Astrobiology
Genetic information

Planets and Astrobiology, Academic Year 2019-2020

Giovanni Vladilo (INAF-OATs)

Genetic information

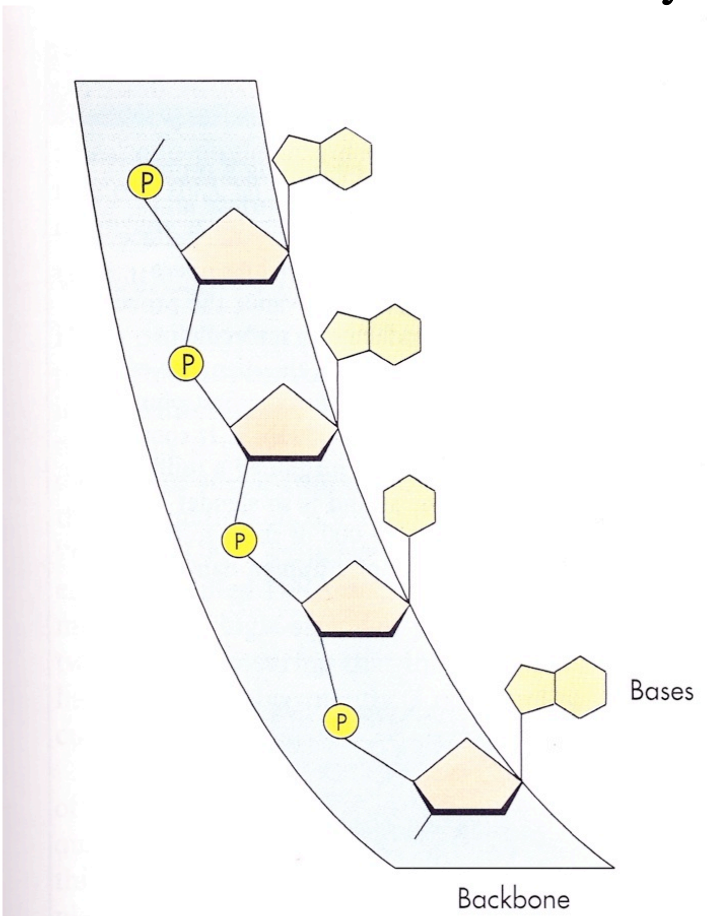
The order of the nucleobases attached to the backbone of the nucleic acids determines the genetic information, which is therefore stored in digital form

The order of the nucleobases is not constrained by chemical laws

We believe this is the result of natural selection at early stages of life evolution

Digital information is more stable than analogic information

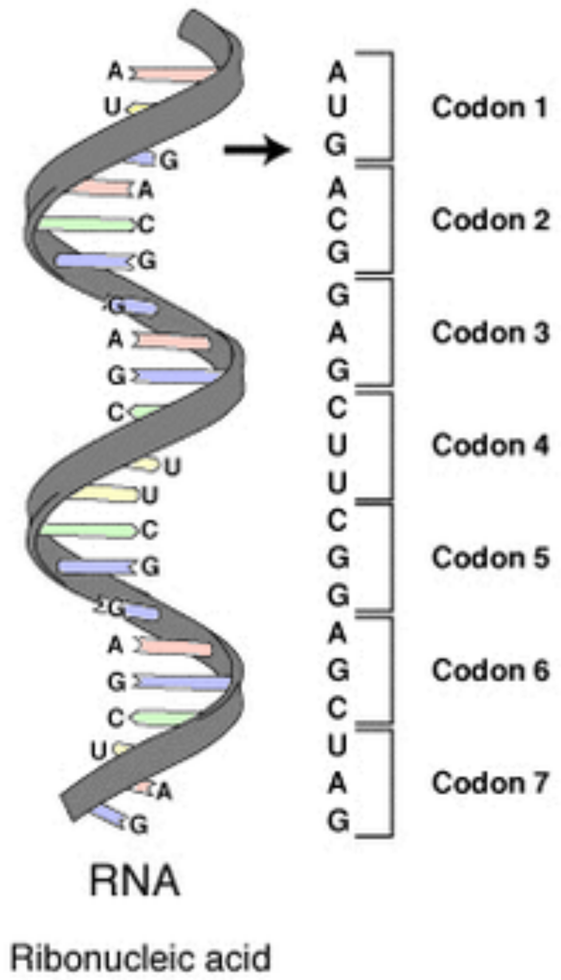
Terrestrial life has started to use digital information more than 3.5 billon years ago



