

Welcome to MIP 280A4, Microbial Sequence Analysis

Tuesdays/Thursdays

9:30 - 10:45

Micro B120

Office hours

By appointment

Please reach out to me anytime!

Math
undergrad

7 years as a
software engineer

PhD in mol.
biology /
biochem.

Postdoc using
microarrays, NGS,
and bioinformatics

Assoc.
Professor at
CSU

1999, Bangkok, Thai Airways test facility



CENTER FOR VECTOR-BORNE
INFECTIOUS DISEASES



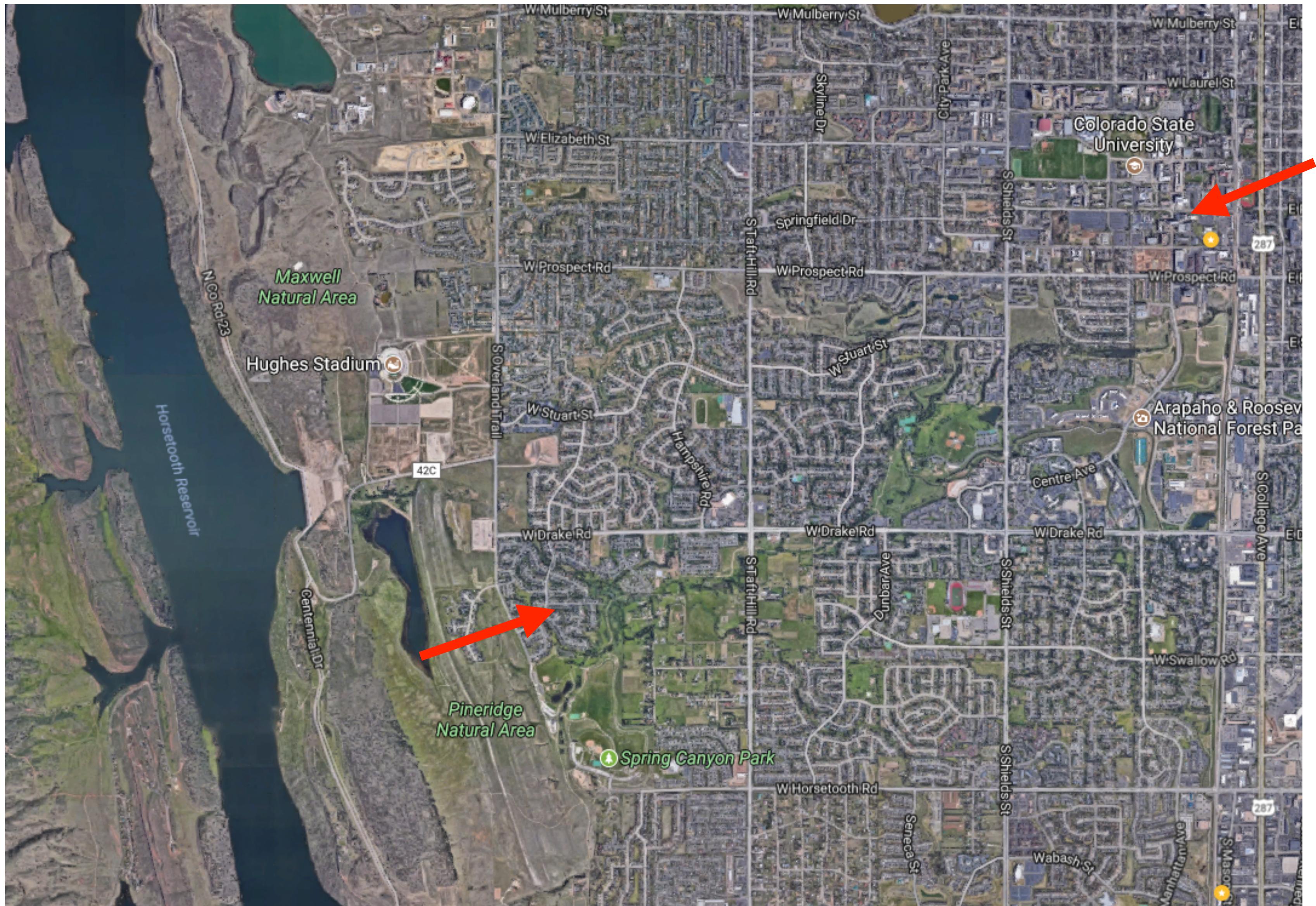
Mark Stenglein, PhD

Associate Professor

Department of Microbiology, Immunology, and Pathology
College of Veterinary Medicine and Biomedical Sciences
Colorado State University

Mark.Stenglein@colostate.edu

StengleinLab.org





A bear, oh my!

From Olander Elementary <noreplyolander@psdschools.org>★

Subject Bear Near Spring Creek and Dixon Creek St

Reply to kbesance@psdschools.org ★

To Mark Stenglein ★

Reply Forward Archive Junk Delete More ▾

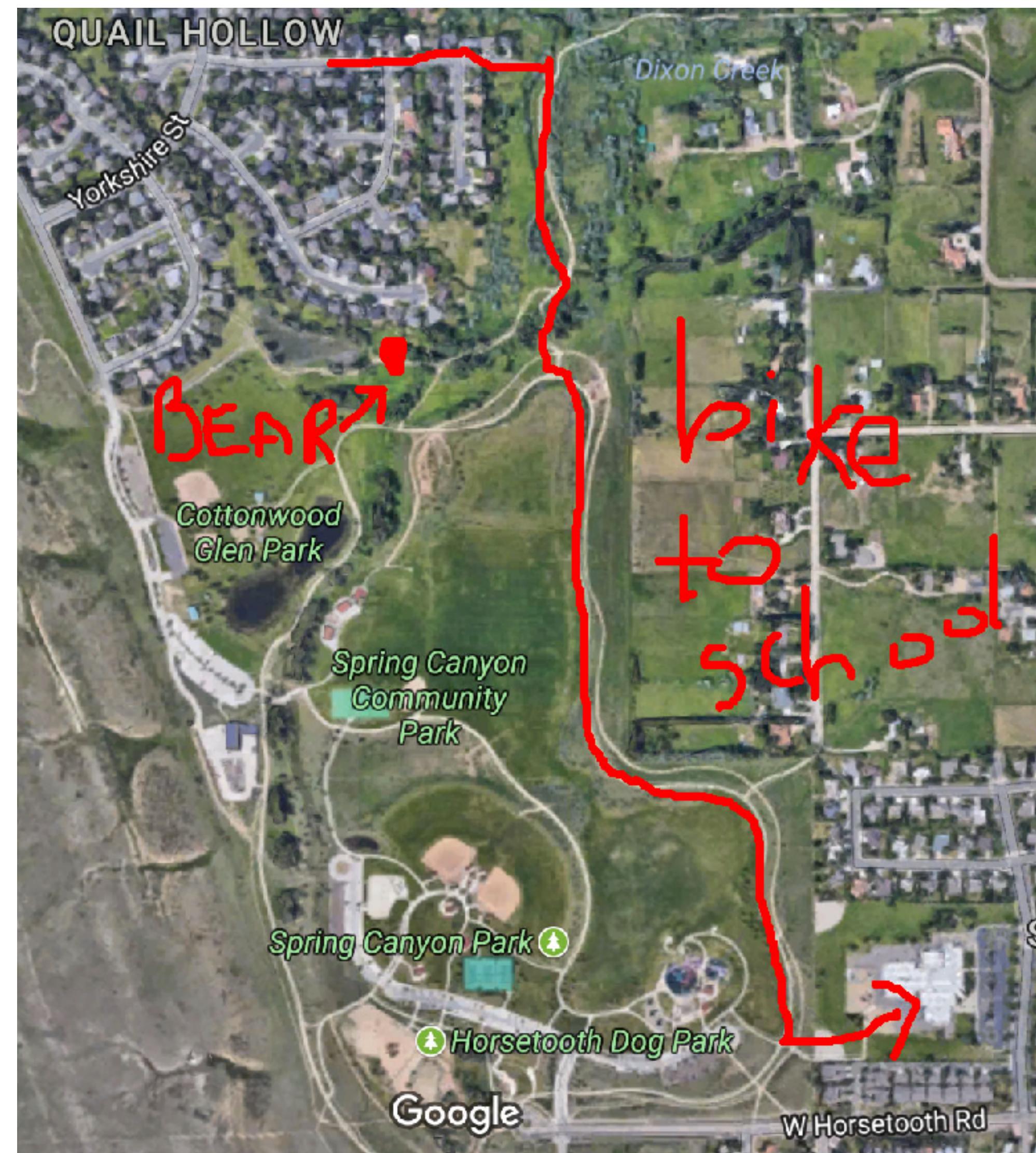
9/20/17, 8:22 AM

Dear Olander Community,

I have received a phone call that a bear has been witnessed 100 yards north of the "Owl Tree" where the path crosses Spring Creek near Quail Hollow and near Dixon Creek Lane. The police and Department of Wildlife are tracking the bear. We would like everyone to be extra cautious if they ride or walk to school this morning. I am in contact with the police and I will be present outside throughout the day. We will update you as we receive new information.

Thank you for your extra safety today.

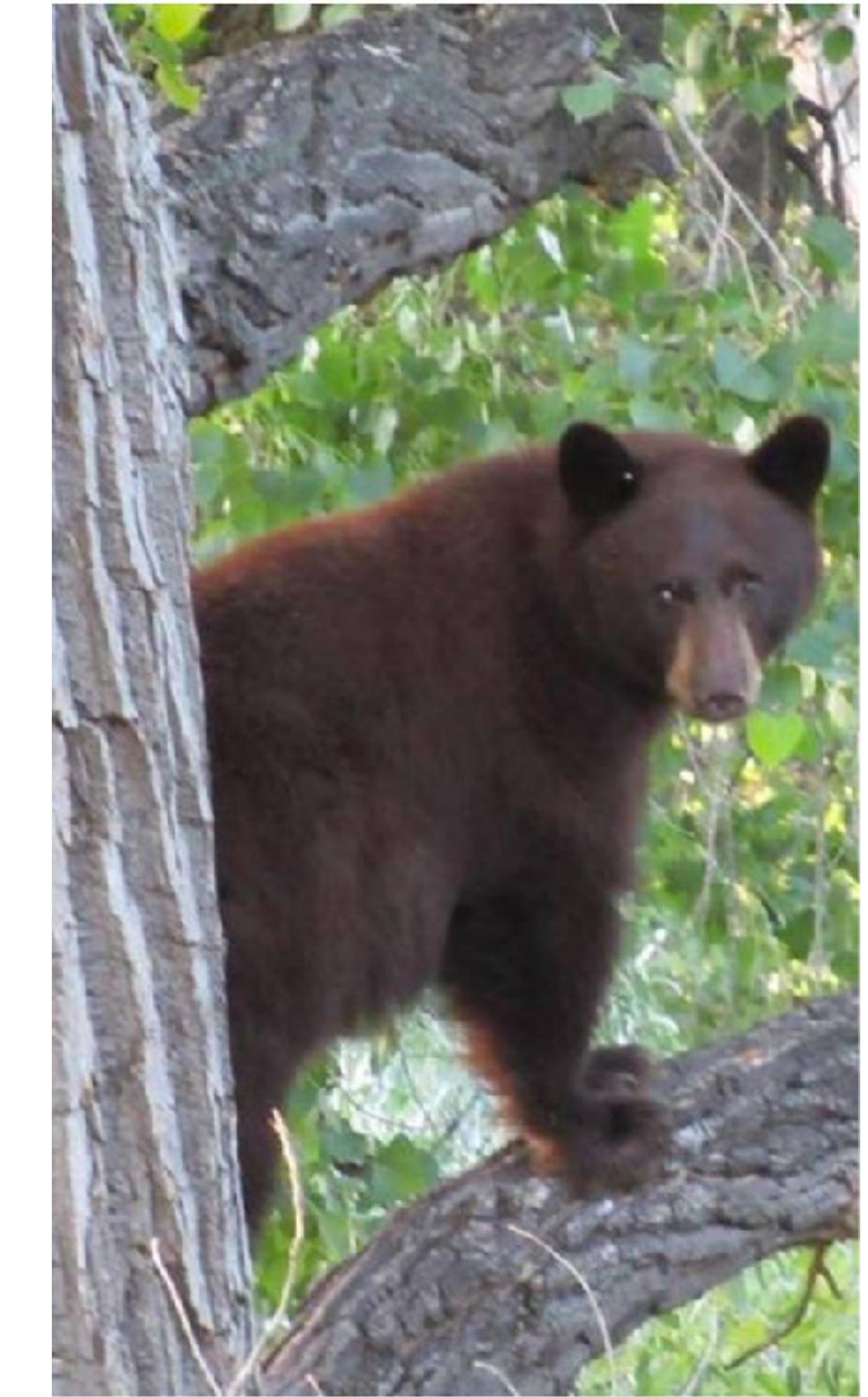
Mark







The bear roamed the neighborhood for a couple days



Is the bear pooping on our neighbors' backyard deck?



The bear and its poop.

Like Comment

9

Jillian Busch I should just say that those crocs are right next to our sliding door to get into the house so I wonder if it tried to open the door?
Like · Reply · 2 · September 20 at 11:38am
Jonny Lorono Pf replied · 1 Reply

Karen Wente Ewwwww
Like · Reply · September 20 at 2:07pm

Jan Warner Oh crap!!
Like · Reply · 2 · September 21 at 6:43am

Hermann Morenc I only see the poo ...
Like · Reply · 2 · September 22 at 6:37am
Hermann Morenc replied · 2 Replies

Ray Cyst The bear was trying to tell you how shitty those crocs are, honey.
Like · Reply · 2 · September 22 at 3:58pm

Write a comment...



The bear came back last night and pooped in the same spot almost on top of Mark's running shoes!

Wow Comment

18

View 8 more comments

Karen Wente I heard people can eat bear meat.
Like · Reply · 1 · September 21 at 2:28pm · Edited
Jillian Busch replied · 1 Reply

Whitney Schneider Vogel You now know the answer is no.

Does a bear shit in the woods?
Like · Reply · 3 · September 21 at 2:40pm

Katie Tjarda Maybe he trying to stop Mark from running away from him!
Like · Reply · 1 · September 21 at 2:53pm

Margaret Otepka what are you going to do?
Like · Reply · September 21 at 4:01pm

Sunny Rae LeMasters Mosal He has an obvious shoe fetish
Like · Reply · September 21 at 6:18pm

Wait, is the 'poopetrator' a bear after all?



Jillian Busch

September 23 · iOS ·

...

We've captured on film what we think is a mountain lion doing this! Not a bear! The scat looks like it could be one of any number of creatures.



A controversy rages in the neighborhood and on social media



Can science help?

Melanie Potyondy Okay, not to be a doubting Thomas, but is there a chance this is a blurry picture of a house cat at a weird angle? I looked at the exit photo, and it looks a lot less mountain lion-y going the other direction.

Like · Reply · 1 · September 24 at 11:42am · Edited

Jillian Busch Yes true, but the piles of poop are ridiculous and not from a house cat. I wonder d about the angle too. Mark's uncle spoke with him and he believes it's a mountain lion (based on all his outdoorsy and bear research)!

Like · Reply · September 24 at 11:35am

9/25/17, 9:56 AM

Mark Smith

I went in to the division of wildlife and talked with a senior biologist there. He said it was "absolutely" a Mtn lion. He also said not to worry about it, as they are actually everywhere and we just don't realize it.

