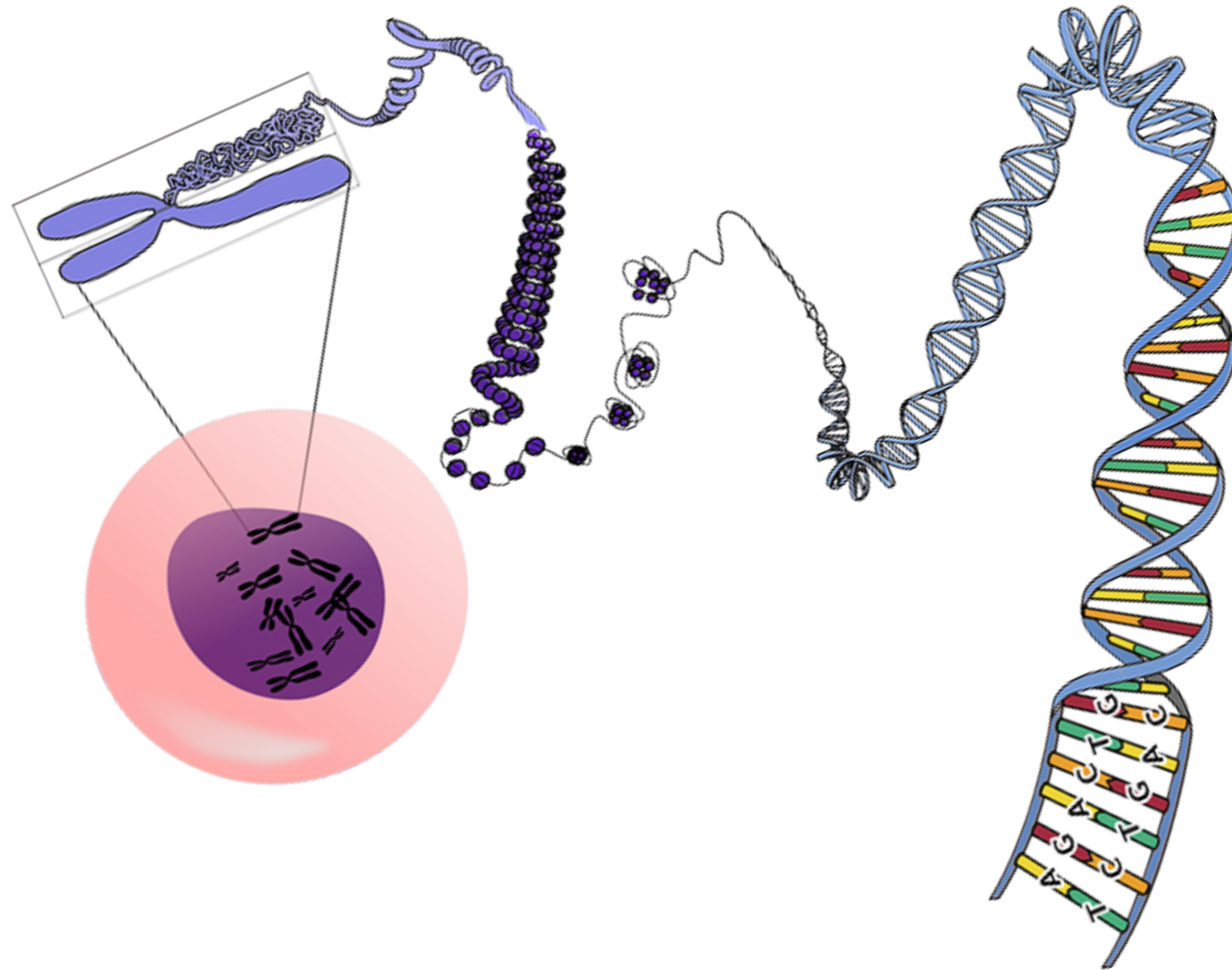




DNA structure

DNA, gene, alleles,
chromosome, trait,
nucleotide

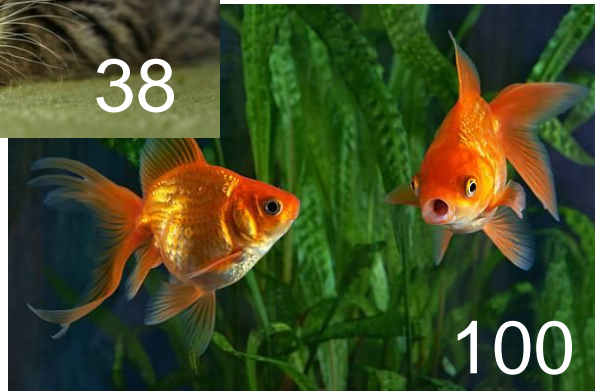


<https://pixabay.com/vectors/genetics-chromosomes-rna-dna-156404/>

Some definitions

- DNA is a molecule shaped as a double helix (twisted, 2 strands) that holds the genetic information of an organism.
- A gene is a section of DNA that codes for a particular trait, e.g. gene for hair colour.
- An allele is an alternative form of a gene, e.g., blond hair and black hair are alleles of the gene for hair colour.
- A trait is a characteristic that an organism shows.
- Chromosome is the 3D shape made of DNA and proteins that is visible before a cell divides through mitosis or meiosis.
- Nucleotide is the basic unit of DNA.

Each species has different numbers of chromosomes



Each human cell:
3.2. billions (3200
millions) nucleotides
(human genome)

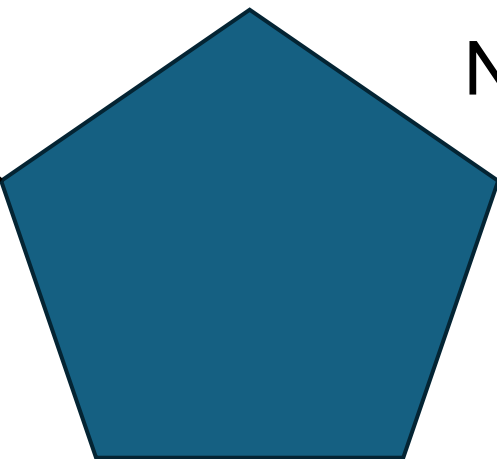
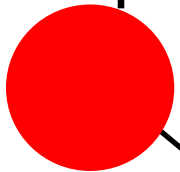
DNA is about 2 m long

About 20 000 protein-
coding genes (1.5% of
the entire genome)

