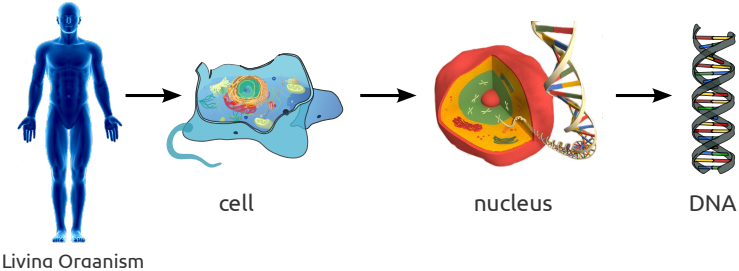
Computational Biology

Warm Up + Cracking the Genetic Code

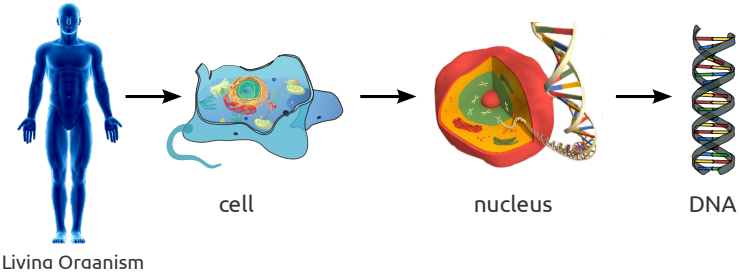
Department of Mathematics
Stockholm University

Fun Facts



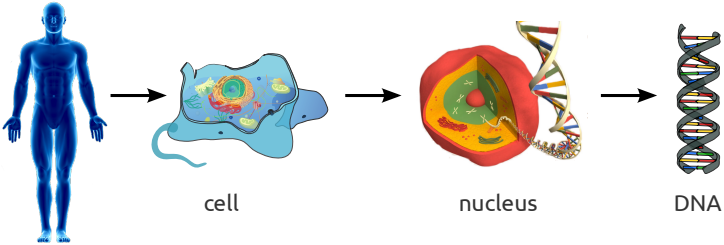
Genetic information about organisms is contained in the DNA
The DNA consist of 4 Basen = **A**denin, **G**uanin, **C**ytosin, **T**hymin,

Fun Facts



DNA = long word of 4 “Letters” A,C,T,G

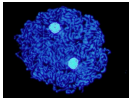
Fun Facts



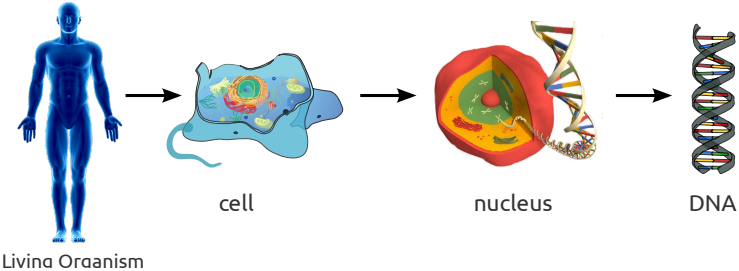
Living Organism

DNA = long word of 4 “Letters” A,C,T,G

Fun Fact 0:

Species	Human	 Carsonella ruddii	 Paris japonica
Genomsize (# “Letters”)	3 270 000 000 (3,27 Billion)	159 662	150 000 000 000 (150 Billion)

Fun Facts



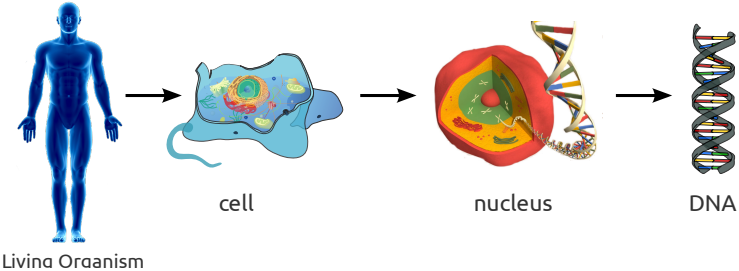
DNA = long word of 4 “Letters” A,C,T,G

Fun Fact 1:

Although tiny, uncoiled human DNA in a single nuclei has length: around 2 meter.

If you uncoil all the DNA in a human and put it end-to-end it would stretch around 150 Mrd. km $\simeq$ 1000times distance earth-sun

Fun Facts

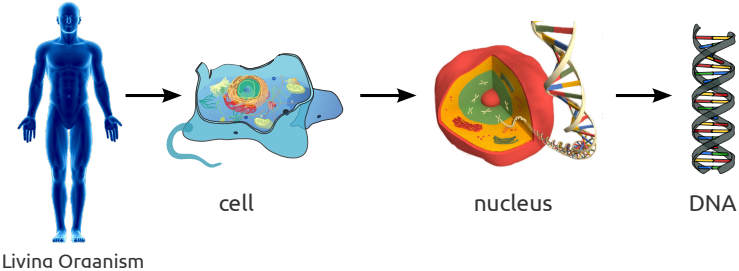


DNA = long word of 4 “Letters” A,C,T,G

Fun Fact 2:

Your genome is only ∼0.5% different from other person’s
Humans share around 96% of their DNA with chimpanzees, 90% with mice and 60% with bananas.

Fun Facts

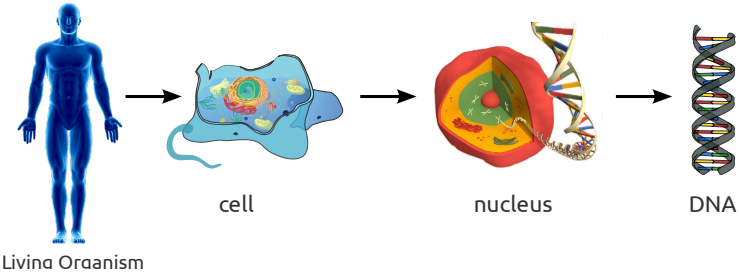


DNA = long word of 4 “Letters” A,C,T,G

Fun Fact 3:

The human DNA would fill $\sim$ 545000 pages (A4, textsize 11)
$\sim$ 545 books each with 1000pages

Fun Facts



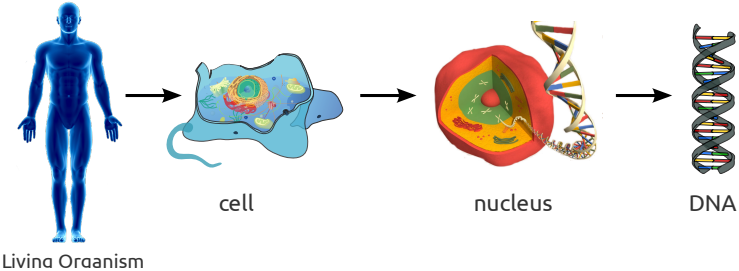
DNA = long word of 4 “Letters” A,C,T,G

Fun Fact 3:

The human DNA would fill $\sim$ 545000 pages (A4, textsize 11)
$\sim$ 545 books each with 1000pages

A change of **a single** letter, say in Book 272 on page 325 replace `A` in (line 17 column 2) by a `T`, may cause a difference in your eye color or a severe disease.

