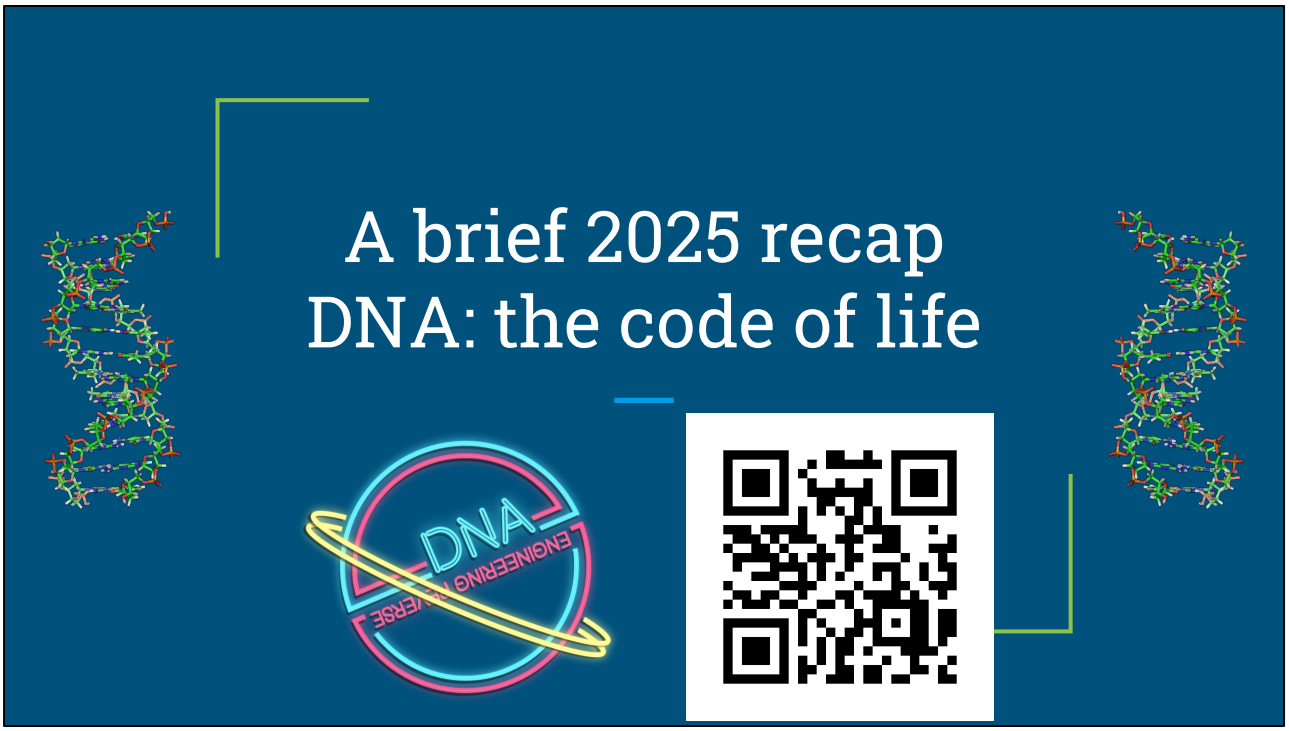


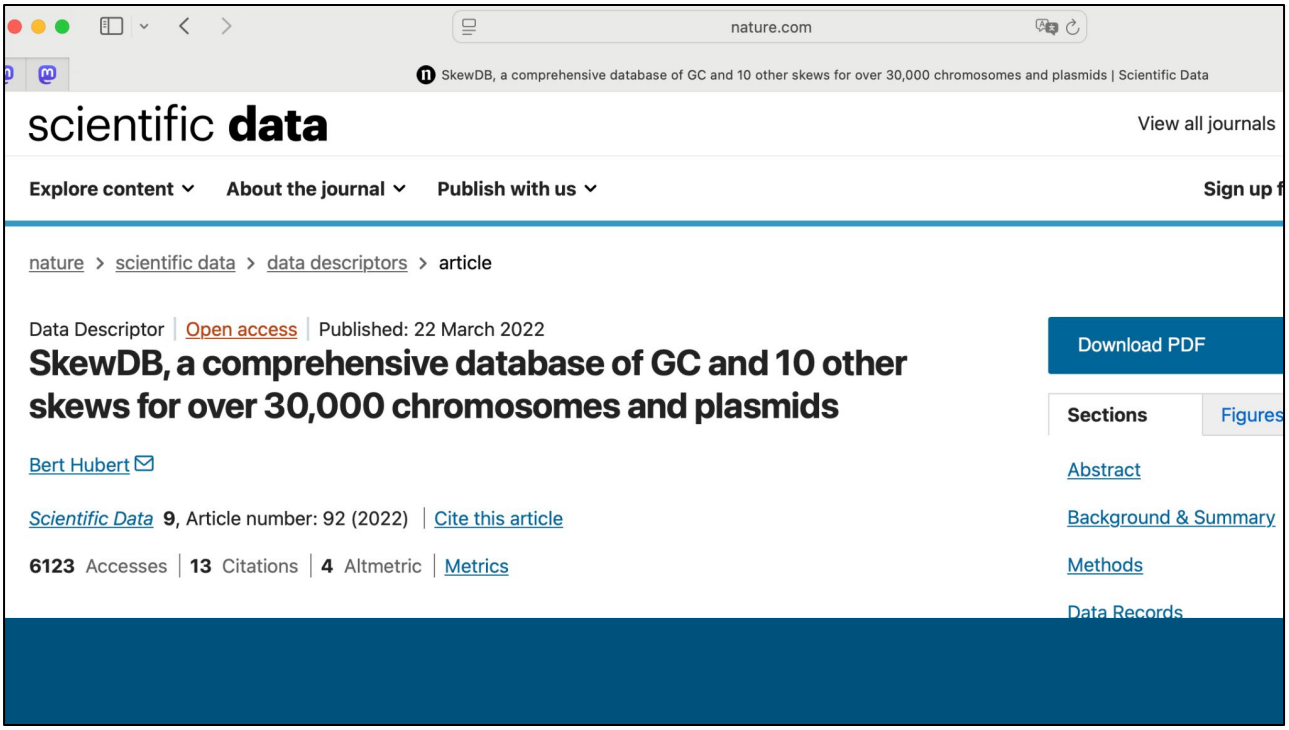
https://berthub.eu/dna https://berthub.eu/articles/posts/mch-dna-and-gps-gnss-talks/

Credits

- This presentation leans heavily on other people's work and graphics
- All credits are available in the **speaker notes** which you should consult to find out who made all these great movies and images
- **Thank you so much Wikipedia Commons in particular!**

https://berthub.eu/whydna

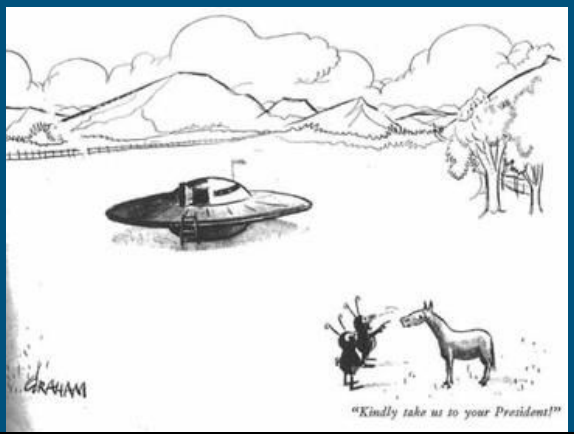
https://www.nature.com/articles/ismej2015107