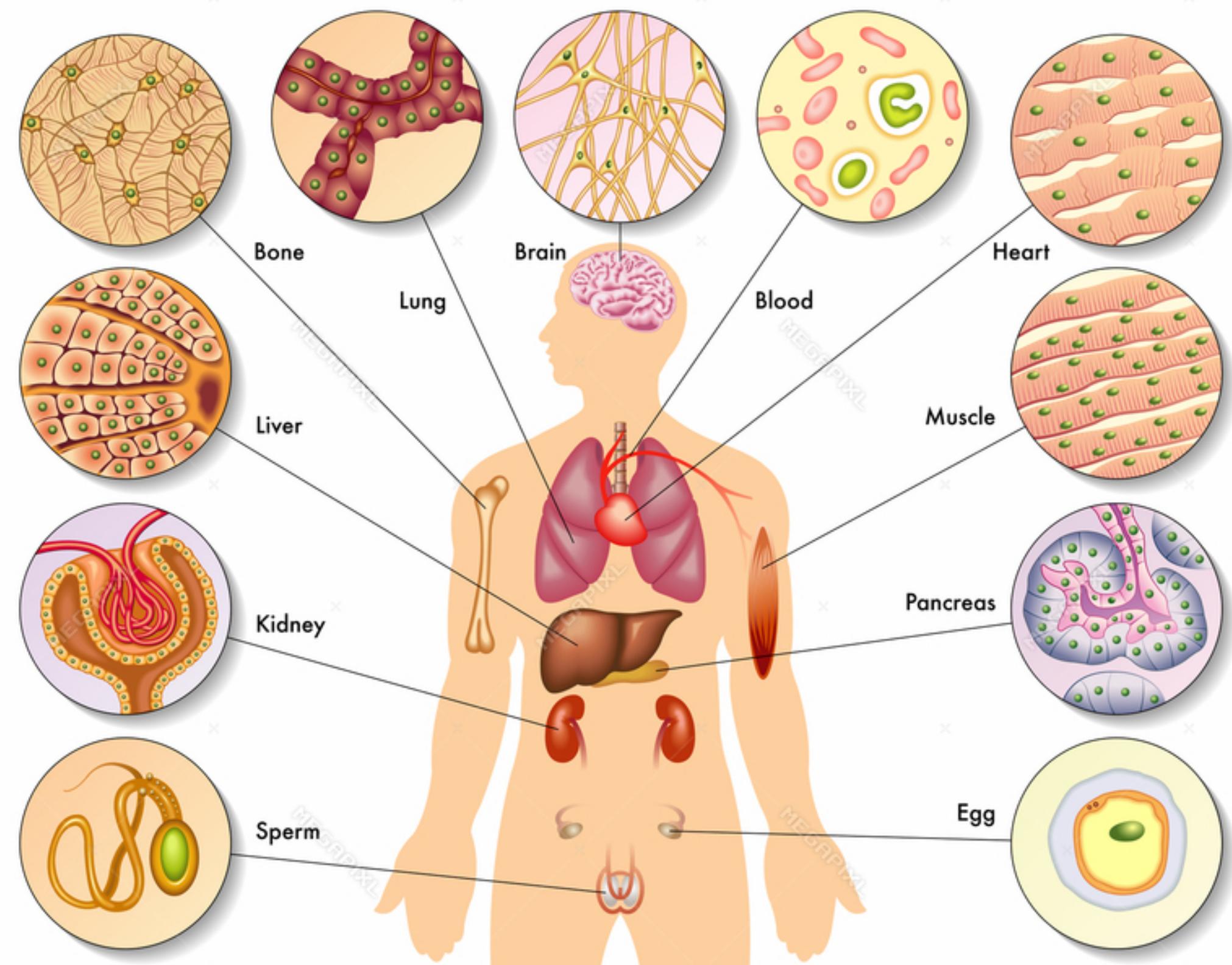
Eric Potvondy

One more text from another wildlife guy I know:

Too small for a mtn lion. Look at size compared to shovel and buckets. I think large domestic cat

DNA sequencing to the rescue!

Living things are made of cells



Each cell contains a copy of the organism's genome

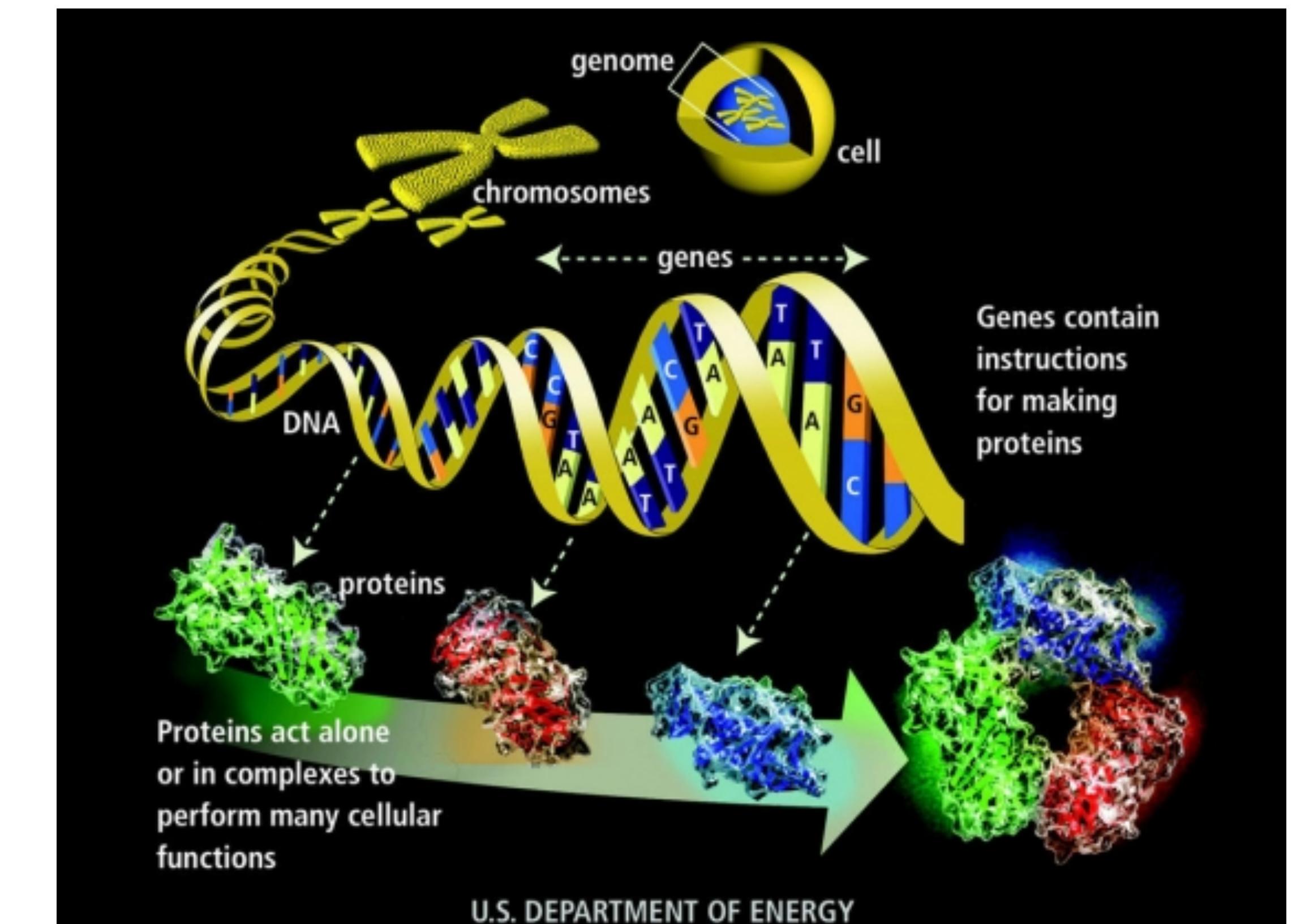


image: US DOE

electron micrograph of poop

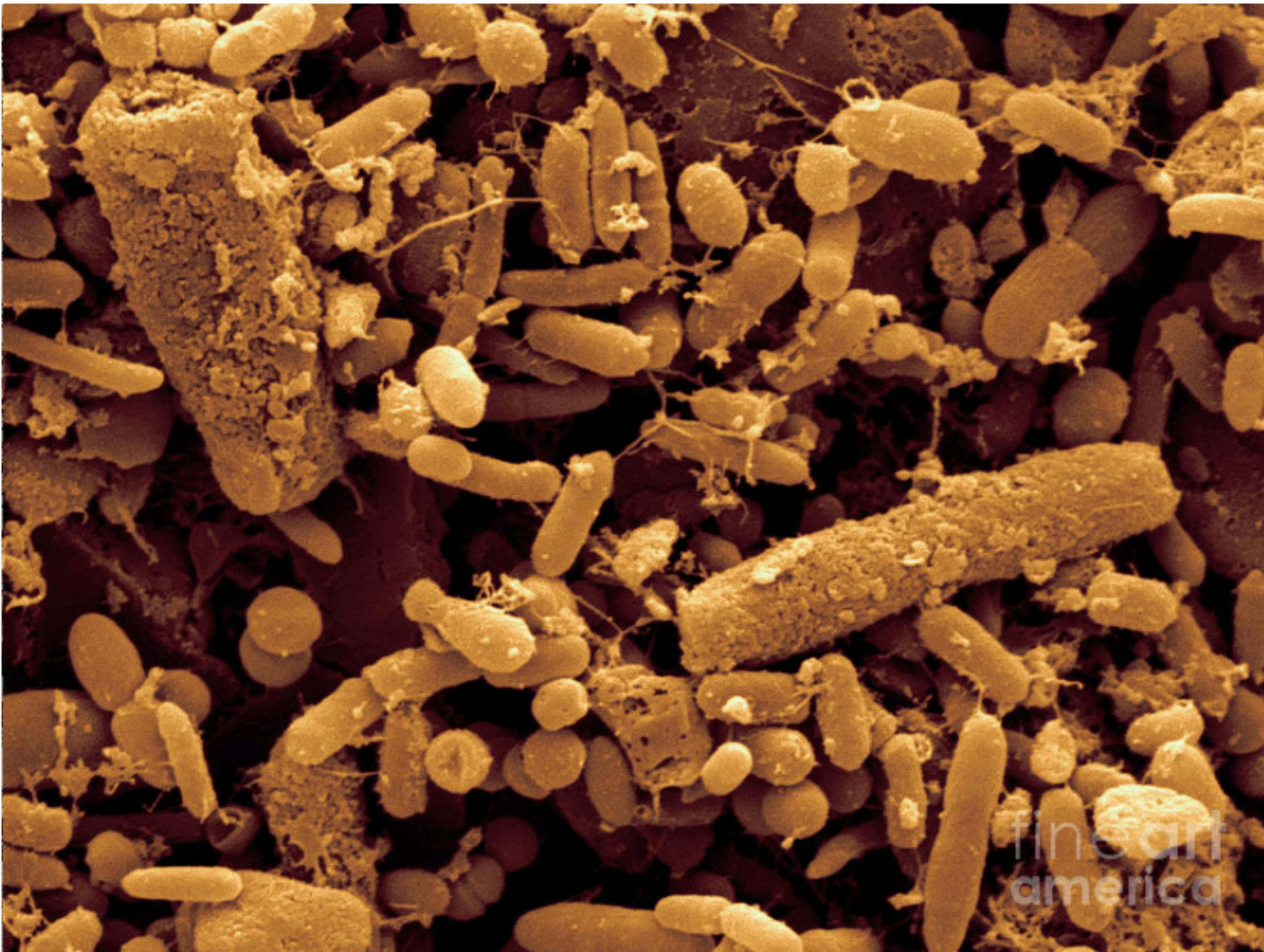
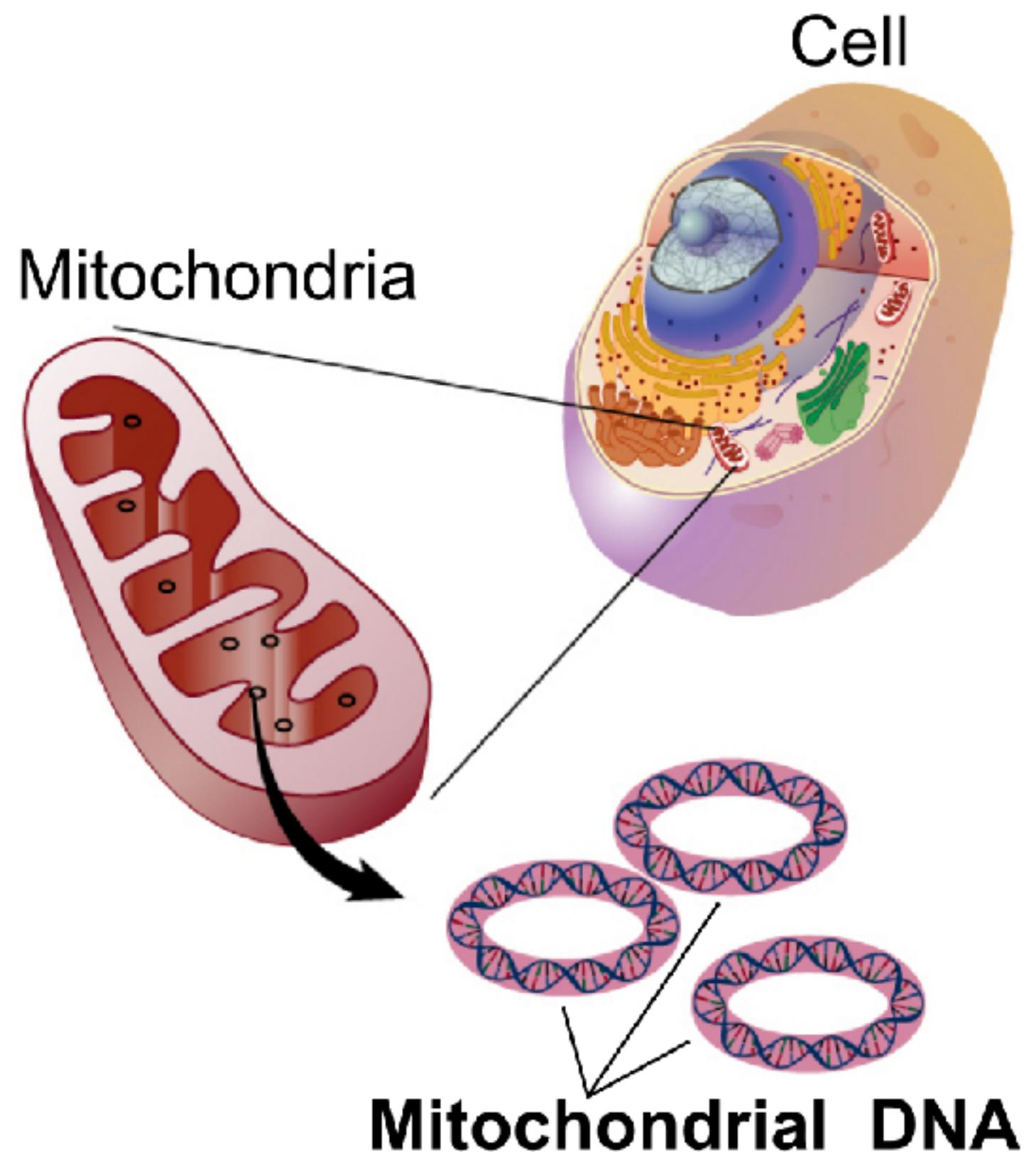
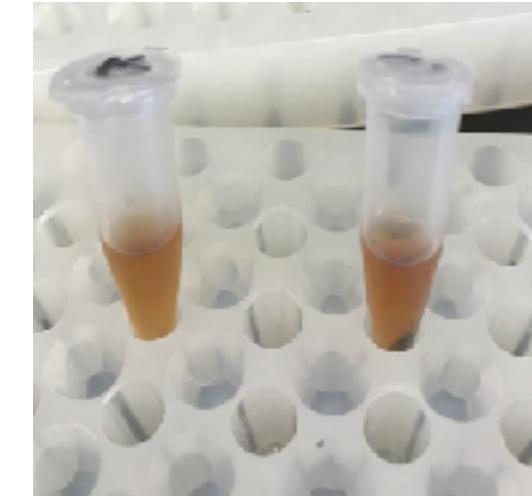


Image: Science Source Images

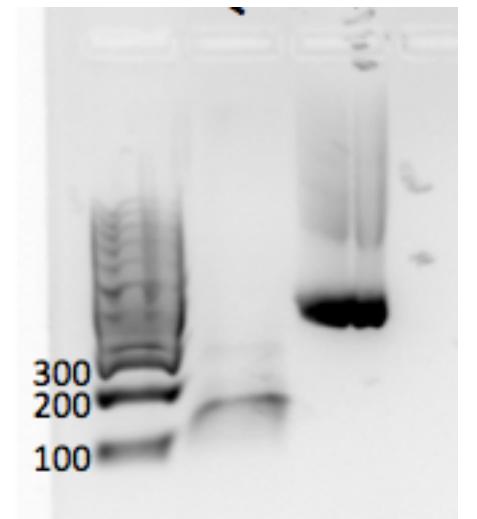
DNA sequencing to the rescue!