Fun Facts



DNA = long word of 4 “Letters” A,C,T,G

Knowledge of these fun facts is based on the knowledge about genetic material.

How do we get this knowledge?

Let us start with a brief history.

Basic Problem: Understand Inheritance & Cracking the Code

1860's Mendel (abstract essentially math. model for "inheritance unit")

1869 Miescher (discovered DNA + Idea: nucleic acids could be involved in heredity)

1883-1949 Kossel, Levene, Chargaff (composition of RNA and DNA)

1928 Griffith's Experiment

(bacteria are capable of transferring genetic information through a process known as transformation.)

1944 Avery, MacLeod und McCarty (1944):

(refined results of Griffith, first clear suggestion that DNA carries genetic information)

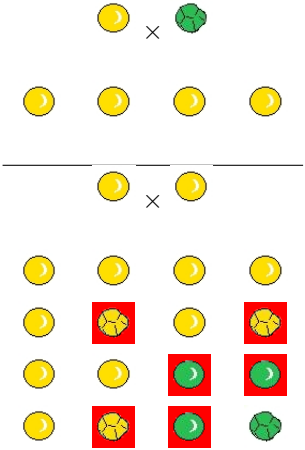
1952 Herschey and Chase (confirmed results of Miescher)

1952 Rosalind Franklin (Photo 51 Xray)

1953 Watson and Crick (double helical structure of DNA)

2003 Human genome is sequenced

1860's Gregor Mendel: inheritance unit



- 1st generation: only smooth and yellow peas
- 2nd generation: all possible combinations between smooth/wrinkled and yellow/green peas

$\Longrightarrow$ "non-observable" information must have been stored somewhere

Mendel gave abstract essentially mathematical model of inheritance: “inheritance unit” that “store” information.

He mentioned that biological variations are inherited from parent organism as specific discrete traits.

1869 Friedrich Miescher discovered DNA

- FM wanted to investigate the composition of cells

  He chose leukocytes (white blood cells) from human pus as his source material, hoping that analysing cells that are not embedded in a tissue would facilitate the identification of the molecular building blocks that make up cells.

  So he collected a lot of pus from bandages at local hospitals

- Through a chemical process, he extracted the nuclei

  (by adding weak alkaline solution to the white blood cells)

- He analysed the nuclei and obsevered that a major component in there was new type of molecule: an acid of large molecular weight and high phosphorus content.

  He called this new type of molecule “nuclein” (now nucleic acids)

- He raised the idea that the nucleic acids could be involved in heredity

https://www.degruyter.com/document/doi/10.1515/hsz-2021-0226/html

1869 Friedrich Miescher discovered DNA

- FM wanted to investigate the composition of cells

  He chose leukocytes (white blood cells) from human pus as his source material, hoping that analysing cells that are not embedded in a tissue would facilitate the identification of the molecular building blocks that make up cells.

  So he collected a lot of pus from bandages at local hospitals
- Through a chemical process, he extracted the nuclei

  (by adding weak alkaline solution to the white blood cells)
- He analysed the nuclei and obsevered that a major component in there was new type of molecule: an acid of large molecular weight and high phosphorus content.

  He called this new type of molecule “nuclein” (now nucleic acids)
- He raised the idea that the nucleic acids could be involved in heredity

https://www.degruyter.com/document/doi/10.1515/hsz-2021-0226/html

1869 Friedrich Miescher discovered DNA

- FM wanted to investigate the composition of cells

  He chose leukocytes (white blood cells) from human pus as his source material, hoping that analysing cells that are not embedded in a tissue would facilitate the identification of the molecular building blocks that make up cells.

  So he collected a lot of pus from bandages at local hospitals
- Through a chemical process, he extracted the nuclei

  (by adding weak alkaline solution to the white blood cells)
- He analysed the nuclei and obsevered that a major component in there was new type of molecule: an acid of large molecular weight and high phosphorus content.

  He called this new type of molecule “nuclein” (now nucleic acids)
- He raised the idea that the nucleic acids could be involved in heredity

https://www.degruyter.com/document/doi/10.1515/hsz-2021-0226/html

1869 Friedrich Miescher discovered DNA

- FM wanted to investigate the composition of cells

  He chose leukocytes (white blood cells) from human pus as his source material, hoping that analysing cells that are not embedded in a tissue would facilitate the identification of the molecular building blocks that make up cells.

  So he collected a lot of pus from bandages at local hospitals
- Through a chemical process, he extracted the nuclei

  (by adding weak alkaline solution to the white blood cells)
- He analysed the nuclei and obsevered that a major component in there was new type of molecule: an acid of large molecular weight and high phosphorus content.

  He called this new type of molecule “nuclein” (now nucleic acids)
- He raised the idea that the nucleic acids could be involved in heredity

https://www.degruyter.com/document/doi/10.1515/hsz-2021-0226/html

1869 Friedrich Miescher discovered DNA

- FM wanted to investigate the composition of cells

  He chose leukocytes (white blood cells) from human pus as his source material, hoping that analysing cells that are not embedded in a tissue would facilitate the identification of the molecular building blocks that make up cells.

  So he collected a lot of pus from bandages at local hospitals
- Through a chemical process, he extracted the nuclei

  (by adding weak alkaline solution to the white blood cells)
- He analysed the nuclei and obsevered that a major component in there was new type of molecule: an acid of large molecular weight and high phosphorus content.

  He called this new type of molecule “nuclein” (now nucleic acids)
- He raised the idea that the nucleic acids could be involved in heredity

https://www.degruyter.com/document/doi/10.1515/hsz-2021-0226/html

1869 Friedrich Miescher discovered DNA

- FM wanted to investigate the composition of cells

  He chose leukocytes (white blood cells) from human pus as his source material, hoping that analysing cells that are not embedded in a tissue would facilitate the identification of the molecular building blocks that make up cells.

  So he collected a lot of pus from bandages at local hospitals
- Through a chemical process, he extracted the nuclei

  (by adding weak alkaline solution to the white blood cells)
- He analysed the nuclei and obsevered that a major component in there was new type of molecule: an acid of large molecular weight and high phosphorus content.