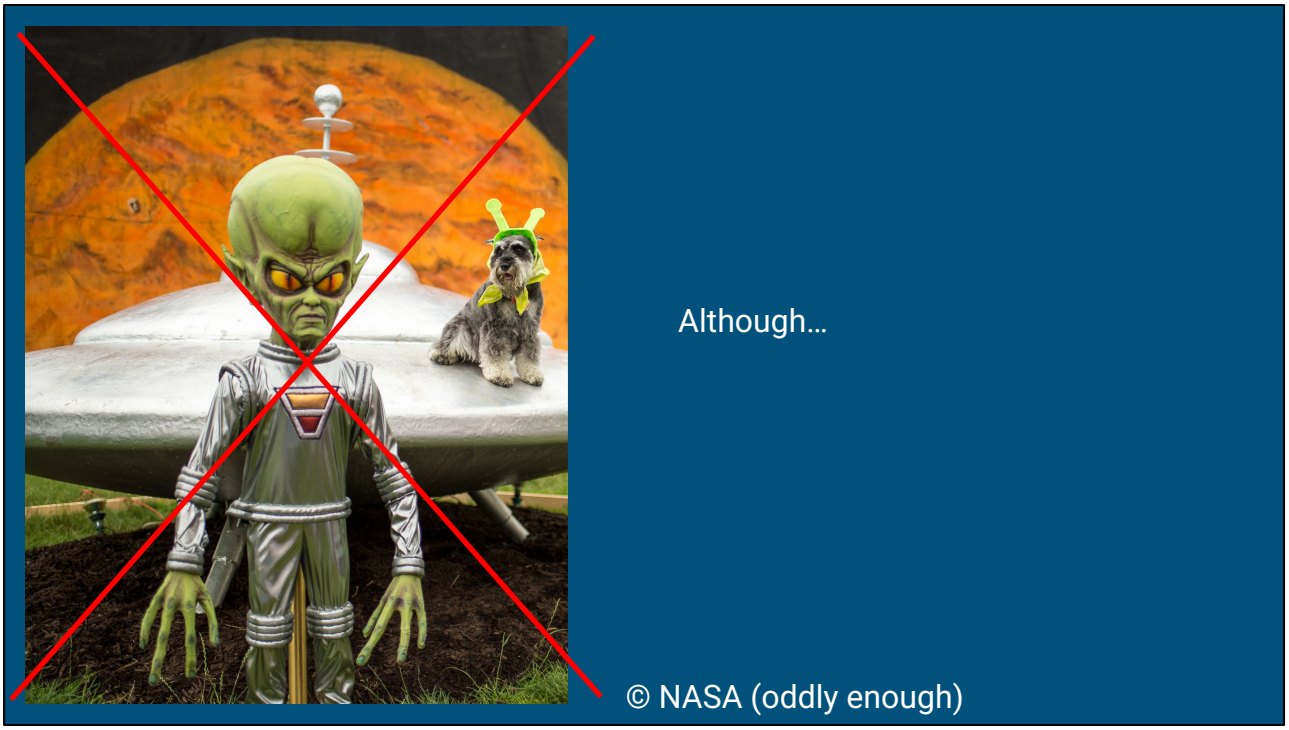
https://skewdb.org/

“Imagine a flashy spaceship lands in your backyard. The door opens and you are invited to investigate everything to see what you can learn. The technology is clearly millions of years beyond what we can make.

This is biology.”

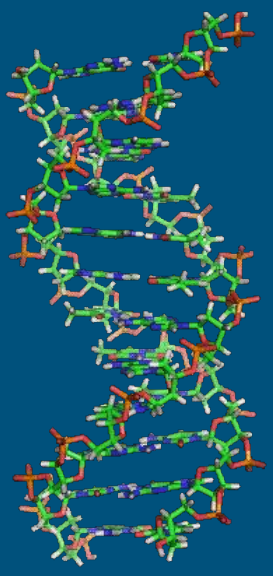


https://jsomers.net/i-should-have-loved-biology/

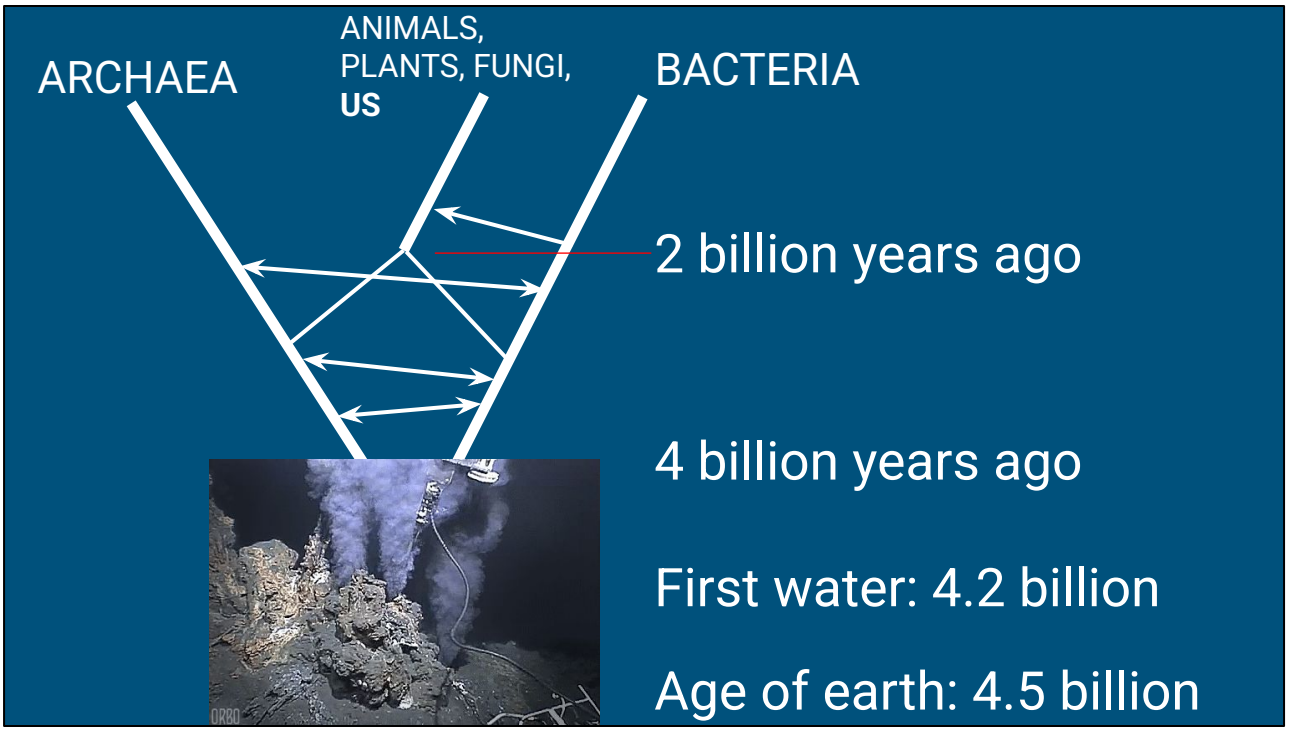


https://en.wikipedia.org/wiki/Little_green_men#/media/File:Mars_New_Year's_Celebration_(201506200007HQ)_(cropped).jpg

1. Information storage
2. The 3D printer of life
3. Algorithms & control
4. Hacks



- DNA: Millions, billions of nucleotides or “bases”:
  - A, C, G, T
- Organized in chromosomes & genes
- Absolutely **atom for atom** universal across all life
  - >4 billion years old
- Stable for 100s of thousands of years

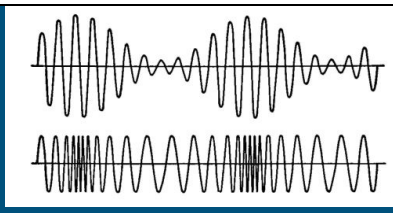


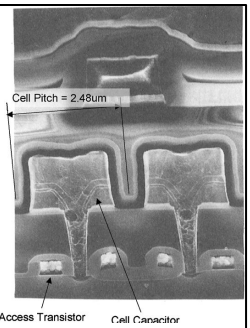
https://giphy.com/gifs/sea-vents-hydrothermal-1bTEQnjArFBy8
https://en.wikipedia.org/wiki/Archaea
https://en.wikipedia.org/wiki/Three-domain_system

Basics: Each nucleotide is **2 (arbitrary) bits**

A	**00**
C	**01**
G	**10**
T	**11**

Basics: DNA



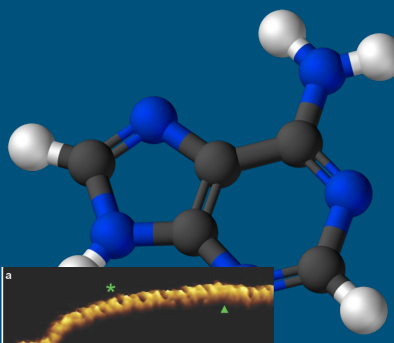


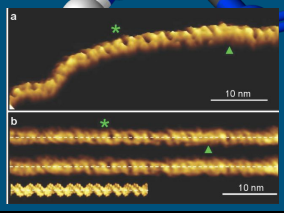
A 00

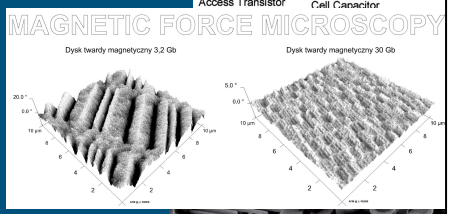
C 01

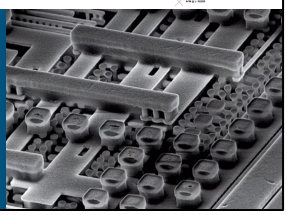
G 10

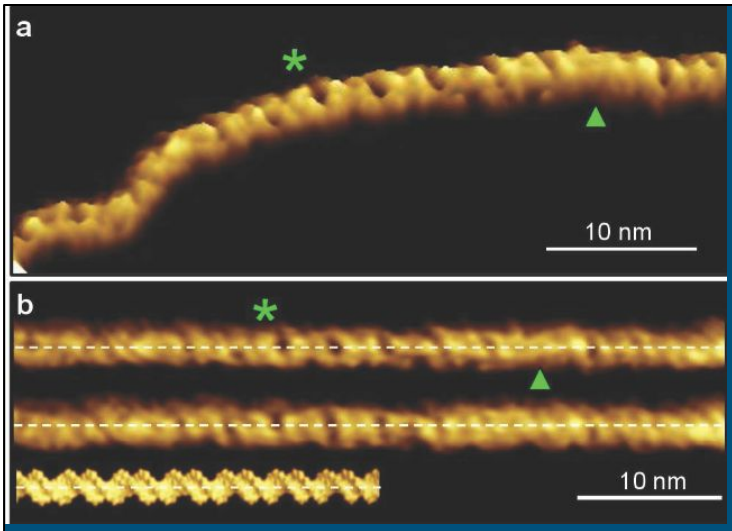
T 11











DNA is very much like tape

Sometimes circular tape - no beginning, no end!

