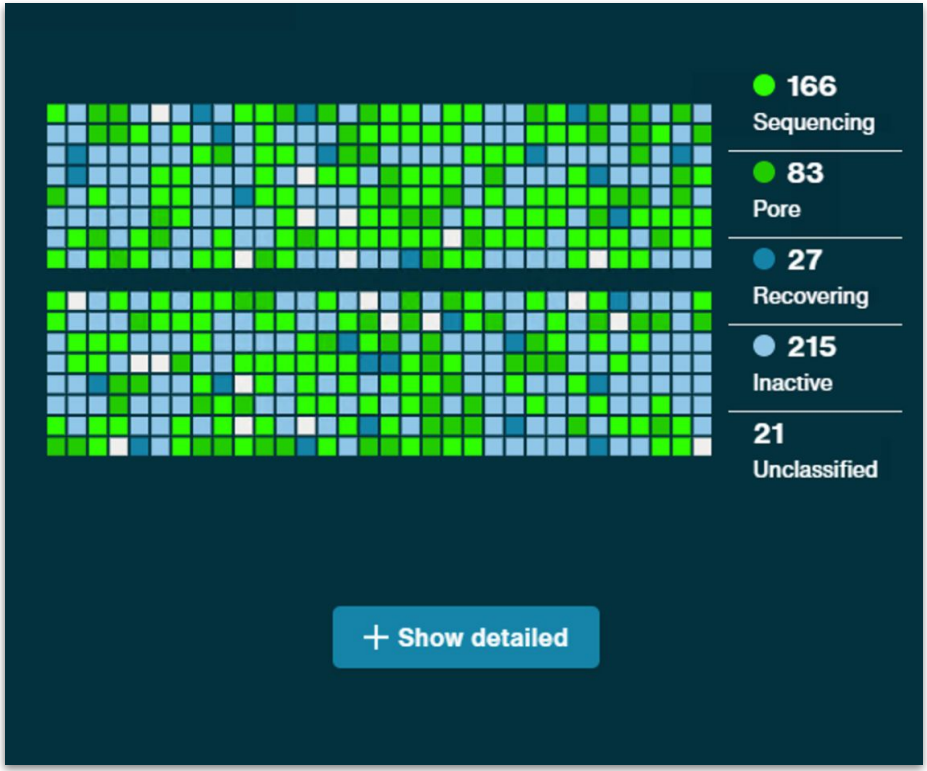



# From Nano to Life

Interpreting graphs from Nanopore sequencing of bacterial genomes

By Carlos C. Goller, Jason Williams, & Carly Sjogren

# Information for instructors

Department/Program: Biotechnology Program

Level: Introductory

Course type: Lab and online asynchronous and synchronous sessions.

Delivery mode: In-person and online

Students: Biology majors and non-majors

Number of students: max enrollment = 24

Estimated duration of activity: 50-75 minutes

# Learning Objectives

- **Identify** specific components of graphs from Oxford Nanopore Technologies sequencing output.
- **Reflect** on your perceptions about using graphs or figures in biology.
- In 3-5 sentences including specific details about the graph, **describe** the major findings and limitations of the Nanopore sequencing dataset depicted.
- **Phrase** a testable hypothesis based on the use of Nanopore sequencing for bacterial genome sequencing.
- Using miro.com, Google Slides, or Jamboard, **design** a simple flowchart describing a future experiment.
- **Reflect** on your perceptions of people who do biology.
- **Compare** your own interests and/or identities to those of people who do biology.

# Outline

1. Rare Earth Elements and Electronic Waste Recycling
2. Bacterial and Rare Earth Elements
3. Overview of Nanopore Sequencing
4. Sample Preparation
5. Interview: Jason Williams
6. Graph Analyses
7. Brainstorming Approaches

# Poll Question 1

You have isolated a microorganism from a unique environment: a heavily contaminated soil. The microbe seems to be capable of recovering commercially-relevant rare earth elements from the soil. You read about rare earth elements: Rare Earths Statistics and Information | U.S. Geological Survey.

Rare earth elements are… [select all that apply]

A. Rare
B. Abundant
C. Used in a variety of applications
D. A group of seventeen metals

**5 min**

# Poll Question 2

Where is this company recovering rare-earth elements from?

A. Trash in landfills
B. Compost
C. Hard drives
D. Computer screens

# Poll Question 3

What are the processes that are needed to recover rare earth elements?

