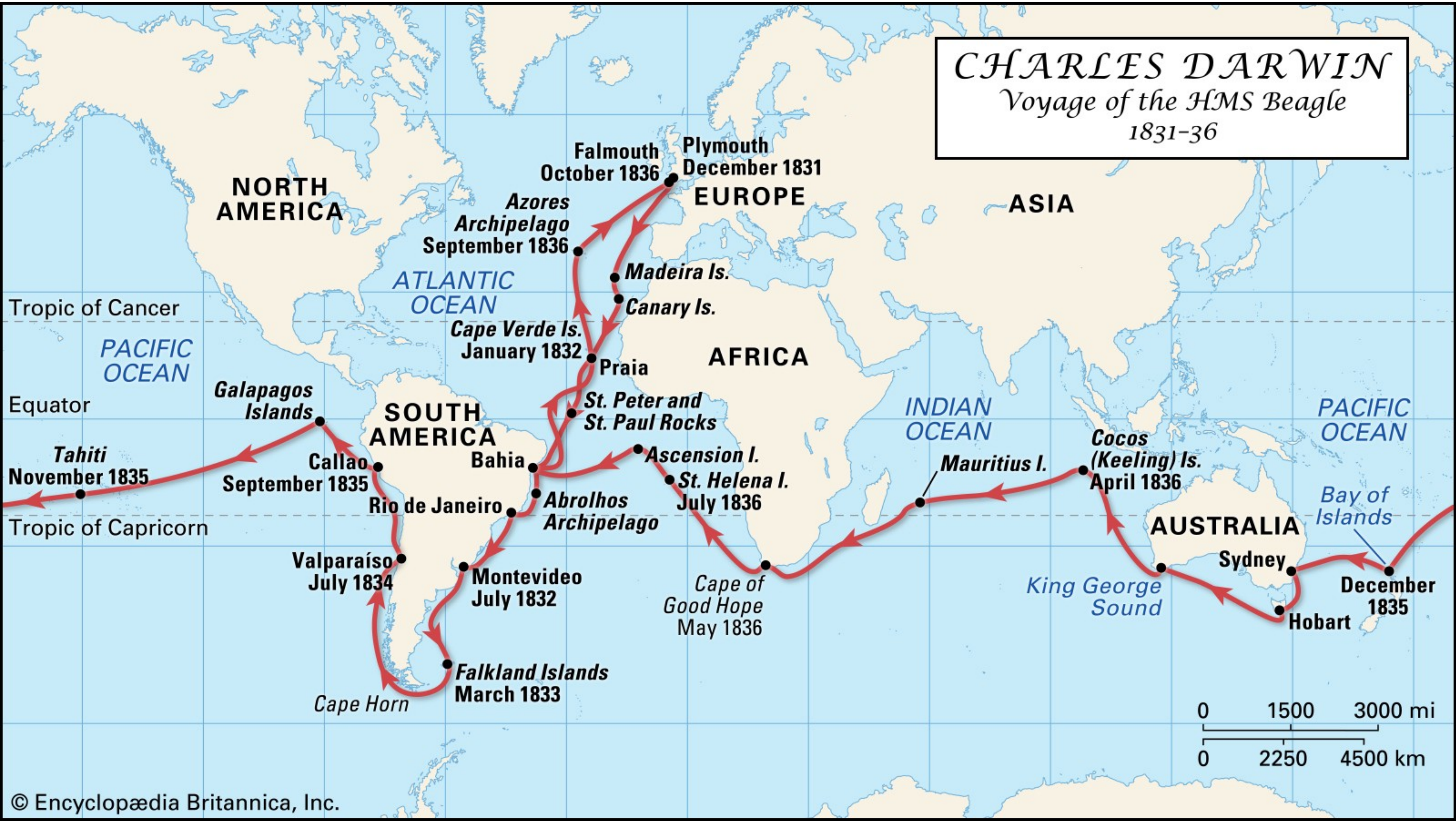


Charles Robert Darwin, 1809–1882

- British geologist, taxonomist, academic and naturalist
- drew significant inspiration from his trip around the world on the *HMS Beagle* (1831–36)
- *The Voyage of the Beagle* (1839)
- ***On the Origin of Species* (1859)**: publishing evidence to support his new **theory of evolution**
- *The Descent of Man* (1871)
- father of the term “natural selection”
- backed the idea of evolution as **descent with modification**
- made a strong case for **common descent** (a single ancestor to all beings)



Second voyage of the *HMS Beagle* (1831–36)