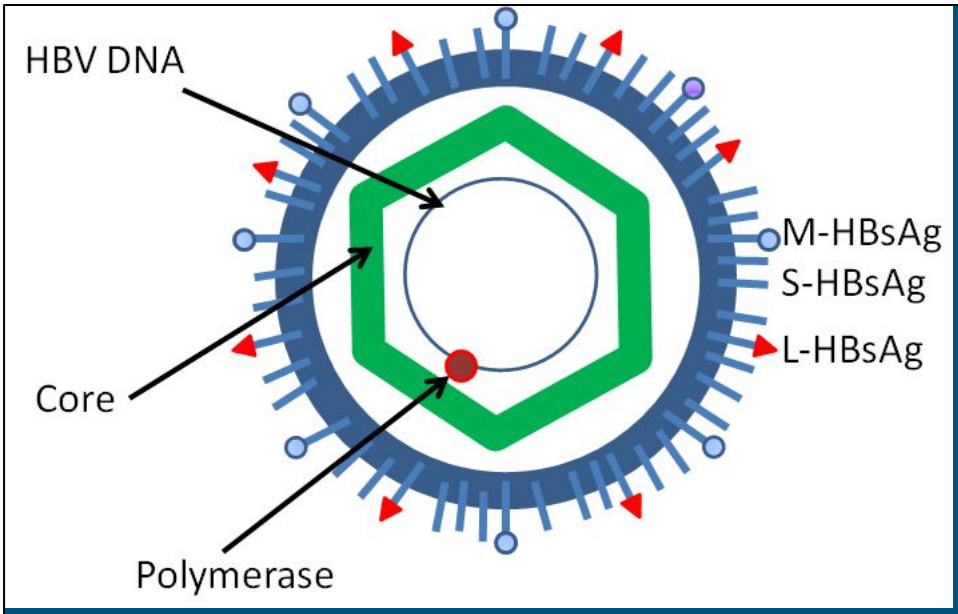
No addressing! No alignment!

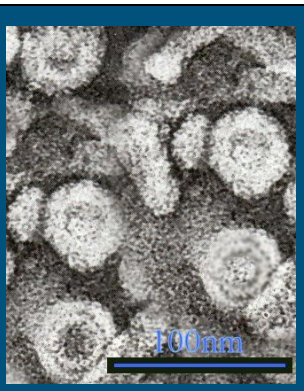
It is a **nucleotide stream** which can be compared to a **bitstream**

It IS however **content addressable**!

And it has FRAMING ~~challenges~~ **possibilities**

http://onlinelibrary.wiley.com/doi/10.1002/smll.201400265/full

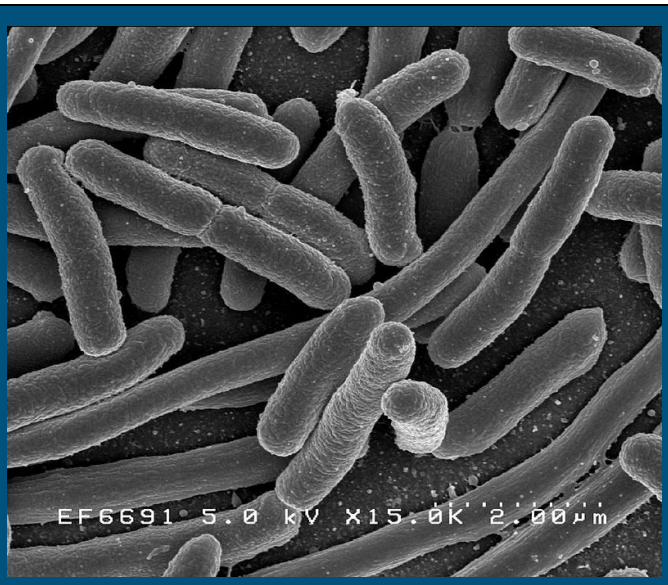


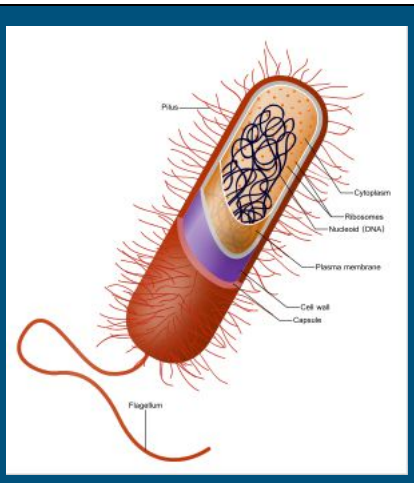


Hepatitis B - 800 bytes (unpleasant)

By Dr Graham Beards - Own work, CC BY-SA 3.0,
https://commons.wikimedia.org/w/index.php?curid=24121844
By GrahamColm at English Wikipedia, CC BY 3.0,
https://commons.wikimedia.org/w/index.php?curid=6032684
https://en.wikipedia.org/wiki/Hepatitis_B_virus

All of Hepatitis-B - 800 bytes

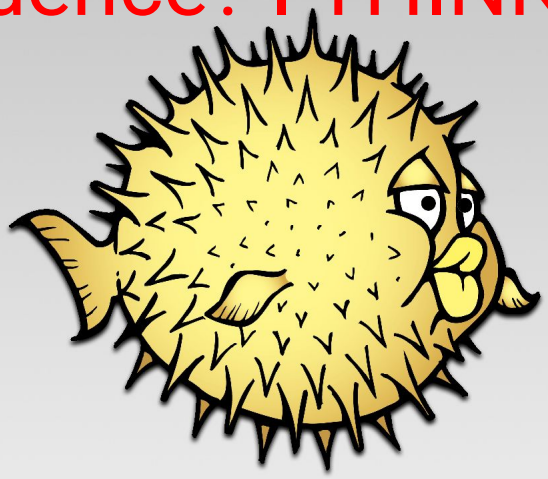




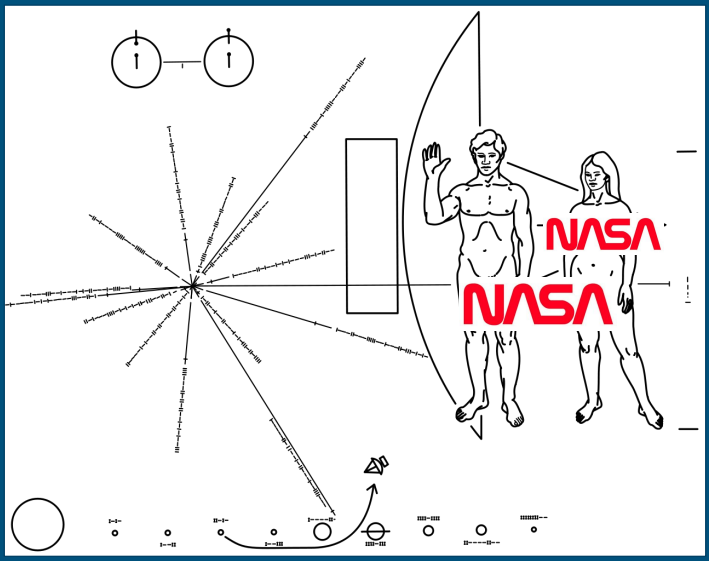
E. Coli: 750KB (can be good, can be unpleasant)

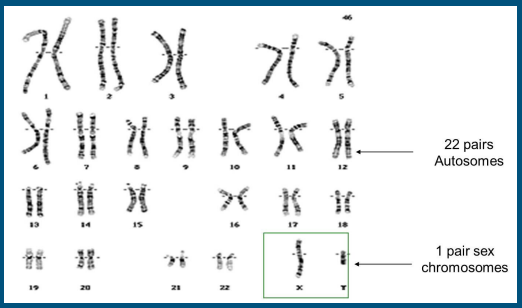
By Credit: Rocky Mountain Laboratories, NIAID, NIH - NIAID: These high-resolution (300 dpi) images may be downloaded directly from this site. All the images, except specified ones from the World Health Organization (WHO), are in the public domain. For the public domain images, there is no copyright, no permission required, and no charge for their use., Public Domain, https://commons.wikimedia.org/w/index.php?curid=104228

Coincidence? I THINK NOT!



