Welcome to Comparative Genomics!

The purpose of this course is to familiarize informatics students with methods and theories behind comparative genomics.

- We will focus on the analysis of genomic sequences, including sequencing technologies, genome assembly, functional and structural predictions, and comparisons at macroevolutionary timescales in a phylogenetic context
- Special emphasis is on building practical skills in computational biology.

Course webpage: https://github.com/marctollis/INF415-515-Comparative-Genomics/

Class Hours: MoWeFri@ 12:40-1:30pm, FrankeCollBusiness Rm 345

Tollis Office Hours: TMoWeFri 2:00-2:30pm
https://nau.zoom.us/j/88304470430
Meeting ID: 883 0447 0430
Password: 816204

In-person format, NAU Flex-enabled.

The work

Readings

- Primary literature, assigned for most classes
- Everyone does one 10 min Presentation, lead class discussion
- If there are too many students, I can assign papers

Lectures

- Based on the module topics

Quizzes

- Based on both lectures and readings
- 12 total (low stakes), drop lowest two grades

Labs

- Focusing on computing and commonly used tools in genomics

Projects and Presentations

- Present a project based on course topics that will help your thesis
- When it's your turn to present, upload your PP to the assignment in the Content Area

Course logistics

Class time will be for:

- Lectures, presentations, discussion
- Labs - instructor goes line by line with the code, you emulate, we work on troubleshooting
  - Schedule is somewhat open, to add some breathing room
  - Labs are due a few weeks so you have time to go back and work on them.

Resources: computing

- Bring laptop to each class (duh)
- Terminal, command line, Monsoon
- Fill out computing survey
- Consider software carpentry link if you have no experience

Prerequisites and Resources

Prerequisites

- No course requirements
  - but general BIO, genetics/evolution will help
- Access to a laptop
- Familiarity with UNIX commands, but we will help

Texts

- no textbooks, we will study primary literature

Other resources:

- Google, SEQanswers, Biostars, StackOverflow
- https://github.com/quinlan-lab/applied-computational-genomics
- http://www.langmead-lab.org/teaching-materials/

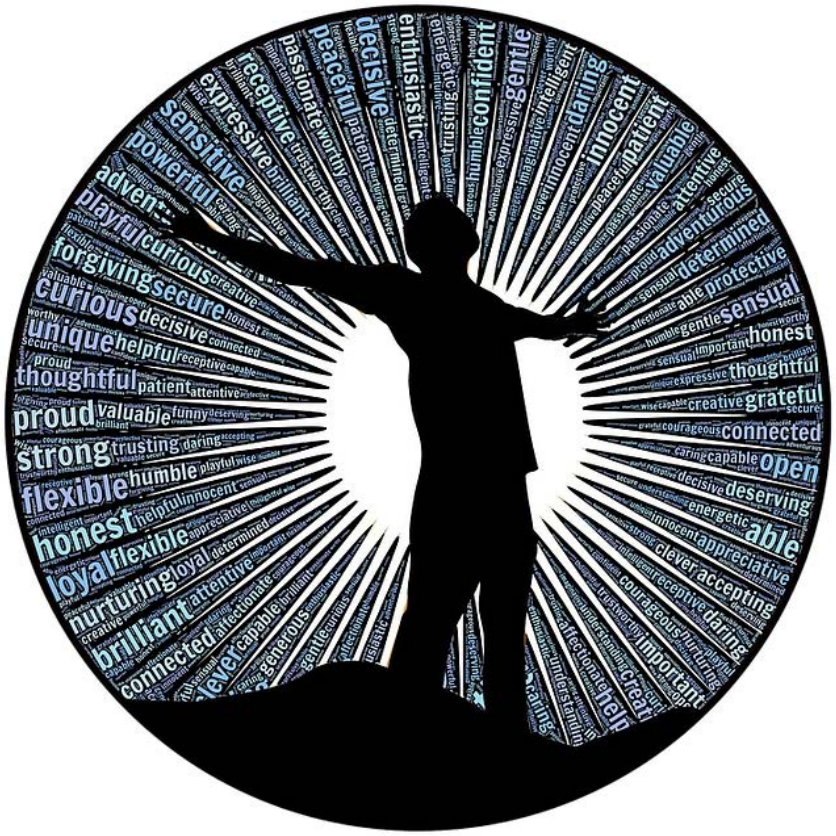
INF 515

Assignment	Weight	Assignment Points (1000 total)
Attendance	10%	100
Participation	10%	100
Paper Discussion	10%	100
Computational Labs	30%	300
Project Proposal	2%	20
Quizzes	20%	200
Final Project	18%	180

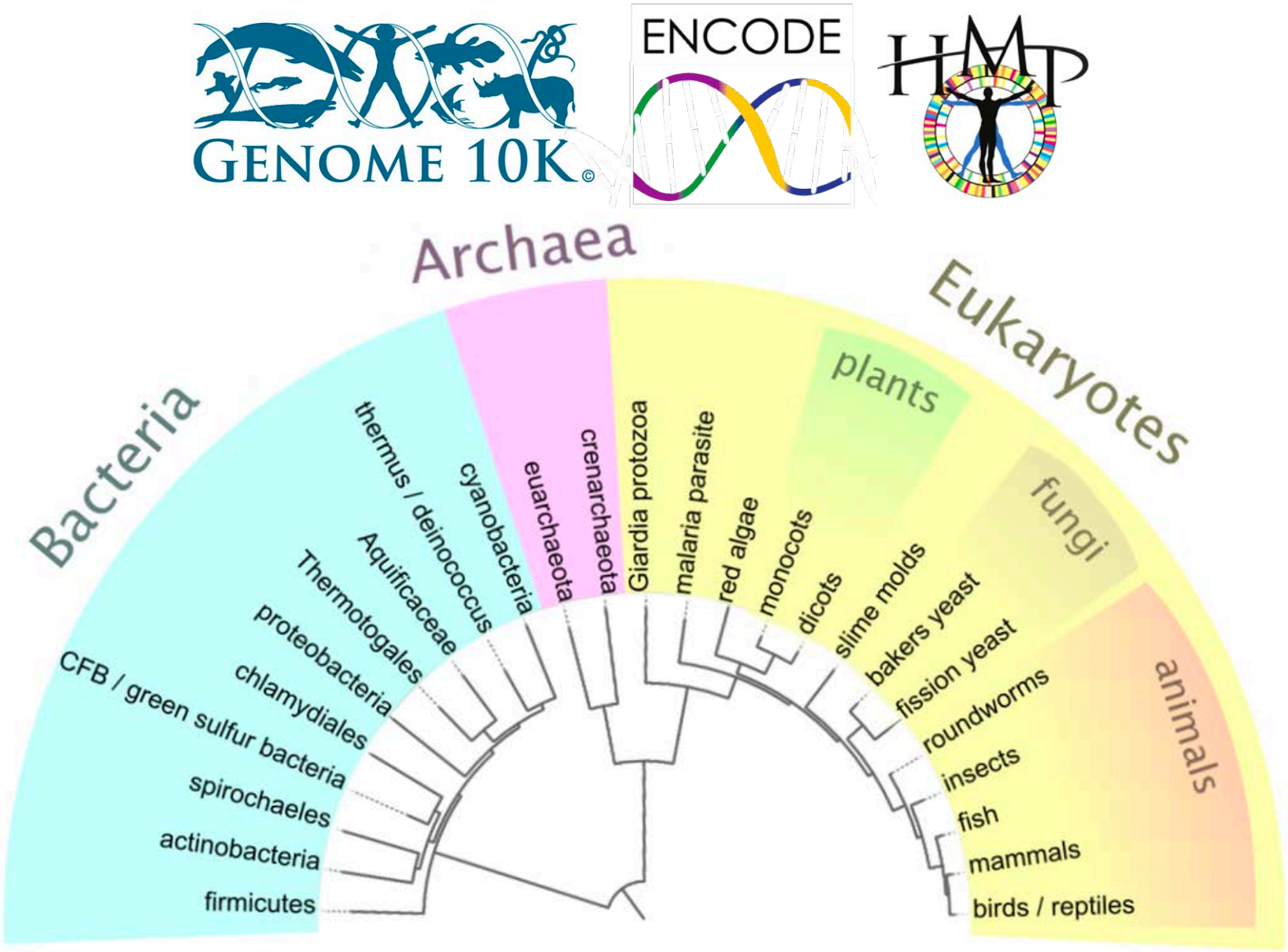
DNA: The Building Blocks of Life

DNA combines with environment to shape who you are

- Height, hair, eye color, skin color, shape
- Disease susceptibility and response to treatment
- Life and death: longevity, cognition