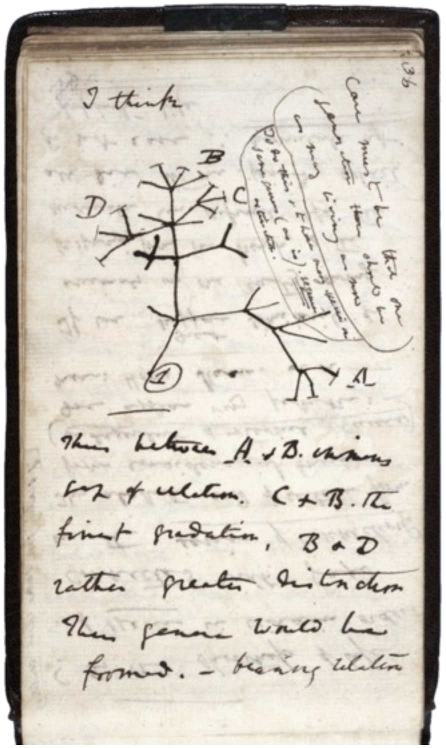
Darwinian evolution: natural selection

The concept of natural selection is key to Darwin's theory of evolution:

"As many more individuals of each species are born than can possibly survive; and as, consequently, there is a frequently recurring **struggle for existence**, it follows that any being, if it **vary however slightly in any manner profitable to itself**, under the complex and sometimes **varying conditions of life**, will have a better chance of surviving, and thus be **naturally selected**. From the strong **principle of inheritance**, any selected variety will tend to propagate its new and modified form."

On the Origin of Species (1859)

Darwinian evolution: transmutation of species



Darwin’s “*B notebook*” on the “transmutation of species” (1837)

- some species are extinct
- species are not fixed
- distinct but close species are differentially distributed e.g. over islands (observations on rheas, Galapagos mockingbirds and finches)
- possible that “one species does change into another” (speculation in Darwin’s *Red Notebook*)
- variation in offspring “to adapt and alter the race to changing world”
- character variation and fixation of changes in a population provide the mechanisms of **speciation**

Darwinian evolution: common descent

Darwin was the first to get the intuition that maybe all species derived from a single ancestor:

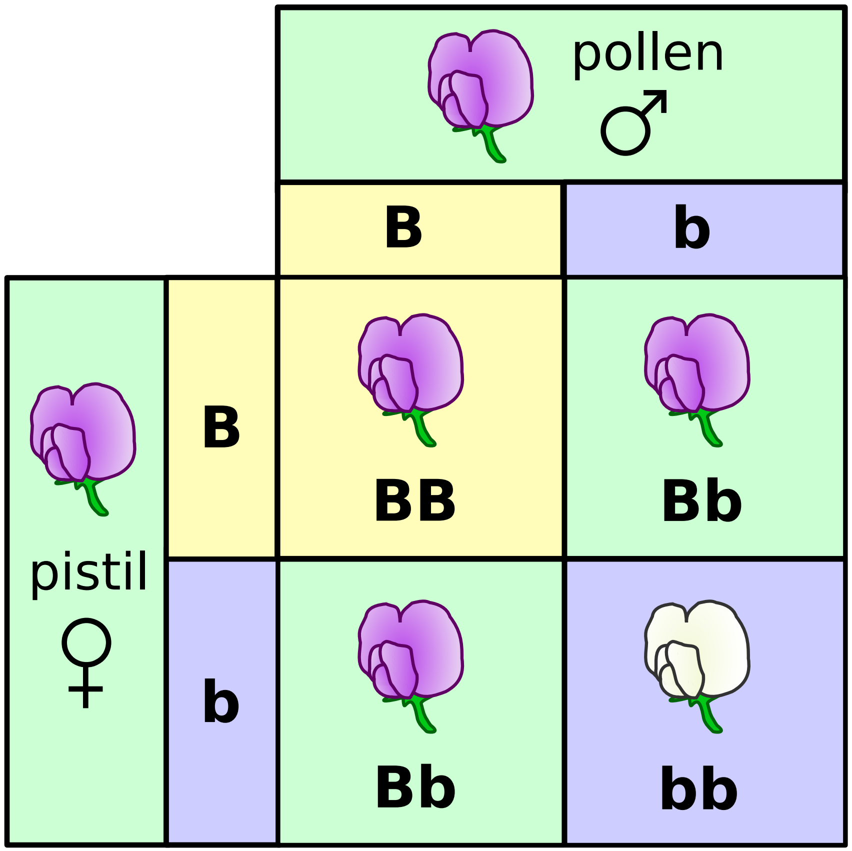
"There is grandeur in this view of **life**, with its several powers, **having been originally breathed into a few forms or into one**; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless **forms** most beautiful and most wonderful **have been, and are being, evolved**."