Pufferfish: 100MB, smallest & tightest genome of all animals (lethal)





Your unique DNA *diff* is around 1-10 megabyte. HOWEVER - could be way more: the pangenome

Homo sapiens: **750MB** (lethal)

Source: NASA

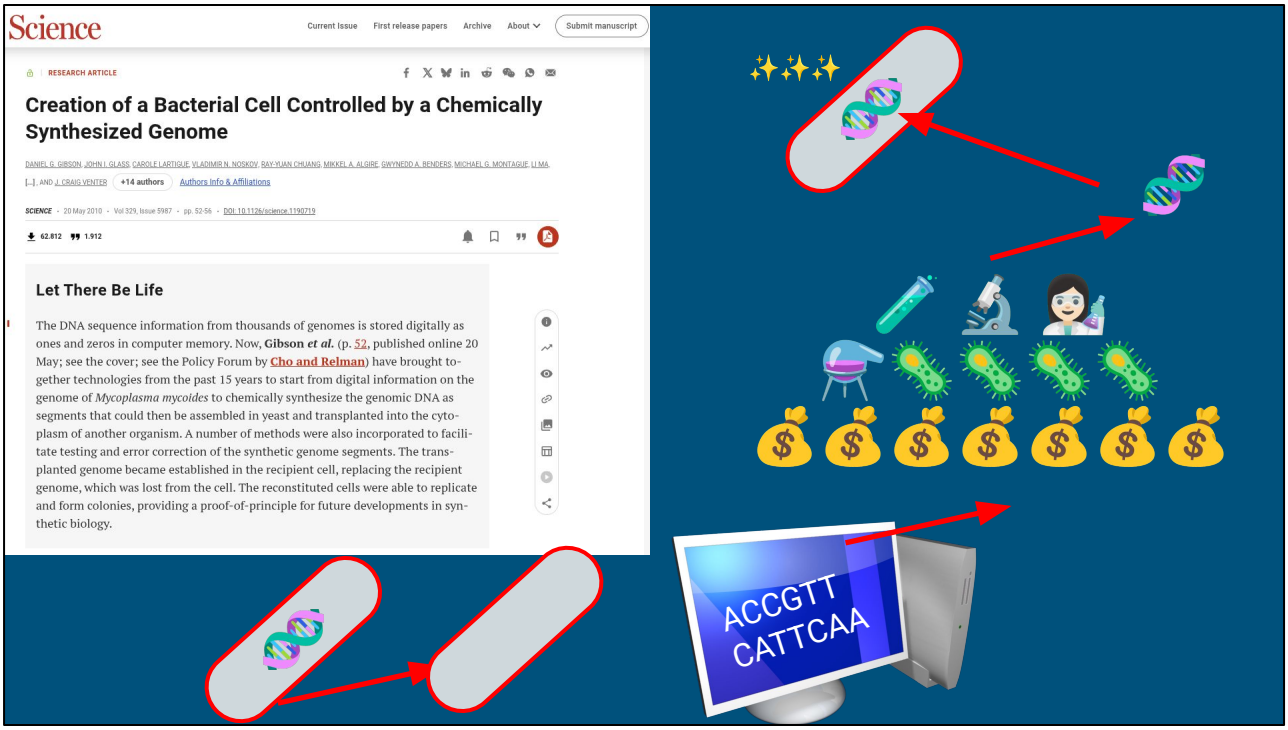
We find that a typical genome differs from the reference human genome at 4.1 million to 5.0 million sites. Although >99.9% of variants consist of SNPs and short indels, structural variants affect more bases: the typical genome contains an estimated 2,100 to 2,500 structural variants (~1,000 large deletions, ~160 copy-number variants, ~915 Alu insertions, ~128 L1 insertions, ~51 SVA insertions, ~4 NUMTs, and ~10 inversions), affecting ~20 million bases of sequence.

Paris Japonica: 37.5GB (quite pleasant)

Marbled Lungfish: 33GB

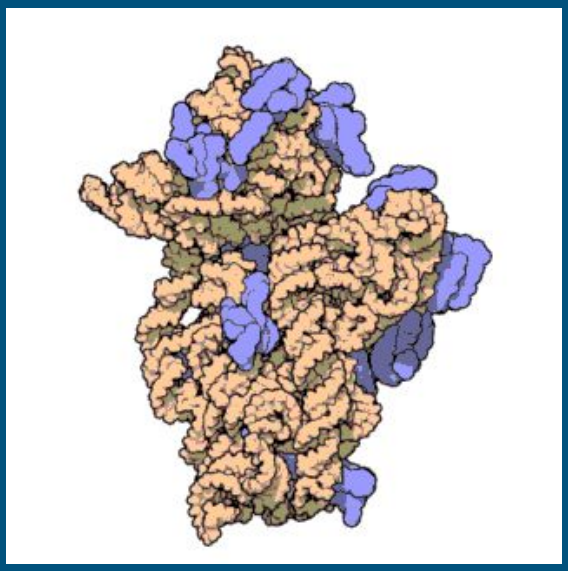
By OpenCage = User:OpenCage - http://opencage.info/pics.e/large_11454.asp, CC BY 2.5, https://commons.wikimedia.org/w/index.php?curid=11750275

But is life REALLY digital?



https://www.science.org/doi/10.1126/science.1190719

Behold! The source of all life & you



~10
nanometers

BEHOLD! Part of the Ribosome, the 1D/3D printer of life!
It printed you or the things that made you!

By Animation by David S. Goodsell, RCSB Protein Data Bank - Molecule of the Month at the RCSB Protein Data Bank, Public Domain, https://commons.wikimedia.org/w/index.php?curid=2839678

