Genetic code
Cells
Xenobiology

Planets and Astrobiology (2018-2019)
G. Vladilo

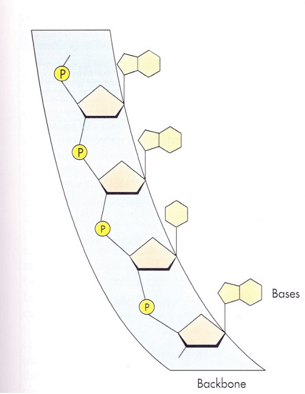
Genetic information

The genetic information is stored in <u>digital form</u> in the <u>sequence</u> of nucleobases attached to the backbone of the nucleic acids

The <u>order of the nucleobases</u> is not constrained by chemical laws;
we believe it is the result of eons of modifications and natural selection

Terrestrial life has started to use digital information more than 3.5 billon years ago

Digital information is more stable than analogic information

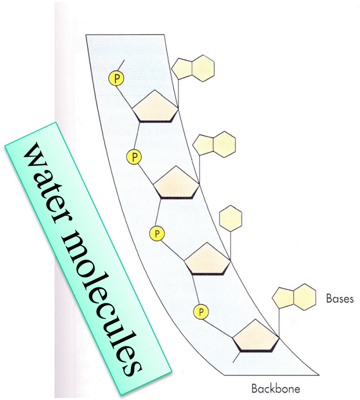


Chemical stability of genetic information

Why is the order of the nucleobases unconstrained by chemical laws?

The repeating charges in the backbone interact with the surrounding water molecules, which are polar

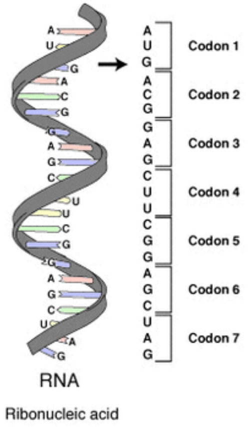
This interaction dominates the physical properties: replacing a nucleobase with another has only a second-order effect on the physical behaviour of the molecule



Genetic information: the codons

The <u>digital information</u> is coded <u>in triplets of nucleobases</u> called <u>codons</u>

Each codon uses 3 of the 4 nucleobases and can express $4^3=64$ possible combinations, equivalent to 6 bits of information ($64=2^6$)



The genetic code

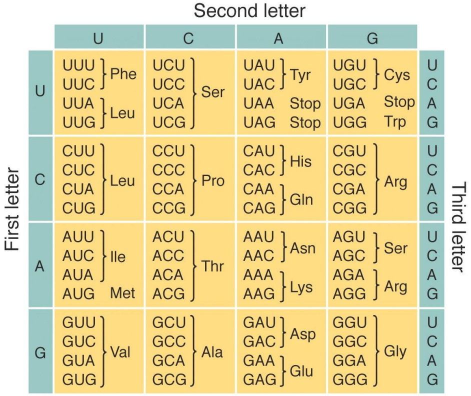
Each codon uniquely identifies a single amino acid

Some aminoacids are coded by more than one codon

Some codons are used as a “stop” signal of the sequence

Correspondence between RNA codons and amino acids

A=Adenine
G=Guanine
C=Citosine
U=Uracyl



Second letter

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

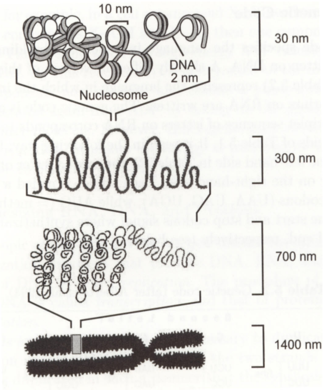
Genes

From the structural point of view, a gene is a sequence of nucleobases along a strand of a nucleic acid