On the Origin of Species (1859)

Darwinian evolution: a few remarks

- **struggle for life** stems from the fact that more individuals come to life than can possibly survive long
- “It is absurd to talk of one animal being higher than another” → evolution does not yield “higher”, “better” beings down a one-way alley. It is a bushy, **omnidirectional process**.
- evolution has **no fixed goal**
- **human beings are animals** (continuity of characters)
- sexual selection plays a role in evolution
- diseases also play a role (negative selection by death)
- Darwin's breakthrough opened the way for modern molecular evolutionary science

Mendelian inheritance (1865/1866)



Punnett square for a single-locus trait: allele *B* is dominant while *b* is recessive

- **laws formulated** by Gregor Johann Mendel after hibridization experiments with pea plants
- birth of modern genetics and the notion of "allele". Wait almost 100 years until Watson & Crick unveil the structure of DNA (1953)
- two alleles for each trait in diploids, which **segregate** during meiosis (First Law)
- alleles are **passed independently** from each other from parents to offsprings (Second Law)
- important observation: no one-to-one association between phenotype (visible traits) and genotype ("invisible" information)

20th century: the advent of computational evolution

L.L. Cavalli-Sforza (standing) & A. Edwards, 1963

- 1950s: numerical clustering methods for classification and phylogenetic inference (Sokal & Sneath)
- 1953: F. Sanger achieves first protein sequencing
- same year: Watson & Crick publish the double-helix structure of DNA
- 1955: discovery of the DNA polymerase (Kornberg, Lehman & al)
- 1960s: phylogenies from gene frequencies, parsimony methods, likelihood methods (Edwards & Cavalli-Sforza)
- 1960s & 1970s: scientific computers and mini-computers become affordable (IBM 7040, PS/900…)

Cavalli-Sforza & Edwards 1967

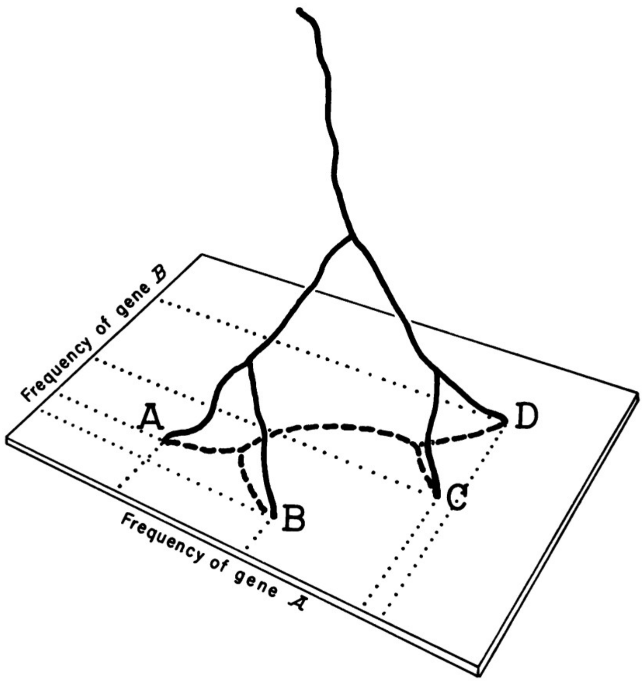


Fig. 1. An evolutionary tree and its projection onto the "now" plane.

from Cavalli-Sforza & Edwards, *American Journal of Human Genetics*, 19(3), May 1967

- studying **gene frequencies** among various populations
- data: current gene frequencies for genes $\mathcal{A}$ and $\mathcal{B}$ in pop. A, B, C and D
- problem: infer the evolutionary tree in space and time
- necessary problem-solving step: make hypotheses about mode and speed of evolution (**model**)
- authors go through a comparison of two approaches: **minimum evolution** and **maximum likelihood**

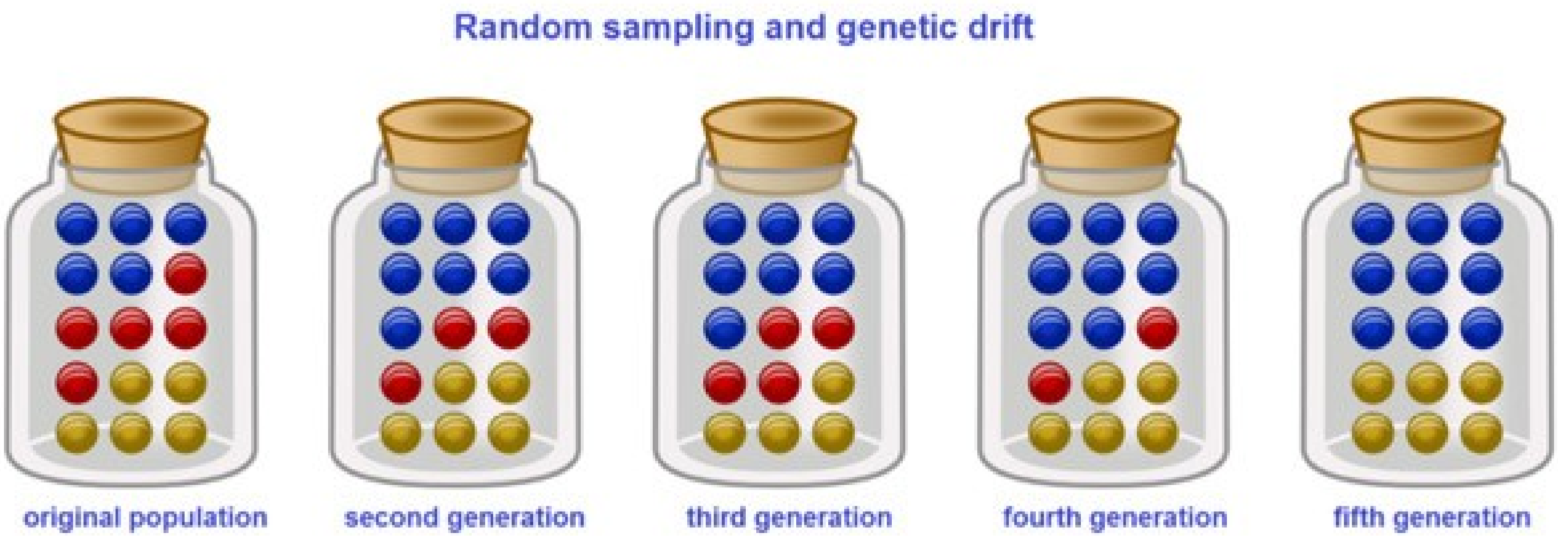
Mechanisms
of molecular evolution
&
speciation