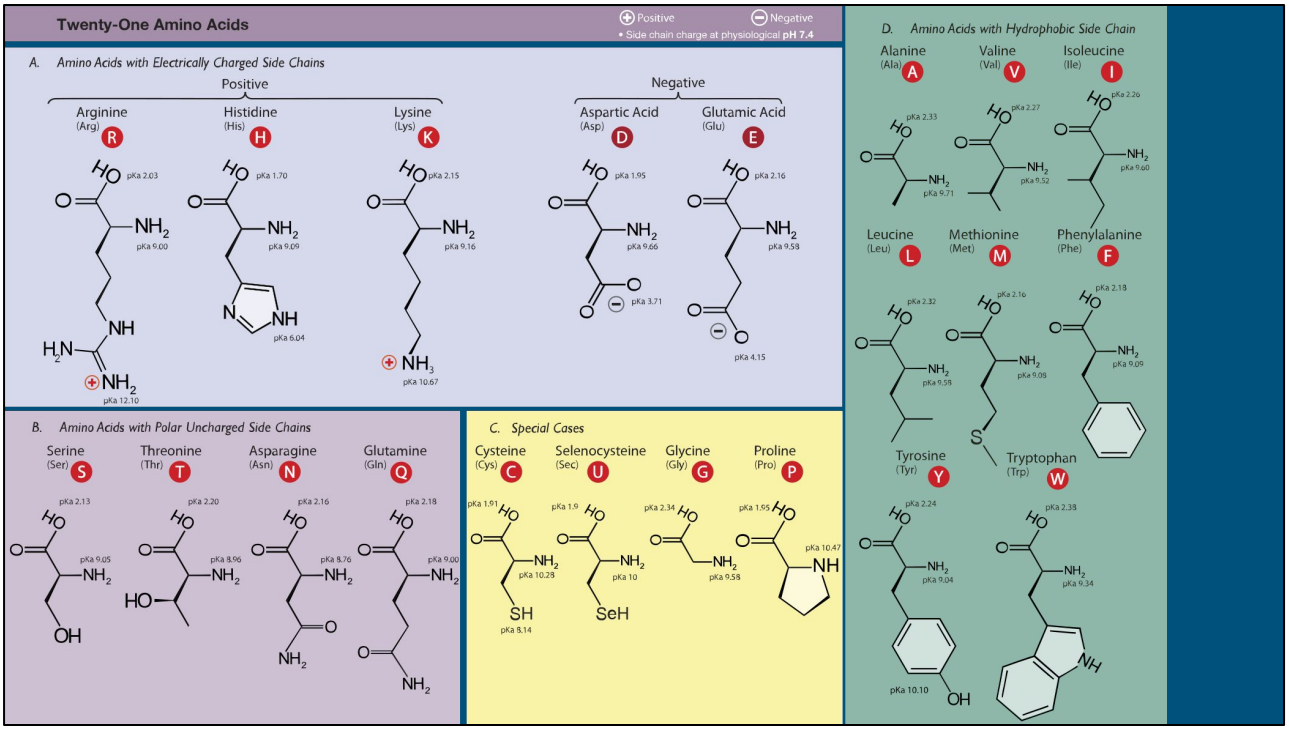
The Central Dogma

Long term storage: **DNA (*/dev/sda*)**

Converts to live form: **RNA (*RAM*)**

RNA converts to: **Proteins**

Proteins: **Sense and Do**



Modified from https://commons.wikimedia.org/wiki/File:Amino_Acids.svg

1st base	2nd base								3rd base
	T		C		A		G		
T	TTT	(Phe/F) Phenylalanine (np)	TCT	(Ser/S) Serine (p)	TAT	(Tyr/Y) Tyrosine (p)	TGT	(Cys/C) Cysteine (p)	T
	TTC		TCC		TAC		TGC		C
	TTA	(Leu/L) Leucine (np)	TCA		TAA	Stop (*Ochre*) *[note 2]	TGA	Stop (*Opal*) *[note 2]	A
	TTG ⇒		TCG		TAG	Stop (*Amber*) *[note 2]	TGG	(Trp/W) Tryptophan (np)	G
C	CTT		CCT	(Pro/P) Proline (np)	CAT	(His/H) Histidine (b)	CGT	(Arg/R) Arginine (b)	T
	CTC		CCC		CAC		CGC		C
	CTA		CCA		CAA	(Gln/Q) Glutamine (p)	CGA		A
	CTG		CCG		CAG		CGG		G
A	ATT	(Ile/I) Isoleucine (np)	ACT	(Thr/T) Threonine (p)	AAT	(Asn/N) Asparagine (p)	AGT	(Ser/S) Serine (p)	T
	ATC		ACC		AAC		AGC		C
	ATA		ACA		AAA	(Lys/K) Lysine (b)	AGA	(Arg/R) Arginine (b)	A
	ATG ⇒	(Met/M) Methionine (np)	ACG		AAG		AGG		G
G	GTT	(Val/V) Valine (np)	GCT	(Ala/A) Alanine (np)	GAT	(Asp/D) Aspartic acid (a)	GGT	(Gly/G) Glycine (np)	T
	GTC		GCC		GAC		GGC		C
	GTA		GCA		GAA	(Glu/E) Glutamic acid (a)	GGA		A
	GTG ⇒		GCG		GAG		GGG		G

Multi-billion year old table!

Multiple codons for same amino acids

This allows for **dialects** and shaping DNA

>gnl|ECOLI|G0-10439 kdpF MONOMER0-12 (complement(728732..728821)) Escherichia coli K-12 substr. MG1655

gtgAGTGCAGGCGTGATAACCGGCGTATTGCTGGTGTTTTTATTACTGGGTTATCTGGTTTATGCCCTGA
TCAATGCGGAGGCGTTCtga

>gnl|ECOLI|G0-10439 kdpF MONOMER0-12 (complement(728732..728821)) Escherichia coli K-12 substr. MG1655

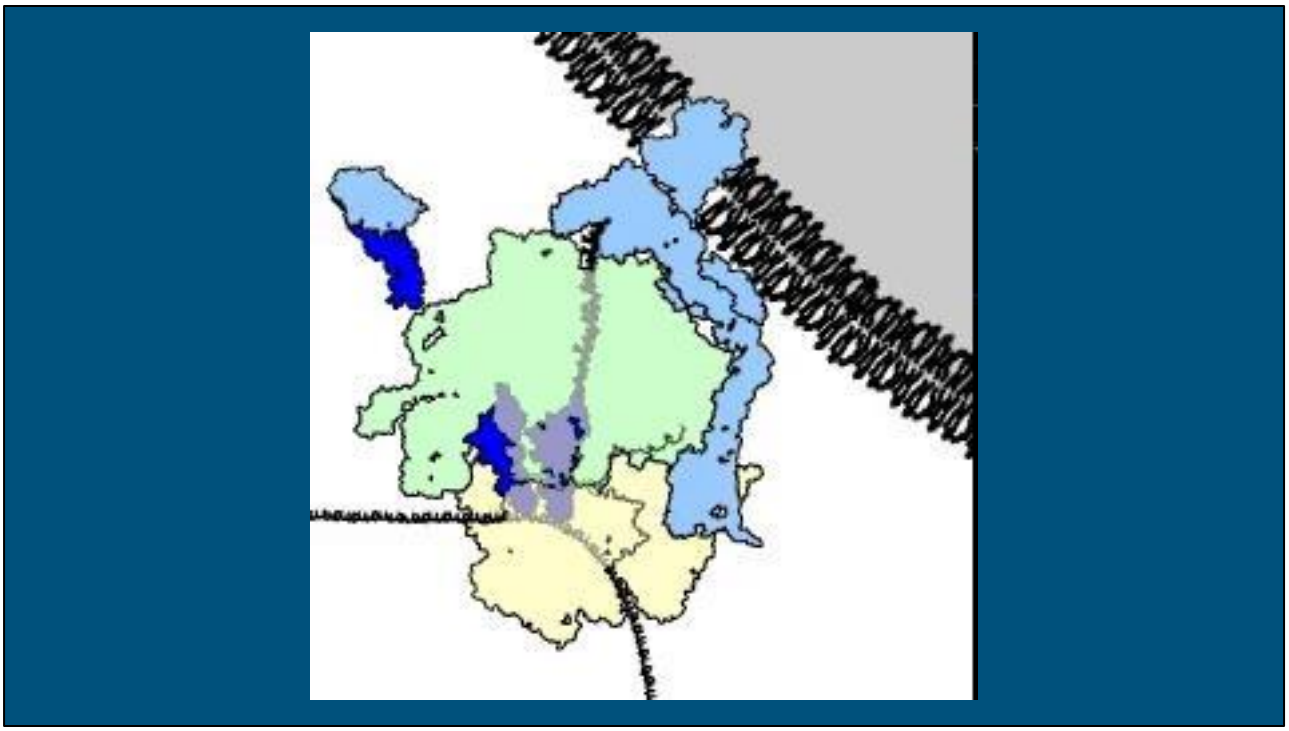
gtg AGT GCA GGC GTG ATA ACC GGC GTA TTG CTG GTG TTT TTA TTA CTG GGT TAT CTG GTT TAT GCC CTG ATC AAT GCG GAG GCG TTC tga

>gnl|ECOLI|MONOMER0-12 gn=kdpF K+ transporting P-type ATPase subunit KdpF (complement(728732..728821)) Escherichia coli K-12 substr. MG1655

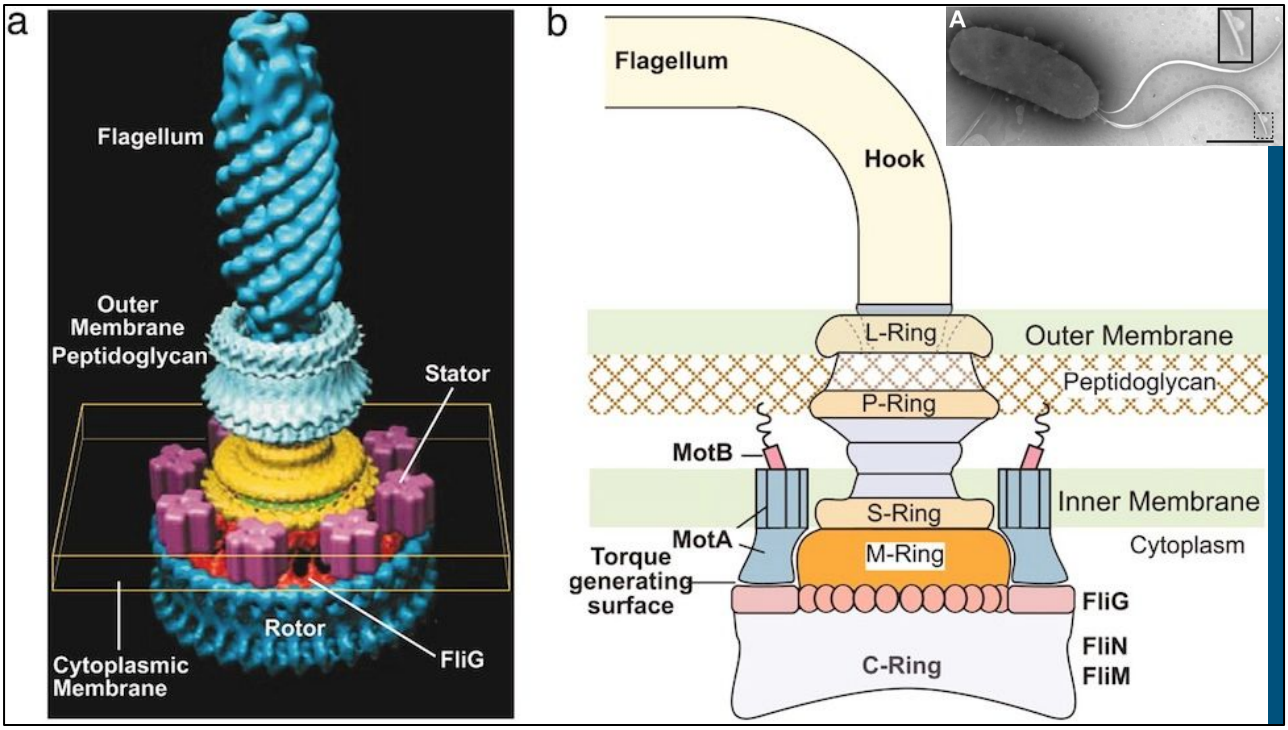
MSAGVITGVLLVFLLLGYLVYALINAEAF<EOP>

"A 1D printer that leads to 3D objects with ATOMIC resolution"