Photos from pixabay.com

Nucleotide

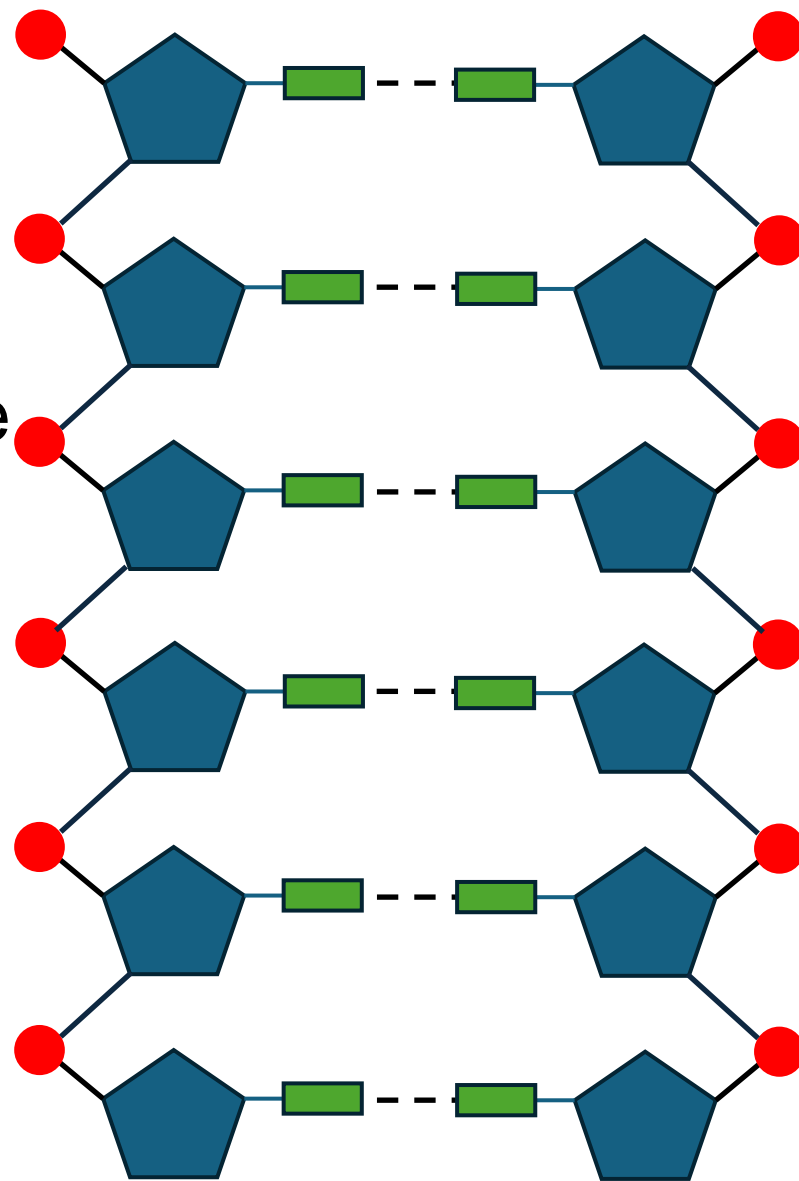
phosphate