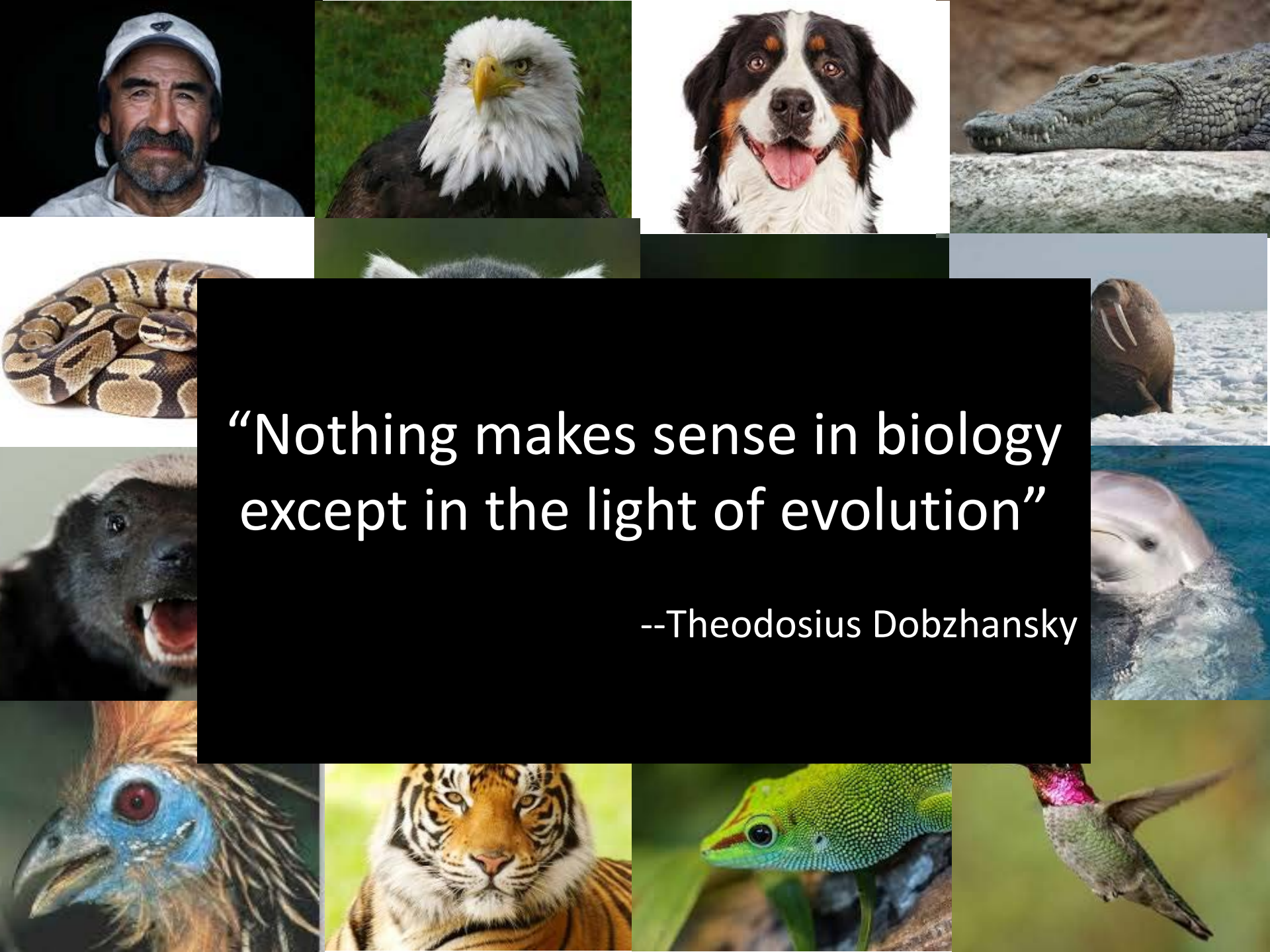


Across the Tree of Life



prokaryotes

eukaryotes



"Nothing makes sense in biology except in the light of evolution"

--Theodosius Dobzhansky

Evolution by Natural Selection

On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection.

Darwin CR, Wallace AR. 1858. *Journal of the Proceedings of the Linnean Society of London.*

On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.

Darwin, CR. 1859. Londn, John Murray.

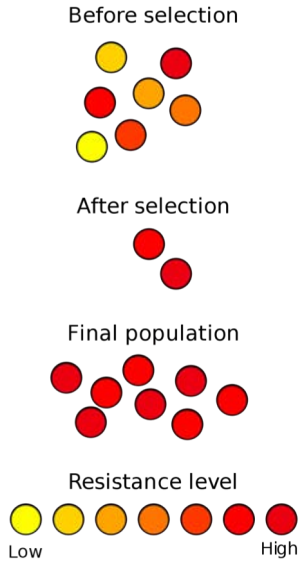
Charles Darwin, age 51

Variations in form are heritable and result in ***differential reproductive success***.

These successful variations are perpetuated and increase in frequency with each subsequent generation.

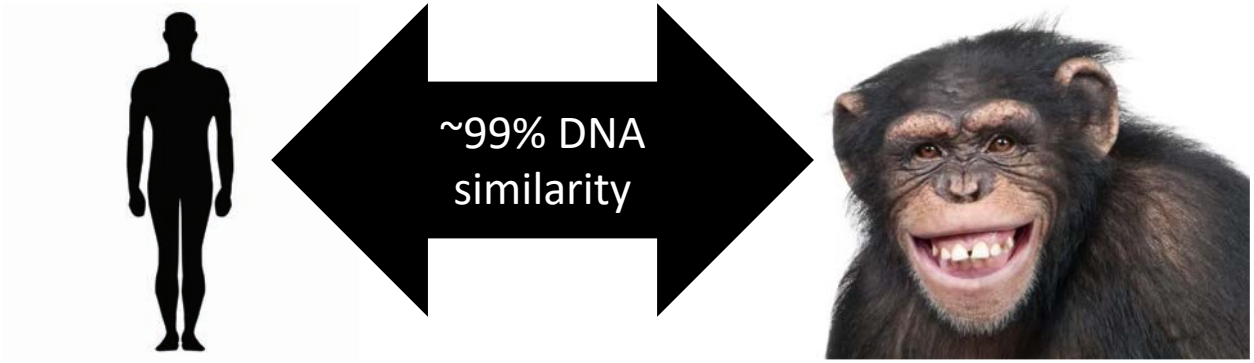
Differences between populations grow over time until they are reproductively isolated – ***forming new species***.

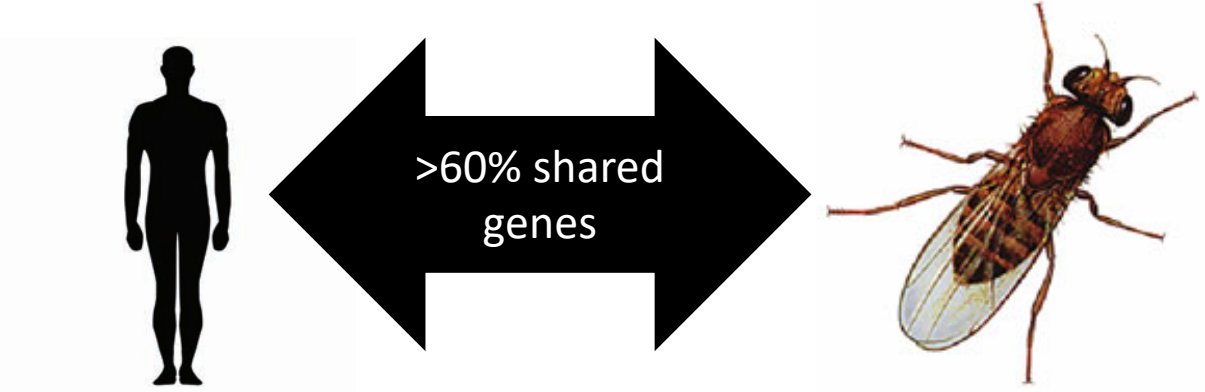
Also key: all lifeforms have descended from a ***common ancestor***.



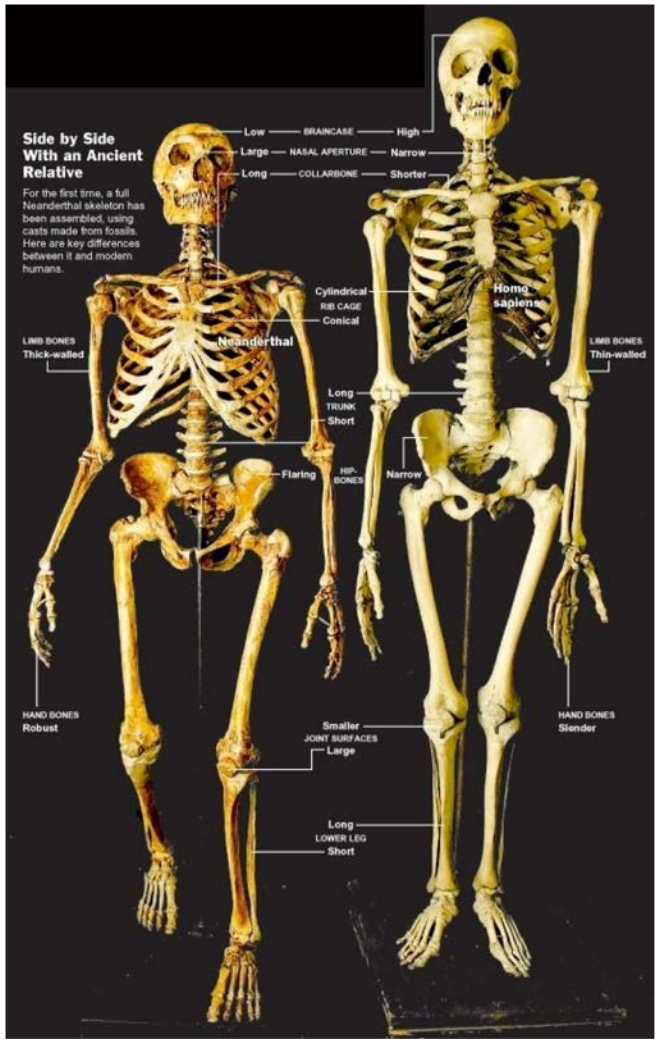
photos from wikipedia

What is Comparative Genomics?