  He called this new type of molecule “nuclein” (now nucleic acids)
- He raised the idea that the nucleic acids could be involved in heredity

https://www.degruyter.com/document/doi/10.1515/hsz-2021-0226/html

1869 Friedrich Miescher discovered DNA

- FM wanted to investigate the composition of cells

  He chose leukocytes (white blood cells) from human pus as his source material, hoping that analysing cells that are not embedded in a tissue would facilitate the identification of the molecular building blocks that make up cells.

  So he collected a lot of pus from bandages at local hospitals
- Through a chemical process, he extracted the nuclei

  (by adding weak alkaline solution to the white blood cells)
- He analysed the nuclei and obsevered that a major component in there was new type of molecule: an acid of large molecular weight and high phosphorus content.

  He called this new type of molecule “nuclein” (now nucleic acids)
- He raised the idea that the nucleic acids could be involved in heredity

https://www.degruyter.com/document/doi/10.1515/hsz-2021-0226/html

Experiments: 1928 Griffith / 1944 Avery, MacLeod, McCarty

Pneumonia was a serious cause of death in the wake of the post-WWI Spanish influenza pandemic, and Griffith was studying the possibility of creating a vaccine.

He used two strains of pneumococcus bacteria to infect mice:

S(mooth)-strain covered itself with a polysaccharide capsule that protected it from the host's immune system, resulting in the death of the host

R(ough)-strain didn't have that protective capsule and was defeated by the host's immune system.

R-strain:	does not harm mice
S-train:	kills mice
killed S-train:	does not harm mice
R-strain + killed S-train:	kills mice

Conclusion?

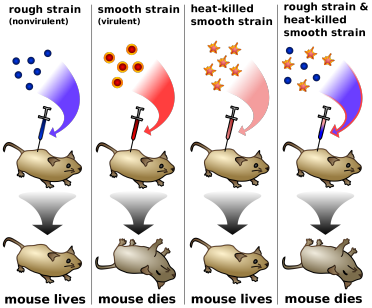
Experiments: 1928 Griffith / 1944 Avery, MacLeod, McCarty

Pneumonia was a serious cause of death in the wake of the post-WWI Spanish influenza pandemic, and Griffith was studying the possibility of creating a vaccine.

He used two strains of pneumococcus bacteria to infect mice:

S(mooth)-strain covered itself with a polysaccharide capsule that protected it from the host's immune system, resulting in the death of the host

R(ough)-strain didn't have that protective capsule and was defeated by the host's immune system.



R-strain:	does not harm mice
S-train:	kills mice
killed S-train:	does not harm mice
R-strain + killed S-train:	kills mice

Conclusion?

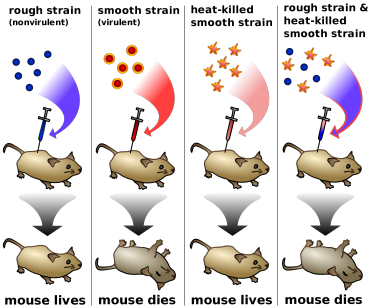
Experiments: 1928 Griffith / 1944 Avery, MacLeod, McCarty

Pneumonia was a serious cause of death in the wake of the post-WWI Spanish influenza pandemic, and Griffith was studying the possibility of creating a vaccine.

He used two strains of pneumococcus bacteria to infect mice:

S(mooth)-strain covered itself with a polysaccharide capsule that protected it from the host's immune system, resulting in the death of the host

R(ough)-strain didn't have that protective capsule and was defeated by the host's immune system.



R-strain:	does not harm mice
S-train:	kills mice
killed S-train:	does not harm mice
R-strain + killed S-train:	kills mice

Conclusion?

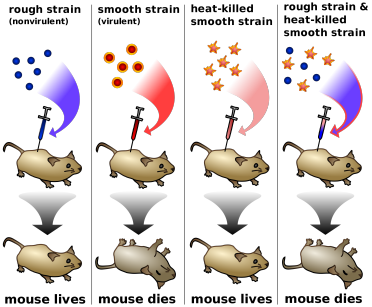
Experiments: 1928 Griffith / 1944 Avery, MacLeod, McCarty

Pneumonia was a serious cause of death in the wake of the post-WWI Spanish influenza pandemic, and Griffith was studying the possibility of creating a vaccine.

He used two strains of pneumococcus bacteria to infect mice:

S(mooth)-strain covered itself with a polysaccharide capsule that protected it from the host's immune system, resulting in the death of the host

R(ough)-strain didn't have that protective capsule and was defeated by the host's immune system.



R-strain:	does not harm mice
S-train:	kills mice
killed S-train:	does not harm mice
R-strain + killed S-train:	kills mice

Conclusion?

Cability to build capsules was transfered from dead S-strains to living R-strains.

Now we know: DNA survived heating process, was "taken up" from *R*-strains and allow *R*-strains to build protective capsule.

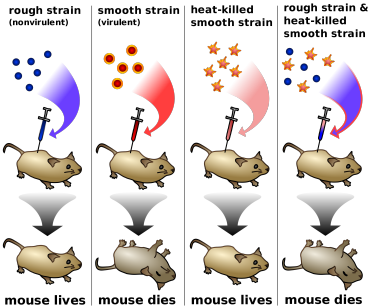
Experiments: 1928 Griffith / 1944 Avery, MacLeod, McCarty

Pneumonia was a serious cause of death in the wake of the post-WWI Spanish influenza pandemic, and Griffith was studying the possibility of creating a vaccine.

He used two strains of pneumococcus bacteria to infect mice:

S(mooth)-strain covered itself with a polysaccharide capsule that protected it from the host's immune system, resulting in the death of the host

R(ough)-strain didn't have that protective capsule and was defeated by the host's immune system.



R-strain:	does not harm mice
S-train:	kills mice
killed S-train:	does not harm mice
R-strain + killed S-train:	kills mice

Conclusion?

Cability to build capsules was transfered from dead S-strains to living R-strains.

Now we know: DNA survived heating process, was "taken up" from *R*-strains and allow *R*-strains to build protective capsule.

Avery-MacLeod-McCarty experiment (1944) reported that DNA is the substance that causes bacterial transformation, in an era when it had been widely believed that it was proteins that served the function of carrying genetic information

Composition of RNA and DNA

1883-1894 Albrecht Kossel discovered the 5 organic compounds present in nucleic acids (bases): adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U)

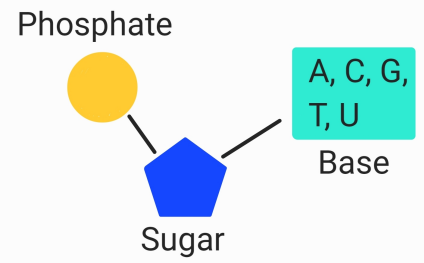
1909-1929 Phoebus Levene discovered the order of the major components of nucleotides: phosphate-sugar-base

and the carbohydrate components of RNA (1909) and DNA (1929): ribose and deoxyribose.