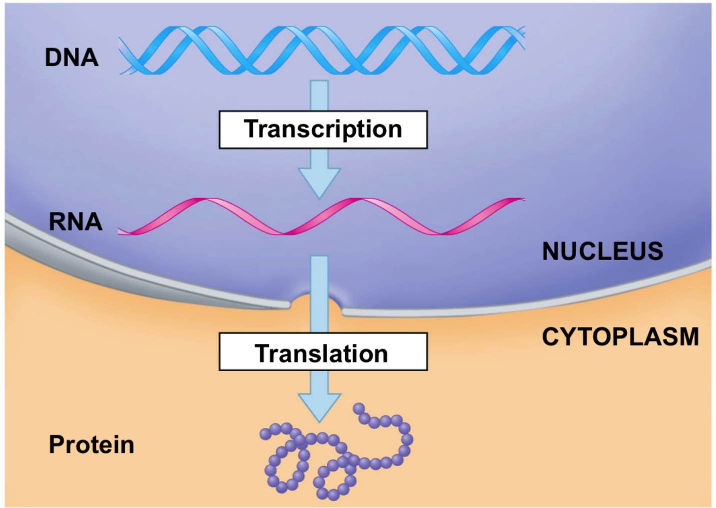
From the point of view of the information content, a gene is a sequence of instructions with a specific function

As an example, a sequence that specifies how to build up a specific amino acid

In complex organisms, the number of genes is extremely high and this is why DNA needs to be stored in very compact structures, such as chromosomes



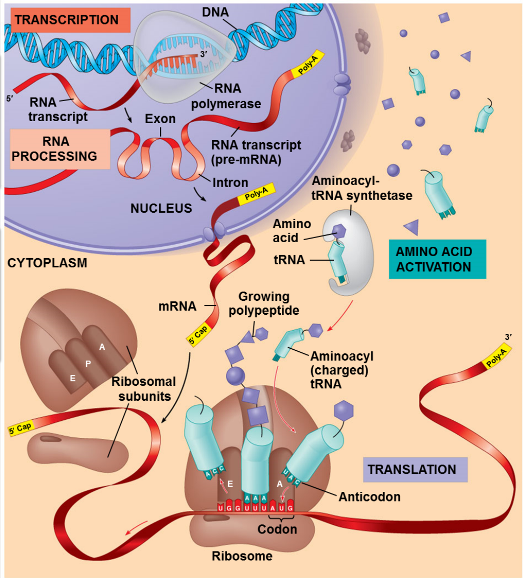
The “central dogma”



A closer look to the molecular machinery

Exons:
coding DNA
Introns:
non-coding DNA

mRNA:
messenger RNA
tRNA:
transfer RNA



The functioning of the molecular machinery requires the presence of nuclei acids and proteins (enzymes)

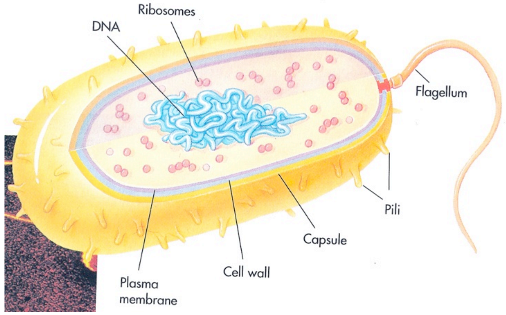
Cells in terrestrial life

- Cells
  - The cell is the basic structural, functional, and biological unit of all known living organisms
  - A cell is the smallest unit of life that can replicate independently

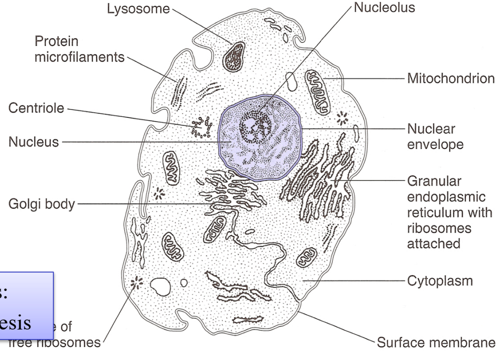
  - The volume of the cell is filled with the cytoplasm, a crowded solution of many different types of molecules embedded in a solution (the cytosol)
  - The cytoplasm is enclosed within a membrane, which contains many biomolecules such as proteins and nucleic acids
  - The membrane is a phospholip bilayer with embedded proteinic structures

  - Terrestrial life is characterized by two different types of cells:
    prokaryotic and eukaryotic