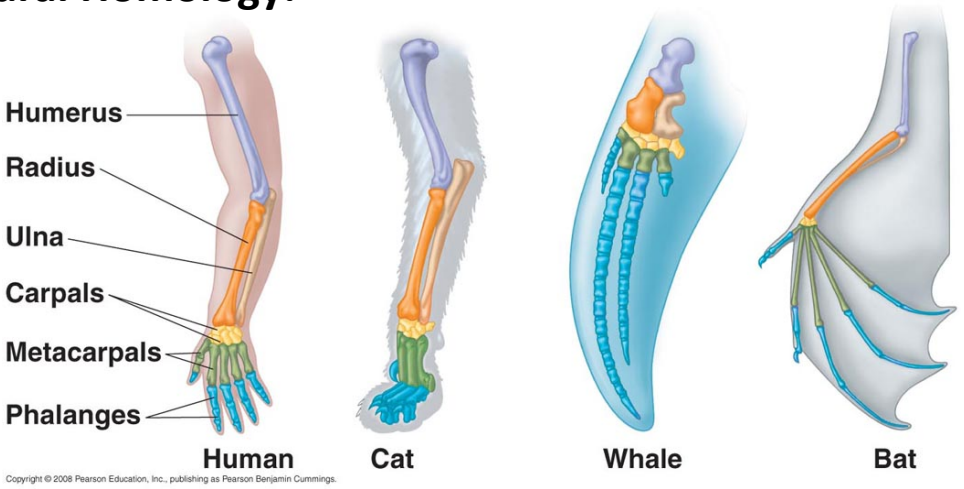


Explaining Diversity: From Morphology to Genomics



Homology and Bioinformatics

Structural Homology:

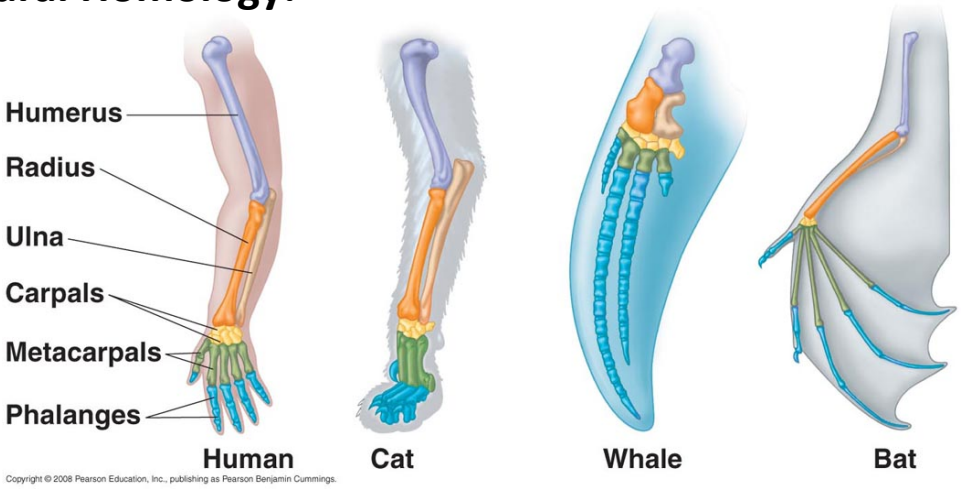


Sequence Homology (Alignment):

```
Human KKASKPKKAASKAPTKKPKATPKATPVKKAKKKLAAT
Mouse KKAAKPKKAASKAPSKKPKATPKATPVKKAKKKPAAT
  Rat KKAAKPKKAASKAPSKKPKATPKATPVKKAKKKPAAT
  Cow KKAAKPKKAASKAPSKKPKATPKATPVKKAKKKPAAT
Chimp KKAAKPKKAASKAPSKKPKATPKATPVKKAKKKLAAT
      ***:**********:******************** ***
```

Homology and Bioinformatics

Structural Homology:

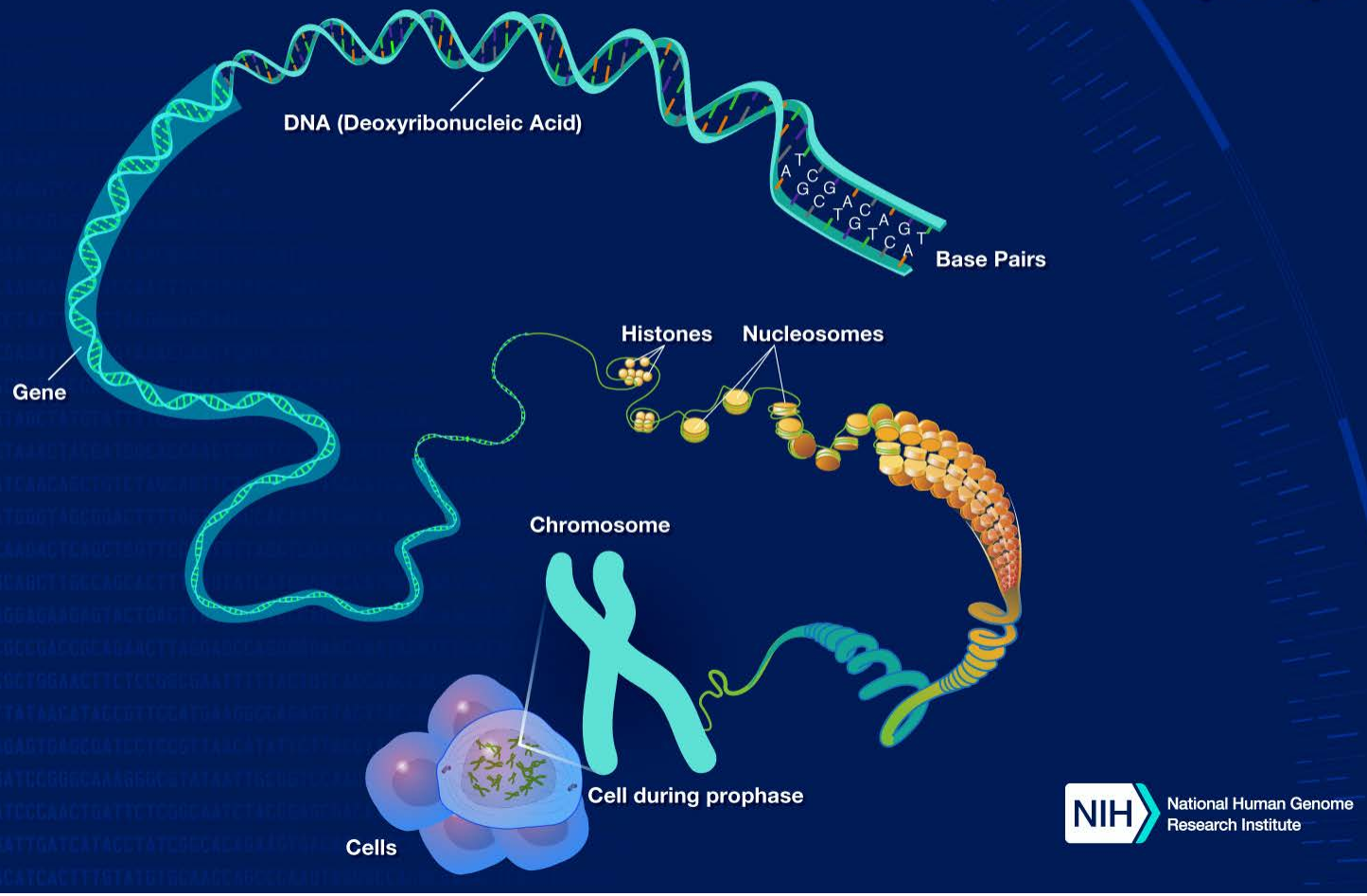


Sequence Homology (Alignment):

```
Human KKASKPKKAASKAPTKKPKATPKATPVKKAKKKLAAT
Mouse KKAAKPKKAASKAPSKKPKATPKATPVKKAKKKPAAT
  Rat KKAAKPKKAASKAPSKKPKATPKATPVKKAKKKPAAT
  Cow KKAAKPKKAASKAPSKKPKATPKATPVKKAKKKPAAT
Chimp KKAAKPKKAASKAPSKKPKATPKATPVKKAKKKLAAT
      ***:**********:******************** ***
```