The genetic information and genetic code

The digital information is coded in triplets of nucleobases called codons

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



Each codon uniquely identifies a single amino acid

Some aminoacids are coded by more than one codon (example of unplanned evolution)

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

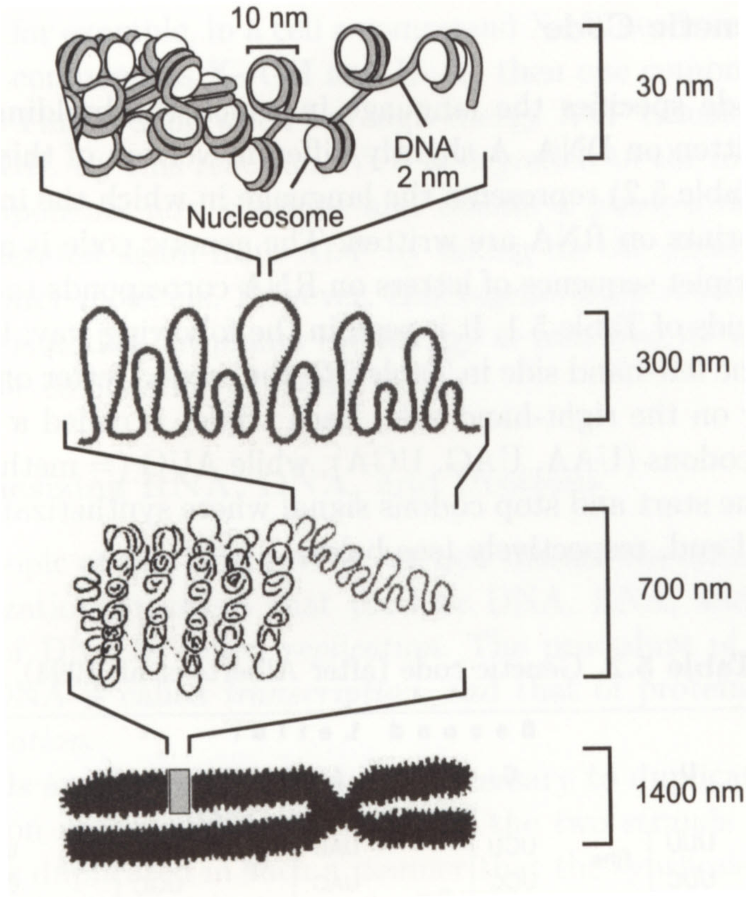
Genes

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

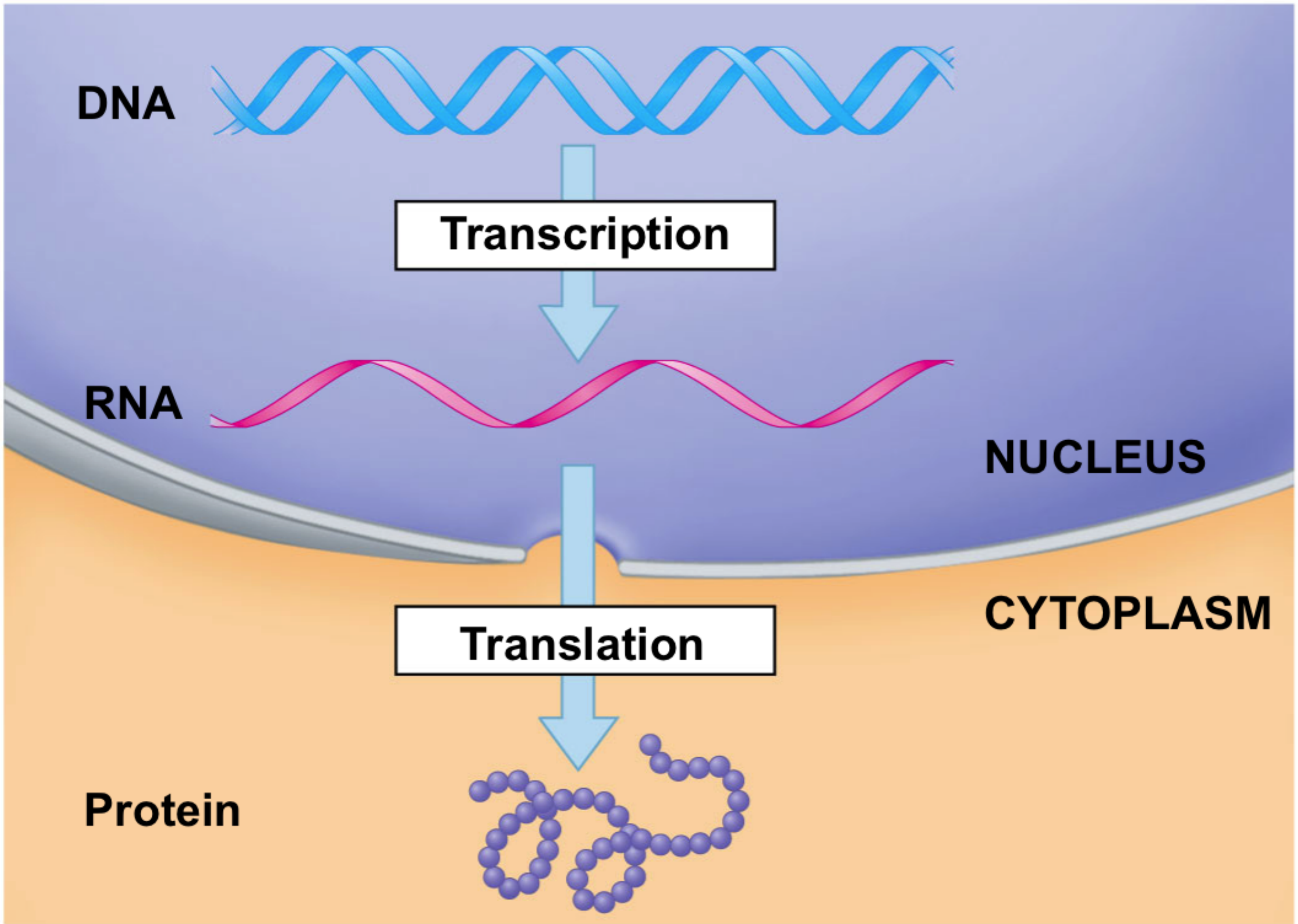
From the point of view of the information content, a gene is a sequence of instructions with a specific function