Poop



DNA



PCR to amplify part
of the mitochondrial
genome

Sanger sequencing

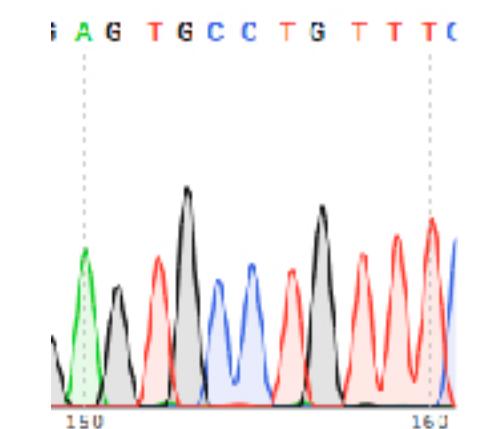


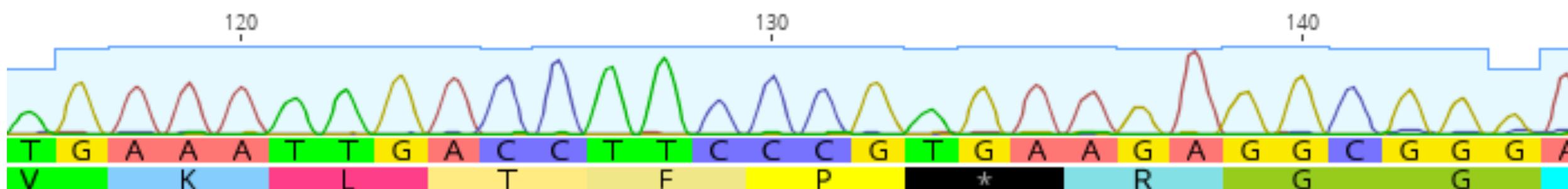
image: wikipedia

Compare to sequences in public databases

Recap



What pooped on our neighbor's shoes?



The answer!

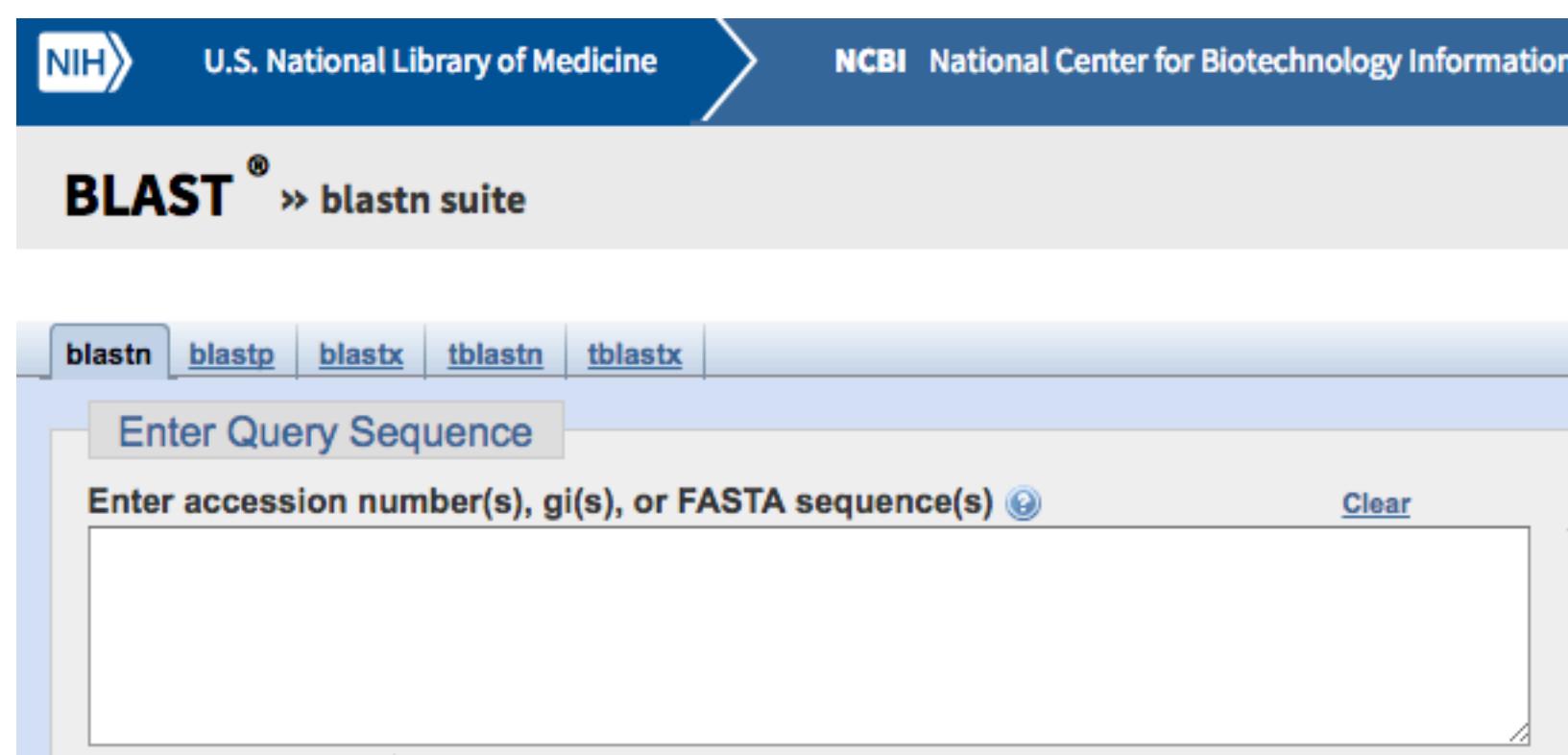
NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information

BLAST® » blastn suite

blastn blastp blastx tblastn tblastx

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)



Sequence
from mystery
poop

>AB462167.1 Procyon lotor mitochondrial gene for 16S ribosomal RNA, complete sequence, specimen_voucher: personal:Tomoharu Tokutomi:NDMC-PL-MIE18010 Length=1586

Score = 433 bits (234), Expect = 1e-117
Identities = 234/234 (100%), Gaps = 0/234 (0%)
Strand=Plus/Plus

Query 1	CAAAAACATCACCTCTAGCATTAAACAGTATTAGAGGCACTGCCTGCCAGTGACATTAGT	60
Sbjct 840	CAAAAACATCACCTCTAGCATTAAACAGTATTAGAGGCACTGCCTGCCAGTGACATTAGT	899
Query 61	TAAACGGCCCGGGTATCCTGACCGTAGCAAAGGTAGCATAATCATTTGTTCTCTAAATAGG	120
Sbjct 900	TAAACGGCCCGGGTATCCTGACCGTAGCAAAGGTAGCATAATCATTTGTTCTCTAAATAGG	959
Query 121	GACTTGTATGAATGCCACACGAGGGTTGACTGTCTCTTACTTCCAACCAGTGAAATTG	180
Sbjct 960	GACTTGTATGAATGCCACACGAGGGTTGACTGTCTCTTACTTCCAACCAGTGAAATTG	1019
Query 181	ACCTTCCC GTGAAGAGGCGGGATAAGAAAATAAGACGAGAAGACCTATGGAG	234
Sbjct 1020	ACCTTCCC GTGAAGAGGCGGGATAAGAAAATAAGACGAGAAGACCTATGGAG	1073

Procyon lotor = raccoon



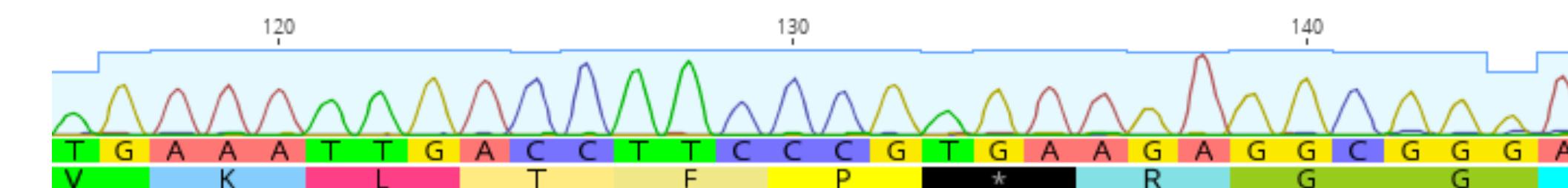
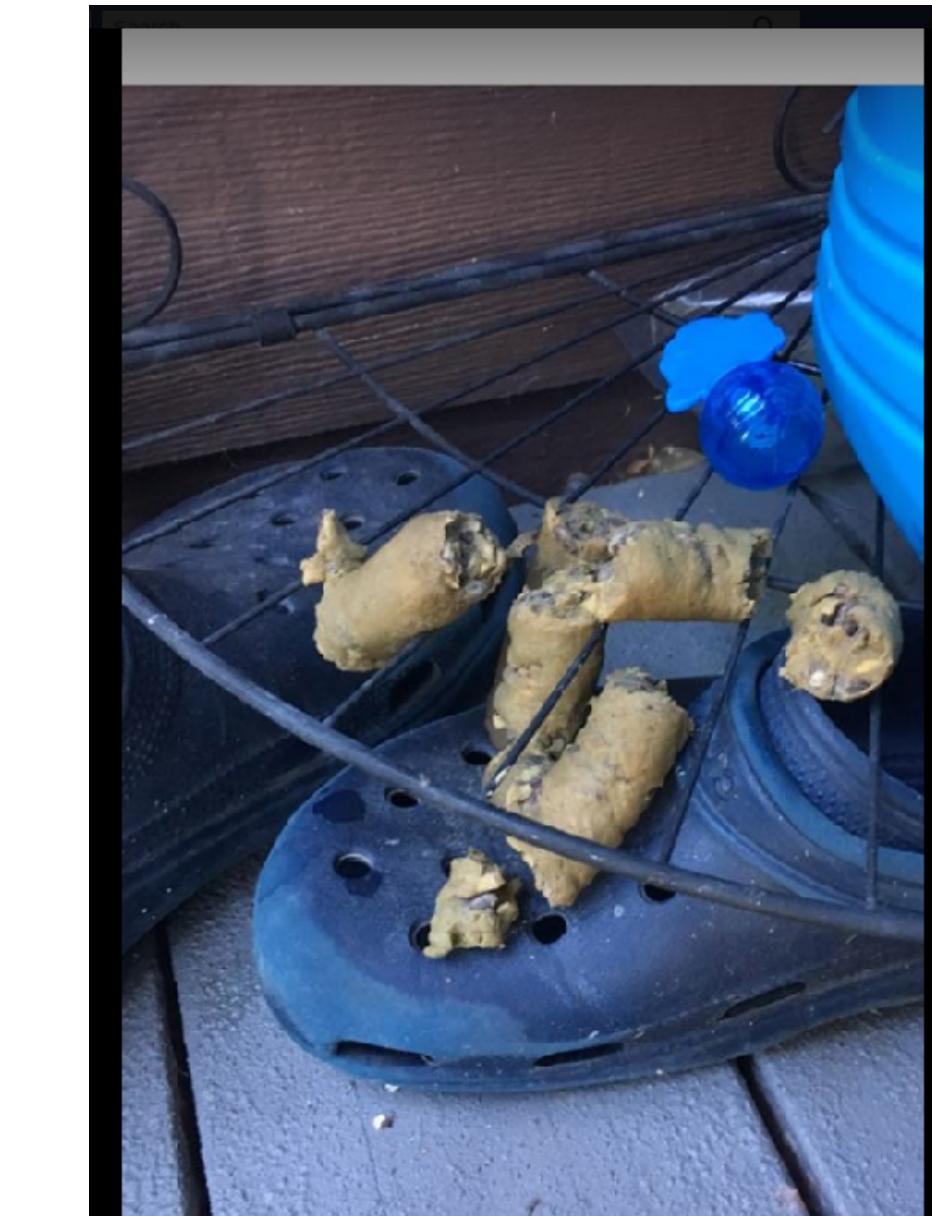
image: wikipedia

So there was raccoon DNA in the poop. From a raccoon? Other possibilities?



image: wikipedia

The available data are consistent with it being raccoon poop



So why and how does this work?

```
>AB462167.1 Procyon lotor mitochondrial gene for 16S ribosomal RNA, complete  
sequence, specimen_voucher: personal:Tomoharu Tokutomi:NDMC-PL-MIE18010  
Length=1586
```

```
Score = 433 bits (234), Expect = 1e-117  
Identities = 234/234 (100%), Gaps = 0/234 (0%)  
Strand=Plus/Plus
```

Query 1	CAAAAACATCACCTCTAGCATTAAACAGTATTAGAGGCAGTGCCCTGCCAGTGACATTAGT	60
Sbjct 840	CAAAAACATCACCTCTAGCATTAAACAGTATTAGAGGCAGTGCCCTGCCAGTGACATTAGT	899
Query 61	TAAACGGCCGCGGTATCCTGACCGTGCAAAGGTAGCATAATCATTGTTCTCTAAATAGG	120
Sbjct 900	TAAACGGCCGCGGTATCCTGACCGTGCAAAGGTAGCATAATCATTGTTCTCTAAATAGG	959
I		
Query 121	GACTTGTATGAATGGCCACACGAGGGTTTGAUTGTCTCTTACTTCCAACCAGTGAAATTG	180
Sbjct 960	GACTTGTATGAATGGCCACACGAGGGTTTGAUTGTCTCTTACTTCCAACCAGTGAAATTG	1019
Query 181	ACCTTCCCGTGAAGAGGGCGGGATAAGAAAATAAGACGAGAAGACCCATGGAG	234
Sbjct 1020	ACCTTCCCGTGAAGAGGGCGGGATAAGAAAATAAGACGAGAAGACCCATGGAG	1073



Why does this work on a biological level?

How does this work from a technical standpoint?