1949 Chargaff observed:

DNA-source	%A	%G	%C	%T
Grasshopper	29.3	20.5	20.7	29.3
Yeast	31.3	18.7	17.1	32.9
Maize	26.8	22.8	23.2	27.2
Octopus	33.2	17.6	17.6	31.6
Wheat	27.3	22.7	22.8	27.1

Any Idea?



https://www.aaas.org/other-discoverers-dna

Composition of RNA and DNA

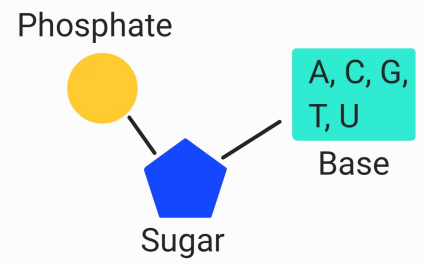
1883-1894 Albrecht Kossel discovered the 5 organic compounds present in nucleic acids (bases): adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U)

1909-1929 Phoebus Levene discovered the order of the major components of nucleotides: phosphate-sugar-base and the carbohydrate components of RNA (1909) and DNA (1929): ribose and deoxyribose.

1949 Chargaff observed:

DNA-source	%A	%G	%C	%T
Grasshopper	29.3	20.5	20.7	29.3
Yeast	31.3	18.7	17.1	32.9
Maize	26.8	22.8	23.2	27.2
Octopus	33.2	17.6	17.6	31.6
Wheat	27.3	22.7	22.8	27.1

Any Idea?



https://www.aaas.org/other-discoverers-dna

Composition of RNA and DNA

1883-1894 Albrecht Kossel discovered the 5 organic compounds present in nucleic acids (bases): adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U)

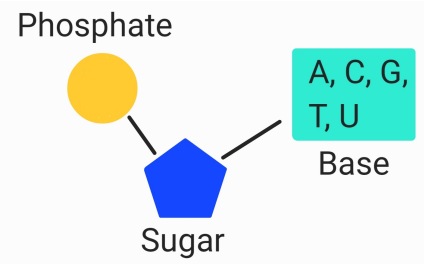
1909-1929 Phoebus Levene discovered the order of the major components of nucleotides: phosphate-sugar-base

and the carbohydrate components of RNA (1909) and DNA (1929): ribose and deoxyribose.

1949 Chargaff observed:

DNA-source	%A	%G	%C	%T
Grasshopper	29.3	20.5	20.7	29.3
Yeast	31.3	18.7	17.1	32.9
Maize	26.8	22.8	23.2	27.2
Octopus	33.2	17.6	17.6	31.6
Wheat	27.3	22.7	22.8	27.1

Any Idea?



https://www.aaas.org/other-discoverers-dna

Composition of RNA and DNA

1883-1894 Albrecht Kossel discovered the 5 organic compounds present in nucleic acids (bases): adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U)

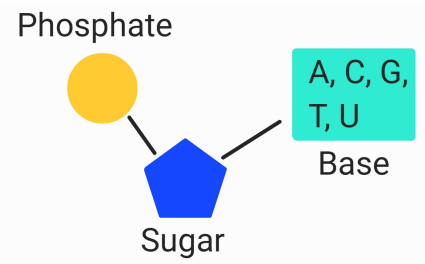
1909-1929 Phoebus Levene discovered the order of the major components of nucleotides: phosphate-sugar-base

and the carbohydrate components of RNA (1909) and DNA (1929): ribose and deoxyribose.

1949 Chargaff observed:

DNA-source	%A	%G	%C	%T
Grasshopper	29.3	20.5	20.7	29.3
Yeast	31.3	18.7	17.1	32.9
Maize	26.8	22.8	23.2	27.2
Octopus	33.2	17.6	17.6	31.6
Wheat	27.3	22.7	22.8	27.1

Any Idea?



https://www.aaas.org/other-discoverers-dna

Composition of RNA and DNA

1883-1894 Albrecht Kossel discovered the 5 organic compounds present in nucleic acids (bases): adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U)

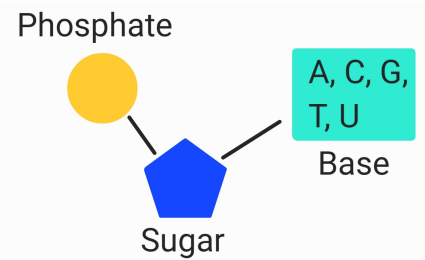
1909-1929 Phoebus Levene discovered the order of the major components of nucleotides: phosphate-sugar-base

and the carbohydrate components of RNA (1909) and DNA (1929): ribose and deoxyribose.

1949 Chargaff observed:

DNA-source	%A	%G	%C	%T
Grasshopper	29.3	20.5	20.7	29.3
Yeast	31.3	18.7	17.1	32.9
Maize	26.8	22.8	23.2	27.2
Octopus	33.2	17.6	17.6	31.6
Wheat	27.3	22.7	22.8	27.1

Any Idea?



Chargaff's rules: Amounts of A & T in DNA were roughly the same, as were the amounts of C & G.

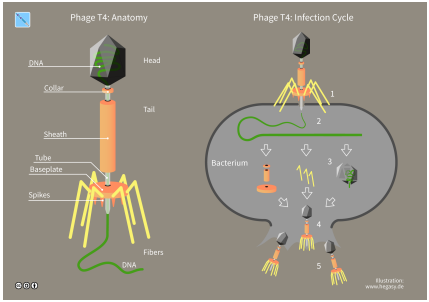
$\Longrightarrow$ Conjecture: bases A,C,G,T always occure as pairs.

https://www.aaas.org/other-discoverers-dna

1952 Alfred Hershey und Martha Chase

At this point, scientists assumed that proteins carried the information for inheritance.

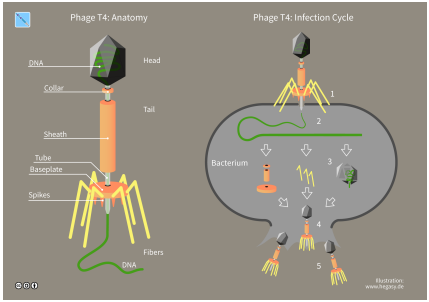
Bacteriophages (viruses that infact bacteria).



1952 Alfred Hershey und Martha Chase

At this point, scientists assumed that proteins carried the information for inheritance.

Bacteriophages (viruses that infact bacteria).