https://www.rcsb.org/structure/6HRB

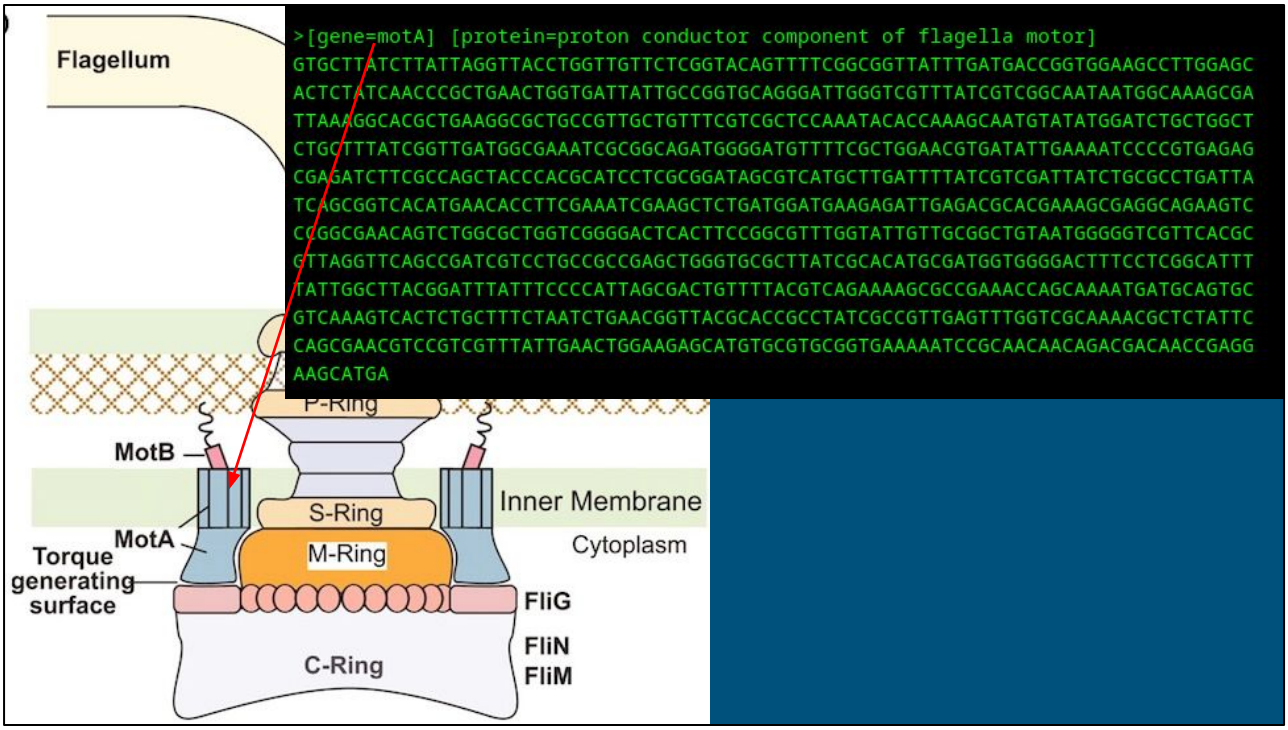


https://commons.wikimedia.org/w/index.php?curid=8287100
https://upload.wikimedia.org/wikipedia/commons/9/94/Protein_translation.gif



Credit: David DeRosie
https://www.thunderbolts.info/wp/2013/07/19/flagellar-motors/
https://elifesciences.org/articles/01579 Caitlin A Brennan, Jason R Hunt, Natacha Kremer, Benjamin C Krasity, Michael A Apicella, Margaret J McFall-Ngai, Edward G Ruby (2014) A model symbiosis reveals a role for sheathed-flagellum rotation in the release of immunogenic lipopolysaccharide eLife 3:e01579

https://doi.org/

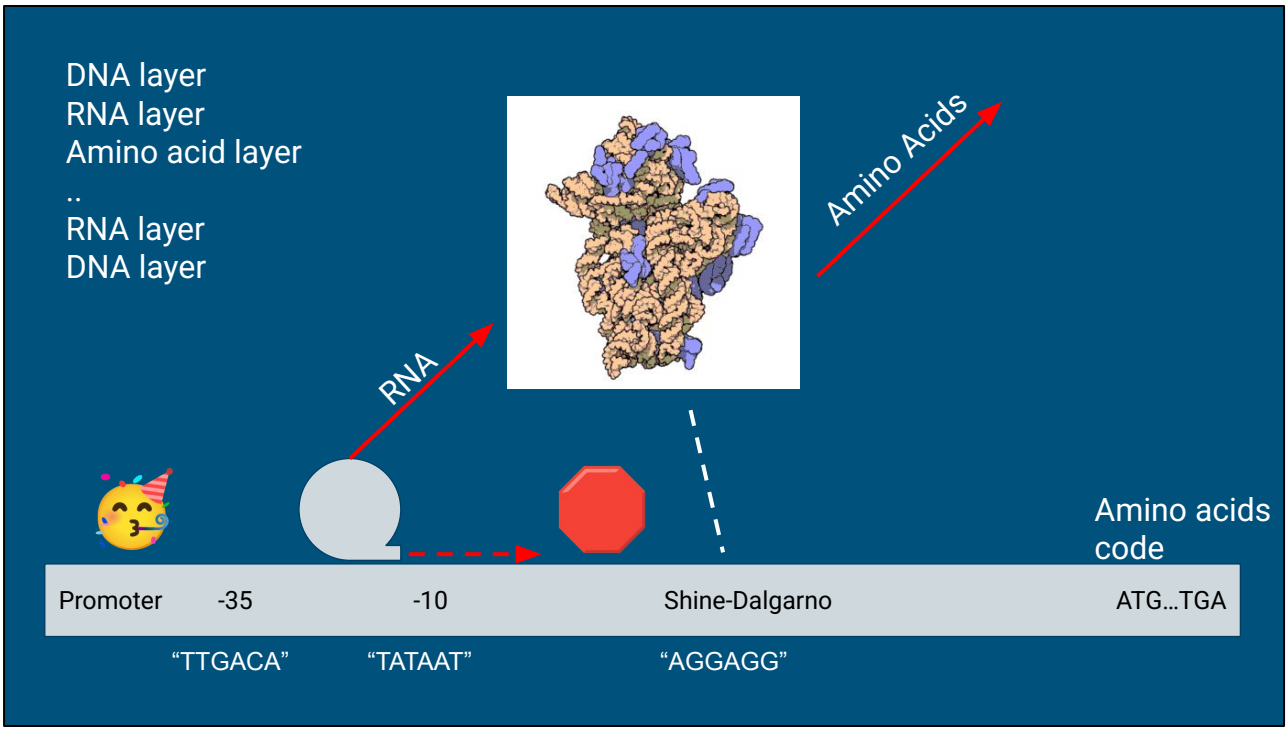


Credit: David DeRosie
https://www.thunderbolts.info/wp/2013/07/19/flagellar-motors/

If you 3D print this model yourself:

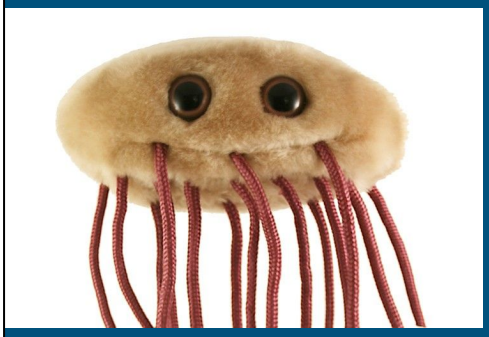
- Your resolution is like 10 million times too coarse
- Your flagellar motor does not actually work
- It looks remarkably silly! (‘honestly, this is for science!’)