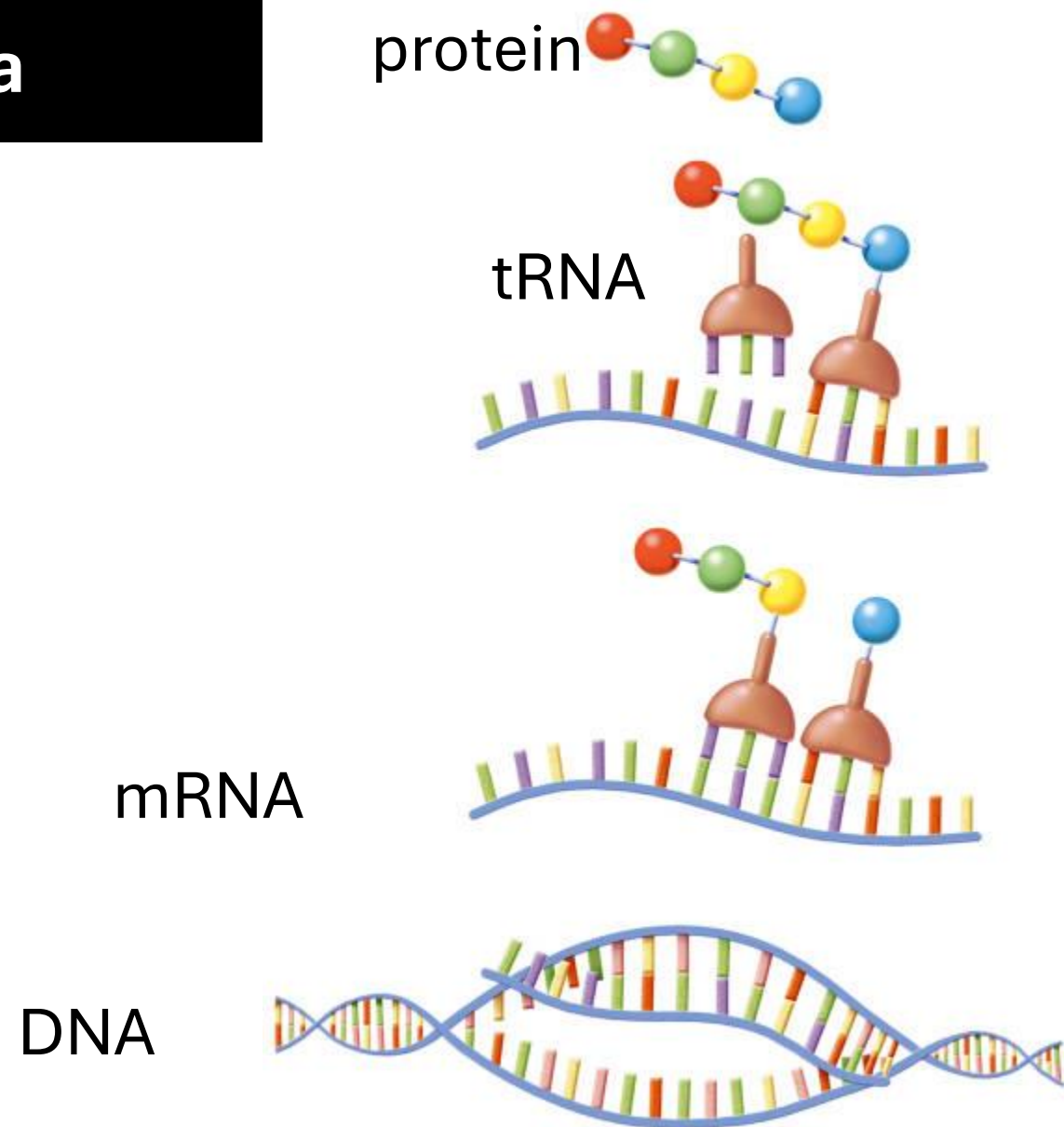
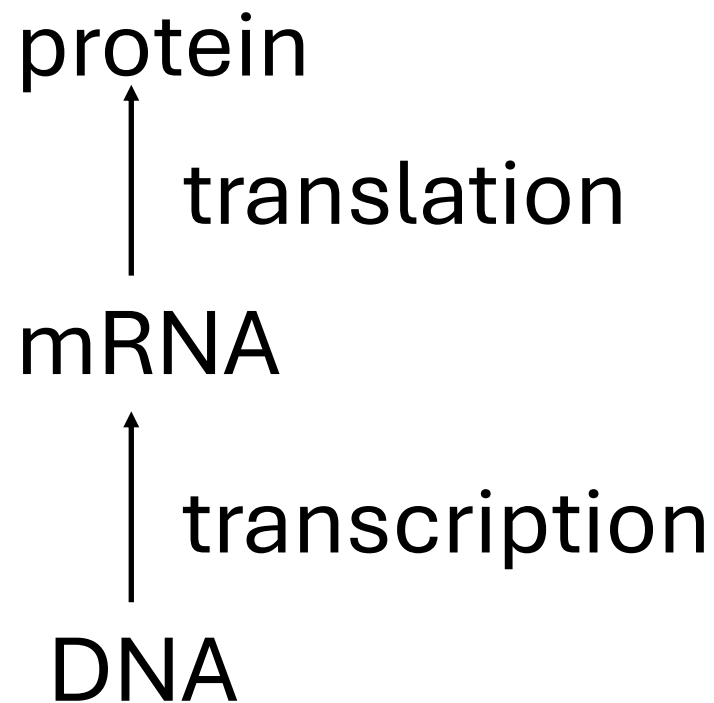
Nitrogenous base