A Brief Guide to Genomics

NHGRI FACT SHEETS
genome.gov

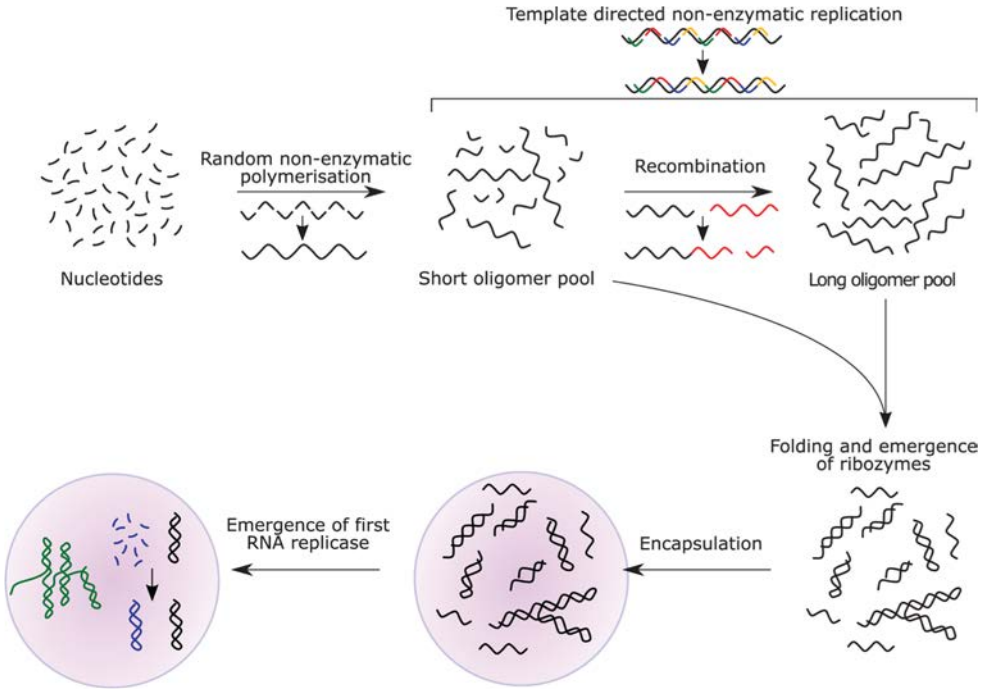


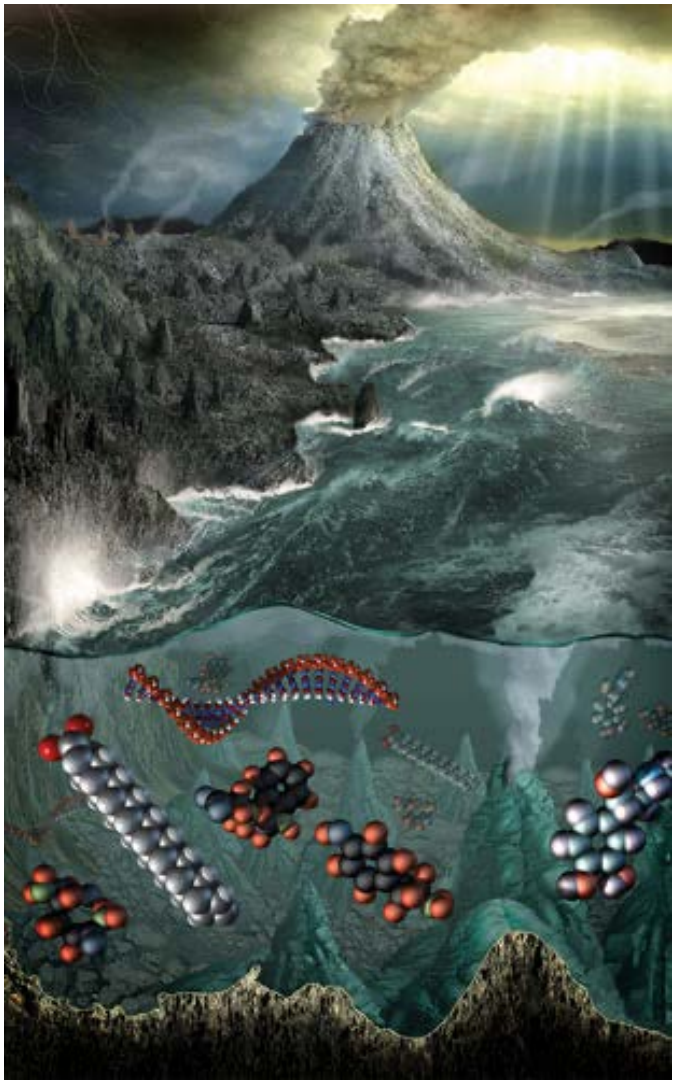
- ***genome***: all the genetic material of an organism
- 1920 by Hans Winkler
- Combination of gene and chromosome

The origins of the genome

RNA World (Rich, 1962)

- RNA stores information
- behaves as a catalyst
- possible to transfer DNA to RNA viruses
- this may have stimulated the move to a DNA world





The origins of the genome

Informational genes process protein synthesis and RNA transcription

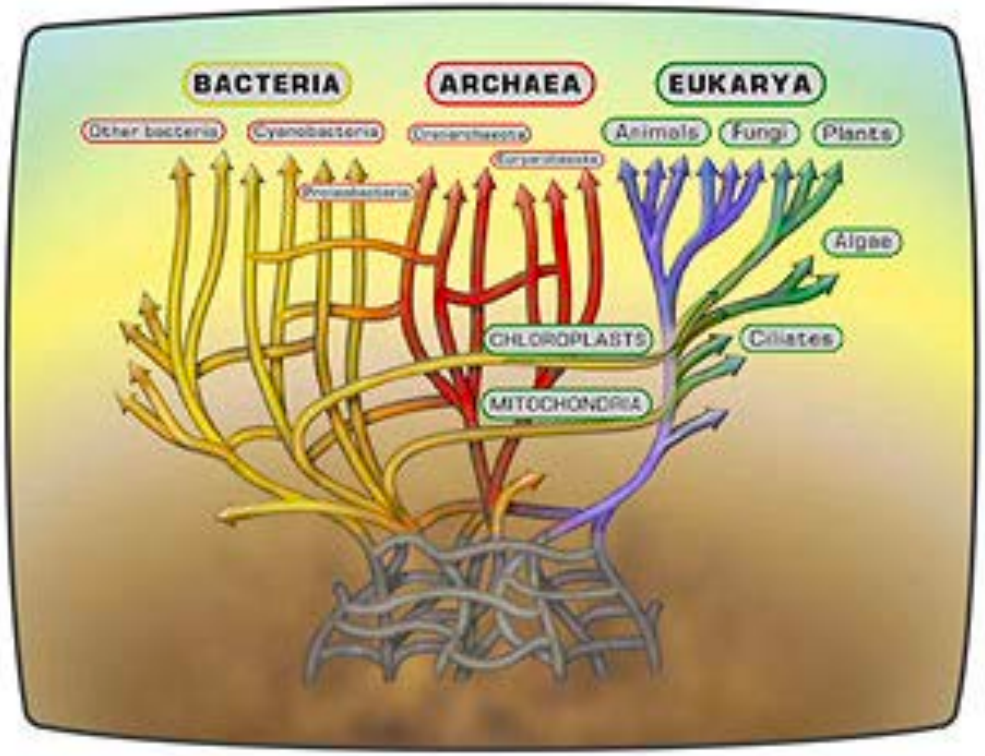
- They may have originated in an RNA world

Operational genes perform cellular functions such as making cell walls and synthesizing amino acids

Both of these types of genes are found in all eukaryotes and prokaryotes

Rivera et al. 1998 Genomic evidence for two functionally distinct gene classes. *PNAS*

The origins of the genome



Lateral gene transfer at the base of the tree of life
Is there even a tree of life?

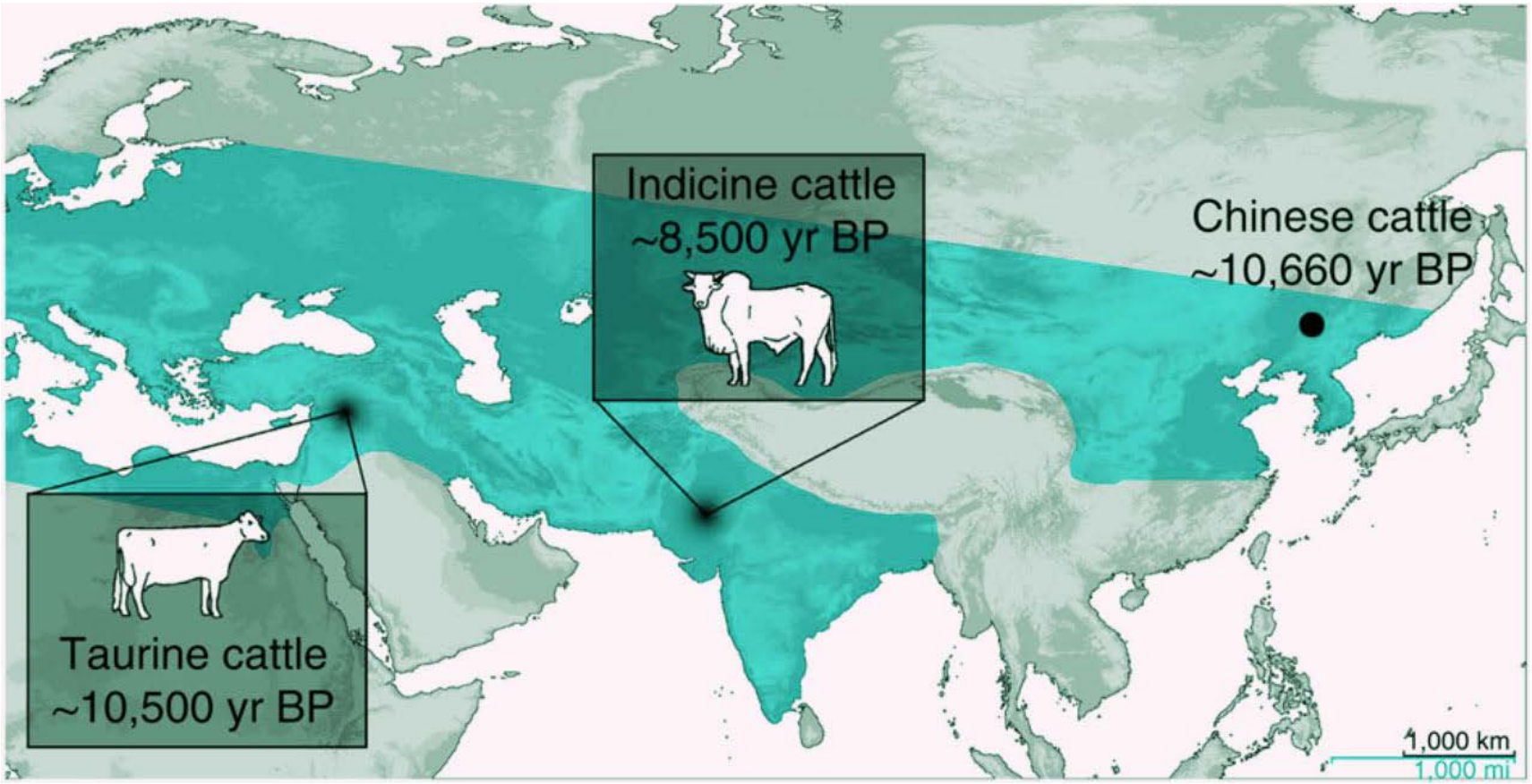
Earliest Genomics Experiments

?