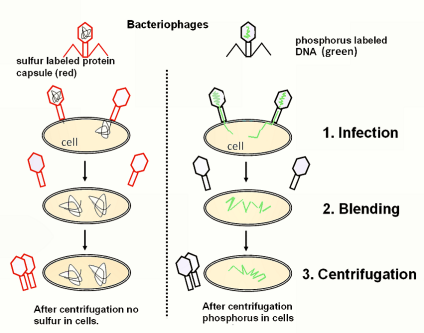


It was known that phages are composed of two major components: proteins and DNA

1952 Alfred Hershey und Martha Chase

At this point, scientists assumed that proteins carried the information for inheritance.

Bacteriophages (viruses that infact bacteria).



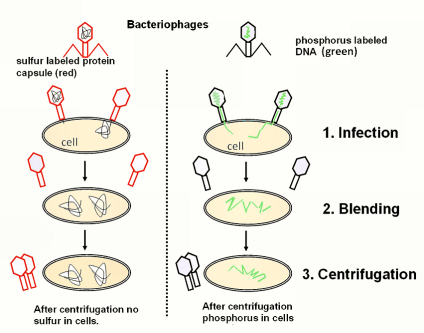
It was known that phages are composed of two major components: proteins and DNA

Hershey and Chase used bacteriophages and were able to “label” proteins and DNA differently.

1952 Alfred Hershey und Martha Chase

At this point, scientists assumed that proteins carried the information for inheritance.

Bacteriophages (viruses that infact bacteria).

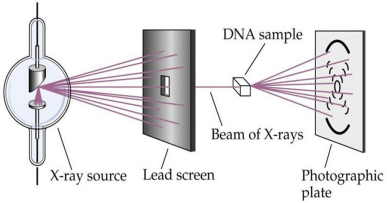


It was known that phages are composed of two major components: proteins and DNA

Hershey and Chase used bacteriophages and were able to “label” proteins and DNA differently.

Conclusion: DNA, not protein, was the genetic material.

1952 Rosalind Franklin / 1953 Francis Crick and James D. Watson



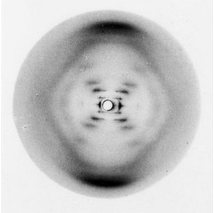
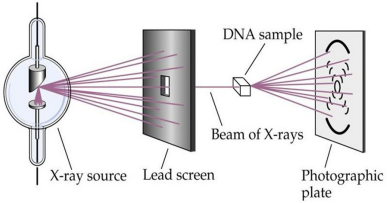


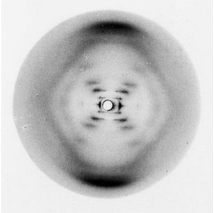
Photo 51

This was the key-stone for Crick&Watson to conclude the double helical structure of DNA (only they received a Nobel-price, not Franklin)

Xray explained: `https://www.youtube.com/watch?v=QjHqzJ7JkPY`

1952 Rosalind Franklin / 1953 Francis Crick and James D. Watson





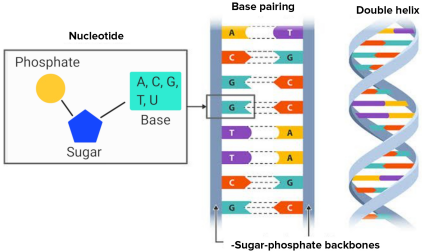


Photo 51

This was the key-stone for Crick&Watson to conclude the double helical structure of DNA (only they received a Nobel-price, not Franklin)

Xray explained: `https://www.youtube.com/watch?v=QjHqzJ7JkPY`

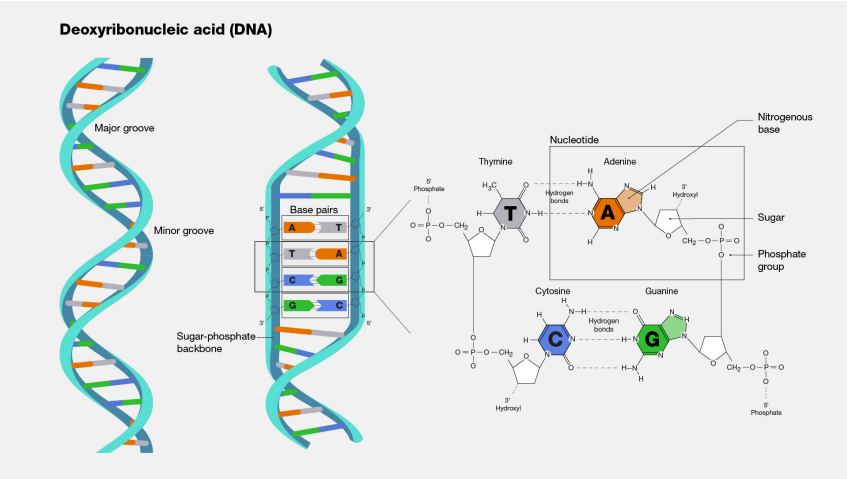
2003 Human Genome Project

The human genome was fully sequenced (i.e., the (order of) base pairs that make up human DNA was determined).

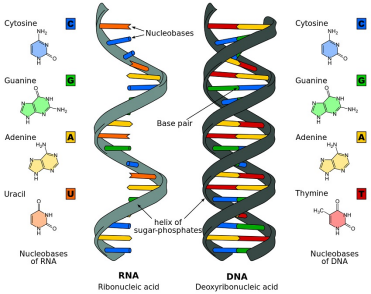
Basic Molecules

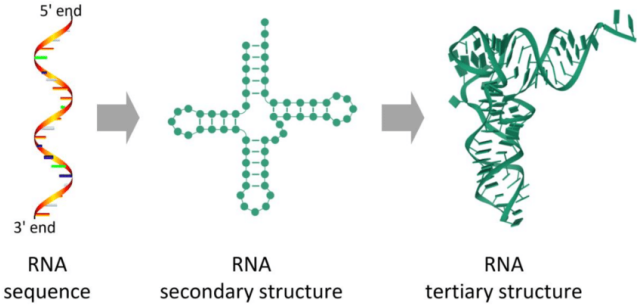
- DNA
  carries genetic information
- RNA
  - mRNA: convey genetic information from DNA to the ribosome
  - tRNA: linking codons to aminoacids
  - snRNA: splicing
  - microRNA: regulation of gene expression
  - RNA can act as genome (virus)
  - ...
- proteins
  perform a vast array of functions within living organisms, including catalyzing metabolic reactions, replicating DNA, responding to stimuli, and transporting molecules from one location to another.