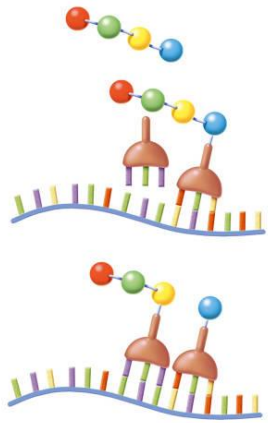
Sugar = deoxyribose



DNA to protein – the central dogma

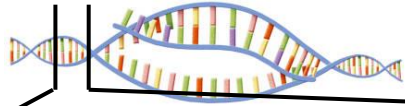


protein

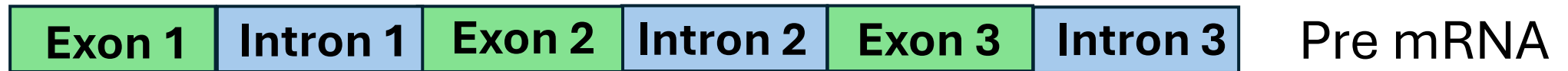
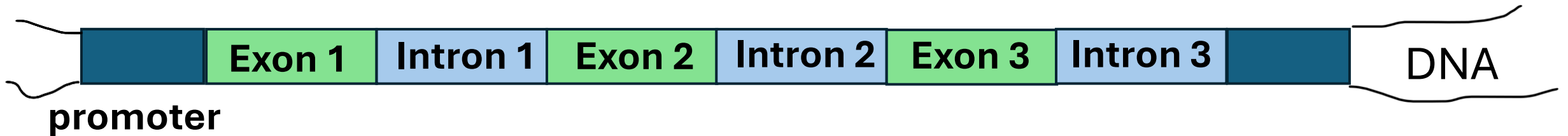