Gene control & how it hangs together



https://www.biorxiv.org/content/10.1101/2025.01.23.634641v2.full

E. coli



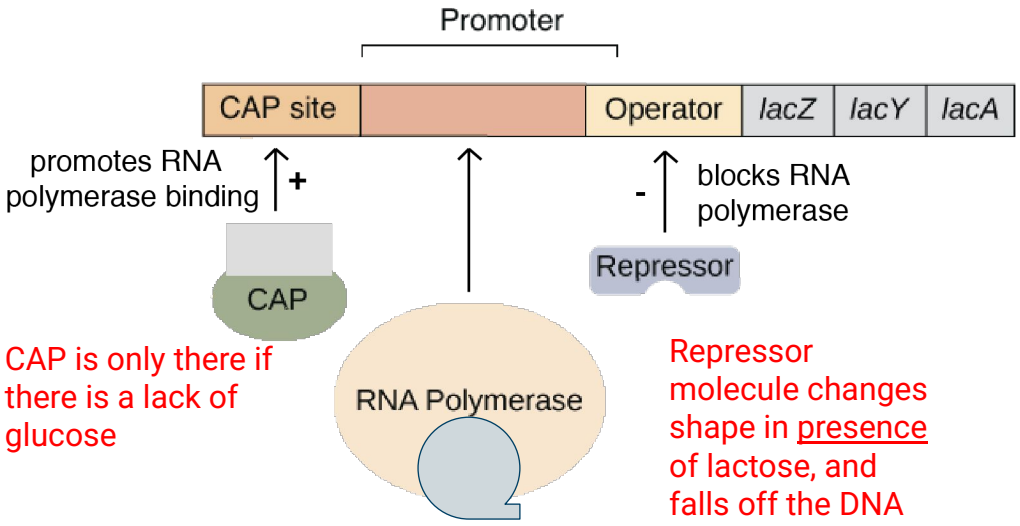
- Needs power to run
  - 0.5pW
- Loves glucose
- Can run on **lactose** if it must
  - Through conversion
  - Less efficient
- Algorithm required:

```
if(!glucosePresent && lactosePresent)

  convertLactoseToGlucose();
```

https://www.khanacademy.org/science/biology/gene-regulation/gene-regulation-in-bacteria/a/the-lac-operon

The *lac* operon:



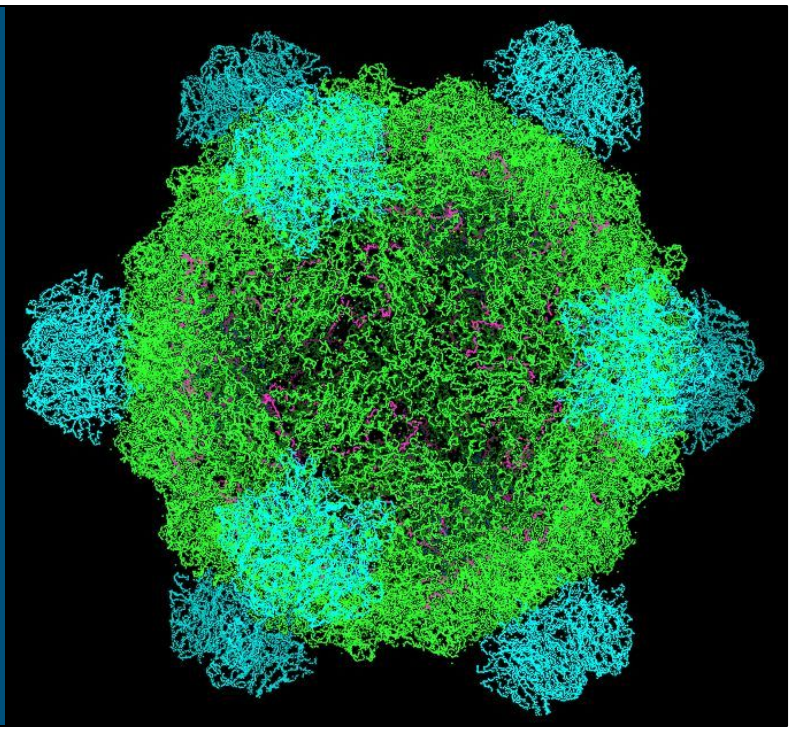
CAP is only there if there is a lack of glucose

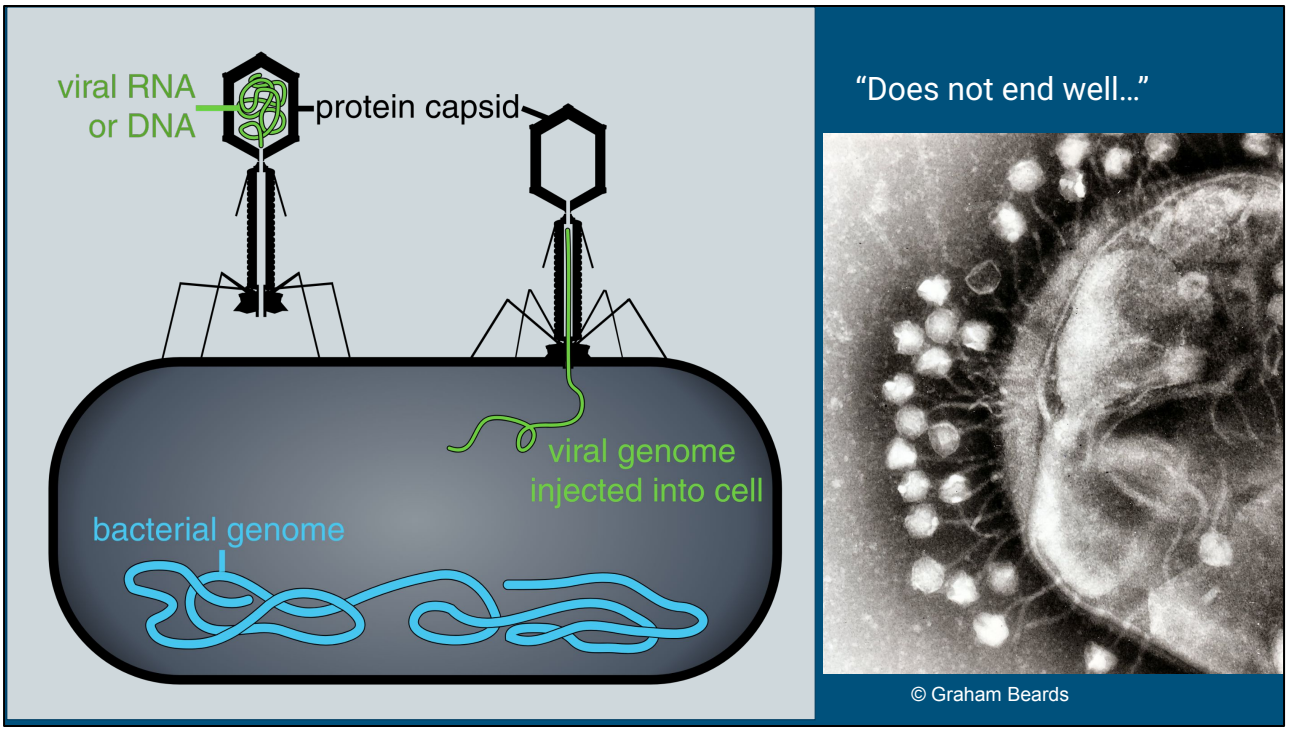
Repressor molecule changes shape in presence of lactose, and falls off the DNA

https://biocyc.org/ECOLI/NEW-IMAGE?type=LOCUS-POSITION&object=G6201&chromosome=COLI-K12
https://en.wikipedia.org/wiki/Lac_operon

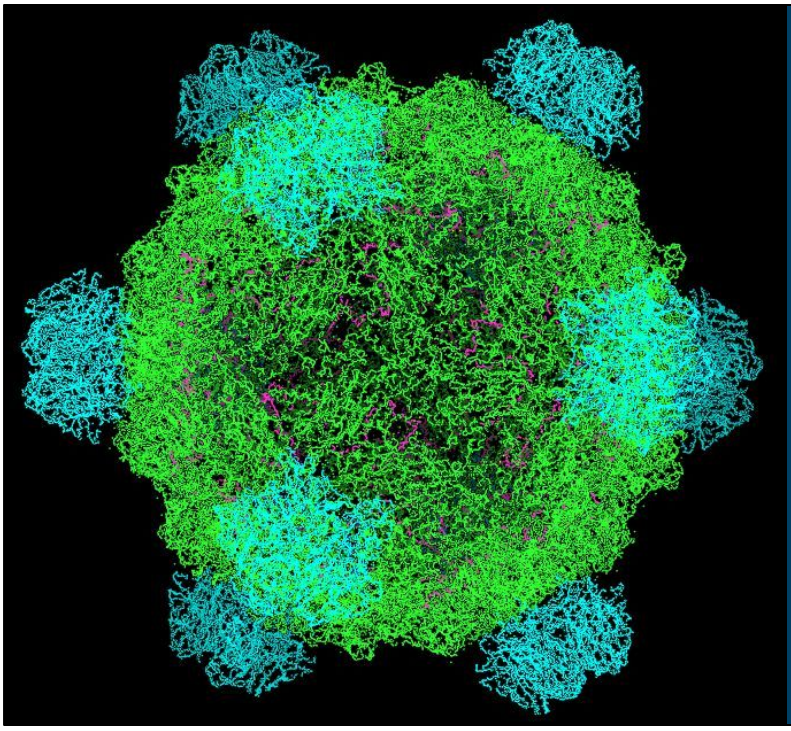
Viruses: the ultimate hackers

“Shell code”





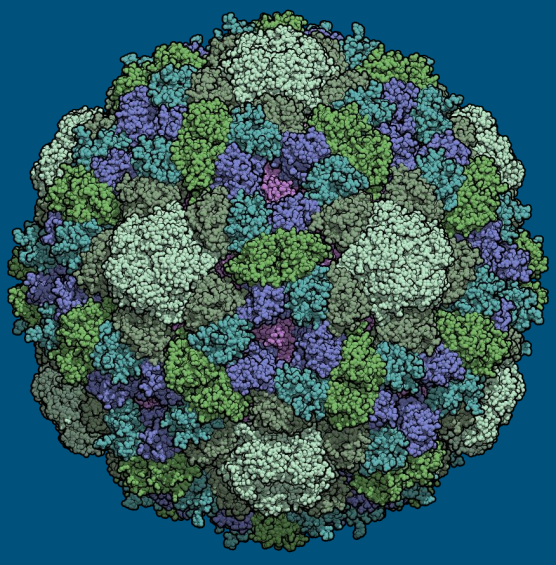
By Professor Graham Beards - en:Image:Phage.jpg, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=5035798
https://commons.wikimedia.org/wiki/File:Phage_injecting_its_genome_into_bacteria.svg



ΦX174 infects e coli.