Human cells contains 2 copies of the human genome,
which is 3 billion ‘bases’ long

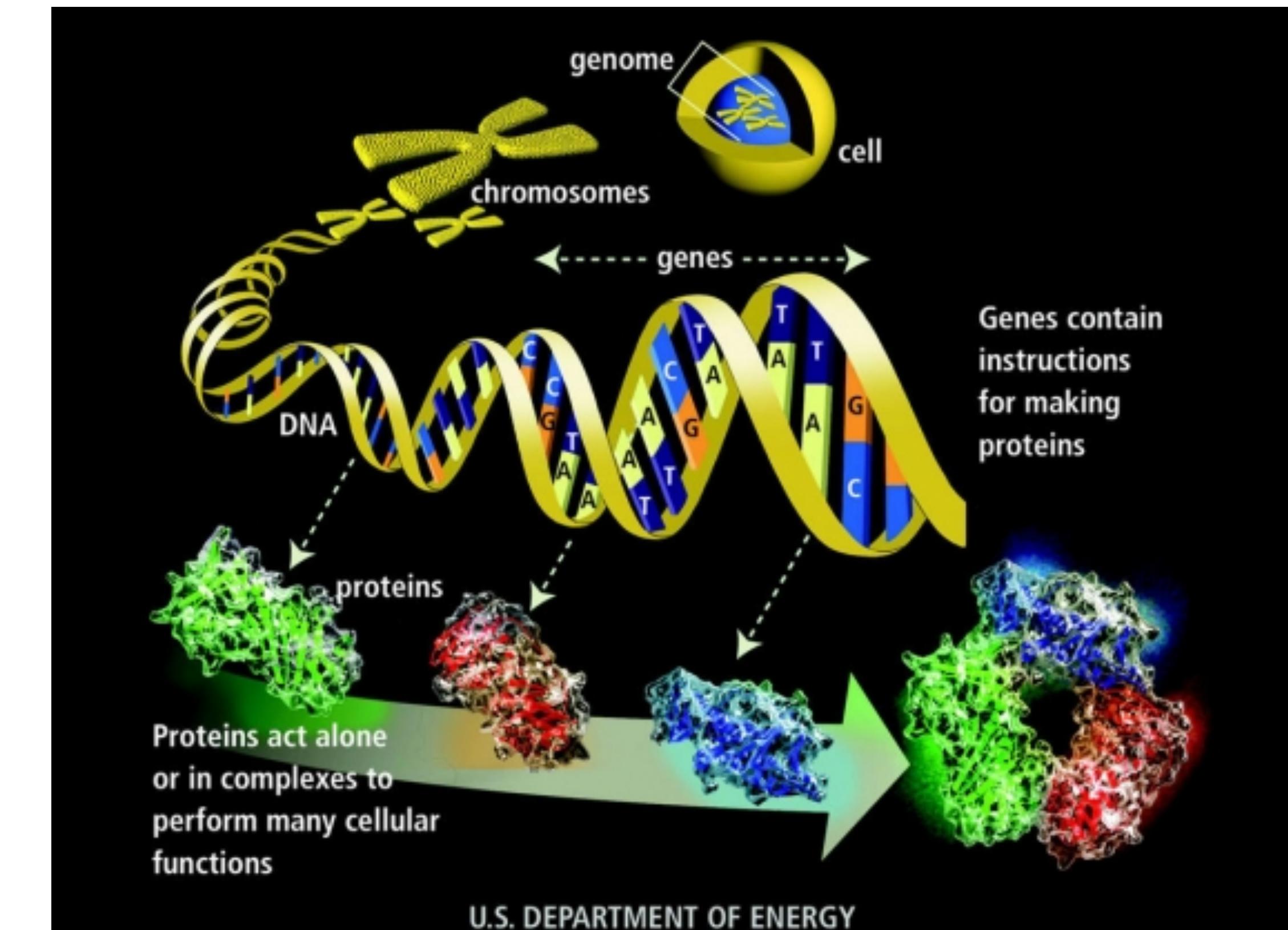
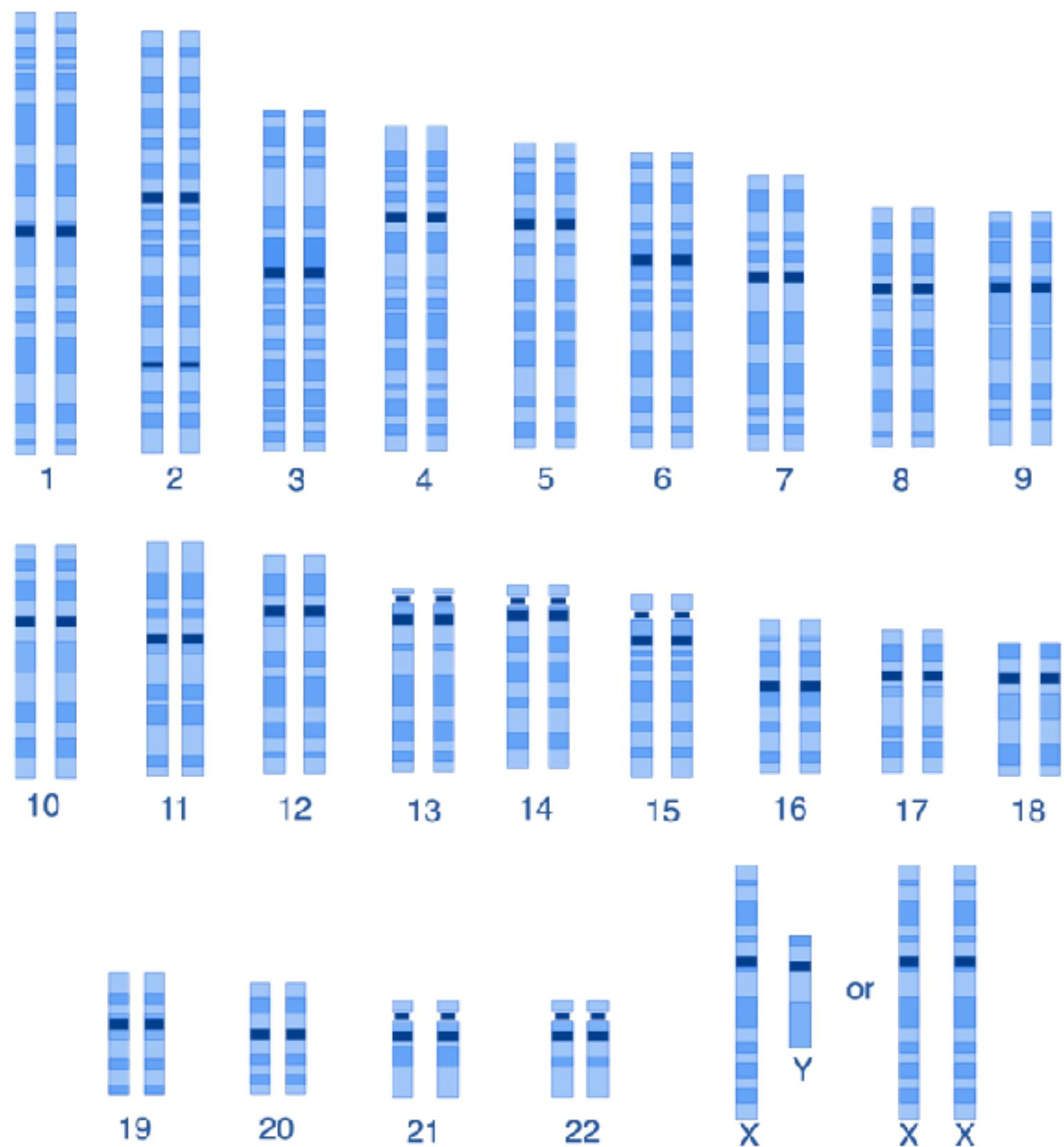


image: wikipedia

image: US DOE

* sperm and egg have 1 copy; red blood cells have 0 copies

Each time a cell divides it has to copy its genome

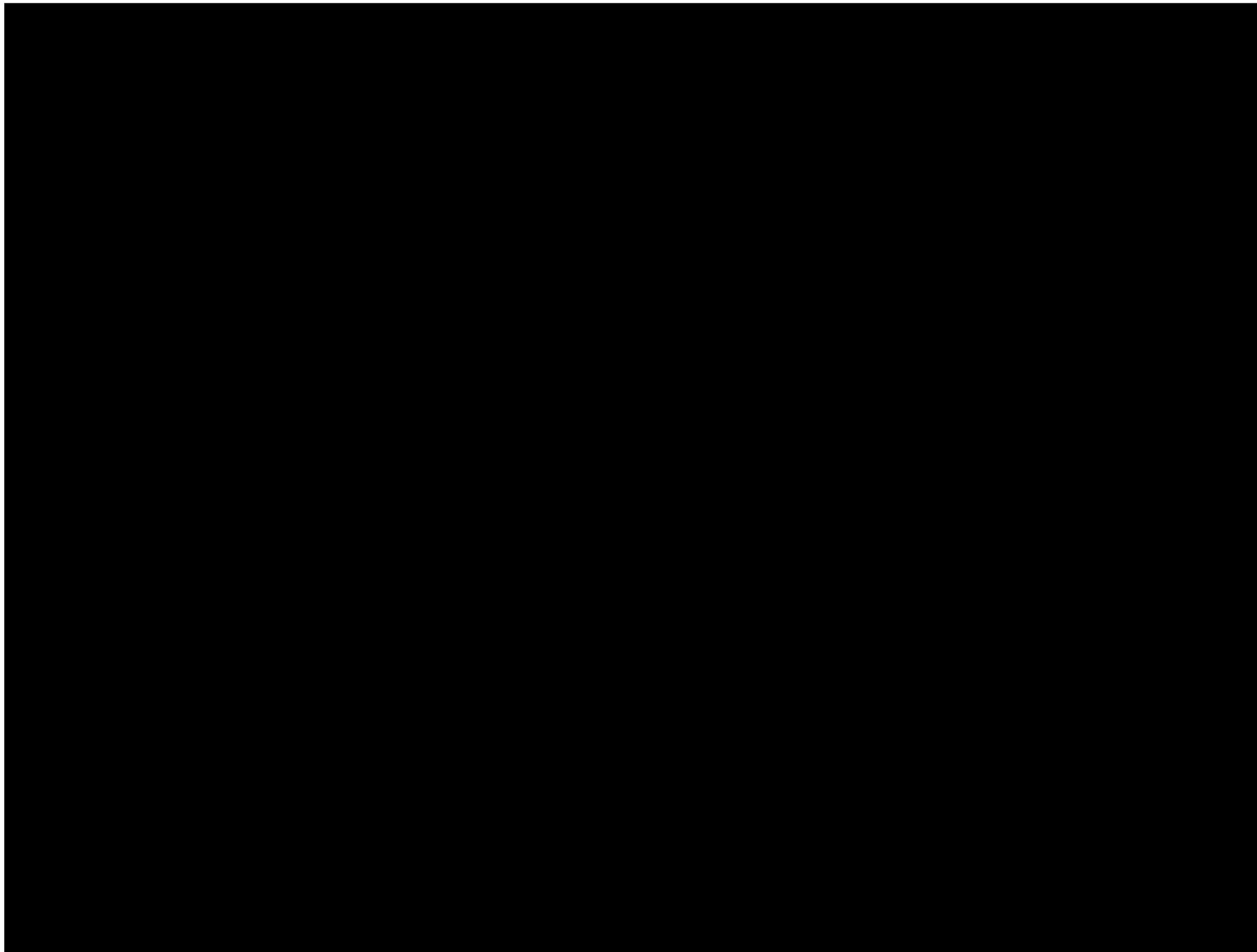
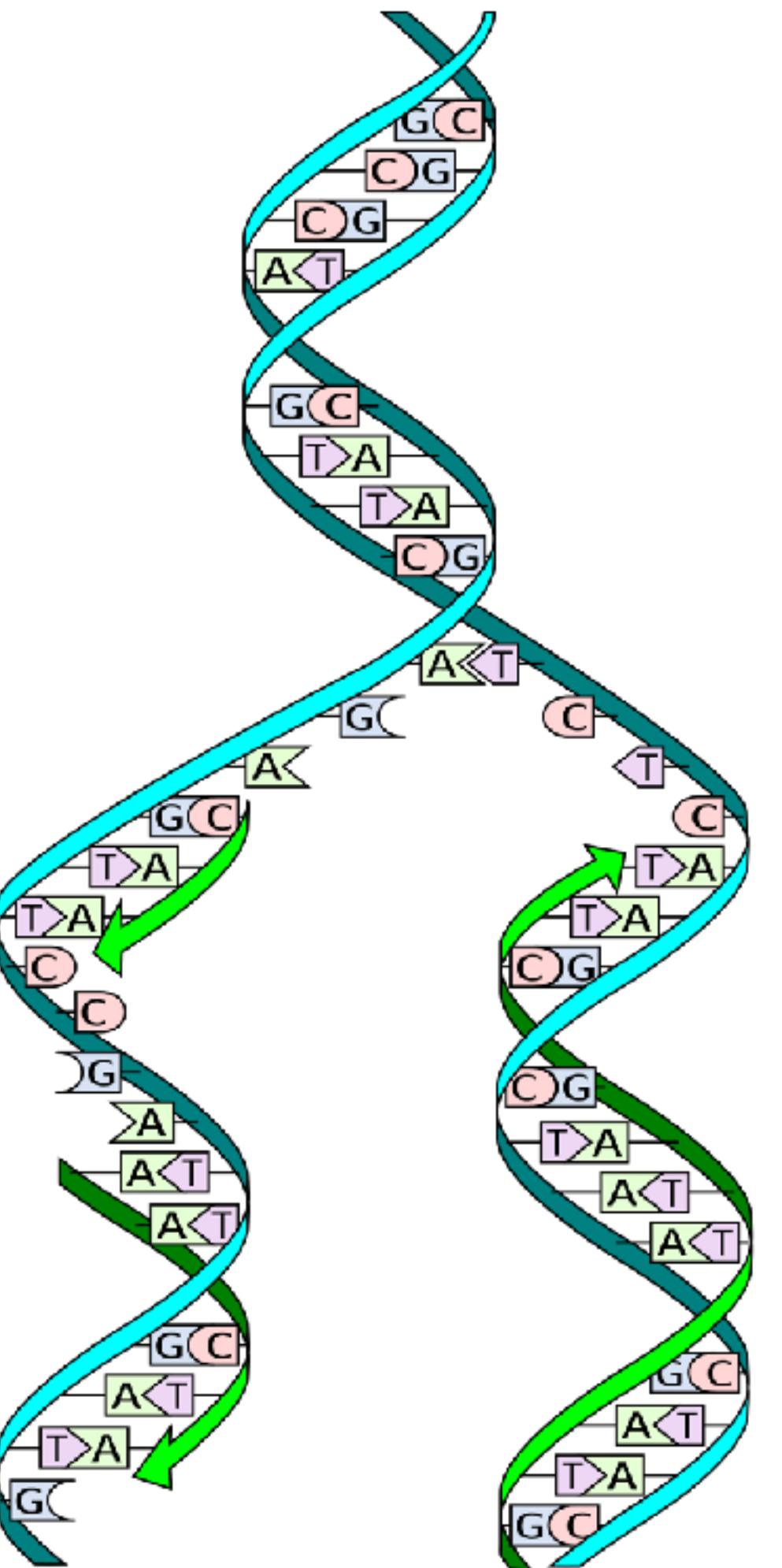


image: wikipedia

video: PBS

Let's try our hands at DNA replication!