- Prokaryotic cells
  - The genetic material is not enclosed in a nucleus
  - The metabolism is carried out directly in the cytosol
  - Sizes are relatively small: 1 – 5 µm
  - They may form colonies, but not multicellular organisms
  - Phylogenetic analysis has revealed the existence of two types of prokaryotic cells: Bacteria and Archaea



- Eukaryotic cells
  - Eukaryotic cells have a much higher level of internal organization
  - The genetic material is enclosed in a nucleus
  - The metabolism is carried out in the organelles, i.e. specialized sub-structures embedded in the cytosol
  - Eukaryotic cells can form multicellular organisms
  - Larger sizes: typically 10 – 100 µm



Mitochndria: Energy production

Ribosomes: Protein synthesis

The diversity of terrestrial life

Terrestrial life shows a great diversity, with a large number of species

The number of species currently living on Earth is in the order of 10^7
(of which ~10^6 are documented)

The total number of extinct species in the course of the evolution of terrestrial life is estimated to be in the order of a few 10^9

The largest diversity is found in the unicellular world,
rather than among multicellular organisms

The unity of terrestrial life at the molecular level

Despite their extremely large biodiversity,
all terrestial organisms show a remarkable unity at the molecular level

The genetic code is shared by all organisms,
from bacteria to men (with rare cases of minor variations)

All terrestrial life uses a well-defined set of biomolecules
selected among countless possibilities provided by organic chemistry
ATP, RNA, DNA, 20 L-aminoacids …

The unity of terrestrial life at the molecular level

The unity of terrestrial life at the molecular level
suggests a common origin of all terrestrial organisms

The unity of terrestrial life may also suggest
that the biomolecules that we know
are the best choice for their specific functions and
are a likely outcome of chemical pathways

Possible types of chemical life

The special properties of water and carbon suggest that they may be essential ingredients also in exobiology

Also the high cosmic abundances of H, O and C favours this possibility

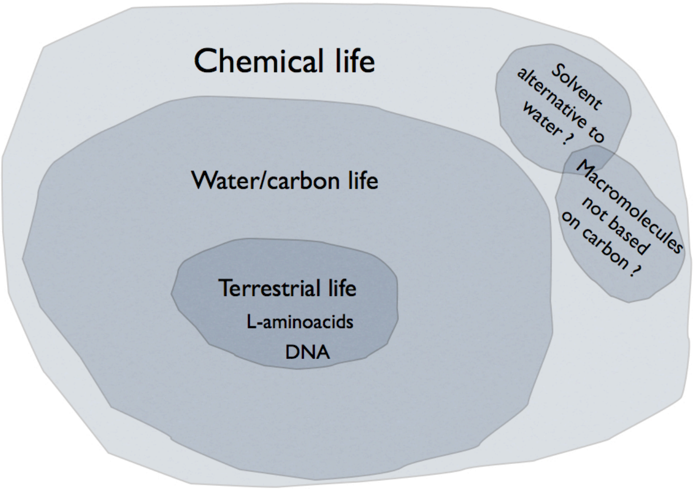
However, we cannot exclude that forms of life not based on carbon and water may have developed in particular regions of the universe

For instance, should the local physical/chemical conditions prevent the use of carbon and water, at variance with the earth's conditions

In any case, the viability of a biochemistry not based on carbon and water needs to be investigated

---

Possible types of chemical life in the Universe



Life based on carbon and water does not need to be equal
to the terrestrial one

we may expect significant differences at the level of
molecular constituents and at higher levels

Chirality

In non terrestrial organisms, biological macromolecules could have a type of chirality different from that of terrestrial life (as an example the amino acids, if present, might have D, rather than L, chirality)

Genetic information

In non terrestrial organisms, the genetic information could be coded using molecules other than the RNA and DNA

Also the genetic code could be different

Testing alternative biochemistries starting from manipulations of terrestrial life

The current advancements in biotechnology provide the possibility to test the viability of alternative forms of chemical life

Synthetic biology

Emerging scientific field at the cross road between
biotechnology and nanotechnology

Artificial design and engineering of biological systems and living organisms for purposes of improving applications for industry or biological research

---

Xenobiology

Xenobiology (XB) is a subfield of synthetic biology that
describes a form of biology that is not found in nature