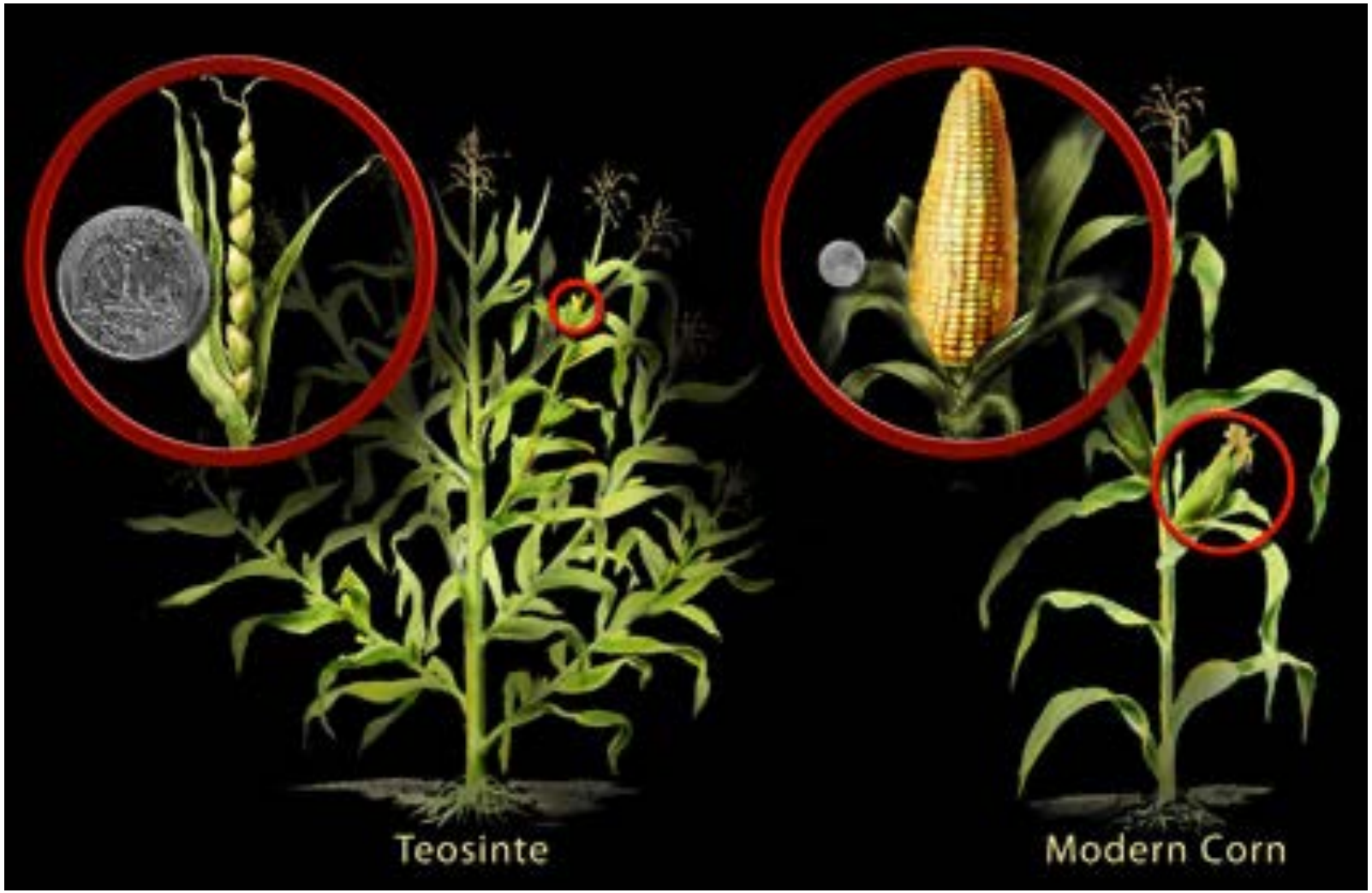
Earliest Genomics Experiments

15,000 to 35,000 years before present

Earliest Genomics Experiments



Earliest Genomics Experiments

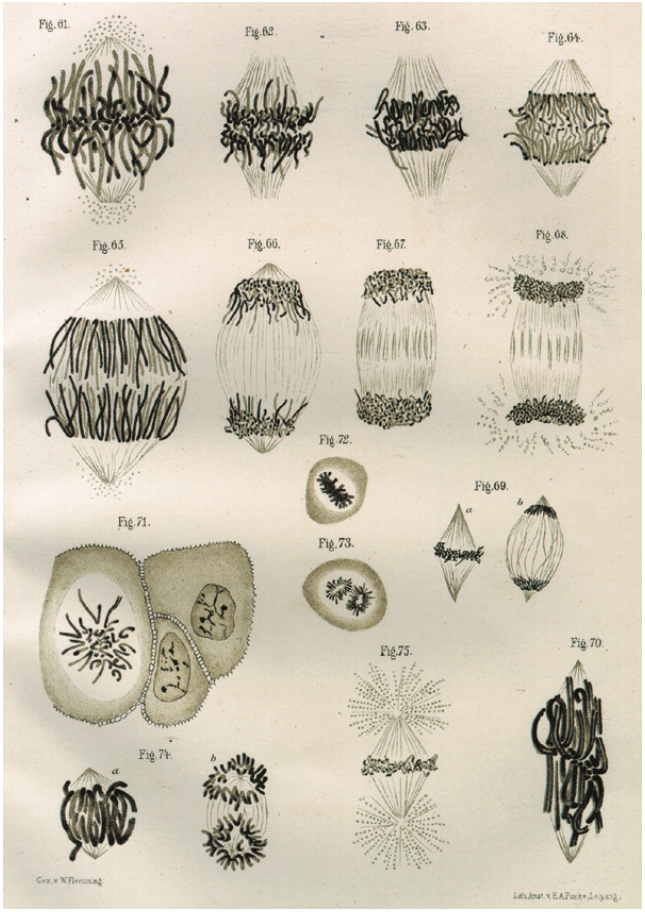


6,000 to 10,000 year before present

Discovery of chromosomes

Mid-19th century: with proper staining techniques, microscopes could observe structures that were involved in cell division.

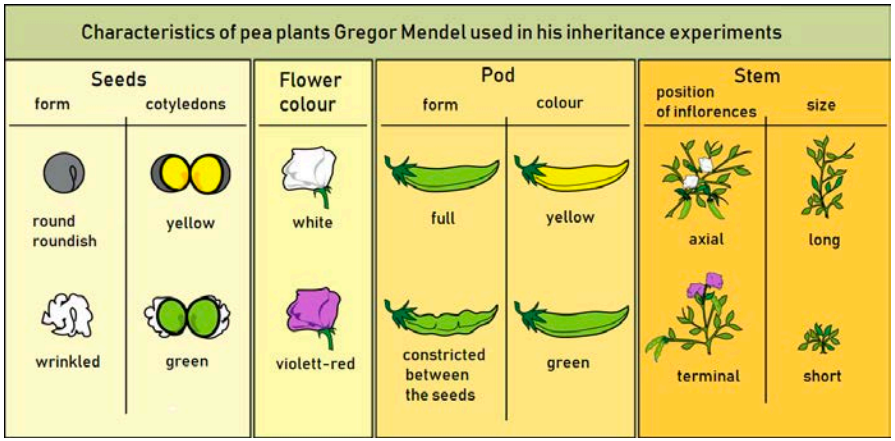
‘Chromosome’ comes from the Greek words for ‘colored body’



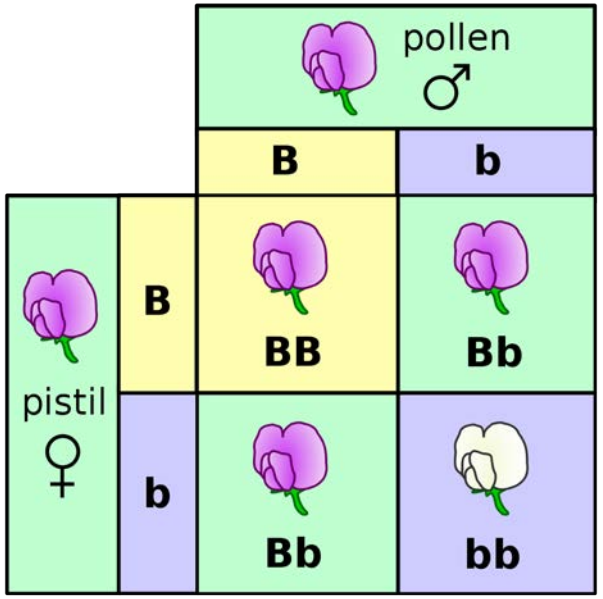
Drawing of mitosis, Walther Flemming.
Flemming. 1882. ***Zellsubstanz, Kern und Zelltheilung*** **(*Cell Substance, Nucleus and Cell Division*)**

Laws of inheritance

1. Law of Dominance
2. Law of Segregation
3. Law of Independent Assortment



Versuche über Pflanzen-Hybriden.Verh. Naturforsch (Experiments in Plant Hybridization)
Mendel, G. (1866).Ver. Brünn 4: 3–47 (in English in 1901, J. R. Hortic. Soc. 26: 1–32).



https://en.wikipedia.org/wiki/Mendelian_inheritance

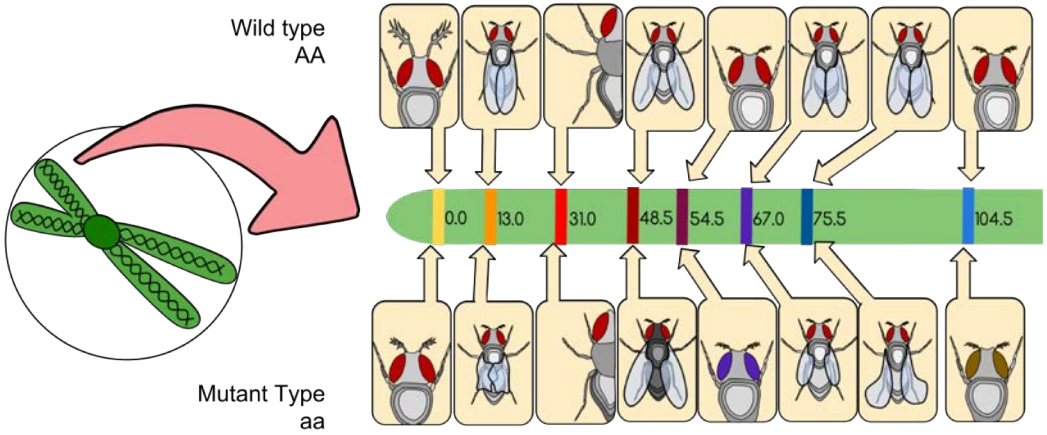
First genetic maps

Mendel's Law of Independent Assortment states that alleles of a gene sort into gametes independently of other genes

Sturtevant and Morgan discovered that the probabilities of co-occurrences of traits are due to the distance between the genes along the chromosome – ***the genes are linked***.