DNA (Deoxyribonucleic acid)

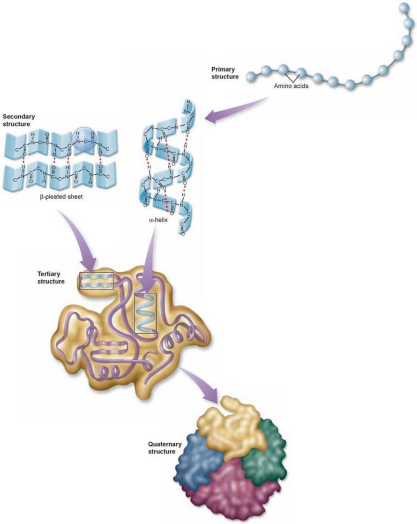


RNA (Ribonucleic acid)





Proteins


Primary structure
Amino acids
Secondary structure
β-pleated sheet
α-helix
Tertiary structure
Quaternary structure

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)

- RNA (Ribonucleic acid)
- Protein

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  - double-stranded helices of two polymers
- RNA (Ribonucleic acid)
- Protein

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  - double-stranded helices of two polymers
  - polymer made of nucleotides+backbone
- RNA (Ribonucleic acid)
- Protein

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  - double-stranded helices of two polymers
  - polymer made of nucleotides+backbone
  - guanine (G), adenine (A), thymine (T), cytosine (C)
- RNA (Ribonucleic acid)
- Protein

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  - double-stranded helices of two polymers
  - polymer made of nucleotides+backbone
  - guanine (G), adenine (A), thymine (T), cytosine (C)
  - alternating sugar (deoxyribose) and
    phospat groups (related to phosphoric acid)
    nucleotides are attached to sugar
- RNA (Ribonucleic acid)
- Protein

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  - double-stranded helices of two polymers
  - polymer made of nucleotides+backbone
  - guanine (G), adenine (A), thymine (T), cytosine (C)
  - alternating sugar (deoxyribose) and
    phospat groups (related to phosphoric acid)
    nucleotides are attached to sugar
  - the nucleotides of two polymers can bind (A-T, C-G)
- RNA (Ribonucleic acid)
- Protein

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  DNA = two sequences s_1, s_2 over the alphabet
  $\mathbb{A} = \{A, C, G, T\}$, where $X \in s_1$ can bind with $Y \in s_2$ if
  $XY \in \mathbb{B} = \{AT, TA, GC, CG\}$ (base pairing rules)
- RNA (Ribonucleic acid)

- Protein

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  DNA = two sequences s_1, s_2 over the alphabet
  $\mathbb{A} = \{A, C, G, T\}$, where $X \in s_1$ can bind with $Y \in s_2$ if
  $XY \in \mathbb{B} = \{AT, TA, GC, CG\}$ (base pairing rules)
- RNA (Ribonucleic acid)
  - single-stranded polymer
- Protein

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  DNA = two sequences s_1, s_2 over the alphabet
  $\mathbb{A} = \{A, C, G, T\}$, where $X \in s_1$ can bind with $Y \in s_2$ if
  $XY \in \mathbb{B} = \{AT, TA, GC, CG\}$ (base pairing rules)
- RNA (Ribonucleic acid)
  - single-stranded polymer
  - polymer made of nucleotides+backbone
- Protein

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  DNA = two sequences s_1, s_2 over the alphabet
  $\mathbb{A} = \{A, C, G, T\}$, where $X \in s_1$ can bind with $Y \in s_2$ if
  $XY \in \mathbb{B} = \{AT, TA, GC, CG\}$ (base pairing rules)
- RNA (Ribonucleic acid)
  - single-stranded polymer
  - polymer made of nucleotides+backbone
  - guanine (G), adenine (A), uracil (U), cytosine (C)
- Protein

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  DNA = two sequences s_1, s_2 over the alphabet
  $\mathbb{A} = \{A, C, G, T\}$, where $X \in s_1$ can bind with $Y \in s_2$ if
  $XY \in \mathbb{B} = \{AT, TA, GC, CG\}$ (base pairing rules)
- RNA (Ribonucleic acid)
  - single-stranded polymer
  - polymer made of nucleotides+backbone
  - guanine (G), adenine (A), uracil (U), cytosine (C)
  - alternating sugar (ribose) and
    phospat groups (related to phosphoric acid)
    nucleotides are attached to sugar
- Protein

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  DNA = two sequences s_1, s_2 over the alphabet
  $\mathbb{A} = \{A, C, G, T\}$, where $X \in s_1$ can bind with $Y \in s_2$ if
  $XY \in \mathbb{B} = \{AT, TA, GC, CG\}$ (base pairing rules)
- RNA (Ribonucleic acid)
  - single-stranded polymer
  - polymer made of nucleotides+backbone
  - guanine (G), adenine (A), uracil (U), cytosine (C)
  - alternating sugar (ribose) and
    phospat groups (related to phosphoric acid)
    nucleotides are attached to sugar
  - the nucleotides of polymer can bind (A-U, C-G, G-U)
- Protein

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  DNA = two sequences s_1, s_2 over the alphabet
  $\mathbb{A} = \{A, C, G, T\}$, where $X \in s_1$ can bind with $Y \in s_2$ if
  $XY \in \mathbb{B} = \{AT, TA, GC, CG\}$ (base pairing rules)
- RNA (Ribonucleic acid)
  RNA = single sequence s over the alphabet
  $\mathbb{A} = \{A, C, G, U\}$, where $X \in s$ can bind with $Y \in s$ if
  $XY \in \mathbb{B} = \{AU, UA, GC, CG, GU, UG\}$
- Protein

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  DNA = two sequences s_1, s_2 over the alphabet
  $\mathbb{A} = \{A, C, G, T\}$, where $X \in s_1$ can bind with $Y \in s_2$ if
  $XY \in \mathbb{B} = \{AT, TA, GC, CG\}$ (base pairing rules)
- RNA (Ribonucleic acid)
  RNA = single sequence s over the alphabet
  $\mathbb{A} = \{A, C, G, U\}$, where $X \in s$ can bind with $Y \in s$ if
  $XY \in \mathbb{B} = \{AU, UA, GC, CG, GU, UG\}$
- Protein
  - large molecule made of amino acids

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  DNA = two sequences s_1, s_2 over the alphabet
  $\mathbb{A} = \{A, C, G, T\}$, where $X \in s_1$ can bind with $Y \in s_2$ if
  $XY \in \mathbb{B} = \{AT, TA, GC, CG\}$ (base pairing rules)
- RNA (Ribonucleic acid)
  RNA = single sequence s over the alphabet
  $\mathbb{A} = \{A, C, G, U\}$, where $X \in s$ can bind with $Y \in s$ if
  $XY \in \mathbb{B} = \{AU, UA, GC, CG, GU, UG\}$
- Protein
  - large molecule made of amino acids
  - order of amino acids determined by order of genes