In practice it describes novel biological systems and biochemistries
that differ from the canonical DNA-RNA-20 amino acid system
(i.e., the classical central dogma in molecular biology)

Originally a research on alternative forms of DNA
was driven by the question of how life evolved on earth
and why RNA and DNA were selected by (chemical) evolution
over other possible nucleic acid structures

Xenobiology

Systematic experimental studies aiming at the diversification
of the chemical structure of nucleic acids
have resulted in completely novel informational biopolymers

For example, instead of DNA or RNA,
xenobiology explores nucleic acid analogues,
termed Xeno Nucleic Acid (XNA) as information carriers

It also focuses on an expanded genetic code and the incorporation of non-proteinogenic amino acids into proteins

So far a number of XNAs with new chemical backbones of the DNA
have been synthesized

Expanding the genetic alphabet

Instead of modifying the backbones,
other experiments target the replacement or enlargement
of the genetic alphabet of DNA with unnatural base pairs

The viability of candidate bases
for possible incorporation in the DNA is being tested

For example, DNA has been designed that has - instead of the four standard bases A,T,G, and C - six bases A, T, G, C, and the two new ones P and Z

New candidate bases may potentially yield a large number of base pairs

Genetic code engineering

One of the goals of xenobiology is to rewrite the genetic code.

The repertoire of 20 canonical amino acids can in principle be expanded.

Existing codons can in principle be reprogrammed

An even more radical approach is the change of a triplet codon to a quadruplet and even pentaplet codon

Experiments of this type are under way and have already shown the feasibility of changes of this type in a limited number of cases

Alternative biomolecules for metabolism

New types of enzymes can in principle be synthesized

This possibility can be expanded by changing the repertoire of 20 canonical amino acids

The experiments of xenobiology may eventually demonstrate that life based on carbon and water can potentially use a broad spectrum of biomolecules

An interesting question related to alternative forms of carbon-water life is whether such forms can interact with life-as-we-know-it

Biosafety

Hypothetical organisms that use XNA, different base pairs and polymerases and has an altered genetic code will hardly be able to interact with natural forms of life on the genetic level

Thus, xenobiological organisms represent a genetic enclave that cannot exchange information with natural cells

Altering the genetic machinery of the cell leads to semantic containment

In analogy to information processing in IT, this safety concept is termed a "genetic firewall"

The concept of the genetic firewall seems to overcome a number of limitations of previous safety systems

Implications for astrobiology:

non-terrestrial forms of life, even if based on carbon and water, may not pose a biological hazard to terrestrial life

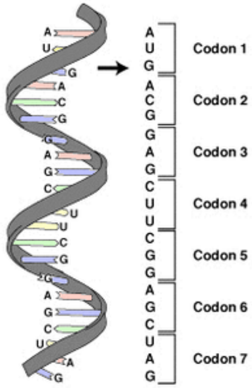
DNA as a digital storage device

A DNA strand with 10^6 nucleotides, typical number in the simple unicellular organisms, can store up to 250 kB of information

Each codon has 6 bits of information (64 combinations=2^6), corresponding to 0.75 Bytes (1 Byte=8 bits)

The largest human chromosome, with ~250 x 10^6 nucleotides, can store up to ~ 62 MB

These numbers are not particularly impressive in absolute value. However, the density of information per unit volume is extremely high. This is why DNA has the potential to be used as a powerful storage device



RNA

Ribonucleic acid

DNA as a digital storage device

- The density of information (MB/unit volume) stored in the DNA can be calculated as follows
- The radius of the helix is $r_h \sim 1$ nm, while the perpendicular distance between adjacent nucleobasis $d_n \sim 0.34$ nm
- The volume occupied by a codon (3 nucleobasis) is therefore $V_{codon} \cong 3\, d_n\, \pi\, r_h^2 \cong 3.2\ \text{nm}^3$
- Each codon has 0.75 bytes
- The density of information is therefore $0.75\ \text{B}/(3.2\ \text{nm}^3) = 0.23\ \text{B/nm}^3 = 2.3\text{x}10^5\ \text{TB/mm}^3$
- This is the maximum density of information that can be obtained by compactified DNA strands
- This density is largest than that of present-day storage devices (e.g. USB pen) by several orders of magnitude