Various mechanisms drive evolution

- **genetic drift**: random sampling from the gene pool of generation *n* to the genetic material of generation *n+1* (rate of the random walk depends on the effective size of the population)
- **selection**: processes of natural (adaptation to the environment or lack thereof) or sexual selection (bias in mating patterns in favor of individuals having certain positive features)
- **migration**: geographic migration of groups of individuals to settle elsewhere, forming subpopulations
- **mutation/recombination**: pseudo-random alteration of the genetic material during the creation of an offspring (inheritance with modification). Examples: nucleotide substitutions, chromosomal crossover, duplications and gene translocation, etc.

Random Genetic Drift

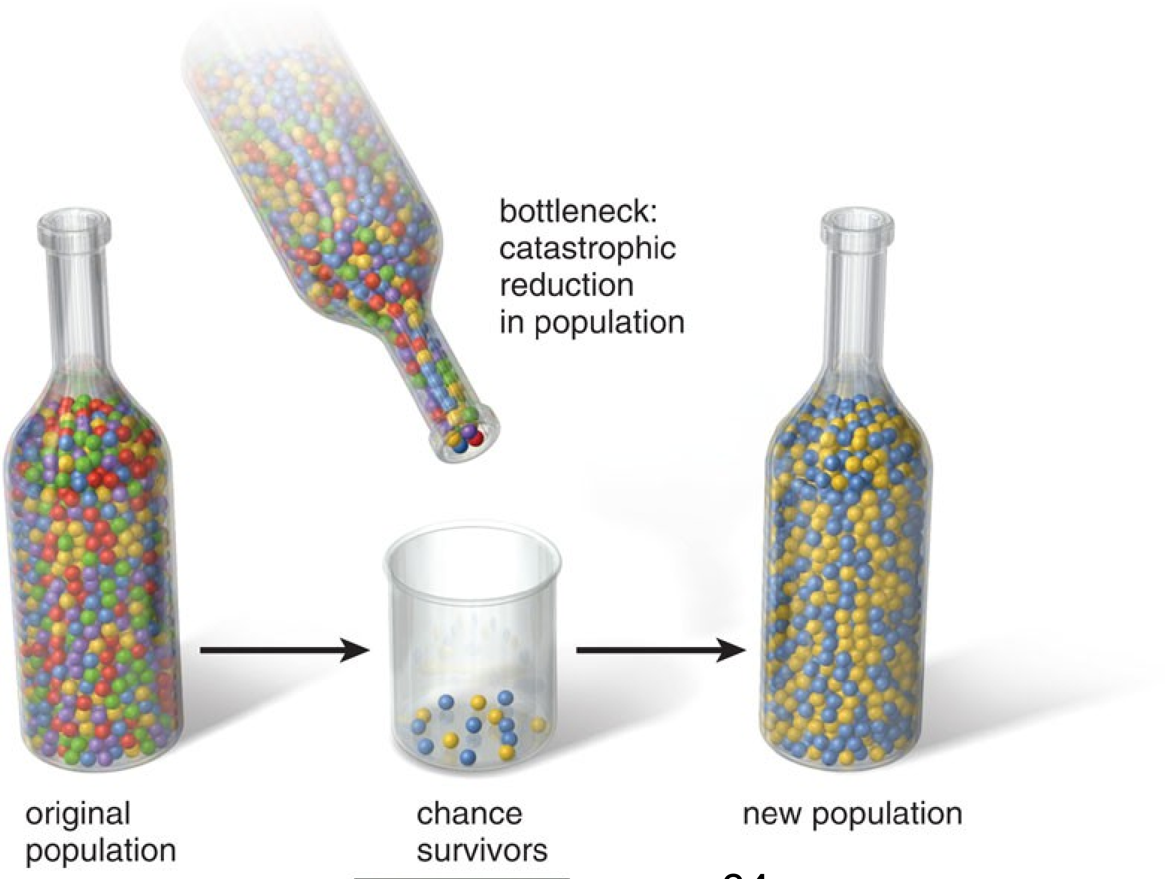
By chance only, in a constant-size population model, some individuals from generation *n* leave more offspring than others to generation *n+1*, hence a modification in allele frequencies over generations.



