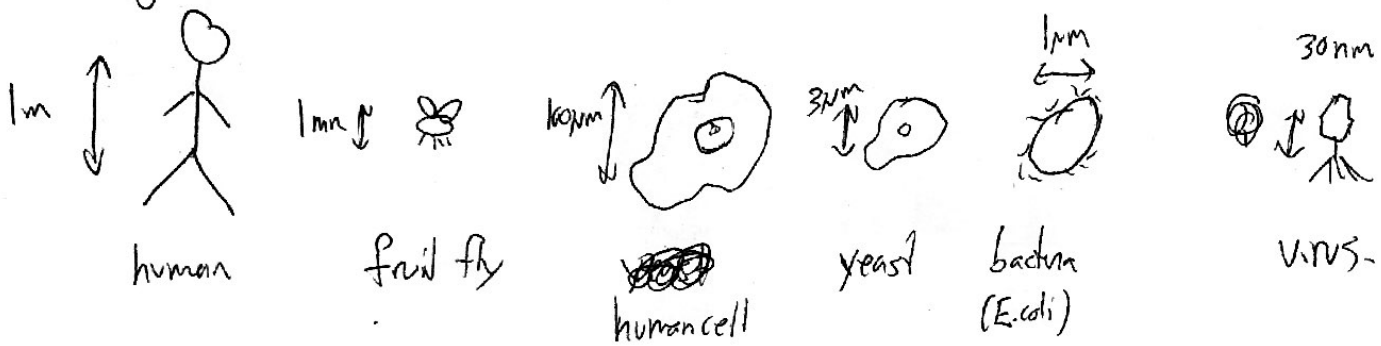
Biology (#s)

I promised that there are no bio prereqs, so I'll quickly review some of the main concepts we'll need for course.

(focus will be on key #'s that you may not have seen before in your bio course in high school)

(1) Organisms: (model?)



these organisms are similar in that the instructions to make them are encoded in a single* long molecule of DNA

information is encoded in sequence of ~~bases~~ "nucleotides" (nitrogenous bases in "base pairs" polymer)

Note: we take this for granted now, but this is pretty crazy from a physicists perspective!

(E. Schroedinger, "What is life")

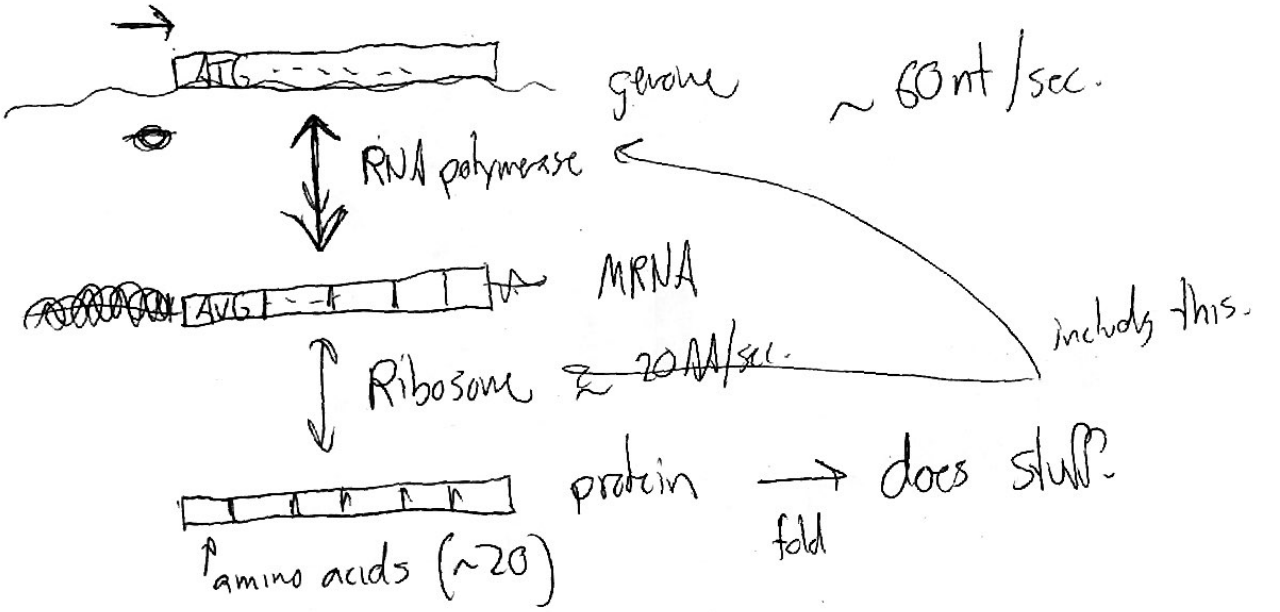
the entire DNA molecule is known as genome

lengths of genomes vary widely across organisms.

human: $\sim 10^9$ bp yeast $\sim 10^7$ bp viruses 10^4-10^5 bp.

fruit fly: $\sim 10^8$ bp bacteria $\sim 10^6$ bp (1Gb, 1Mb, 1kb = 10^9 bp, 10^6 bp, 10^3 bp)

information often encoded in genes (make proteins)



how does Ribosome do it? ATT = codon → 1 amino acid. (isoleucine)

64 different codons → 20 amino acid + "start" codon + "stop" codon. "genetic code"

* degeneracy (will revisit)

* real life is of course a little more complicated.