mRNA

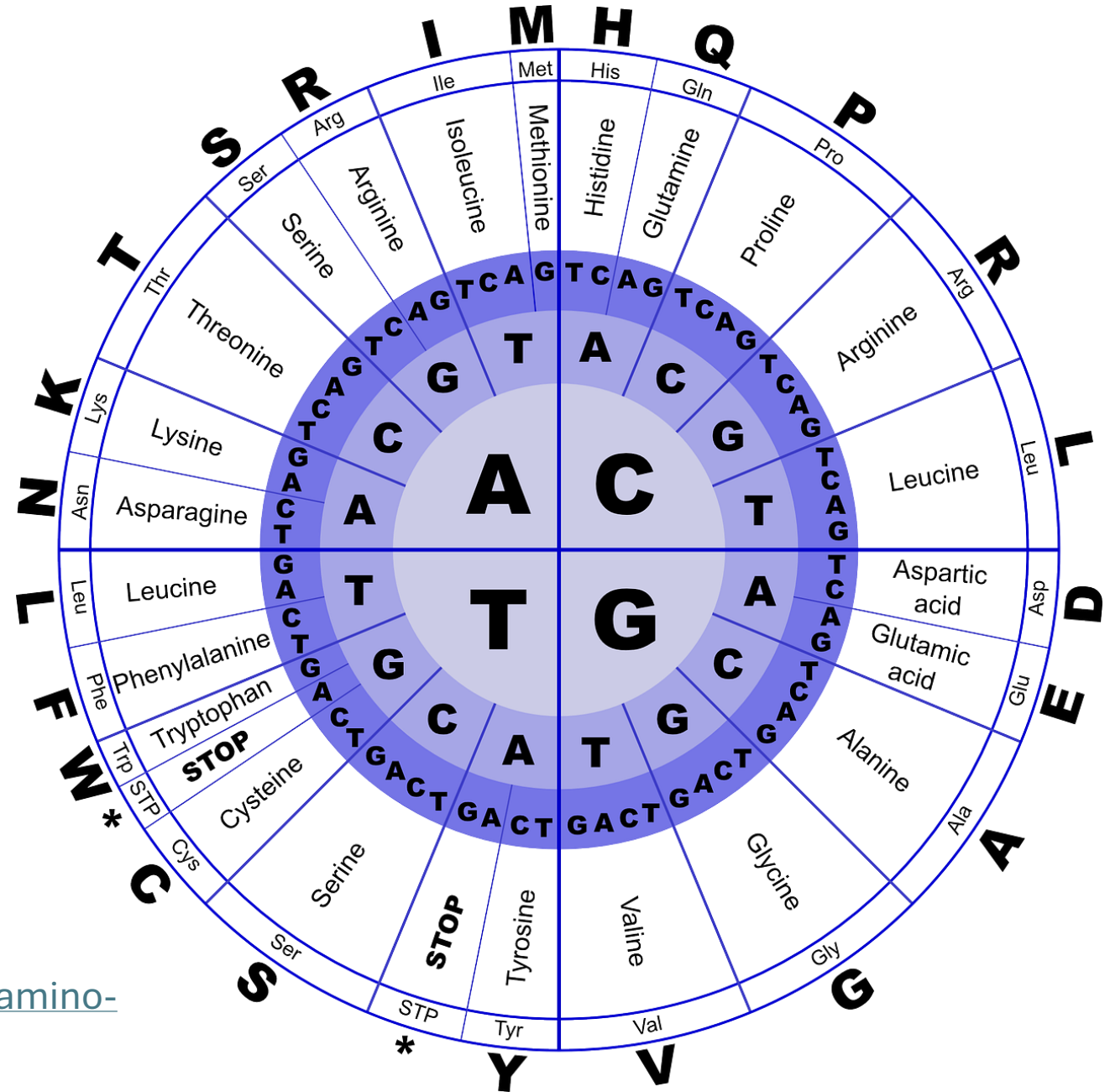
DNA



Gene - section of DNA
Genome – all genes
contained in an organism



Codon chart

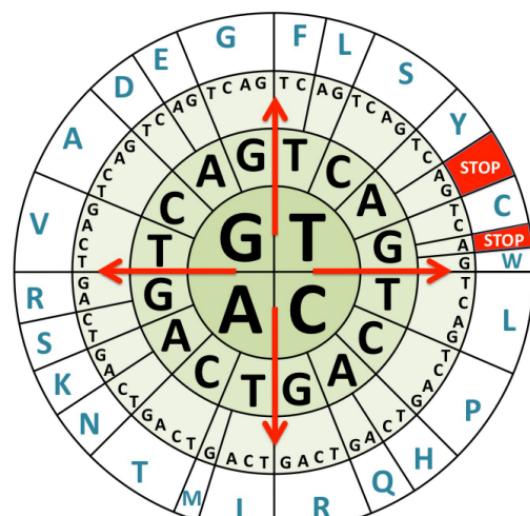


<https://pixabay.com/vectors/dna-amino-acids-biology-code-152135/>

Dr Thierry Lints



Make secret messages using life's code




Use the codon chart to convert your first or last name into a sequence of bases. If your name has the letters **B, J, O, U, X** or **Z** in it you will need to select another word instead (maybe a family member's, pet's or famous person's name, e.g. Einstein).