source: http://www.darwinwasright.org/genetic_drift.html

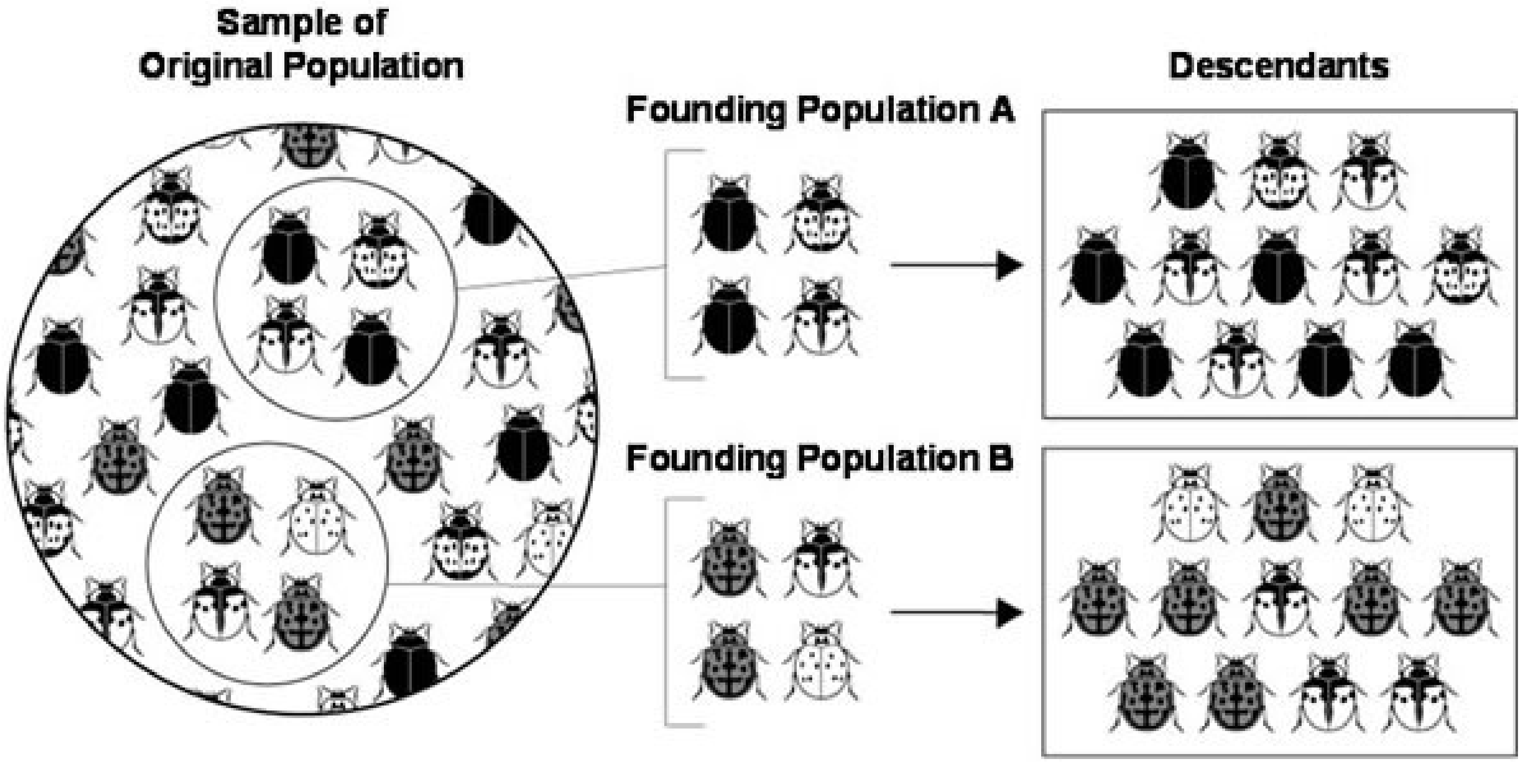
Genetic Drift: Bottleneck Effect

When a drastic decrease in population size occurs (e.g. massive hunting, ice age, severe environmental change), only a few individuals survive, hence a loss in genetic diversity.



Genetic Drift: Founders Effect

When a small subgroup parts to found a new population. The new pop has less genetic diversity than the original pop.