Having **less DNA** means faster copying, smaller virion, **evading defences**

“Shell code shaping”



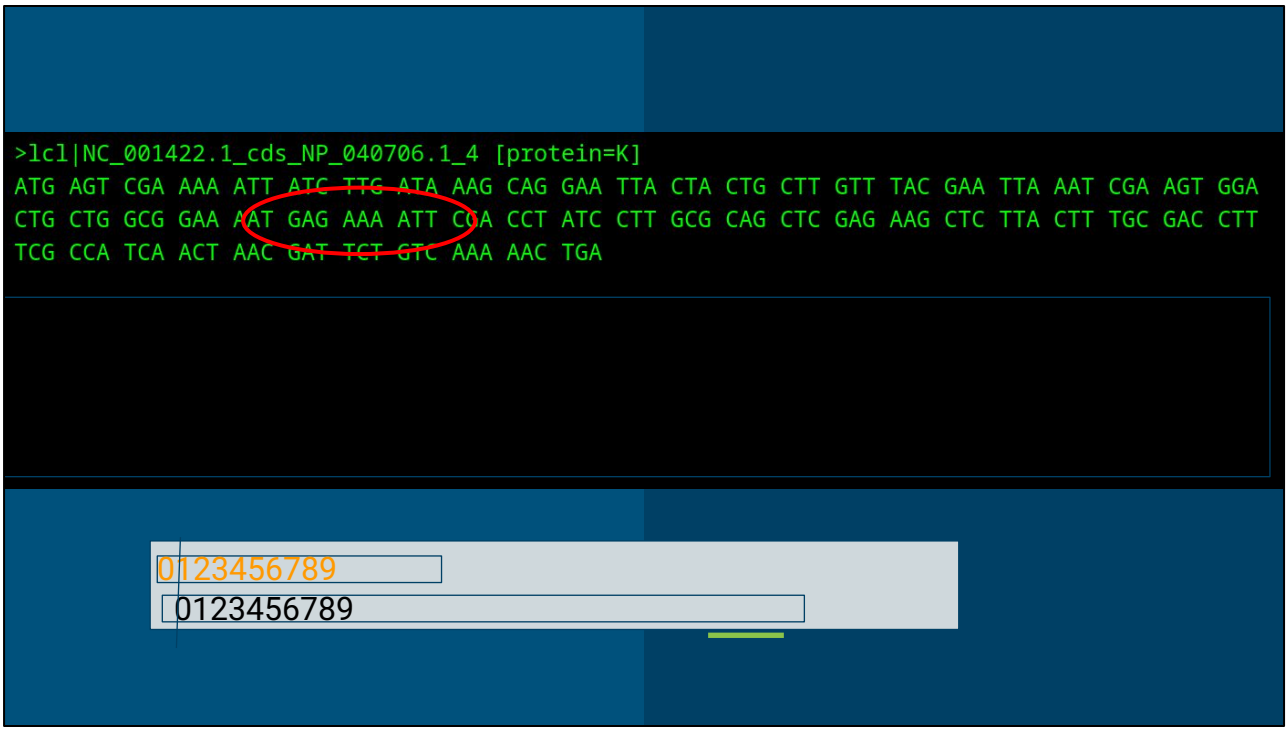
https://commons.wikimedia.org/wiki/File:Phi_X_174_procapsid.png

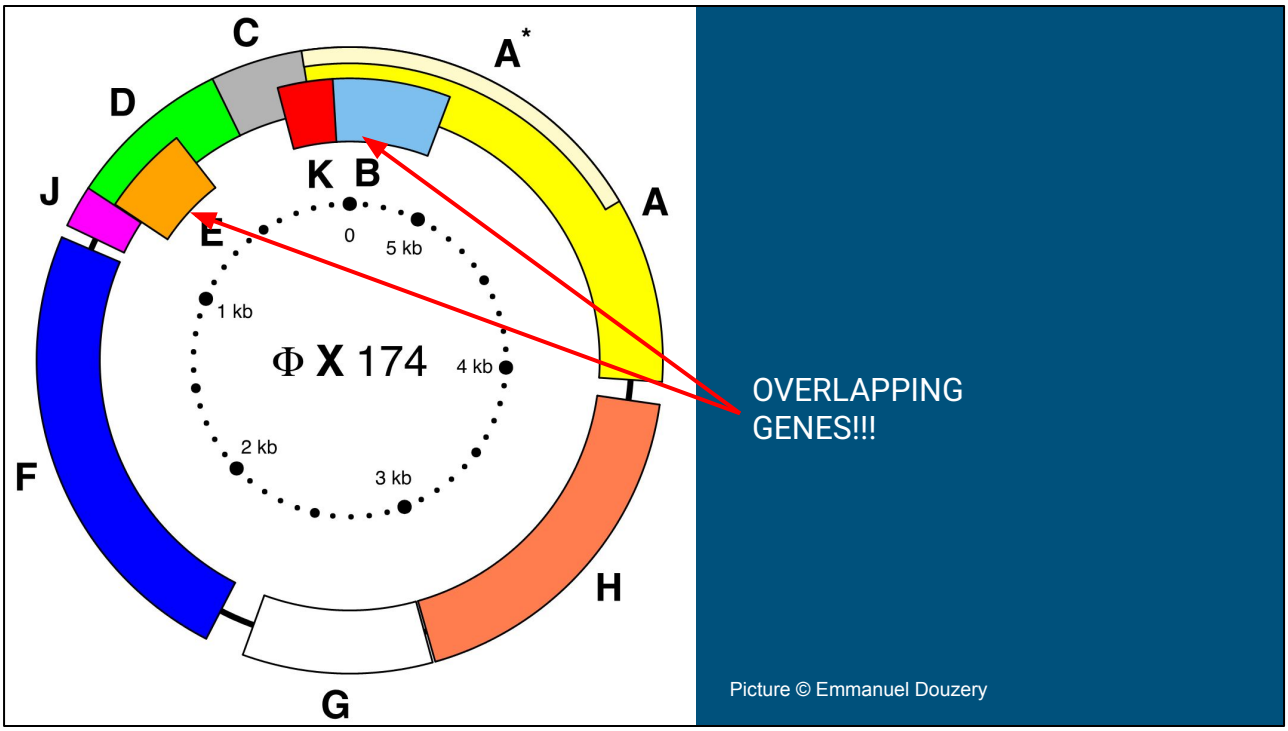
ΦX174 has 5386 DNA letters

Enough for 5386/3 = **1795** amino acids, TOPS

… it has 2334 amino acids!

?????


>lcl|NC_001422.1_cds_NP_040706.1_4 [protein=K]
ATG AGT CGA AAA ATT ATC TTG ATA AAG CAG GAA TTA CTA CTG CTT GTT TAC GAA TTA AAT CGA AGT GGA
CTG CTG GCG GAA AAT GAG AAA ATT CGA CCT ATC CTT GCG CAG CTC GAG AAG CTC TTA CTT TGC GAC CTT
TCG CCA TCA ACT AAC GAT TCT GTC AAA AAC TGA
0123456789
0123456789



https://en.wikipedia.org/wiki/Overlapping_gene
https://en.wikipedia.org/wiki/Phi_X_174#/media/File:Genome_map_of_the_bacteriophage_%CE%A6X174_showing_overlapping_genes.svg
By This picture is a work by Emmanuel Douzery.Please credit this with: Picture : Emmanuel Douzery in the immediate vicinity of the picture.If you use one of my works, please email me (account needed) or leave me a short message on my discussion page. It would be greatly appreciated!Do not copy this picture illegally by ignoring the terms of the license below, as it is not in the public domain.If you would like special permission to use, license, or purchase the picture, please contact me to negotiate terms. - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=46682223

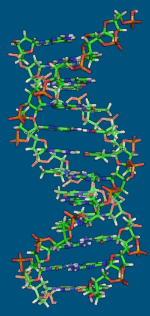
Even more DNA!!

Tomorrow, Sunday, 20:00, Andromeda (this) tent:

Reverse Engineering Life: What we can learn from the DNA

Monday, 15:00, Cassiopeia:

Afterparty for the talks, more Q&A and some bonus content.



A brief 2025 recap
DNA: the code of life