Traits that are commonly observed together are literally near each other.

The Linear Arrangement of Six Sex-Linked Factors in Drosophila as shown by their mode of Association
Sturtevant, A. H. (1913) Journal of Experimental Zoology, 14: 43-59



https://en.wikipedia.org/wiki/Genetic_linkage

Jumping genes

Early geneticists thought genes were stable entities arranged linearly along a chromosome like beads on a string

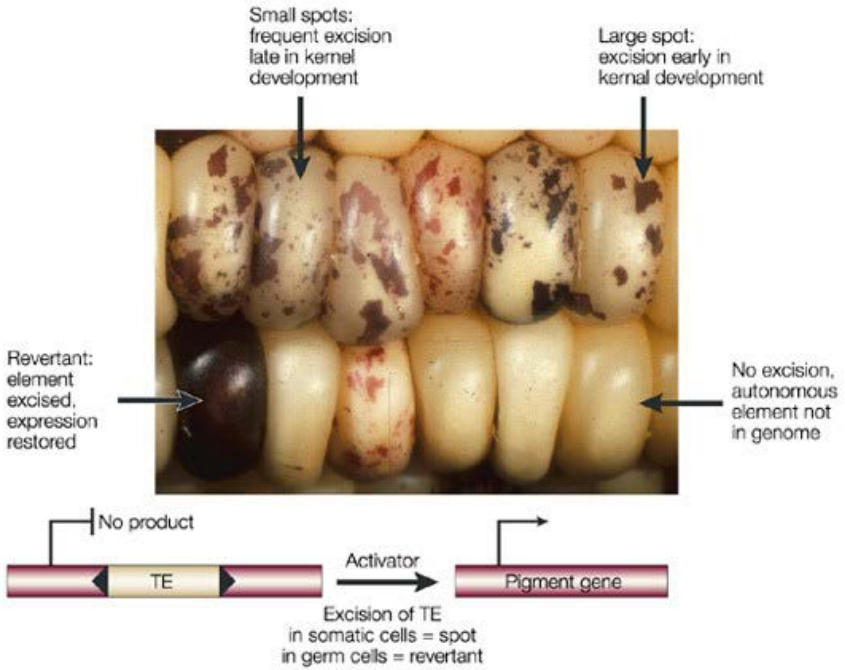
The discovery of transposable elements showed that some genes can move or copy themselves

The movement of these elements during meiosis had an effect on many traits – ***and revealed genetic regulation for the first time.***

More than half of the human genome is comprised of transposable elements.

Barbara McClintock, Winner of Nobel Prize in Physiology or Medicine (1983)

(wikipedia)



The origin and behavior of mutable loci in maize.
McClintock, B. (1950) PNAS. 36(6):344–355.

DNA

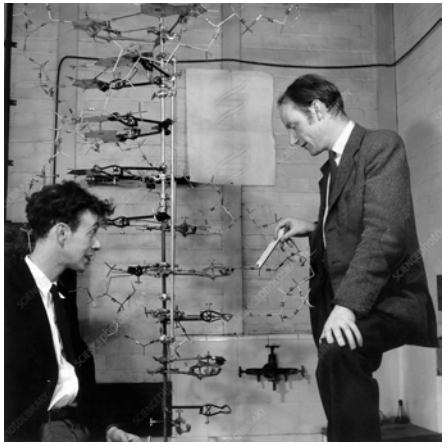
Identified in 1869 by Friedrich Miescher in human WBCs

1944 Avery – hereditary units (***genes***) are made of DNA

Chargaff (1950) described a "chemistry of heredity" – found that DNA varies across species

Watson and Crick – initially credited with discovery of the ***double helix***

Rosalind Franklin took the ***X-ray crystallography*** photographs that proved it



Watson and Crick, 1953

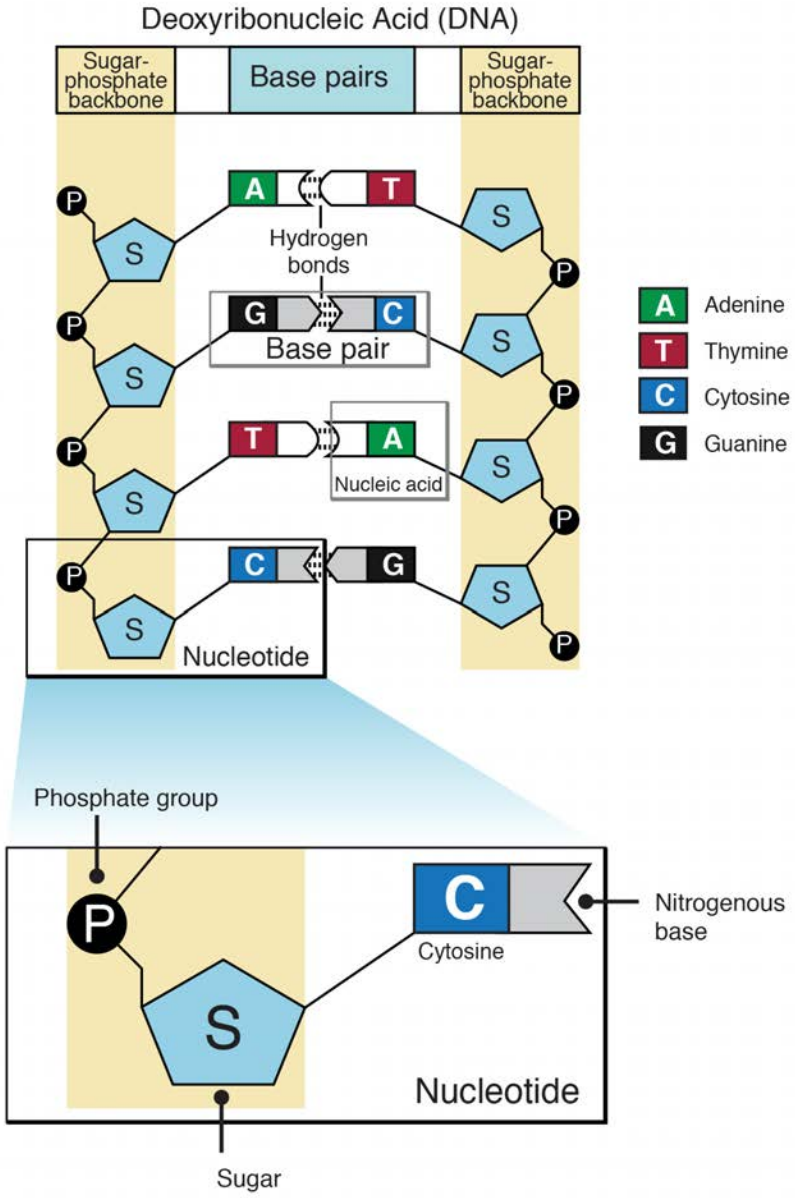
Rosalind Franklin, 1958

Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid
Watson JD, Crick FH (1953). Nature 171: 737–738.

Nobel Prize in Physiology or Medicine in 1962

DNA

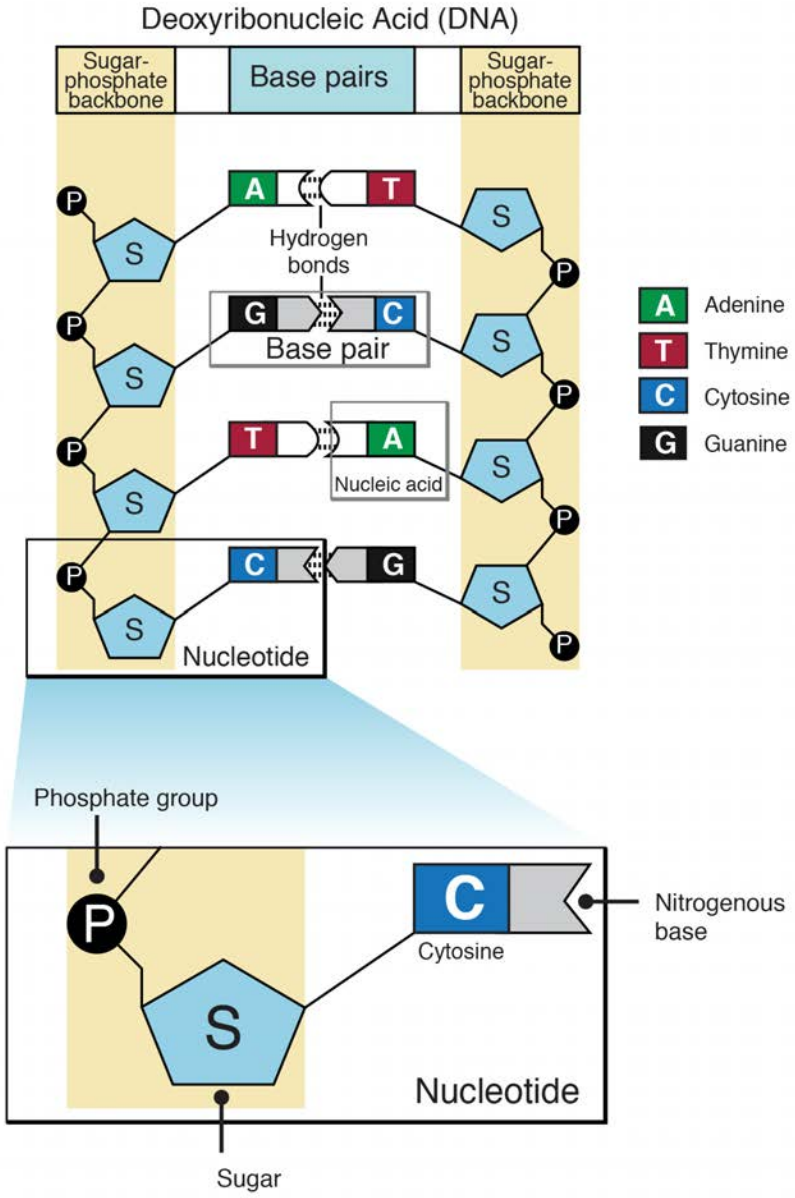
- Deoxyribonucleic acid
  - A molecule made of two nucleotide chains coiled around each other (double helix)
  - each nucleotide contains a sugar attached to a phosphate group, and a nitrogen containing base.
    - adenine (A)
    - cytosine (C)
    - guanine (G)
    - thymine (T)