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  DNA = two sequences s_1, s_2 over the alphabet
  $\mathbb{A} = \{A, C, G, T\}$, where $X \in s_1$ can bind with $Y \in s_2$ if
  $XY \in \mathbb{B} = \{AT, TA, GC, CG\}$ (base pairing rules)
- RNA (Ribonucleic acid)
  RNA = single sequence s over the alphabet
  $\mathbb{A} = \{A, C, G, U\}$, where $X \in s$ can bind with $Y \in s$ if
  $XY \in \mathbb{B} = \{AU, UA, GC, CG, GU, UG\}$
- Protein
  - large molecule made of amino acids
  - order of amino acids determined by order of genes
  - in general, genetic code specifies 20 standard amino acids

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  DNA = two sequences s_1, s_2 over the alphabet
  $\mathbb{A} = \{A, C, G, T\}$, where $X \in s_1$ can bind with $Y \in s_2$ if
  $XY \in \mathbb{B} = \{AT, TA, GC, CG\}$ (base pairing rules)
- RNA (Ribonucleic acid)
  RNA = single sequence s over the alphabet
  $\mathbb{A} = \{A, C, G, U\}$, where $X \in s$ can bind with $Y \in s$ if
  $XY \in \mathbb{B} = \{AU, UA, GC, CG, GU, UG\}$
- Protein
  Protein = sequence over the alphabet $\mathbb{A}$ =set of 20 aminoacids

Understand Inheritance - Math. Framework

- DNA (Deoxyribonucleic acid)
  DNA = two sequences s_1, s_2 over the alphabet
  $\mathbb{A} = \{A, C, G, T\}$, where $X \in s_1$ can bind with $Y \in s_2$ if
  $XY \in \mathbb{B} = \{AT, TA, GC, CG\}$ (base pairing rules)
- RNA (Ribonucleic acid)
  RNA = single sequence s over the alphabet
  $\mathbb{A} = \{A, C, G, U\}$, where $X \in s$ can bind with $Y \in s$ if
  $XY \in \mathbb{B} = \{AU, UA, GC, CG, GU, UG\}$
- Protein
  Protein = sequence over the alphabet $\mathbb{A}$ =set of 20 aminoacids

What is the genetic code?
How is the information on DNA used to code proteins?

Some More History:
Cracking the genetic code - The magic number 20

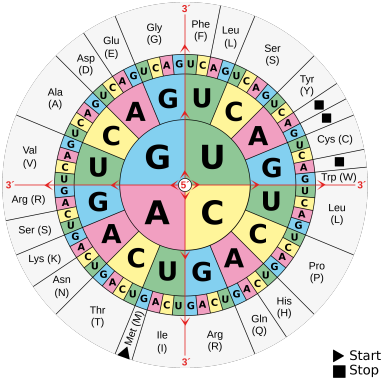
Question: How can a 4-letter alphabet code for 20 aminoacids?

- Garmov - Diamond Code
- Crick - Non-Overlapping Commafree Code
- Nirenberg - Matthaei - Experiment

→ board

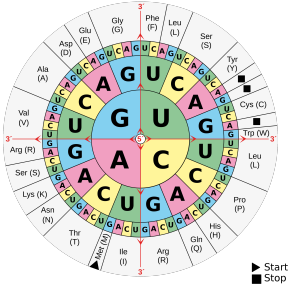
https://www.chemistryviews.org/details/ezine/11312121/Deciphering_the_Genetic_Code_The_Most_Beautiful_False_Theory_in_Biochemistry__Pa/

Genetic code is simply a map $f : C \to A$ where, $C = \{(x_1 x_2 x_3) \mid x_i \in \{A, C, G, U\}\}$ and $A =$ set of aminoacids and start/termination codon.



Amino acid	Genetic code	Abbr
Alanine	GCA GCC GCG GCU	Ala
Arginine	AGA AGG CGA CGC CGG CGU	Arg
Asparagine	AAC AAU	Asn
Aspartic acid	GAC GAU	Asp
Cysteine	UGC UGU	Cys
Glutamine	CAA CAG	Gln
Glutamic acid	GAA GAG	Glu
Glycine	GGA GGC GGG GGU	Gly
Histidine	CAC CAU	His
Isoleucine	AUA AUC AUU	Ile
Leucine	CUA CUC CUG CUU UUA UUG	Leu
Lysine	AAA AAG	Lys
Methionine	AUG	Met
Phenylalanine	UUC UUU	Phe
Proline	CCA CCC CCG CCU	Pro
Serine	AGC AGU UCA UCC UCG UCU	Ser
Threonine	ACA ACC ACG ACU	Thr
Tryptophan	UGG	Try
Tyrosine	UAC UAU	Tyr
Valine	GUA GUC GUG GUU	Val
STOP sign	UAA UAG UGA	

Genetic code is simply a map $f : C \to A$ where, $C = \{(x_1 x_2 x_3) \mid x_i \in \{A, C, G, U\}\}$ and $A =$ set of aminoacids and start/termination codon.



Amino acid	Genetic code	Abbr
Alanine	GCA GCC GCG GCU	Ala
Arginine	AGA AGG CGA CGC CGG CGU	Arg
Asparagine	AAC AAU	Asn
Aspartic acid	GAC GAU	Asp
Cysteine	UGC UGU	Cys
Glutamine	CAA CAG	Gln
Glutamic acid	GAA GAG	Glu
Glycine	GGA GGC GGG GGU	Gly
Histidine	CAC CAU	His
Isoleucine	AUA AUC AUU	Ile
Leucine	CUA CUC CUG CUU UUA UUG	Leu
Lysine	AAA AAG	Lys
Methionine	AUG	Met
Phenylalanine	UUC UUU	Phe
Proline	CCA CCC CCG CCU	Pro
Serine	AGC AGU UCA UCC UCG UCU	Ser
Threonine	ACA ACC ACG ACU	Thr
Tryptophan	UGG	Try
Tyrosine	UAC UAU	Tyr
Valine	GUA GUC GUG GUU	Val
STOP sign	UAA UAG UGA	

From a math. POV, this code is not elegant and does not seem to follow a systematic way.

Crick called this code "frozen accident"

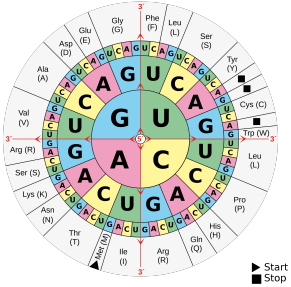
In 1990's, changes in in genetic code were observed:

stop codon: UGA (usually) $\to$ Try (in some plants)

stop codon: UAA (usually) $\to$ Tyr (flatworms)

$\implies$ there are changes (not frozen)!

Genetic code is simply a map $f : C \to A$ where, $C = \{(x_1 x_2 x_3) \mid x_i \in \{A, C, G, U\}\}$ and $A =$ set of aminoacids and start/termination codon.