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



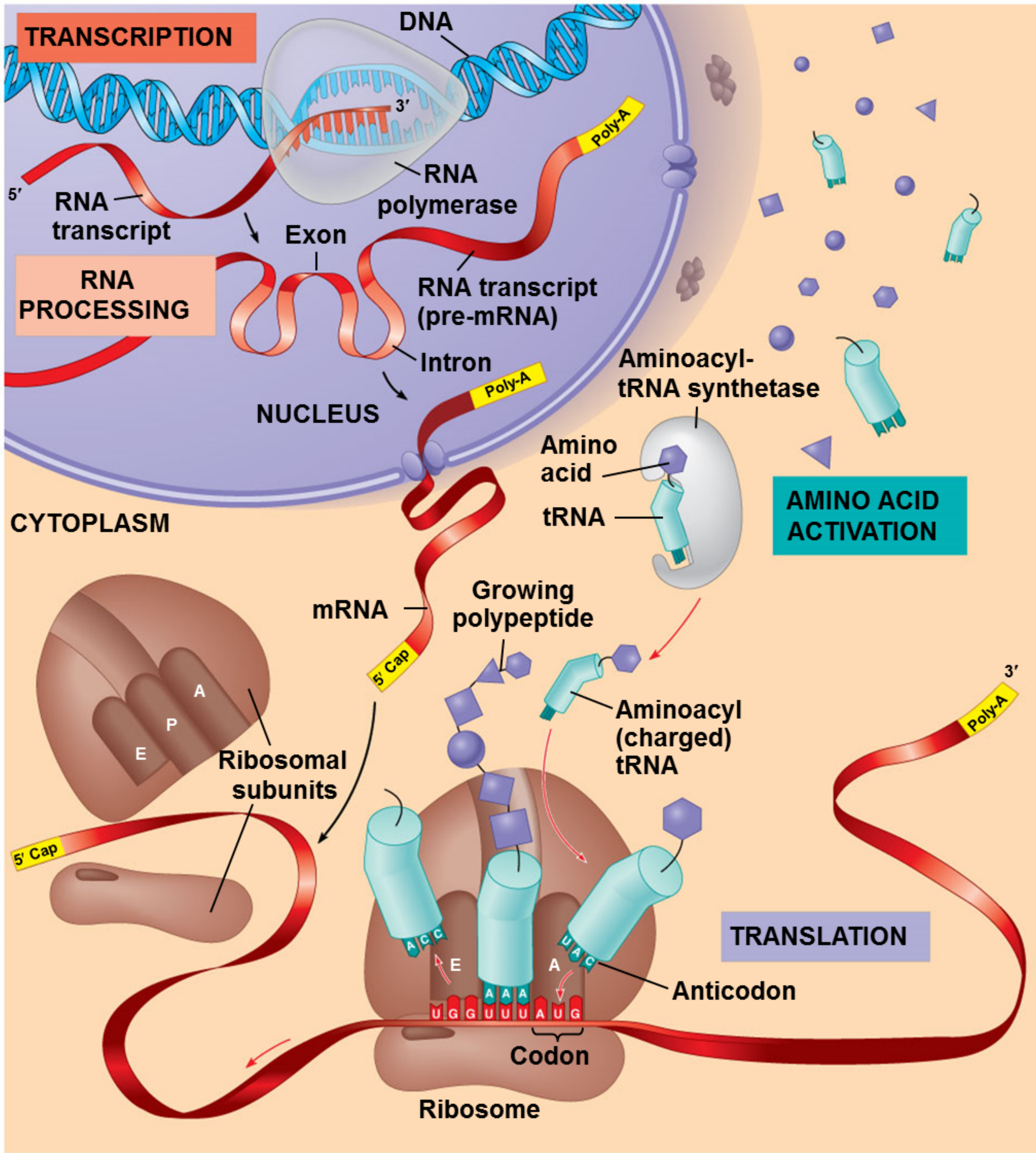
The "central dogma"



A closer look to the molecular machinery

Exons:
coding DNA
Introns:
non-coding DNA

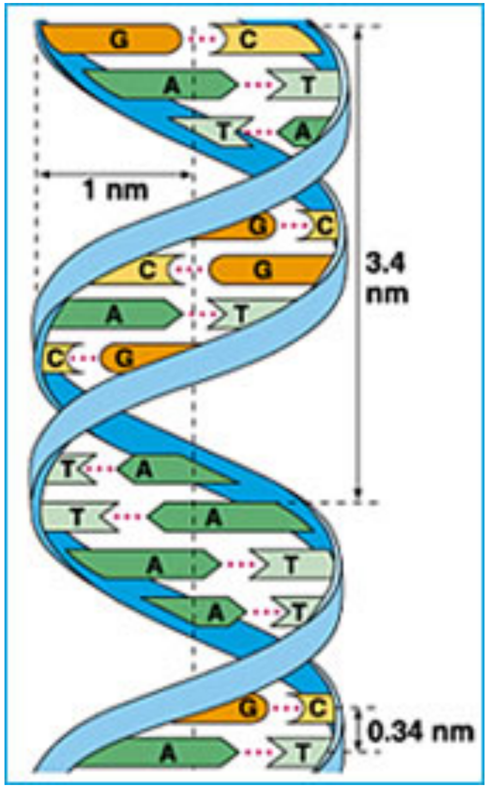
mRNA:
messenger RNA
tRNA:
transfer RNA



The functioning of the molecular machinery requires the presence of nuclei acids <u>and</u> proteins (enzymes)

Appendix:
the 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 $V_{codon} \cong 3\, d_n\, \pi\, r_h^2 \cong 3.2\ nm^3$
- Each codon has 6 bits of information (64 combinations=2^6), corresponding to 0.75 bytes
- The density of information is therefore
  $0.75\ B/(3.2\ nm^3) = 0.23\ B/nm^3 = 2.3x10^5\ TB/mm^3$
- This is the maximum density of information that can be obtained by compactified DNA strands



- Exercise: check that this density of information is largest than that of present-day storage devices by several orders of magnitude