genome.gov

DNA

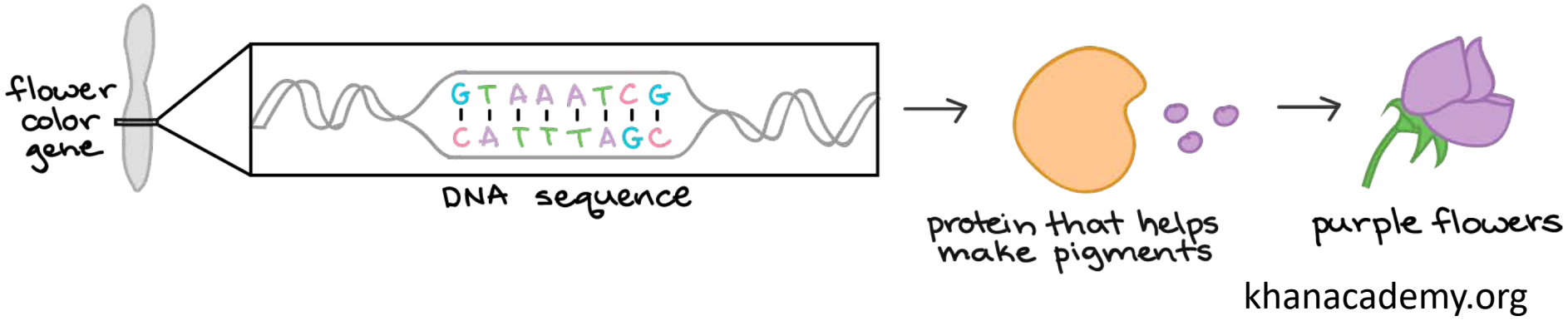
- RNA is composed of nucleotides too.
  - uracil replaces thymine
- each nucleotide has a complement that it fits into like a lock-and-key with a hydrogen bond in order to complete the double helix.
  - A <-> T
  - G <->C
- This is called base pairing.
  - base vs base pair (bp)



genome.gov

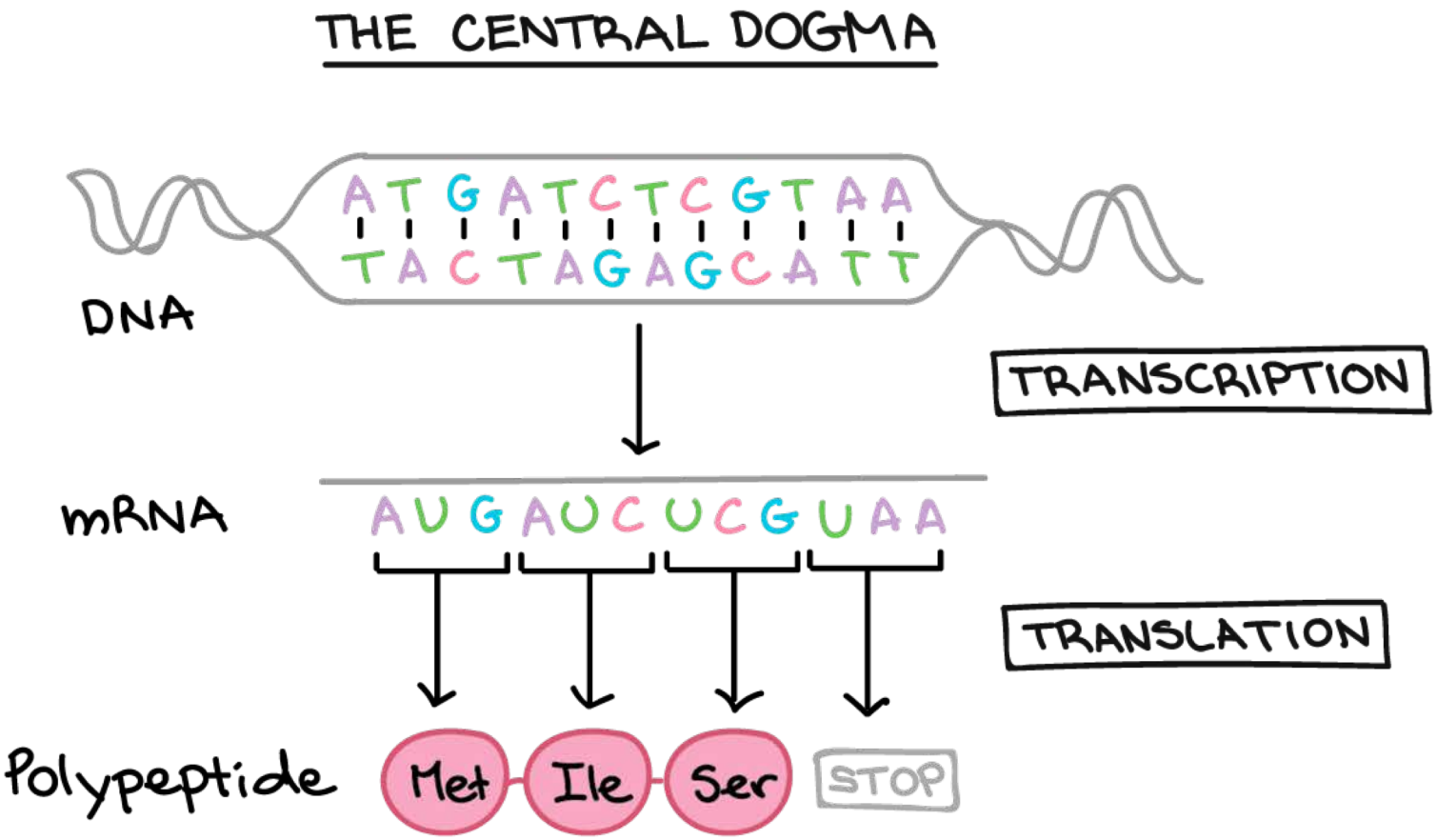
The Central Dogma of Molecular Biology

"Once 'information' has passed into protein it cannot get out again. In more detail, the transfer of information ***from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible***, but transfer from protein to protein, or from protein to nucleic acid is impossible. Information means here the precise determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein"

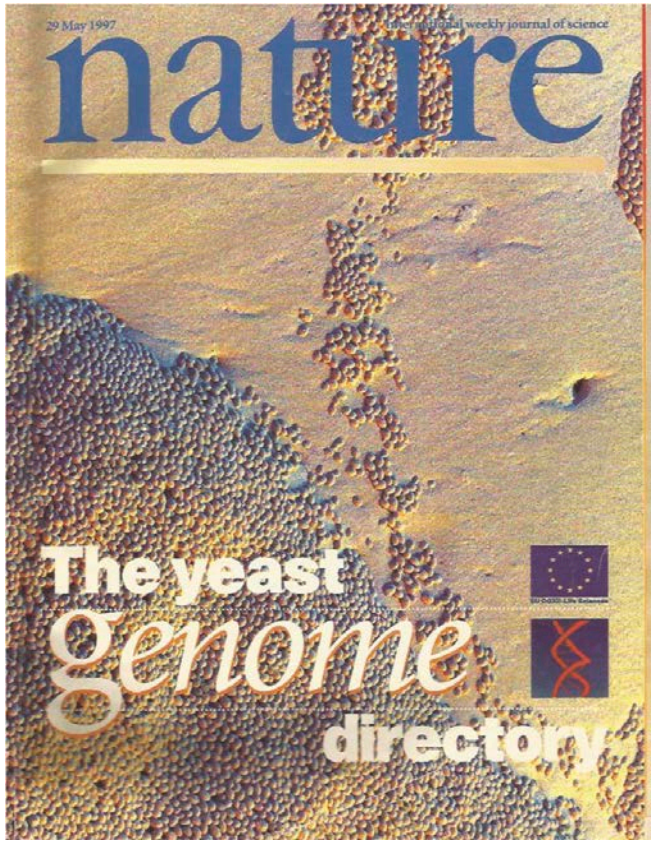


On Protein Synthesis
Crick, F.H.C. (1958). Symposia of the Society for Experimental Biology pp. 138–163.