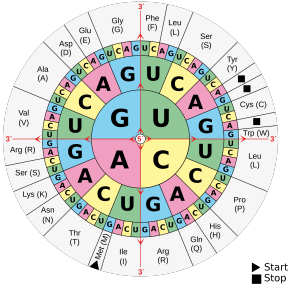
Amino acid	Genetic code	Abbr
Alanine	GCA GCC GCG GCU	Ala
Arginine	AGA AGG CGA CGC CGG CGU	Arg
Asparagine	AAC AAU	Asn
Aspartic acid	GAC GAU	Asp
Cysteine	UGC UGU	Cys
Glutamine	CAA CAG	Gln
Glutamic acid	GAA GAG	Glu
Glycine	GGA GGC GGG GGU	Gly
Histidine	CAC CAU	His
Isoleucine	AUA AUC AUU	Ile
Leucine	CUA CUC CUG CUU UUA UUG	Leu
Lysine	AAA AAG	Lys
Methionine	AUG	Met
Phenylalanine	UUC UUU	Phe
Proline	CCA CCC CCG CCU	Pro
Serine	AGC AGU UCA UCC UCG UCU	Ser
Threonine	ACA ACC ACG ACU	Thr
Tryptophan	UGG	Try
Tyrosine	UAC UAU	Tyr
Valine	GUA GUC GUG GUU	Val
STOP sign	UAA UAG UGA	

Could this code be a result of evolutionary "optimization" processes?

Freeland and Hurst (1998): If genetic code is result of evol. optimization,
then it must dominate/outperform other possible codes.

What does outperform mean? (a measure is needed!)

Genetic code is simply a map $f : C \to A$ where, $C = \{(x_1 x_2 x_3) \mid x_i \in \{A, C, G, U\}\}$ and $A =$ set of aminoacids and start/termination codon.



Amino acid	Genetic code	Abbr
Alanine	GCA GCC GCG GCU	Ala
Arginine	AGA AGG CGA CGC CGG CGU	Arg
Asparagine	AAC AAU	Asn
Aspartic acid	GAC GAU	Asp
Cysteine	UGC UGU	Cys
Glutamine	CAA CAG	Gln
Glutamic acid	GAA GAG	Glu
Glycine	GGA GGC GGG GGU	Gly
Histidine	CAC CAU	His
Isoleucine	AUA AUC AUU	Ile
Leucine	CUA CUC CUG CUU UUA UUG	Leu
Lysine	AAA AAG	Lys
Methionine	AUG	Met
Phenylalanine	UUC UUU	Phe
Proline	CCA CCC CCG CCU	Pro
Serine	AGC AGU UCA UCC UCG UCU	Ser
Threonine	ACA ACC ACG ACU	Thr
Tryptophan	UGG	Try
Tyrosine	UAC UAU	Tyr
Valine	GUA GUC GUG GUU	Val
STOP sign	UAA UAG UGA	

2 extremes

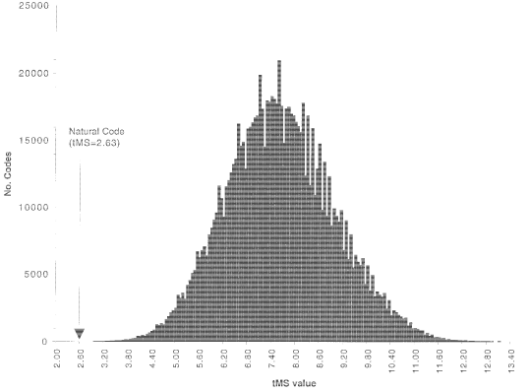
"worst" case: Mutation of single nucleotide in DNA results in new aminoacid that then leads to new but useless protein = death of organism

("low" error tolerance)

"best" case: Mutation of single nucleotide in DNA may result in new aminoacid but preserves functionality of protein = organism can survive

("high" error tolerance)

Genetic code is simply a map $f : C \to A$ where, $C = \{(x_1 x_2 x_3) \mid x_i \in \{A, C, G, U\}\}$ and $A =$ set of aminoacids and start/termination codon.



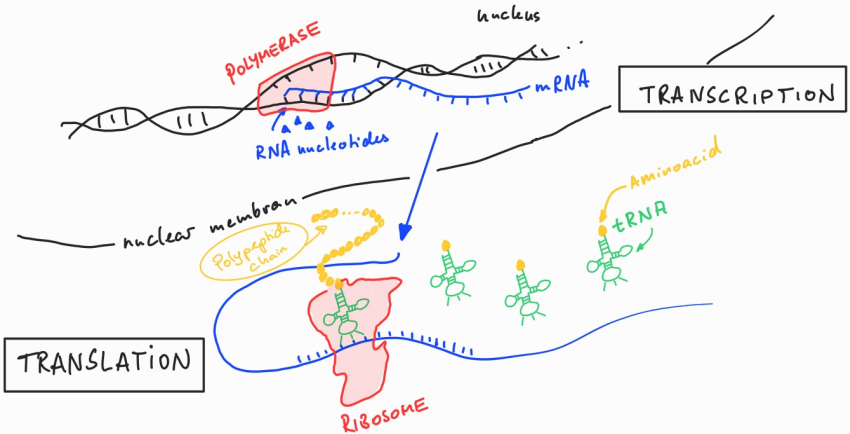
tMS value = error-proneness / No of codes

Based on the latter idea (and many more), Freeland and Hust quantified possible "meaningful" genetic codes and sampled among the $\sim 2,5 \times 10^{18}$ hypothetical codes $\sim 10^6$

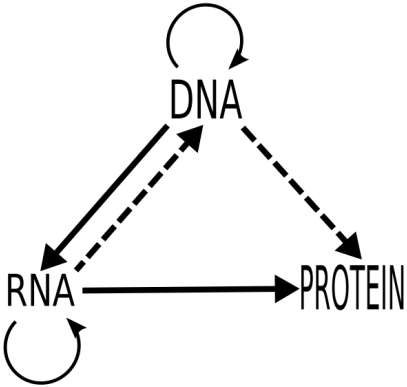
Among them only one was slighlty "better" (clear hint for evol. opt.)

Evolution is still running!

Protein Synthesis - what we know now nowadays



Central Dogma - what we know now nowadays



DNA->DNA	DNA Replication
DNA->RNA	Transciption
RNA->Protein	Translation
RNA->DNA	Reverse Transcription (e.g. eukaryotes[a] or retroviruses (as HIV))
RNA->RNA	RNA replication (e.g. in many viruses)
DNA->Protein	Direct Translation (in vitro)

[a] organisms whose cells have a membrane-bound nucleus (in contrast to Prokaryotes)