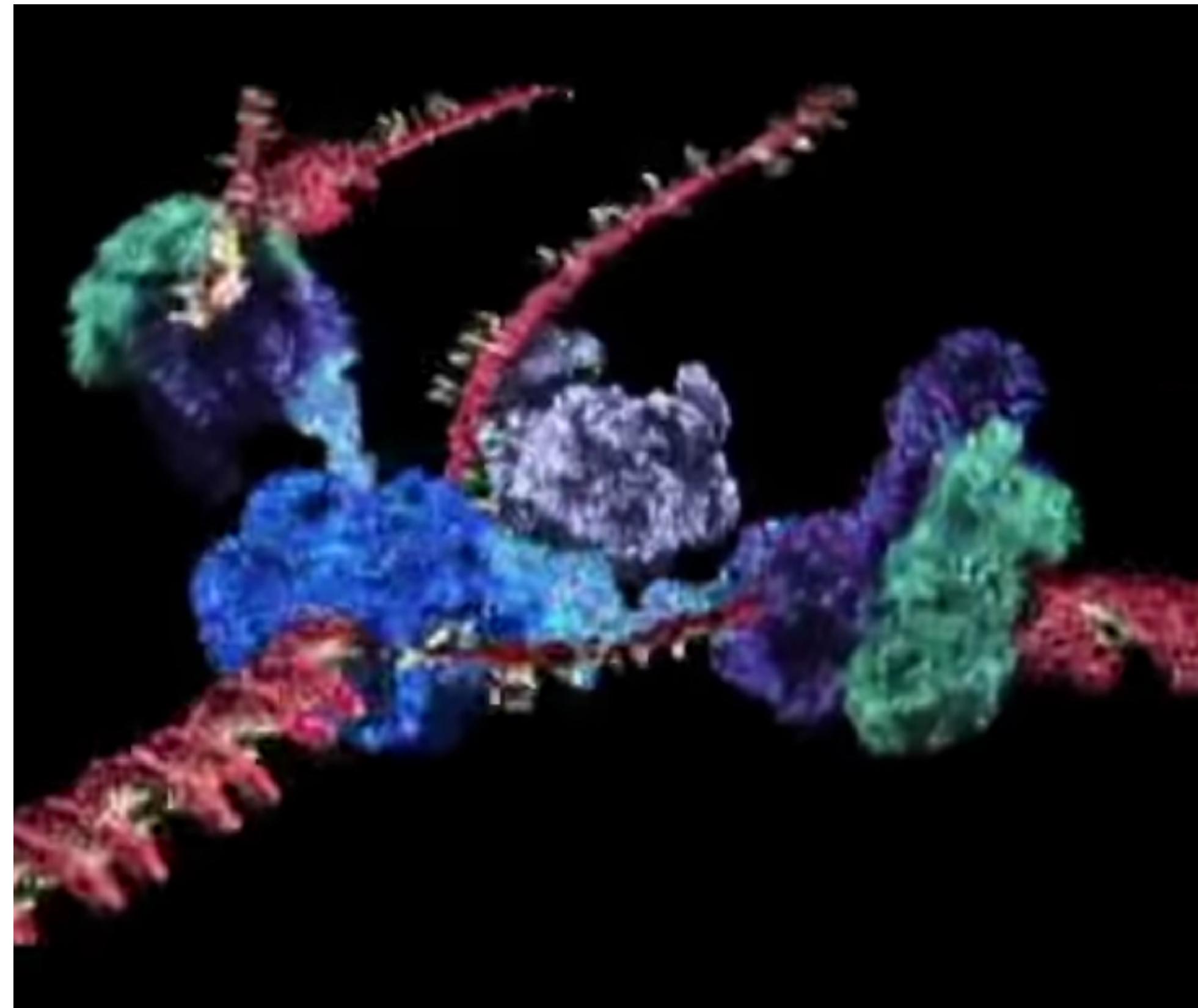
Can you accurately copy 100 bases in 1 minute?

AAAAAGGATTTCTCAATGTTGGT
GCACCATTTATTGCTATGAGTC
TGGATAGAGACGGAACCCCGCTCTC
CAATGACCACTCTATAGCTGAAAGA

A human replication fork would have copied ~3000 bases in 1 minute

GAACAAACCAAGCTACAACAAATCTACAATATTGTCAATTACATTACTACTTCCATTTAACATGTCTAGTGCAATGTATGAAACAATTATCAAATCGAAGTCCTCA
ATCTGGGGAACACATCTCGGGTAAAGCAGTAGATAGTATTGGATTGATCAATCTCCGGAAAGAAGTTGGTCGAAGCTCAACTCTATTCTGACTCCAG
GAGCAAGACCAGTTCTGTTACACTGGTAAAGTTGGCTTCTCCAACAGAAGAAAAAGAAATTATAGTGAGATGTTGTGCCTATTGGATGACATTGATCTGA
ATTTCTCCTTTCAGGGAAATGTTGTCGAAATTCTGGTCAGATCTAACACAAACACAAACAGGTGTTAACATCAAGGTCAACTCAAAGTGTATCCTCTCAGTTG
CTCAGAATGCTTGAAGAGCAAATAGCAGTGCCTGAAATTACTCAAGATTGGCTGAAAGAATCTGACATCTCCCTCCAAATAATTCAATTGAAGCTGCAAATAA
AGGATCATTGTCTTGTCAAAGAAGTCCTTTGATGTCAAGTATTCAAACACCAATCCATGGCAAAGTCAGTGTCTTCTCCTACCAGAAGTGTGATTGAAT
GGCTGTACACACTTAAGCCTGTTAACCAATCCCAGACCAACACAGGACAGTAAACACTTGGCTGAAAATCACTGGCAATGTCTGCAACTCTGATTAAATG
TCAGATACTCATCGTTGTCAGGCTCAATAACAAGCCTTTAAAATCAGCCTTGGATGCGCATCCCTAAAATAATGAAATCAAACACATAAGCCGGTTCTT
CACCCGTCTGATGAATCTTCTCCTAAAGAGTATTATAAGCATTCAATGTCTTCCGAATCACACAAATGTTGAAACAGTCATTGAATATAACTTGATCAGTCAA
ACCTCTTCTGAATCAACTCCTCTAGCAGTGATTCAAAAATTGAGATGAATTGGATCTAAAAGAACCTTACAATGTTATCCATGATATGCGTATCCTCAA
AGAATTGTCATTCACTCTTGAAATCCACACAGAACTGCTCAAACGTCTGTGACAGTGTGAGCAAGACATGATTGTTCACTATAATGAGCCAGATCTAAA
GCCAAAAAAAGTTGAGCTAGGGAAAAAGACTTAAATTATTAGAAGATGGTTATGGGAGAAAATATTCCTTCTCAGACCTTGAAAAGTCTCCGAGAAACTCAC
AAACAAATGTCTATTGGATAGCATCCAGATGCCGATTGGAAATTGACTATGCTGCAGGTGAAATAAAAATTCTCCTAGATCAGAGGATGTTGAAAGCTATT
TCTAAATTAGATTAAACCTGGTAAACTGTCCTAAGTAAAGTTGTTACATGCATTAGATCAGATTAAACAAATCTAACAGATAAACCAAAAC
AATCATATGAAATAAACATAACATAAAATATAAAAAATACAAAAAAATCATAAAATAAAACCAAAAGGATGGCCTCGGGCACAAATTGGTTGC
TTAATAATGCTTAAAAGAATGTATTAGTAAATTAAACTTTAAATCCAATCTACTCACAAATTGGCAAAAATTGTATTGTTTTGTTTTGTTTTGTT
TTGTTTTGTTTTATTGTTTTATTGTTTTATTGTTTTATTGTTTTATTGTTTTATTGTTTTATTGTTTTATTGTTTTATTGTTTTATTGTT
TTATTATTGTTAGCTTTACTGTTATTTCACACGCAAACACACTTCAAGTTATATATTAAACACACATTAAACTTAAATGCATTAAACACACAAAG
ACACTTAATACACTCAAACAATAATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAAATGCATTAAACACACACAAAG
CAAACCAAGCTCAAATCTCTTAAATAGAATCATTTCACACGCAAACACACTTCAAGTTATGTTAGCATTAAACATGCTGAAATGGATGTAAGCCCTTCTTGTAGGGCATTG
CAGCAAGTCCTTAGCTTCGGACTACAAGCCTTAGTATCTGCATATTGTTAGCCTTGCAATTCAACAGAGTCATGCTATATCCTTGTGTTAGAACT
GTGCACACTTCCAACTGCCTTTAGCTGCTAAACTTAGACATGTCAATTCAAGCTCAACATGTTAGCATTGATAAATGCCGAAACTAGTGAGCTATT
AAAATTCACTGCTATCAGAGGAAGACTCCTCCTAACAGAGAACACCCAAGACACAGGATTCAAATCTGTGGTTGCAAGACCATAAGGCAATCAGAGGGT
GACTTGGAAAGGCTATTAGCTTCAGAGCAGATCCATTGCTTATTCAACAAAGATGAACCTCACCATTGCATCAAGTCTCGGAAAGTCATATCA
TTGACCCCAACTCTTCTGAATTGTTCTTAATTGTGACTGATCCAAAAGTGAAGTCAGCAGCTTAATGACTCTCATTATAGATTGCCTATTCTGAG
GAAGGATAGGCAGGATGCAGTAGTCATGTTCTGAATCTTCAACGGTTGGTAAAGAAGTCAGTGAAATTGAAAGACCCCTCATTGAGTTCCCTCAAATTCTA
AGGAATCAGATTGAGTCAAAGCTGACTATGTTCTCCTGGTAATCTTGCTTGTTCATCTGATCTGCTGACTTTACTAAAGCTTAAAGTGTCAAAT
TACTAAATAGTACTTGCCTTAAAGTAGTATTGGTAAATTGTAATTTCAGTTCTAGCTTGGATTATGATGTTATTCGTGACACAATTGCTCT

DNA replication has remarkably low error rate



Inherent error rate of copying enzymes:
1 error per 100,000,000 bases (10^{-8})

And, only 1% of these errors are not fixed
post replication (10^{-2})

Effective error rate of genome replication:
1 error per 10,000,000,000 bases (10^{-10})

Most of the time, our cells copy their
genome perfectly

Nevertheless, errors (mutations) occur,
and they provide the raw material for evolution



Kids have ~100 mutations
relative to their parents

Mutations need not be bad.
They could be neutral or
advantageous.

Mutations arise from other
causes too, e.g. irradiation.

As populations diverge into distinct species,
they have distinct, differentiating genome sequences

ring-tailed cat



kinkajou



raccoon



ringtail	GAGGTCAC	A	CGCATGGTC	A	T	CATCATGGTCAT	T	G	CATTCTGAT	C	T	GCTGGG	G	TGCCCT
raccoon	GAGGTCAC	G	CGCATGGTC	A	T	CATCATGGTCAT	T	G	CATTCTGAT	C	T	GCTGGG	G	TGCCCT
kinkajou	GAGGTCAC	A	CGCATGGTC	G	T	CATCATGGTCAT	C	G	CATTCTGAT	T	T	GCTGGG	T	TGCCCT

DNA sequences can be used to infer evolutionary history

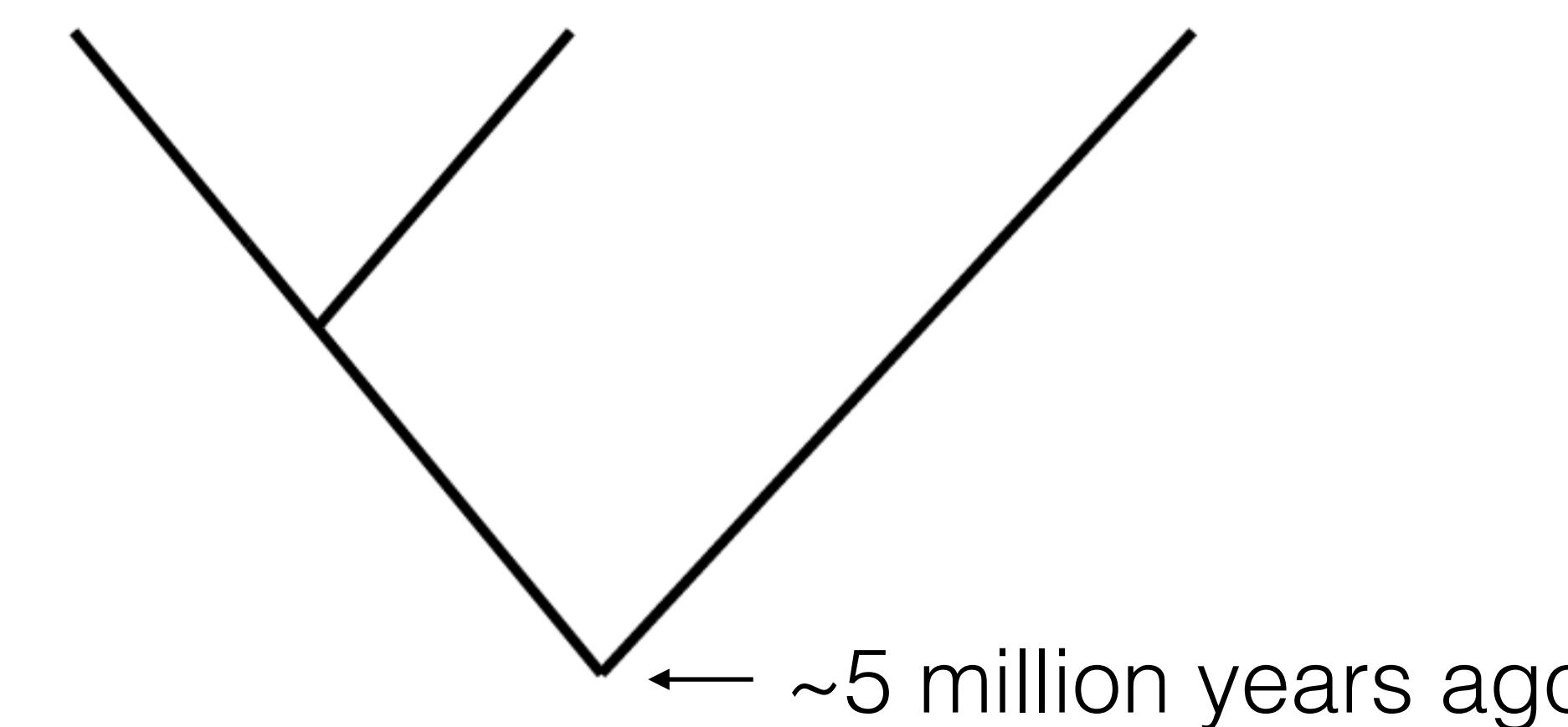
ring-tailed cat



raccoon

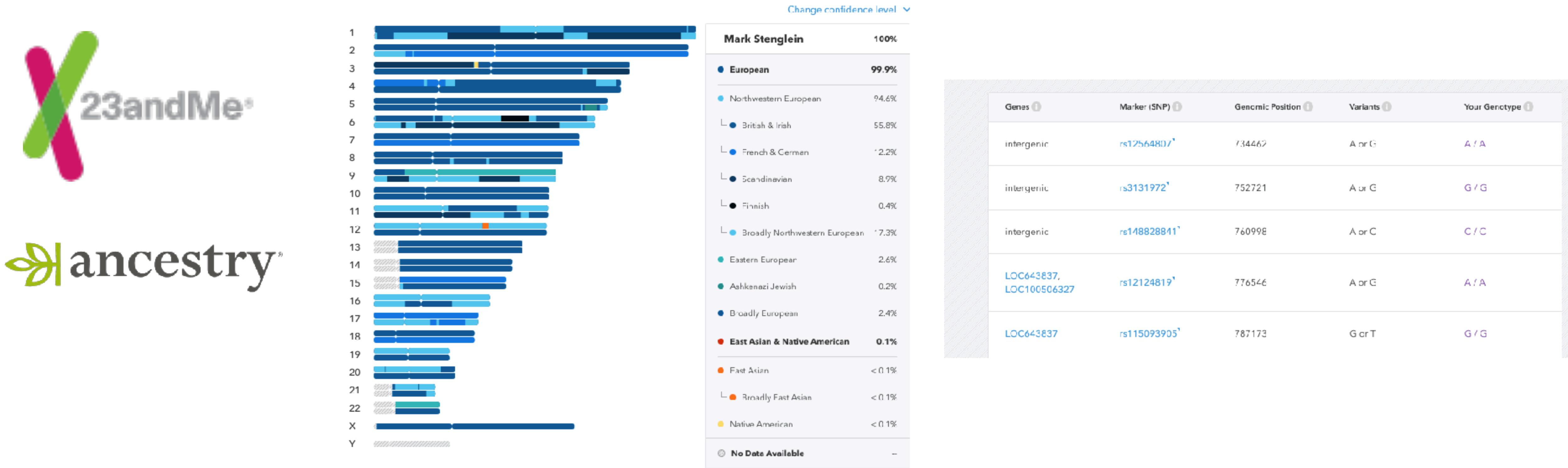


kinkajou



ringtail	GAGGTACAC	A	CGCATGGTC	A	T	CATCATGGTCAT	T	GCATTCTGAT	C	TGCTGG	G	TGCCCT
raccoon	GAGGTACAC	G	CGCATGGTC	A	T	CATCATGGTCAT	T	GCATTCTGAT	C	TGCTGG	G	TGCCCT
kinkajou	GAGGTACAC	A	CGCATGGTC	G	T	CATCATGGTCAT	C	GCATTCTGAT	T	TGCTGG	T	TGCCCT