Type each letter of your name into a box

--	--	--	--	--	--	--	--	--

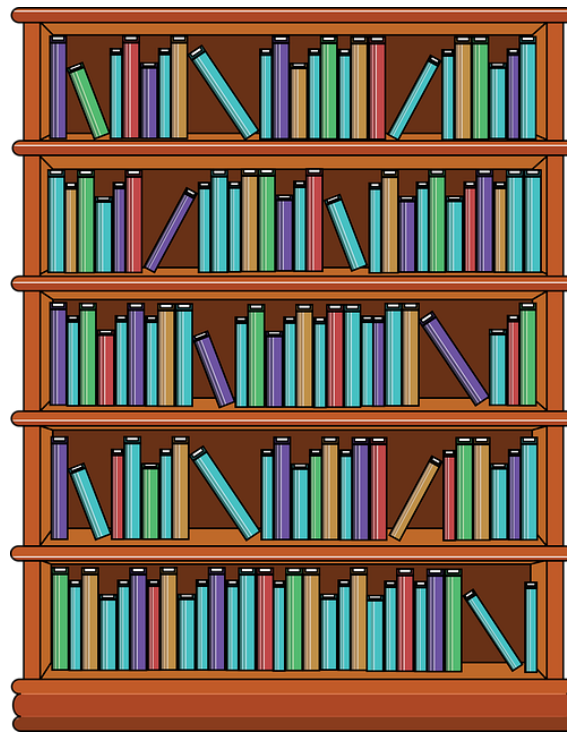
Then type a set of three bases corresponding to each letter

--	--	--	--	--	--	--	--	--



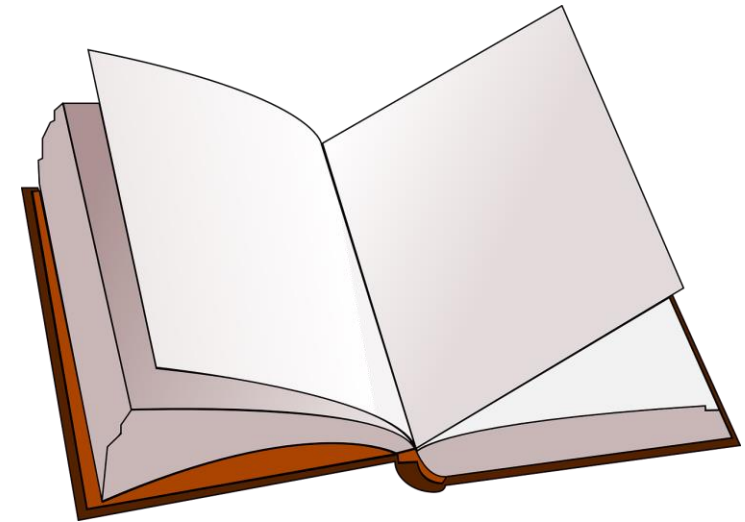
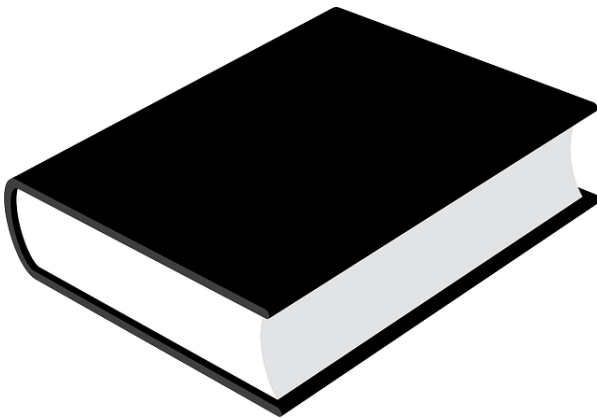
Understanding gene versus genome and where their information is stored