Natural Selection Principles

Hypotheses:

- many more individuals come to life than can survive
- competition for resources, mating partners, etc
- intra-specific genetic diversity
- the **fitness** of an individual as a function of their genetic makeup and their environment

Conclusions:

- the individuals will have offspring (generation *n+1*) in proportion to their own fitness
- in generation *n+1*, the gene pool of a species is more biased towards increased fitness genes than in generation *n*

Natural Selection in Action (1/3)

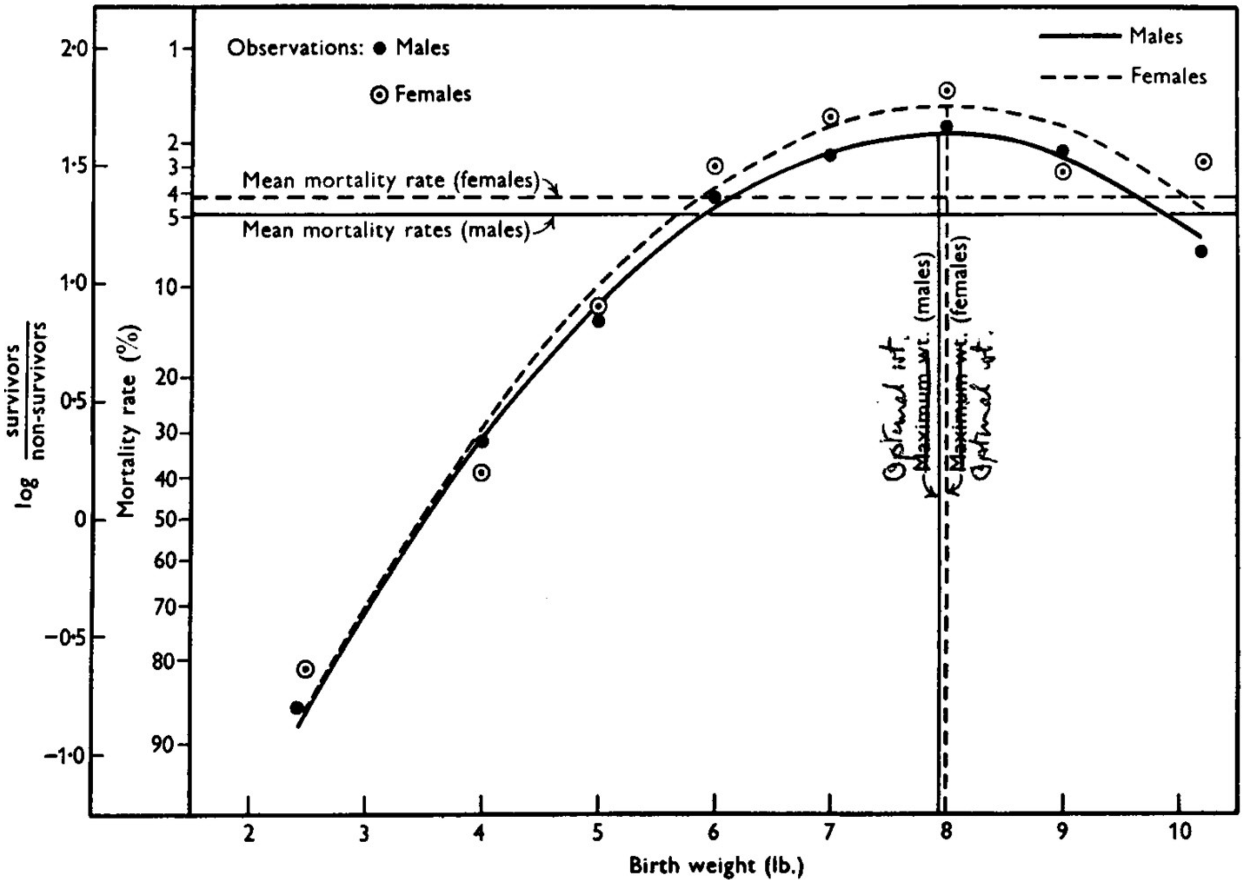
Purifying selection: “A résumé of investigations of the evolution of melanism in the *Lepidoptera*” (HBD Kettlewell, PRSL, 1956)

Biston betularia typica (left) and *B. b. carbonaria* (right)

- birds eat both forms of moths
- better camouflage means increased fitness
- **natural selection** patterns depend on the nature of the **environment** (relative darkness of resting places)
- the industrialization of parts of England led to increased fitness for the melanic form
- revesible selection pattern

Natural Selection in Action (2/3)

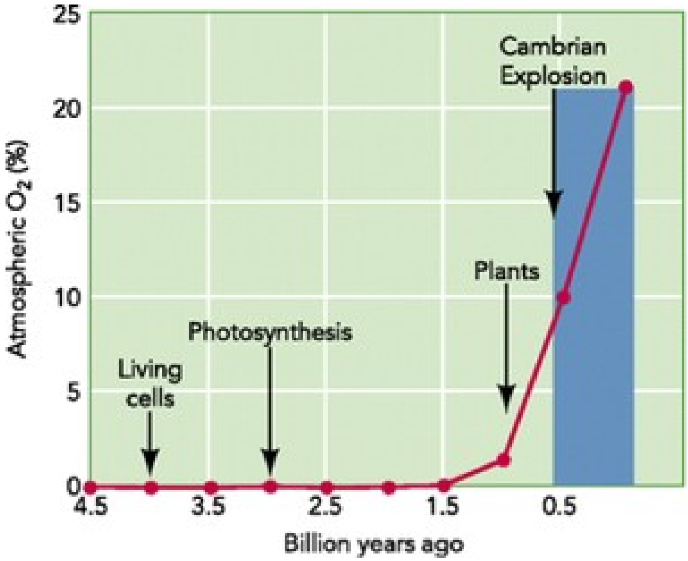
Stabilizing selection: birth weight and infant survival (ML Karn and LS Penrose, Annals of Eugenics 1951)



