

What microorganisms live in Bloomington lakes/ponds? (and what do they do?)

Gene sequencing and CS/DS module

Luddy summer camp, 2022

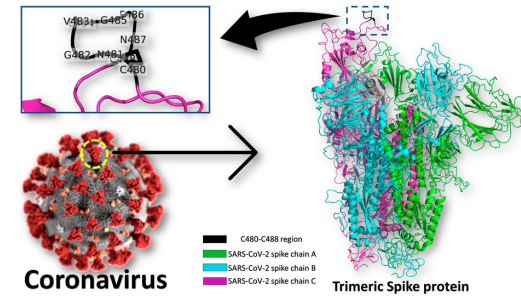
Yuzhen Ye, Computer Science Department

DNA and DNA sequencing

SARs-Cov-2 genome

A long string of about 30 thousands of characters in the alphabet of ATCG

ATTAAAGGTTTATACCTTCC....



Viruses constantly change through mutation

SARs-Cov-2 variants:

B.1.1.7 (Alpha), B.1.351 (Beta), ..., and B.1.617.2 (Delta), Omicron etc

Omicron

- a large number of mutations (> 50 mutations)
- many of the mutations are in the genes that encode spike protein
- the spike protein is where the virus attaches human cells;
- the spike protein is also recognized by human immune system recognizes

(SARS-CoV-2 is single-stranded RNA virus)

Several generations of DNA sequencers

First Generation



Sanger Sequencing
Maxam and Gilbert
Sanger Chain-termination

- Infer nucleotide identity using dNTPs then visualize with electrophoresis
- 500-1000 bp fragments

Second Generation (Next Generation Sequencing)



454, Solexa,
Ion Torrent
Illumina

- High throughput from the parallelization of sequencing reactions
- ~50-500 bp fragments

Third Generation



PacBio
Oxford Nanopore

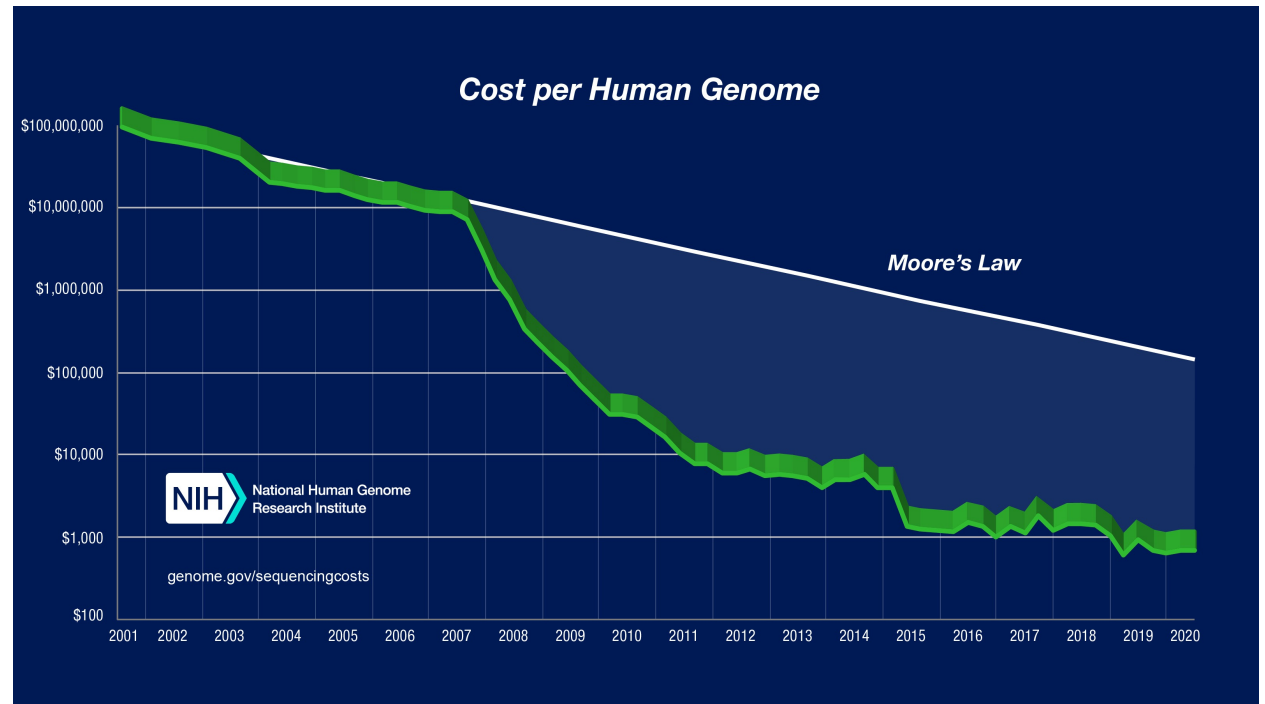
- Sequence native DNA in real time with single-molecule resolution
- Tens of kb fragments, on average