Genbank



One chapter = one chromosome

One book = Organism = genome



One page = one gene

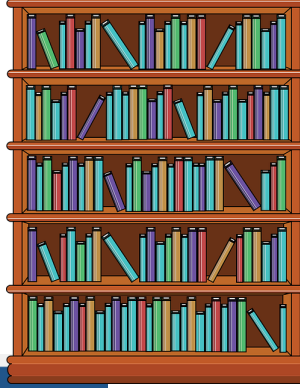
Genbank = library of books


**Book = (organism with own)
genome**

One page in book = gene

**Complete set of genes in a
cell or organism = genome**

<https://www.ncbi.nlm.nih.gov/genbank/>



 **National Library of Medicine**
National Center for Biotechnology Information

GenBank

GenBank ▾

Submit ▾

Genomes ▾

WGS ▾

Metagenomes ▾

TPA ▾

TSA ▾

INSDC ▾

Documentation ▾

Other ▾

GenBank Overview

What is GenBank?

GenBank[®] is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences ([Nucleic Acids Research, 2013 Jan;41\(D1\):D36-42](#)). GenBank is part of the [International Nucleotide Sequence Database Collaboration](#), which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis.

A GenBank release occurs every two months and is available from the [ftp site](#). The [release notes](#) for the current version of GenBank provide detailed information about the release and notifications of upcoming changes to GenBank. Release notes for [previous GenBank releases](#) are also available. GenBank growth [statistics](#) for both the traditional GenBank divisions and the WGS division are available from each release.

An [annotated sample GenBank record](#) for a *Saccharomyces cerevisiae* gene demonstrates many of the features of the GenBank flat file

GenBank Resources

- [GenBank Home](#)
- [Submission Types](#)
- [Submission Tools](#)
- [Search GenBank](#)
- [Update GenBank Records](#)

GenBank

GenBank

GenBank

What is GenBank

GenBank® is the

[Research, 2011](#)

the DNA Data Bank

BLAST®

Home Recent Results Saved Strategies Help

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

NEWS

BLAST+ 2.15.0 is here!

We have included two exciting new features in the latest BLAST+ release

Tue, 28 Nov 2023 [More BLAST news...](#)

Web BLAST

Nucleotide BLAST

nucleotide ► nucleotide

blastx

translated nucleotide ► protein

tblastn

protein ► translated nucleotide

Protein BLAST

protein ► protein

BLAST Genomes

Enter organism common name, scientific name, or tax id

Search

Standard Nucleotide BLAST

blastn

blastp

blastx

tblastn

tblastx

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)Query subrange [?](#)From To

Or, upload file

Choose File

No file chosen



Job Title

Enter a descriptive title for your BLAST search [?](#)☐Align two or more sequences [?](#)

Choose Search Set

Database

- ☒ Standard databases (nr etc.): ☐ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus ☐ Experimental data
- ☐ Core nucleotide database **NEW**

Nucleotide collection (nr/nt) [?](#)Organism
Optional☐

exclude

[Add organism](#)Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

BLAST® » blastn suite

Standard Nucleotide BLAST

blastn

blastp

blastx

tblastn

tblastx

BLASTN programs search nucleotide databases using a nucleotide query sequence

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) ? Clear

```
accgccgaga ccgcgtccgc cccgcgagca cagagcctcg ccttgccga tccgccgcc
61 gtccacaccc gccgccagct caccatggat gatgatatcg ccgcgctcgt cgtcgacaac
121 ggctccggca tgtcaaggc cggcttcgcg ggcgacgatg cccccgggc cgtcttccc
181 tccatcgtgg ggcgccccag gcaccagggc gtgatggtgg gcatgggtca
```

Query subrange ?

From

To

Or, upload file

Choose File

No file chosen

?

Job Title

Enter a descriptive title for your BLAST search ?

☐ Align two or more sequences ?