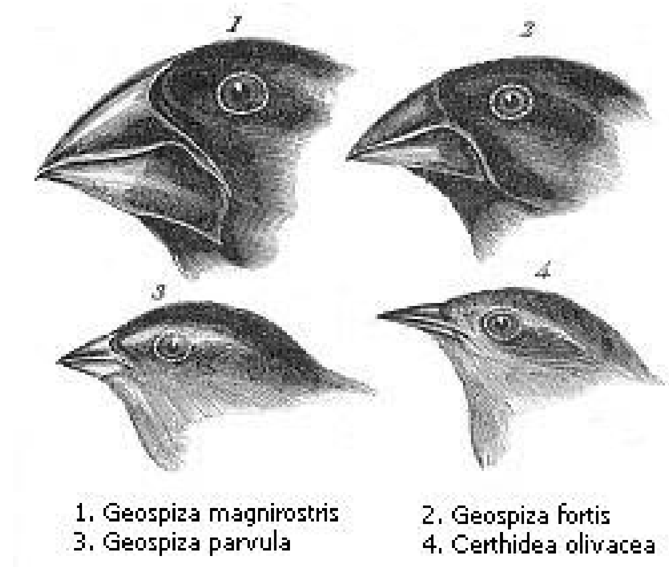
- mortality includes stillbirths and 28day non-survivors
- mean birth weight differs from optimal value
- **natural selection** favours large enough, healthy babies
- infants too heavy are selected against (higher risks of birth injury or maternal death)
- normalizing selection force shown to be relaxed in wealthy **environments** with improved med. care (Ulizzi & Terrenato 1992)

Natural Selection in Action (3/3)

Adaptive radiation: in certain conditions (e.g. after a bottleneck), rapid emergence of species adapted to different environments



The Cambrian explosion led to the advent of most metazoans (animals) in a relatively short period of time.

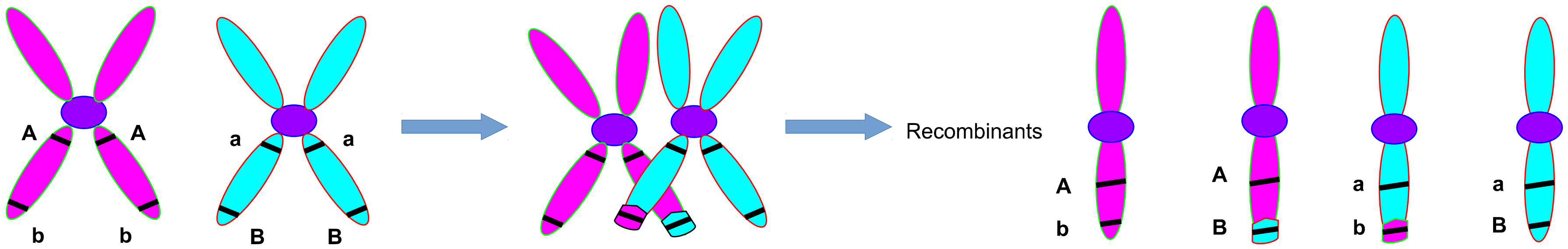


Finches from Galapagos Archipelago

Observed by Darwin during his voyage on the Beagle, these birds evolved into separate species to better adapt to different environmental conditions.

Chromosomal crossover during meiosis

- key mechanism to ensure that organisms in generation *n+1* are made of a unique makeup from the genetic material of their two parents (genetic diversity)
- takes place during meiosis in diploid organisms using sexual reproduction, when haploid cells (gametes) are formed from diploid cells
- genetic diversity is key to adaptation (plasticity of species)



By Abbyprovenzano - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=26261468

What is a species?