"DNA makes RNA makes protein"



khanacademy.org



1997: *Saccharomyces cerevisae* (12Mb)

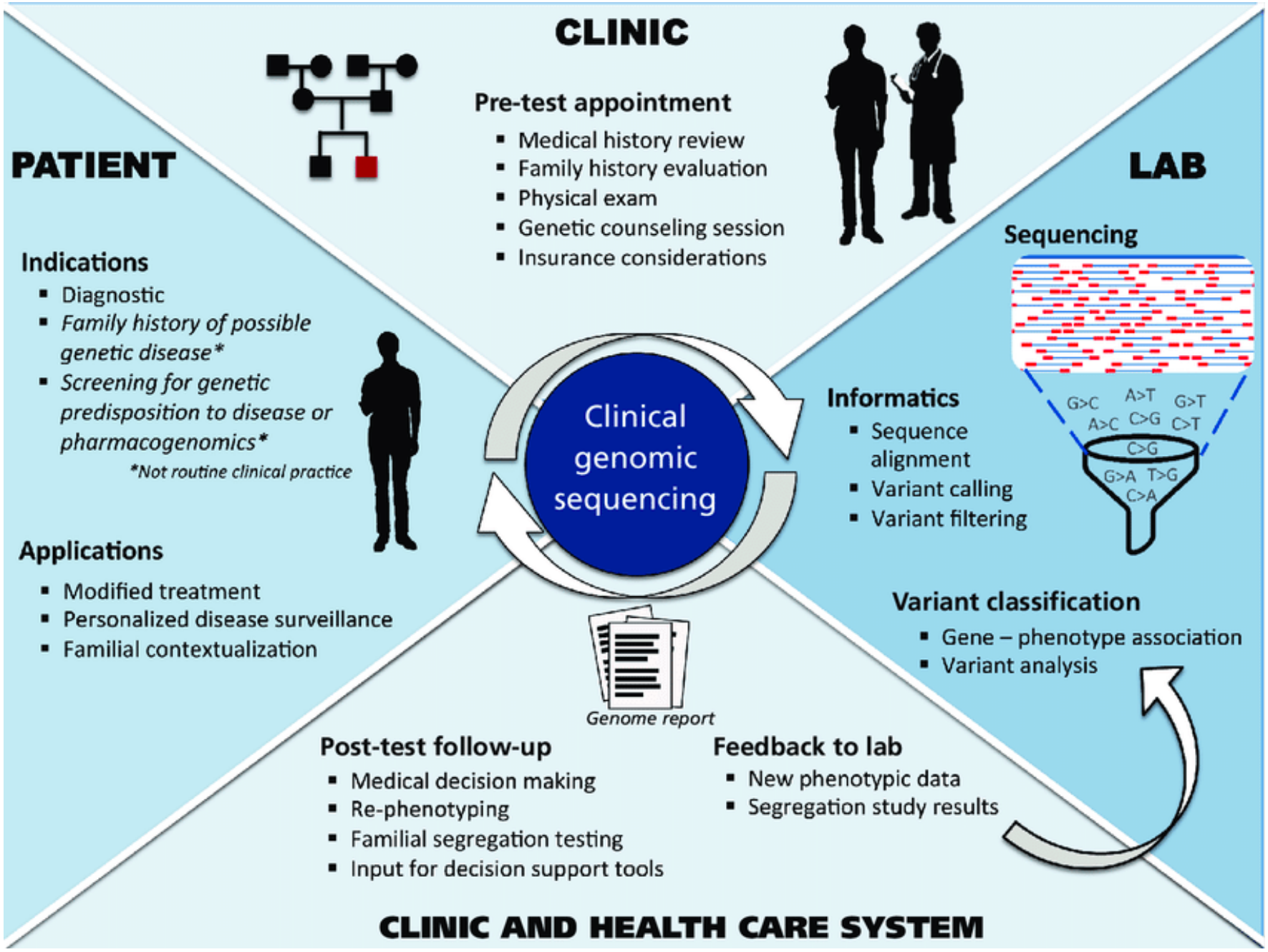
1998: *Caenorhabditis elegans* (97Mb)

2000: *Arabidopsis thaliana* (115 Mb)

2000: *Drosophila melanogaster* (120 Mb)

Lander et al. 2001

Venter et al. 2001


CLINIC
Pre-test appointment
Medical history review
Family history evaluation
Physical exam
Genetic counseling session
Insurance considerations
PATIENT
Indications
Diagnostic
Family history of possible genetic disease*
Screening for genetic predisposition to disease or pharmacogenomics*
*Not routine clinical practice
Applications
Modified treatment
Personalized disease surveillance
Familial contextualization
Clinical genomic sequencing
LAB
Sequencing
G>C A>T G>T
A>C C>G C>T
C>G
G>A T>G
C>A
Informatics
Sequence alignment
Variant calling
Variant filtering
Variant classification
Gene – phenotype association
Variant analysis
Genome report
Post-test follow-up
Medical decision making
Re-phenotyping
Familial segregation testing
Input for decision support tools
Feedback to lab
New phenotypic data
Segregation study results
CLINIC AND HEALTH CARE SYSTEM

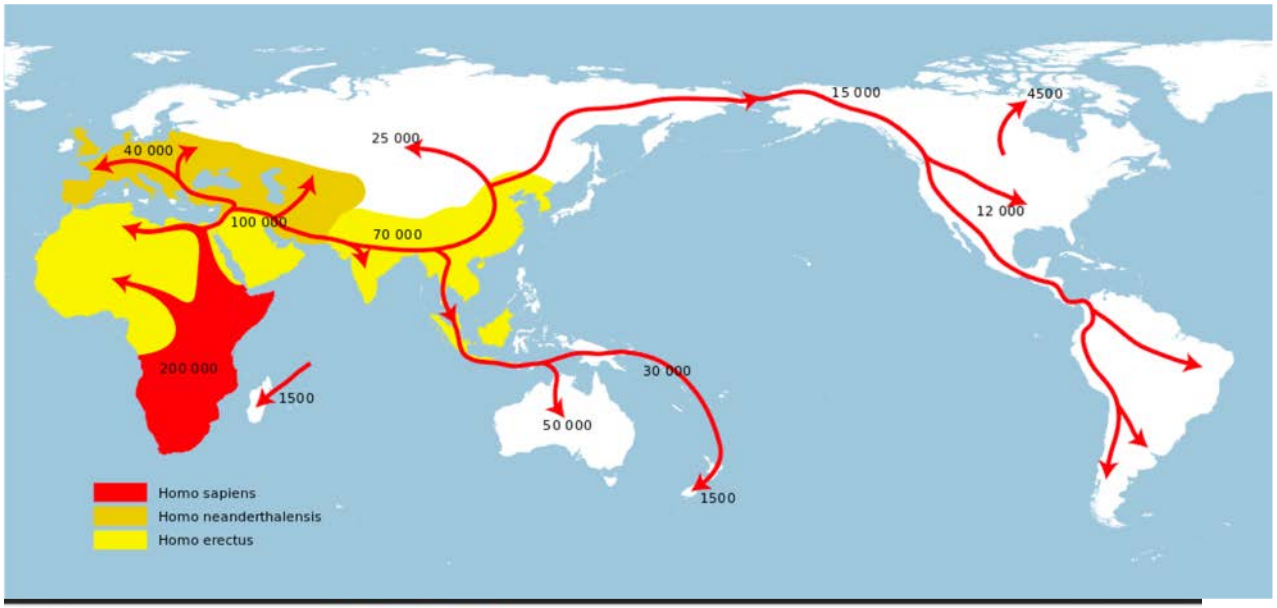
De novo mutation discovery and validation

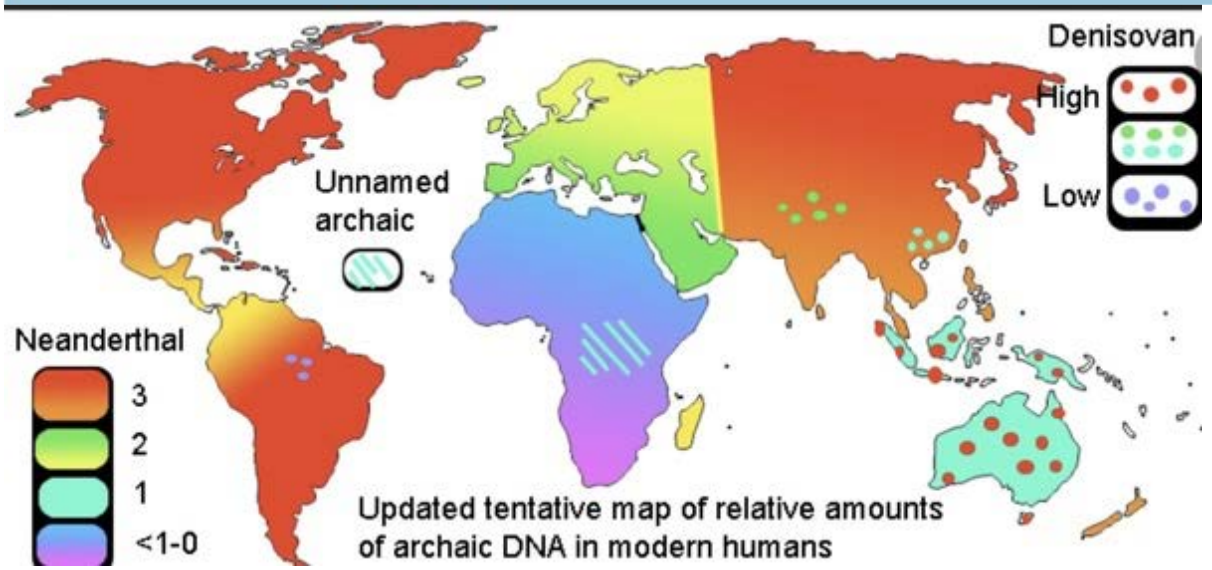
De novo mutations are sequences not inherited from your parents.

```
Reference:    ...TCAAATCCTTTTAATAAAGAAGAGCTGACA...

Father(1):    ...TCAAATCCTTTTAATAAAGAAGAGCTGACA...
Father(2):    ...TCAAATCCTTTTAATAAAGAAGAGCTGACA...

Mother(1):    ...TCAAATCCTTTTAATAAAGAAGAGCTGACA...
Mother(2):    ...TCAAATCCTTTTAATAAAGAAGAGCTGACA...

Sibling(1):   ...TCAAATCCTTTTAATAAAGAAGAGCTGACA...
Sibling(2):   ...TCAAATCCTTTTAATAAAGAAGAGCTGACA...

Proband(1):   ...TCAAATCCTTTTAATAAAGAAGAGCTGACA...
Proband(2):   ...TCAAATCCTTTTAAT****AAGAGCTGACA...
```

4bp heterozygous deletion at chr15:93524061 CHD2

Out of Africa: we contain Neanderthal DNA in our genomes





For the rest of this week

1. Reflect on the fact that we all share a bit of Earth's history within ourselves.
2. Sign up for presentation.
3. Add a thread to the Introductions Forum.
4. Fill out computing survey.
5. Read Shendure et al. 2017 – I will present.
6. Quiz 1 – will cover lecture and paper