Short-read sequencing

Long-read sequencing

DNA sequencing costs

Human genome: a complete set of nucleic acid sequences for humans, encoded as DNA within 23 chromosome pairs.



TTCCGGAGAGGGAGCCTGAGAAATGGCTACCACATCCACGGAGAGG

GCCTGAGAAATGGCTACCACATC

CCACATCCACGGAGAGG

TTCCGGAGAGGGAGCCTGAG

Human second genome

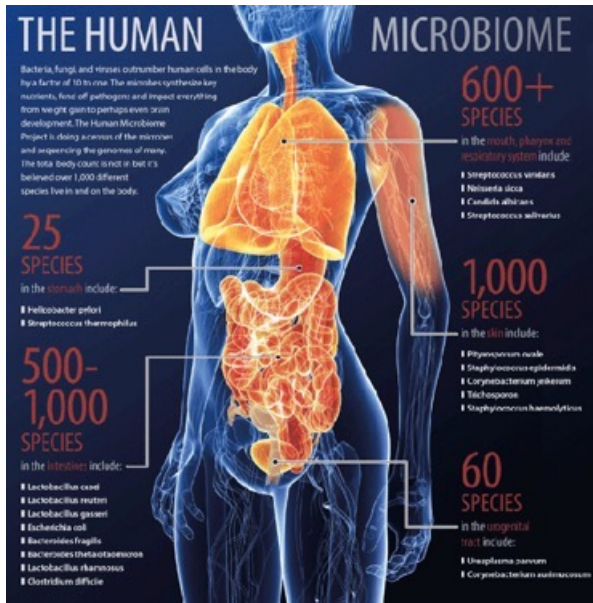


image credit: <https://i0.wp.com/allergiesandyourgut.com/wp-content/uploads/2015/06/gut-microbiome.gif>