* typical proteins are ~ 300 AA long (1000 bp of DNA)

⇒ # of genes also vary across organisms,

humans: ~ 20,000 genes. yeast ~ 6000 genes,

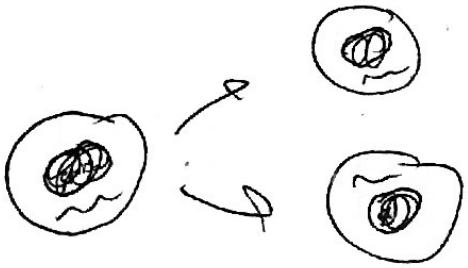
E coli ~ ~4000 genes viruses ~ 10 genes.

↘ 1000x bigger genome, but only 5x as many genes.

⇒ rest of genome is noncoding (regulation / junk?)

End of Lecture 1

Net effect of all this dumb stuff is that organism makes a copy of itself:



① needs to build a new cell wall, all other proteins. ↑ ribosomes!

② needs to copy its DNA.

↑ typically rate limiting factor.

* Viruses are tricky: they have to hijack a host cell's machinery!

doubling time varies across organisms:

human ~ 20yrs.

human cells ~ 1 day

E. coli ~ 20 mins - 1hr

Prochlorococcus (ocean bacteria) ~ 1 day.

when genome gets copied, there's a chance for introducing an error (since it's a single molecule).

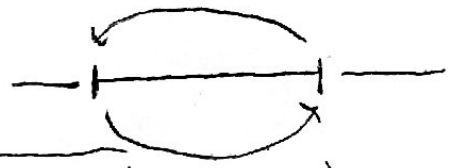
"mutation" ATGCCA → ATGTCA

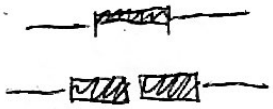
simplest kinds of mutations are <u>single nucleotide mutations</u> (also "point mutations" "substitutions")

A→T, T→C, etc.

but can also have bigger things: insertions ATGTTTCA (+TTT)

deletions: ATGTCA (slippage of DNA pol.)

rearrangements 

duplications 

pretty complicated!

↳ often via special genes that jump in & out of genomes ("transposons")

cells have sophisticated machinery for detecting & fixing errors.

⇒ (point) mutation rates vary widely across organisms:

eg. Humans: $\mu \sim 10^{-8}$ bp/gen
human cells: $\mu \sim 10^{-10}$/bp/division

E.coli – $\mu \sim 10^{-10}$/bp/division.
viruses ~ $\mu \sim 10^{-5}$/bp/division

- worth pausing & thinking about these #s a bit:

Humans: genome is 3×10^9 bp long → so 30 mutations per genome introduced every generation.

~ 10^{10} humans on earth, so every single bp mutation is present in ~ 100 individuals. that or alive today.

⇒ all double mutations are real $(10^{10}\times10^{-8}\times10^{-8})$ → sequence space is big!

E. coli: genome is 4×10^6 bp long, so 4×10^{-4} mutations per/gen

⇒ > 1000 replications before e.coli produces a single error!

in gut, ecoli are at 10^9 cells, so almost every bp mutated 1+/in us every day.

× 10^9 billion guts → almost all double mutants around in worldwide pop.

(but not triple...)

---

what do mutations do? (genotype ⇒ phenotype map)

⇒ in general, we don't know. (even for e.coli)

⇒ but can make some guesses based on what we know about "central dogma"

e.g. if mutation occurs in middle of a gene, it will change the codon: ATC → ATT

① because of degeneracy, might not change AA (therefore, doesn't change protein) "synonymous" mutation.

② might change AA to something else "nonsynonymous"

↳ other AA (small change?) ("missense")

↳ stop codon → truncates gene (big change) ("nonsense") "loss-of-function"

these 3 classes of mutations are easiest to think about because we have some prior expectations.

⇒ but lots of mutations occur outside of genes as well.

⇒ some missense mutations will be more important than others.

That's enough background for now. will introduce new stuff as needed.