Historical definition by Ernst Mayr (1942), widely accepted:

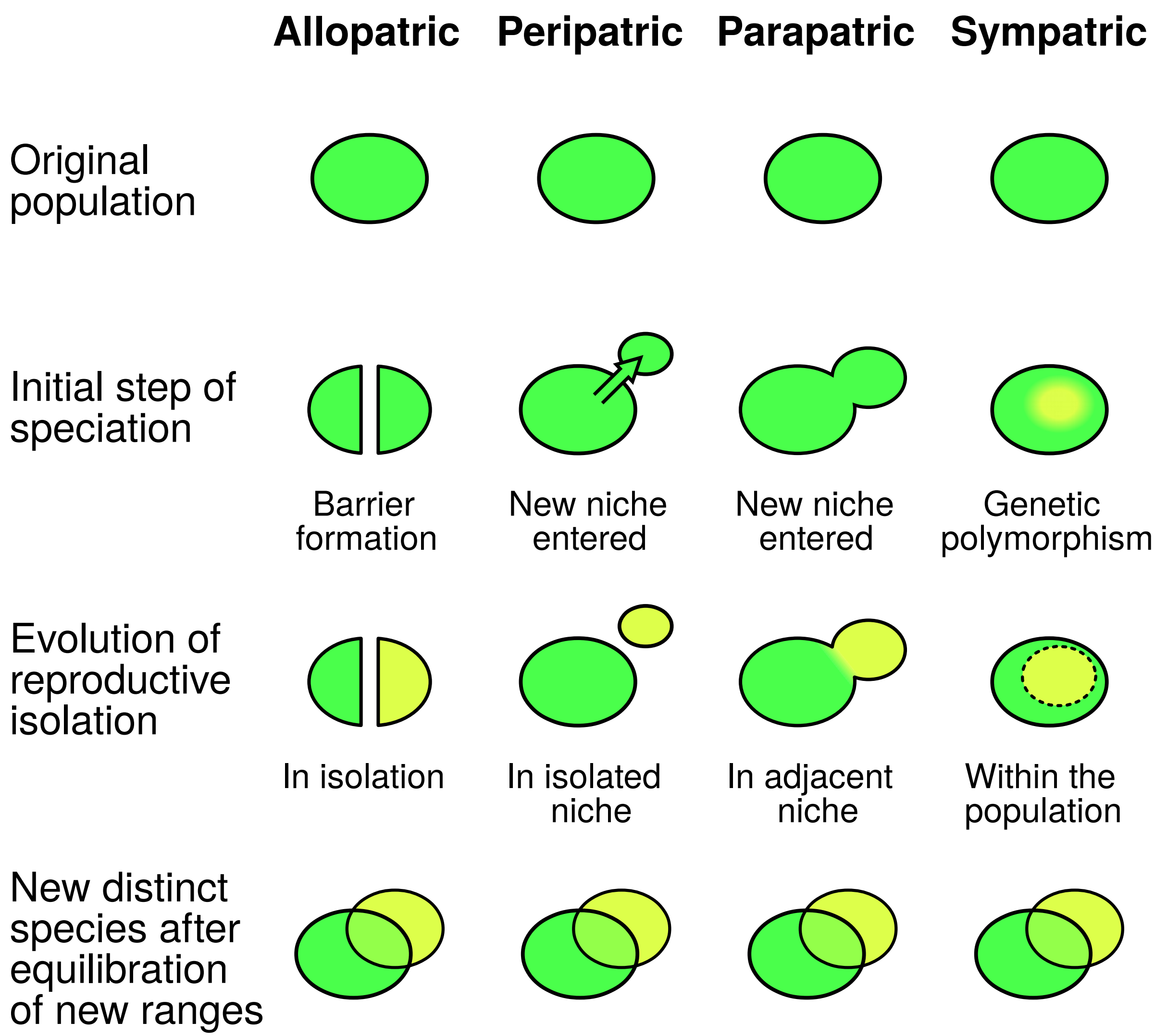
- A species is formed of a set of individuals who can **interbreed** to give **viable and fertile offspring**, while not being able to breed and give such offspring with members of other species.

  → key concept of **reproductive isolation**

When geographical isolation or artificial selection of traits provide significant phenotypic variation without preventing viable interbreeding, we talk about **subspecies (animals) or varieties (plants)**.

Example: wolf (*Canis lupus*) and dog (*Canis lupus familiaris* or *Canis familiaris*) do not (yet) form two distinct species.

How does speciation occur?



(source: Wikipedia)

Science

REPORTS

Cite as: S. Lamichhaney *et al.*, *Science* 10.1126/science.aao4593 (2017).

Rapid hybrid speciation in Darwin's finches

Sangeet Lamichhaney,[1]* Fan Han,[1] Matthew T. Webster,[1] Leif Andersson,[1,2,3]† B. Rosemary Grant,[4] Peter R. Grant[4]

[1]Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden. [2]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden. [3]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX, USA. [4]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA.

*Present address: Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA.

†Corresponding author. Email: leif.andersson@imbim.uu.se

Homoploid hybrid speciation in animals has been inferred frequently from patterns of variation, but few examples have withstood critical scrutiny. Here we report a directly documented example from its origin to reproductive isolation. An immigrant Darwin's finch to Daphne Major in the Galápagos archipelago initiated a new genetic lineage by breeding with a resident finch (*Geospiza fortis*). Genome sequencing of the immigrant identified it as a *G. conirostris* male that originated on Española >100 km from Daphne. From the second generation onwards the lineage bred endogamously, and despite intense inbreeding, was ecologically successful and showed transgressive segregation of bill morphology. This example shows that reproductive isolation, which typically develops over hundreds of generations, can be established in only three.