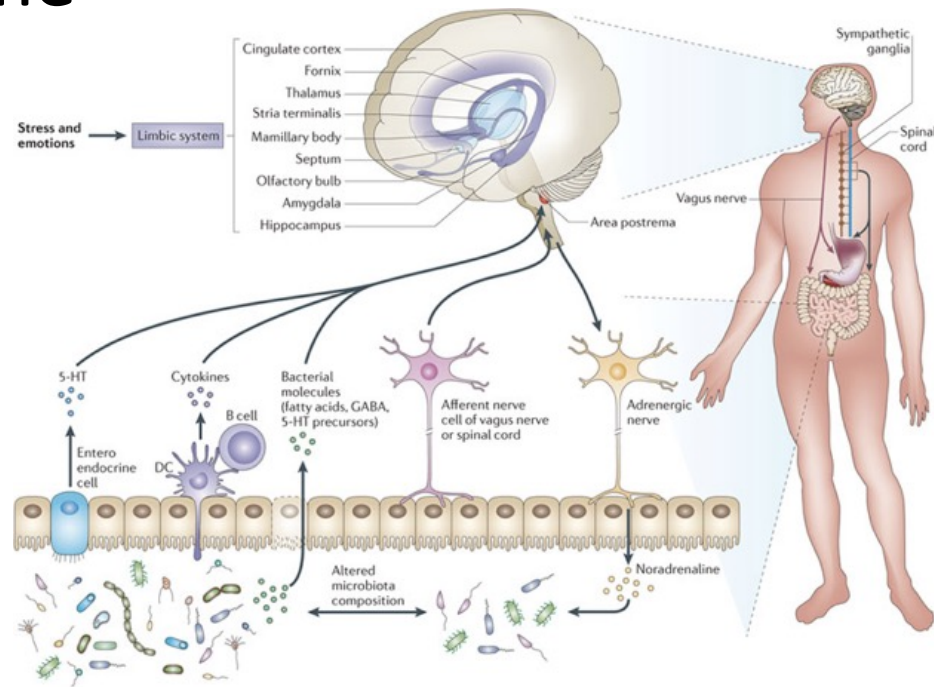


image credit: Nature Reviews Microbiology 10, 735-742

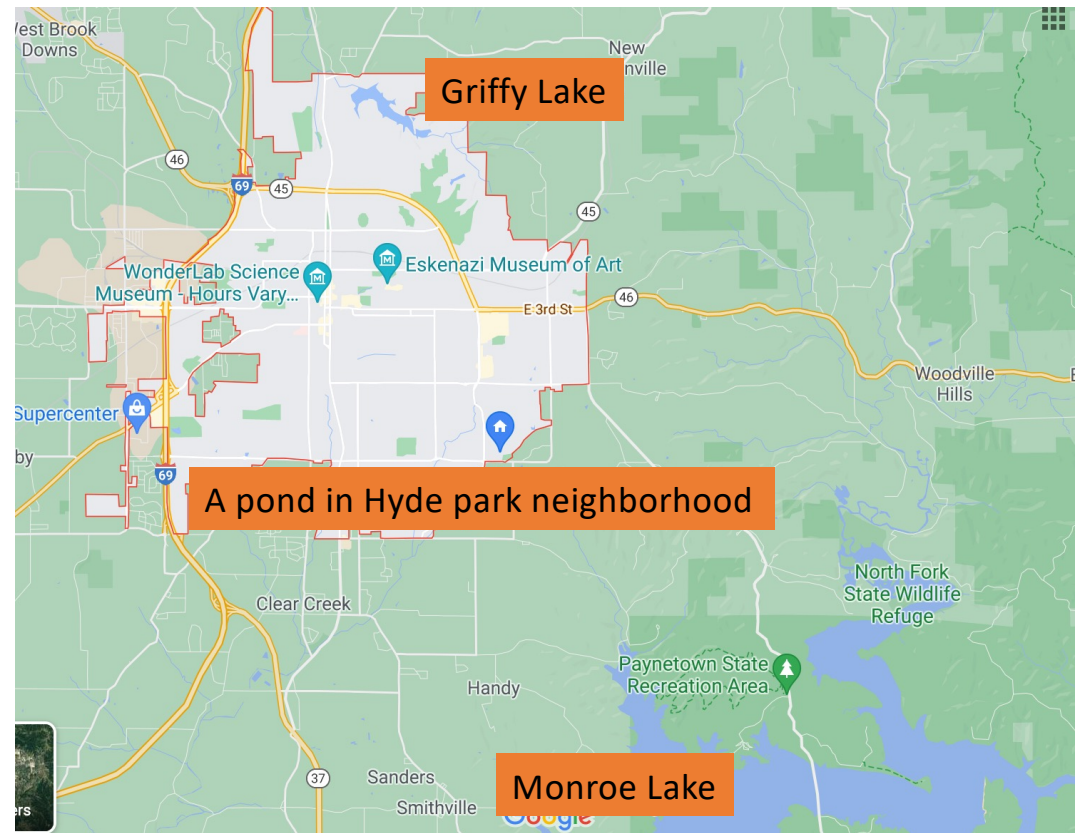
Nature Reviews | Microbiology

Microbes play many important roles in health of humans, animals, and the environment.

Microbiome: A characteristic community of different microorganisms that populate a particular environment (e.g., the human small intestine or Griffy Lake).

Human microbiome (human second genome): the set of genomes carried by the human microbiota.

Our Bloomington lake/pond microbiome project -- from sample collecting to data analysis



Lots of DNA sequences..
...CGGCGGCAGAAAATACATTTTCAACTTTCAAAGAAATGAA...

Tasks:

- Inference of spike protein from DNA sequence
- Exploratory analysis of the Bloomington water microbiome