Choose Search Set

Database

☒ Standard databases (nr etc.): ☐ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus

Result of sequence alignment

Descriptions										
Graphic Summary										
Alignments										
Taxonomy										
Sequences producing significant alignments										
Download Select columns Show 100 ?										
select all 100 sequences selected										
GenBank Graphics Distance tree of results MSA Viewer										
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	
<input checked="" type="checkbox"/>	Homo sapiens actin beta (ACTB), mRNA	Homo sapiens	1330	1330	100%	0.0	100.00%	1812	NM_001101.5	
<input checked="" type="checkbox"/>	Homo sapiens mRNA for beta actin variant, clone: HRC08987	Homo sapiens	1330	1330	100%	0.0	100.00%	1656	AK225414.1	
<input checked="" type="checkbox"/>	Homo sapiens mRNA for beta actin variant, clone: KAT00430	Homo sapiens	1330	1330	100%	0.0	100.00%	1643	AK223055.1	
<input checked="" type="checkbox"/>	Homo sapiens mRNA for beta actin variant, clone: HRC07191	Homo sapiens	1330	1330	100%	0.0	100.00%	1833	AK222925.1	
<input checked="" type="checkbox"/>	Homo sapiens cDNA FLJ26647 fis, clone MPE04710, highly similar to Actin, cytoplasmic 1	Homo sapiens	1330	1330	100%	0.0	100.00%	1805	AK130157.1	

Homo sapiens actin beta (ACTB), mRNA

Sequence ID: [NM_001101.5](#) Length: 1812 Number of Matches: 1

Range 1: 1 to 720 [GenBank](#) [Graphics](#)

[▼ Next Match ▲](#) [P](#)

Score		Expect	Identities	Gaps	Strand
1330 bits(720)		0.0	720/720(100%)	0/720(0%)	Plus/Plus
Query	1	ACCGCCGAGACCGCGTCCGCCCCGCGAGCACAGAGCCTCGCCTTTGCCGATCCGCCGCC			60
Sbjct	1	ACCGCCGAGACCGCGTCCGCCCCGCGAGCACAGAGCCTCGCCTTTGCCGATCCGCCGCC			60
Query	61	GTCCACACCCGCCGCCAGCTACCATGGATGATGATATCGCCGCGCTCGTCGTCGACAAC			120
Sbjct	61	GTCCACACCCGCCGCCAGCTACCATGGATGATGATATCGCCGCGCTCGTCGTCGACAAC			120
Query	121	GGCTCCGGCATGTGCAAGGCCGGCTTCGCGGGCGACGATGCCCCCGGGCCGTCTTCCCC			180
Sbjct	121	GGCTCCGGCATGTGCAAGGCCGGCTTCGCGGGCGACGATGCCCCCGGGCCGTCTTCCCC			180
Query	181	TCCATCGTGGGGCGCCCCAGGCACCAGGGCGTGATGGTGGGCATGGGTGAGAAGGATTCC			240
Sbjct	181	TCCATCGTGGGGCGCCCCAGGCACCAGGGCGTGATGGTGGGCATGGGTGAGAAGGATTCC			240
Query	241	TATGTGGGCGACGAGGCCAGAGCAAGAGAGGCATCCTCACCTGAAGTACCCATCGAG			300
Sbjct	241	TATGTGGGCGACGAGGCCAGAGCAAGAGAGGCATCCTCACCTGAAGTACCCATCGAG			300
Query	301	CACGGCATCGTCACCAACTGGGACGACATGGAGAAAATCTGGCACCACACCTTCTACAAT			360
Sbjct	301	CACGGCATCGTCACCAACTGGGACGACATGGAGAAAATCTGGCACCACACCTTCTACAAT			360
Query	361	GAGCTGCGTGTGGCTCCCGAGGAGACCCCGTGCTGCTGACCGAGGCCCCCTGAACCCC			420
Sbjct	361	GAGCTGCGTGTGGCTCCCGAGGAGACCCCGTGCTGCTGACCGAGGCCCCCTGAACCCC			420
Query	421	AAGGCCAACC GCGAGAAGATGACCCAGATCATGTTTGAGACCTTCAACACCCAGCCATG			480
Sbjct	421	AAGGCCAACC GCGAGAAGATGACCCAGATCATGTTTGAGACCTTCAACACCCAGCCATG			480
Query	481	TACGTTGCTATCCAGGCTGTGCTATCCCTGTACGCCTCTGGCCGTACCACTGGCATCGTG			540

Bank of variants' data

- MYBPC3 R

NIH National Library of Medicine
National Center for Biotechnology Information

GenBank ClinVar Search

GenBank Submit Genomes WGS Metagenomes TPA TSA INSDC Documentation Other

GenBank Overview

What is GenBank?

GenBank® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences ([Nucleic Acids](#))

GenBank Resources

- [GenBank Home](#)
- [Submission Types](#)
- [Submission Tools](#)