This works for *populations* of individuals of the same species too



human genomes are on average 99.9% identical
0.1% of 3 billion = 3,000,000 variant sites between any 2 people
~600,000 characterized by 23 and me

Individuals can be identified by their unique genetic signature

The New York Times

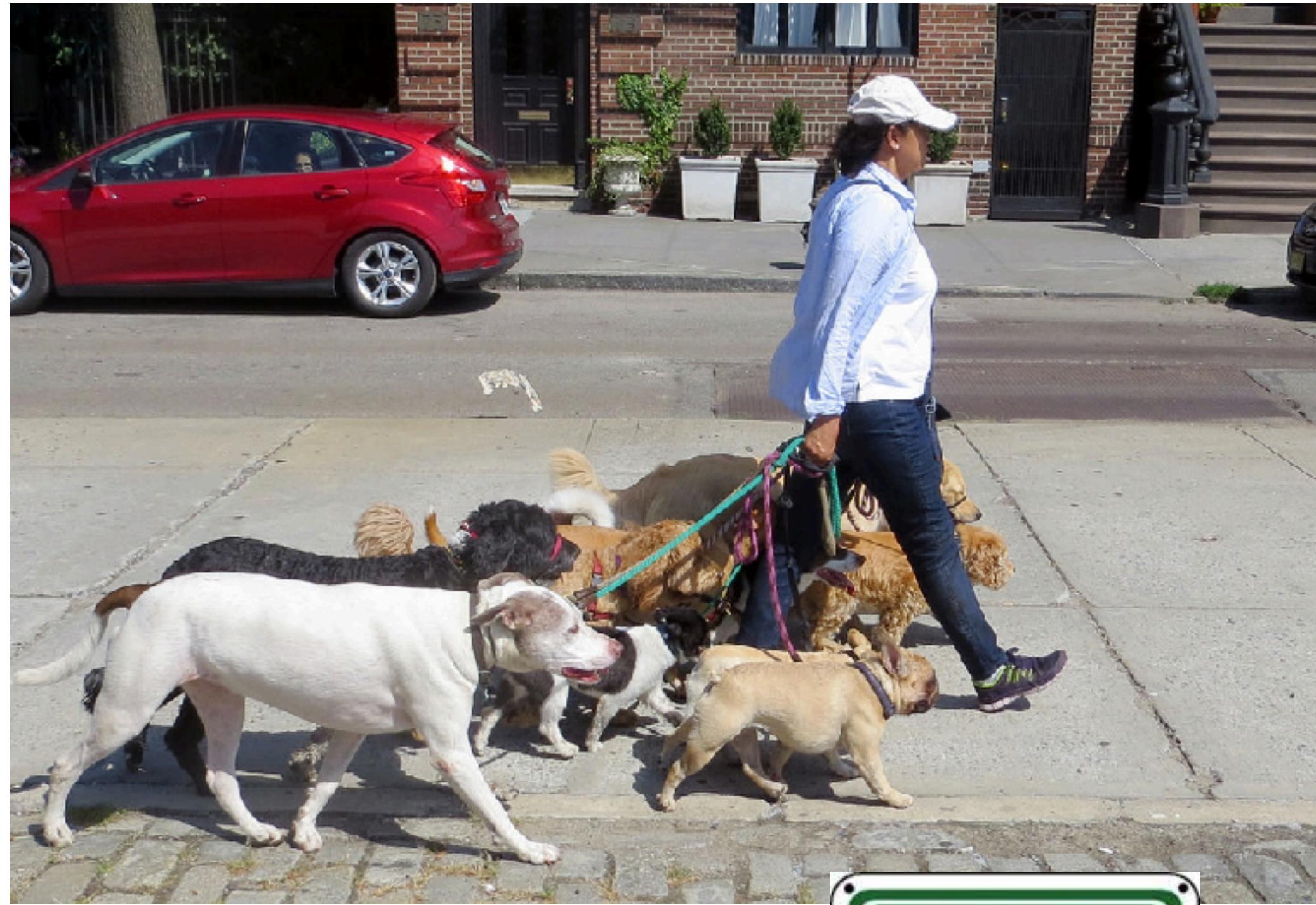
Using DNA to Fight Dog Owners' Discourtesy in Brooklyn

Big City

By GINIA BELLAFANTE JULY 30, 2015



Same idea for DNA forensics in criminal cases



images: Alan Kotok

Researchers have used a similar approach to identify seafood fraud

The New York Times

ENVIRONMENT

Tests Reveal Mislabeling of Fish

By ELISABETH ROSENTHAL MAY 26, 2011



iStockphoto/Illustration by The New York Times



A mako shark fillet, above, is often passed off as swordfish, top. NOAA

images: NY Times / NOAA

Sequencing one piece of DNA is powerful!

Sequencing many pieces of DNA is even more powerful!

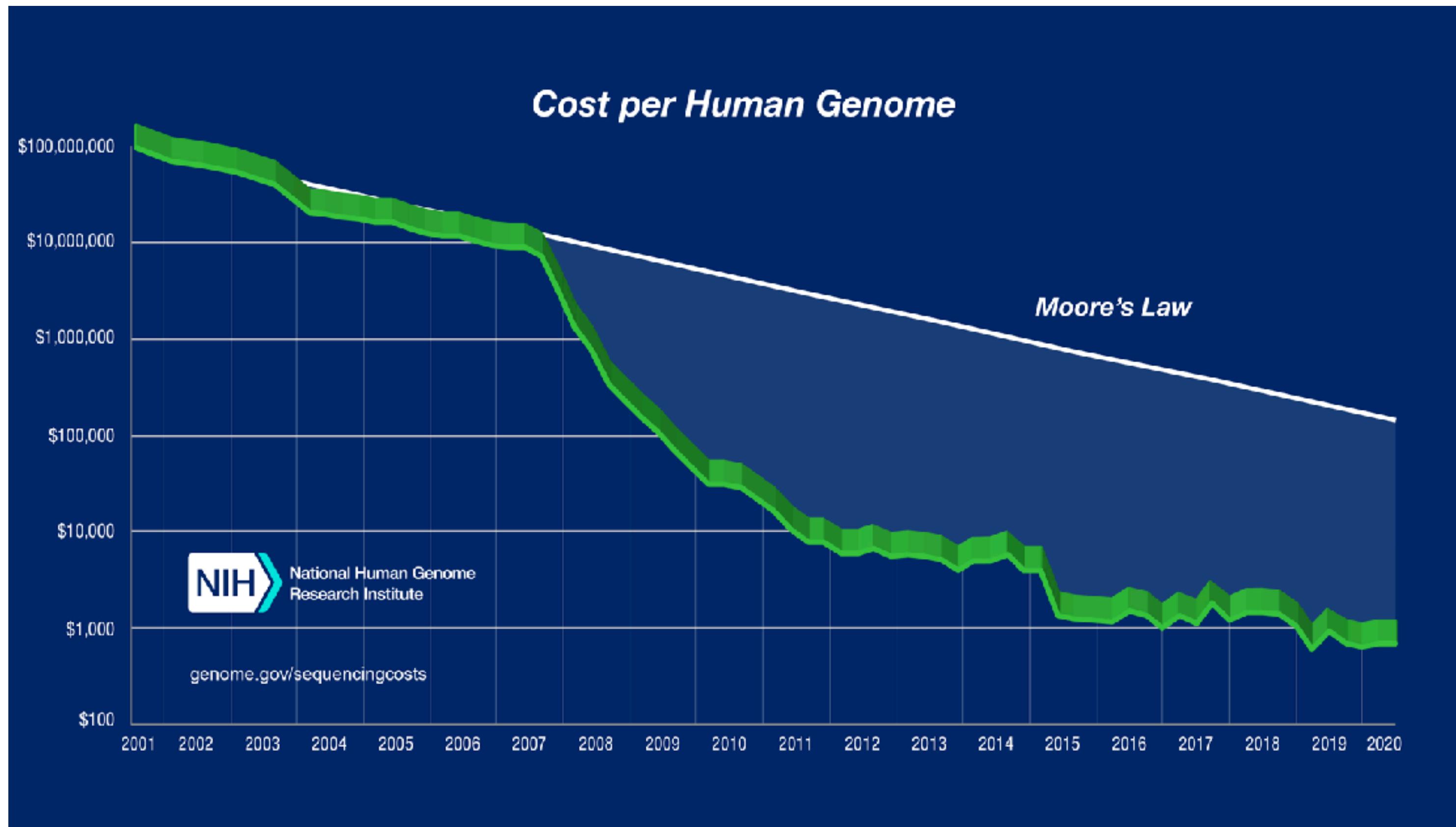
Illumina MiSeq sequencer



Illumina sequencers can generate
100s of millions of sequences in a single run!!

This type of “high throughput sequencing” is not standard in biology research and medicine
It is critical that you have computer skills to be able to analyze and understand this data!

The cost of sequence data generation has fallen 100,000x since 2001!



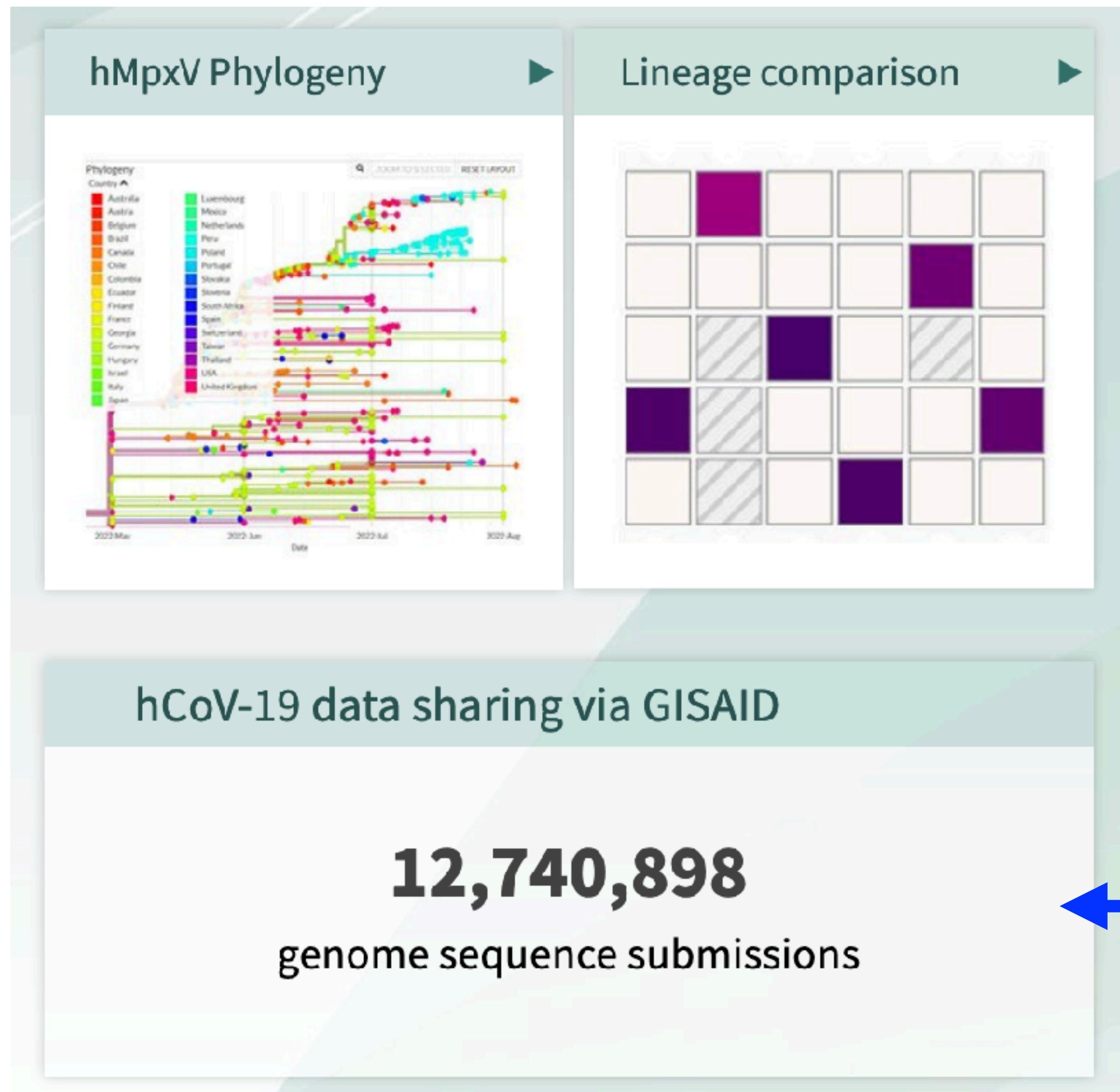
2001 Toyota Corolla MSRP \$14,198



2021 Mercedes S-Class MSRP \$0.14

I wish I could buy a car for \$0.14!!

Over **12 million** SARS-CoV-2 genomes have been sequenced since the pandemic started!

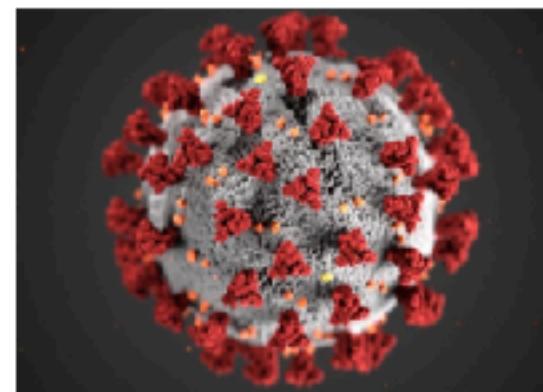


You need computers to analyze these massive amounts of sequence data

Raw DNA sequence data files contain millions of sequences

Hey Life Science Undergraduates! Did you ever wonder...

How do you identify a coronavirus variant?



CDC.gov

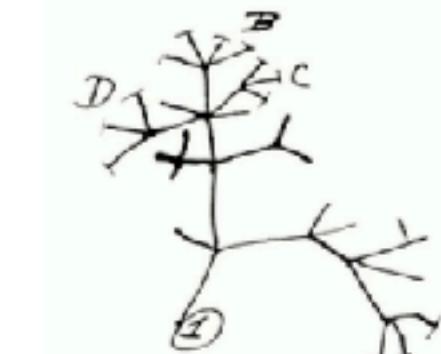
How does BLAST actually work?

Enter Query Sequence
Enter accession number(s), gi(s), or FASTA sequence(s)

What pooped on my neighbor's shoe?



How do you read a phylogenetic tree?

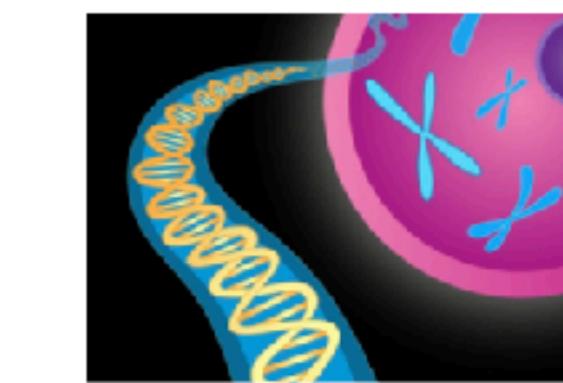


Where did SARS-CoV originate?



"Pangolin Model" by veracous jess is licensed with CC BY-NC-SA 2.0.

How could you create a brand new genome sequence?



National Institutes of Health, CC BY-NC 2.0.

Learn to answer these questions, and more!

Bioinformatics, applied to bacteria and viruses!

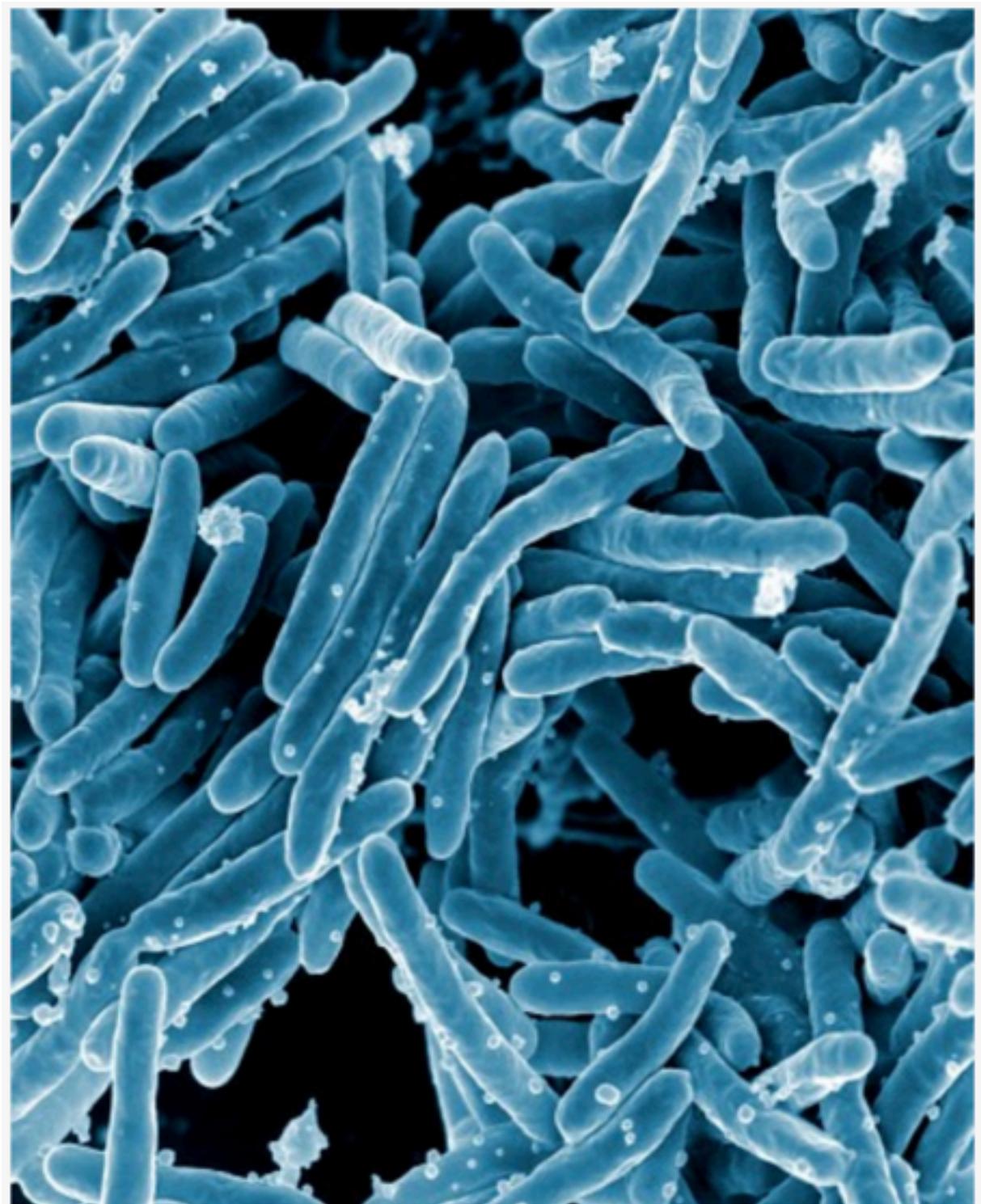
Develop the computational skills that will help you get a job and thrive in the new era of big data in biology!

Microbial Sequence Analysis,
MIP 280A4
Tuesdays Thursdays 9:30-10:45
Fall 2022

For more information
Including prerequisite questions
Contact:
Mark.Stenglein@ColoState.edu

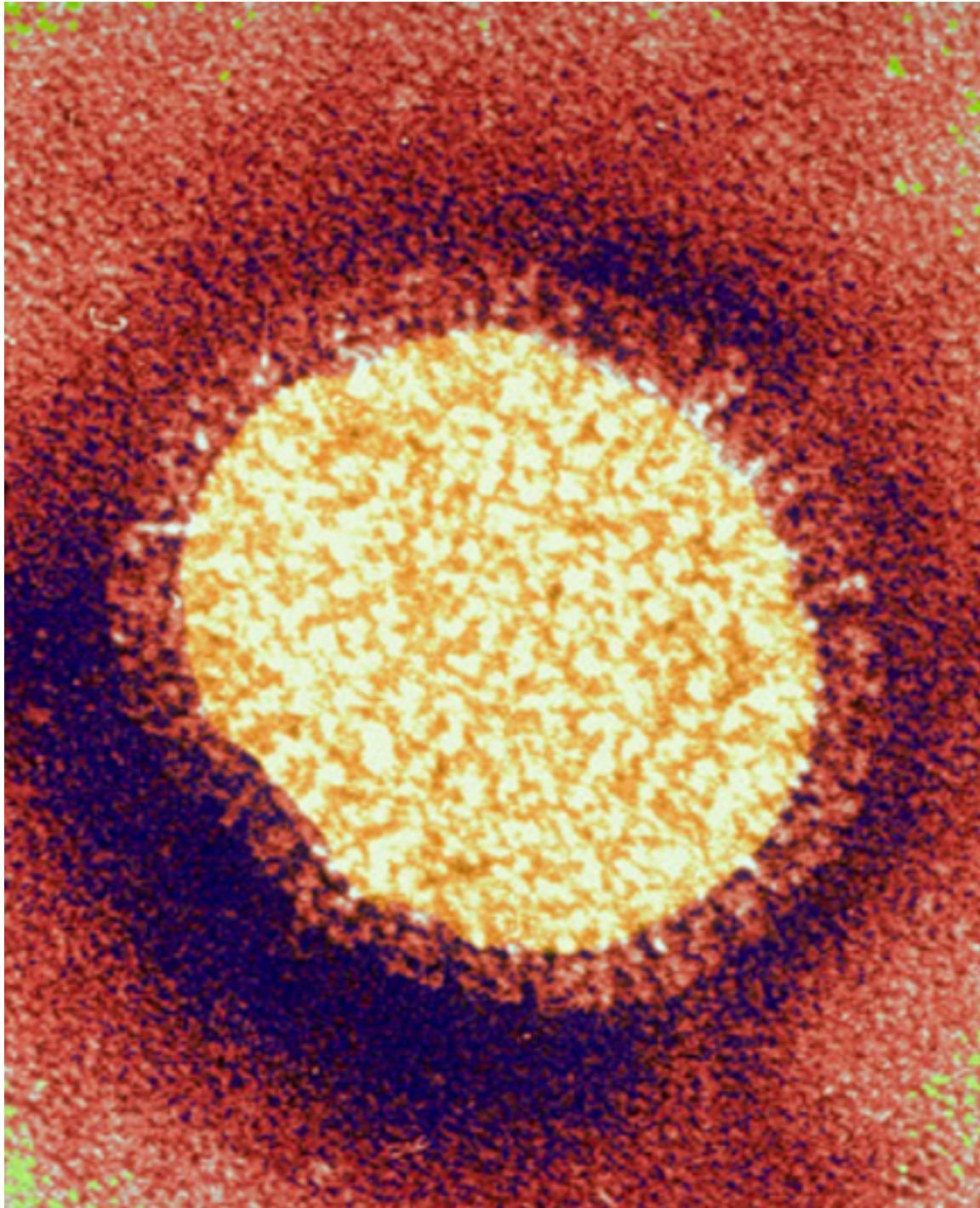
In this class you will learn: “The theory and practice of computational biology applied to bacteria and viruses”

Mycobacterium tuberculosis



["Mycobacterium tuberculosis Bacteria"](#) by [NIAID](#)
is licensed under [CC BY 2.0](#)

Influenza virus



["Influenza virus"](#) by [Sanofi Pasteur](#) is licensed
under [CC BY-NC-ND 2.0](#)

Why viruses and bacteria?

- Most pathogens are viruses and bacteria.
- They have relatively small genomes.
- The lessons learned will be broadly applicable to other organisms, including eukaryotes.
- Many MIP researchers study viruses and bacteria: opportunity to use these skills in an MIP lab.

Schedule of topics we will cover

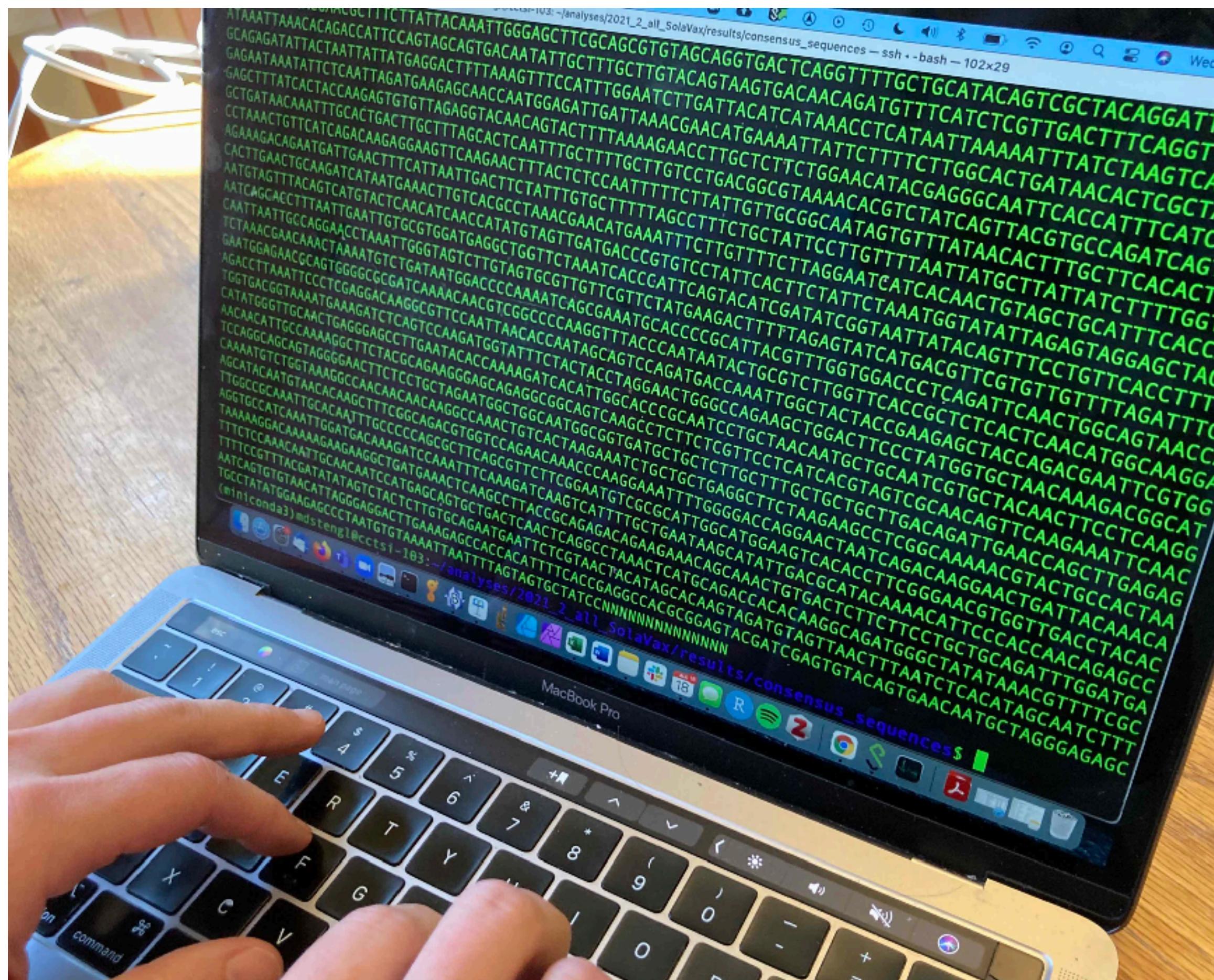
Day	Date (Fall 2022)	Topic
1	8/23	Introduction
2	8/25	Molecular biology review
3	8/30	Genome diversity and structure
4	9/1	Sequence formats and databases
5-6	9/6, 9/8	Pairwise sequence alignments
7	9/13	Alignment-based searches (BLAST)
8	9/15	Multiple sequence alignments
9	9/20	Alignment of entire genomes
10	9/22	Buffer day / work on problem set #1
11	9/27	Phylogenetic tree creation and interpretation Problem set #1 due 9/27
12	9/29	Phylogenetic trees, continued.
13	10/4	Computer architecture and remote computing
14-16	10/6, 10/11, 10/13	Command line computing
17	10/18	Next generation sequencing technology and data
18	10/20	QC and cleaning of NGS data
19-20	10/25, 10/27	Mapping
21	11/1	Buffer day / work on problem set #2
22-23	11/3, 11/8	Assembly Problem set #2 due 11/3
24	11/10	No class
25	11/15	Variant calling
26	11/17	Reproducible research practices
-	11/22, 11/24	Fall break
27-29	11/29, 12/1, 12/6	Final projects preparation
30	12/8	Final projects presentations and final reports due

Learning objectives

- Find, retrieve, and organize various types of microbiological sequence data
- Use basic command line tools
- Contrast typical genomic features of viruses, bacteria, and eukaryotes
- Perform local and global sequence alignments manually and using a computer.
- Perform a BLAST search to identify the most closely related database sequence and critique the credibility and usefulness of the results.
- Differentiate between categories and use-cases of sequence alignment algorithms.
- Map reads to a genome, generate downstream metrics, and evaluate mapping quality.
- Create a genome assembly and evaluate its quality.
- Generate variant calls relative to a reference sequence and determine their functional impact.
- Assess the quality of publicly available sequence data and genome assemblies.
- Differentiate between computational approaches designed for use with an existing reference genome and approaches appropriate for a not-yet sequenced organism.
- Generate and interpret a phylogenetic tree.
- Illustrate how commonly used bioinformatics tools and techniques can produce incorrect or misleading results if not applied correctly. Illustrate common pitfalls of commonly-used bioinformatics tools.
- Devise appropriate positive and negative controls for use in computational experiments.
- Implement a reproducible computational workflow.
- In a final project, integrate knowledge about mapping, assembly, and sequence search to design a computational workflow that will answer a specific question using real sequencing data. An example of such a project would perform a metagenomic characterization of a publicly available sequence dataset.

This class will emphasize active learning

Computer based



Not computer based



Grading breakdown

In-class exercises <i>Written answers due at 11:59 PM the day of the following class</i>	40%
(In-class) Participation	10%
Take home problem set #1 <i>Due 9/27/2022</i>	10%
Take home problem set #2 <i>Due 11/3/2022</i>	10%
Final project report and presentation <i>Due 12/8/2022</i>	30%

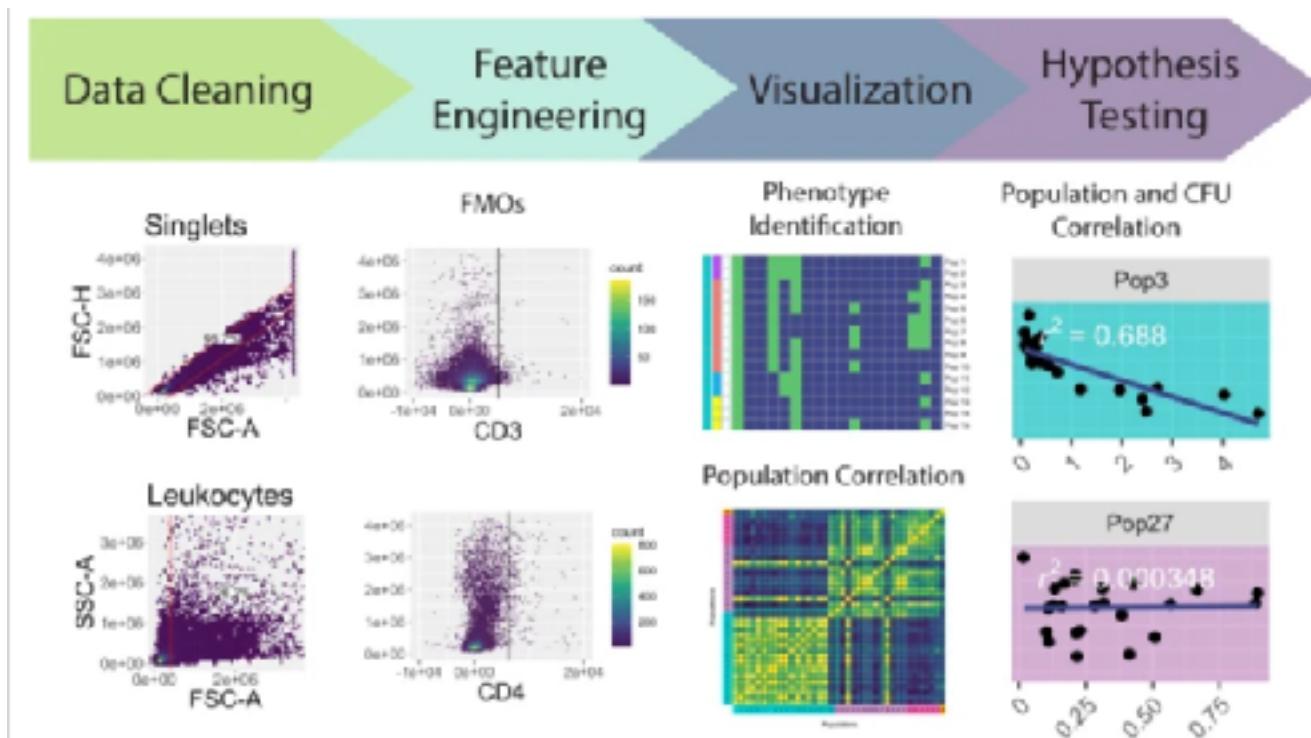
Grading Scheme

A	< 100% to 90%
B	< 90% to 80%
C	< 80% to 70%
D	< 70% to 60%
F	< 60% to 0%

Other types of biological “big data” require computers for analysis

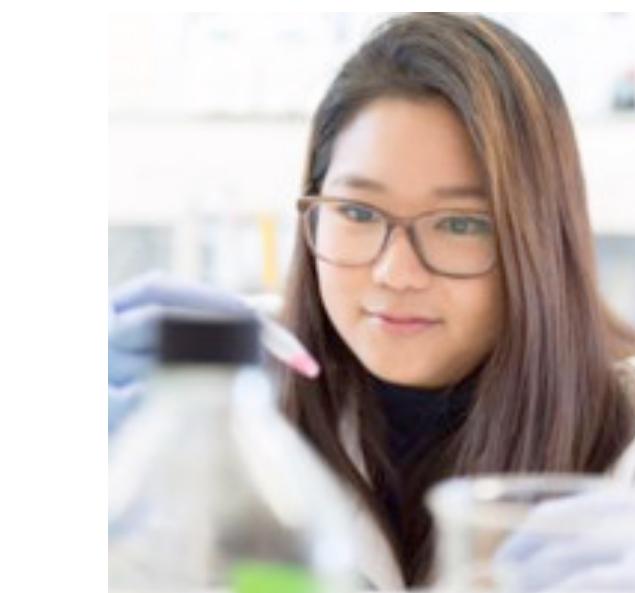
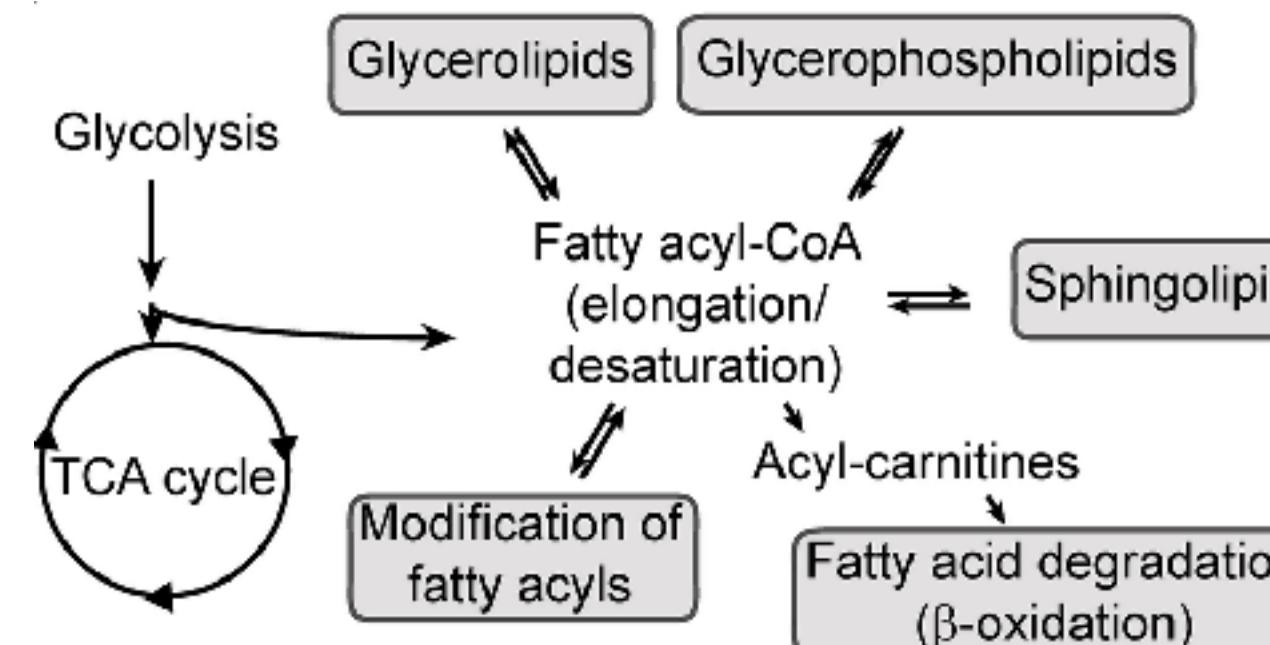
Flow Cytometry

Amy Fox / Marcela Henao Tamayo
*Pipeline to identify and characterize
immune cell populations*



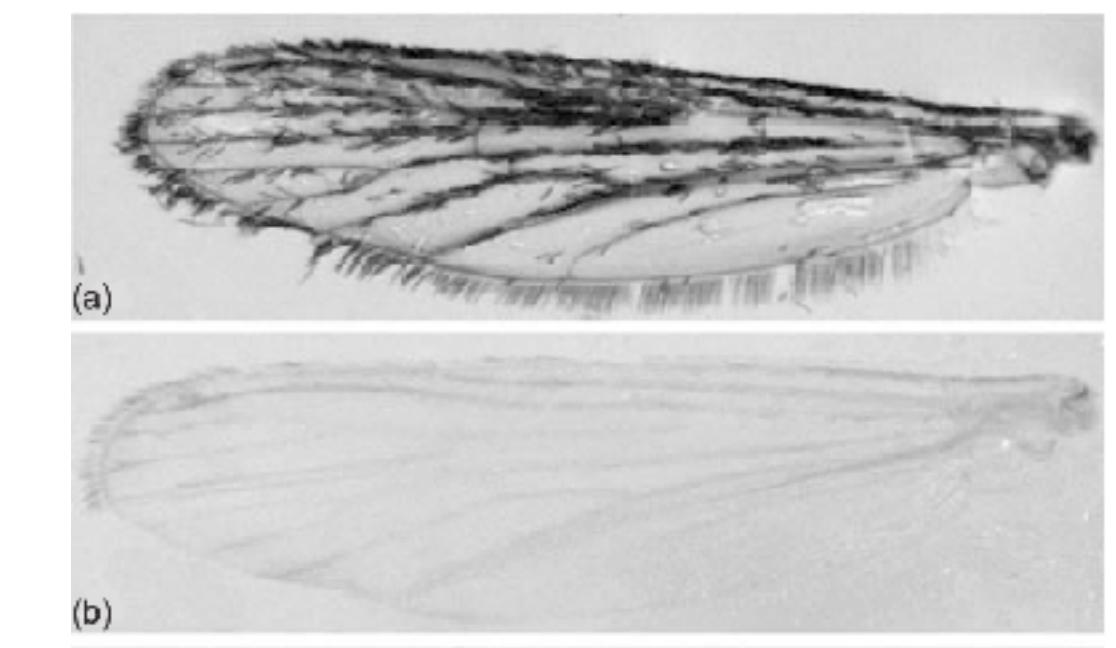
Proteomics and Metabolomics

Nunya Chotiwan / Rushika Perera
*Metabolomics of dengue-infected
mosquitoes*



Machine Learning and Image Analysis

Lyndsey Gray / Brian Foy
*ML image analysis to
determine mosquito age*



The computer skills you will learn will be broadly useful in many careers
Including fast-growing, high-paying careers

Plus, computing is fun!



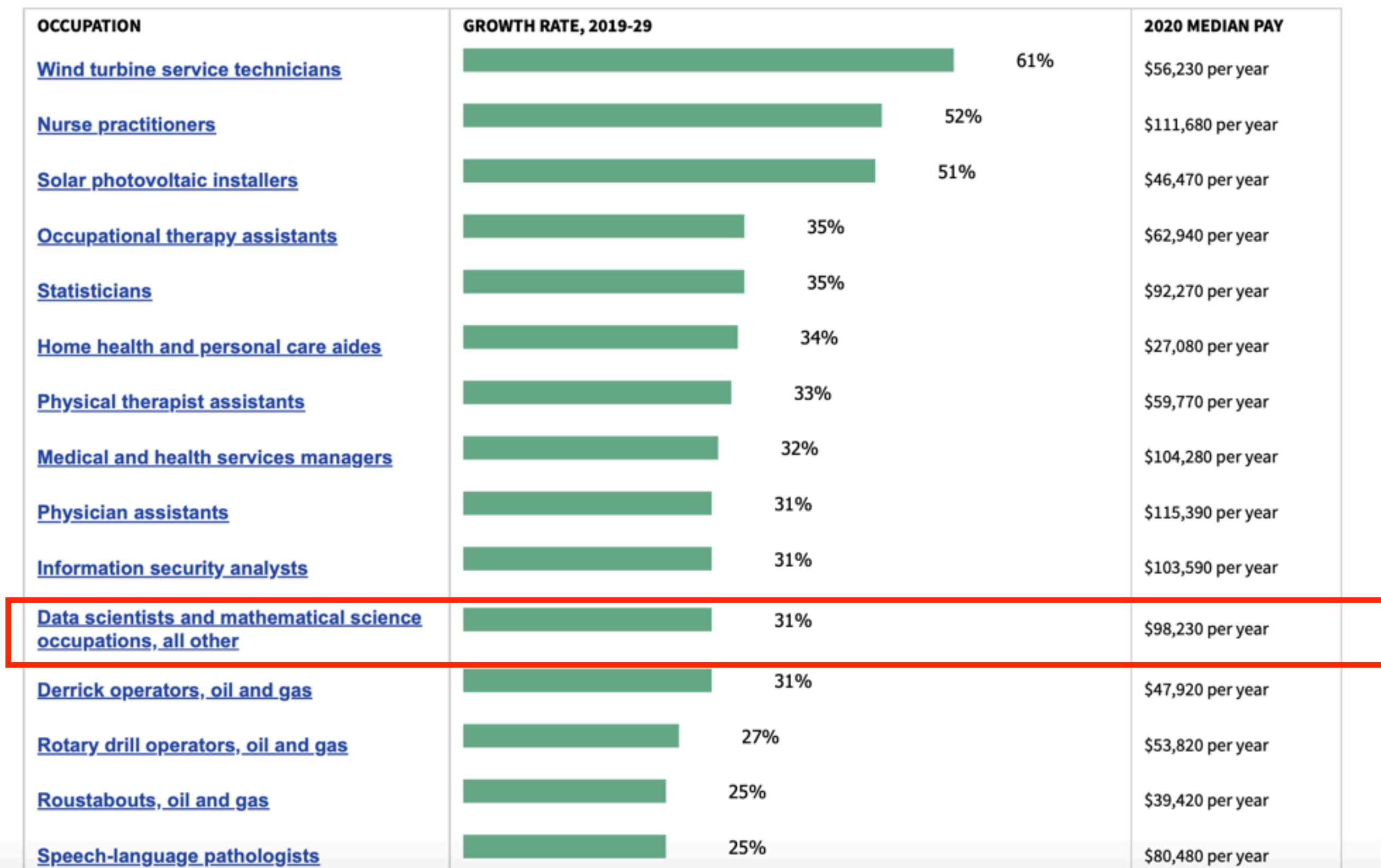
Data scientist jobs growing annually at 31%

Fastest Growing Occupations

PRINTER-FRIENDLY

Fastest growing occupations: 20 occupations with the highest percent change of employment between 2019-29.

Click on an occupation name to see the full occupational profile.



Source: <https://www.bls.gov/ooh/fastest-growing.htm#>

MIP graduates are landing sweet jobs thanks to their computational skills



Amy Fox (She/Her) · 2nd
Principal Risk Specialist at Capital One
Houston, Texas, United States · [Contact info](#)



Steven Lakin · 2nd
Veterinary Medical Officer at USDA National Bio and Agro-Defense Facility (NBAF)
Manhattan, Kansas, United States · [Contact info](#)



Bryce Asay (He/Him) · 2nd
Founder-CEO at Viden Technologies L.L.C.
Fort Collins, Colorado, United States · [Contact info](#)

A hot market for data scientists means starting salaries of \$125K and up this year

BY DAWN RZENIKIEWICZ
February 24, 2022, 7:22 AM

It's a great time to have a degree in data science thanks to a hot job market for these roles. Data scientists made a median salary of \$164,500 in 2020, according to a 2021 survey of engineering professionals by the Institute of Electrical and Electronics Engineers (IEEE). That's a nearly 8% increase from the \$152,500 median salary in 2019.



Andrew Lane (He/Him) · 1st
Machine Learning Engineer at Cash App
San Francisco, California, United States · [Contact info](#)