A. Mechanical
B. Chemical
C. Mechanical and Chemical

# Poll Question 4

How does Dr. Schmitz use bacteria to extract rare earth elements?

**A.** High pressures
**B.** Bioleaching
**C.** Harsh chemicals

Video link: [Rising Women Innovator Award - Alexa Schmitz, Ph.D.] (2 min)

# Poll Question 5

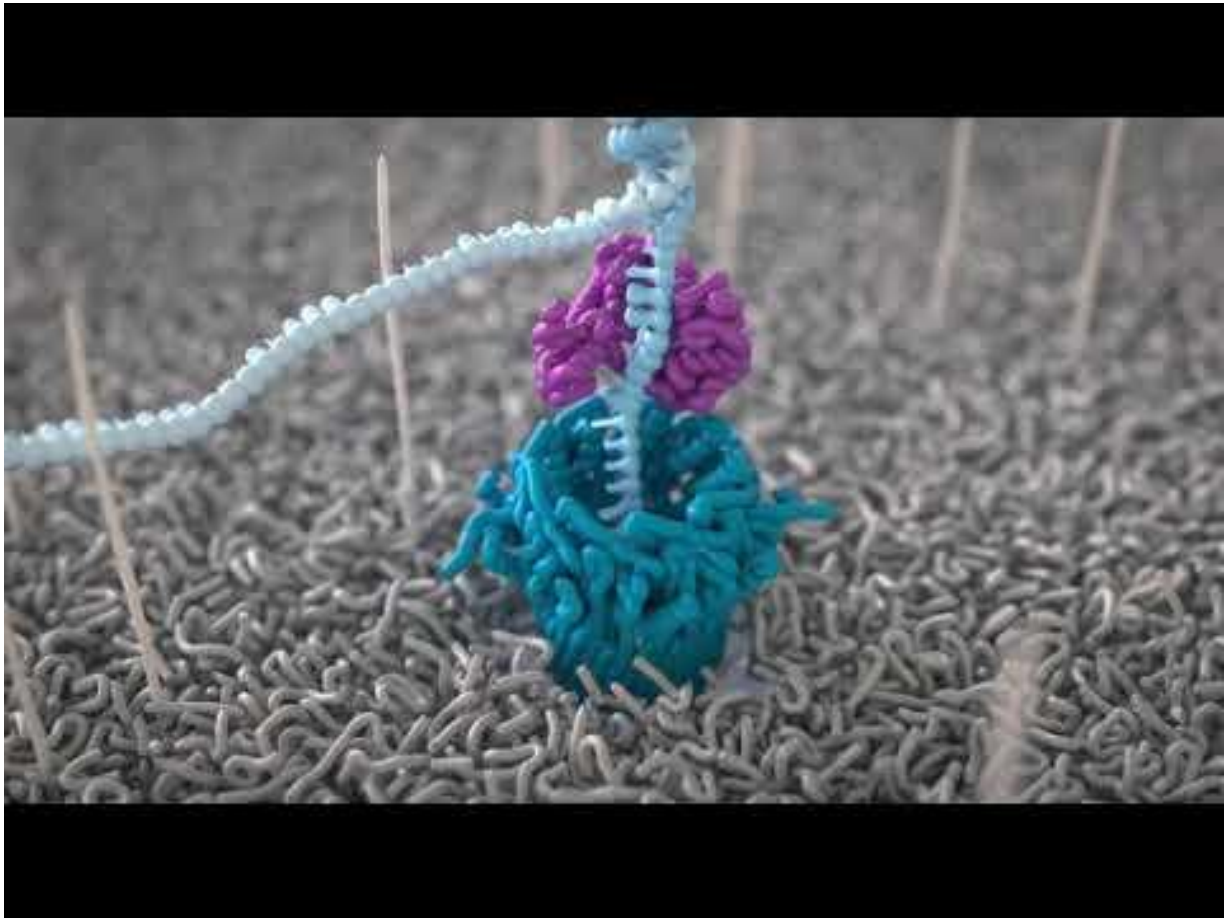
Why are they knocking out genes?

**A.** Determine genes involved in REE recovery.
**B.** Reduce harmful growth of bacteria
**C.** Make bacteria bigger

Video link: [Rising Women Innovator Award - Alexa Schmitz, Ph.D.] (5 min)

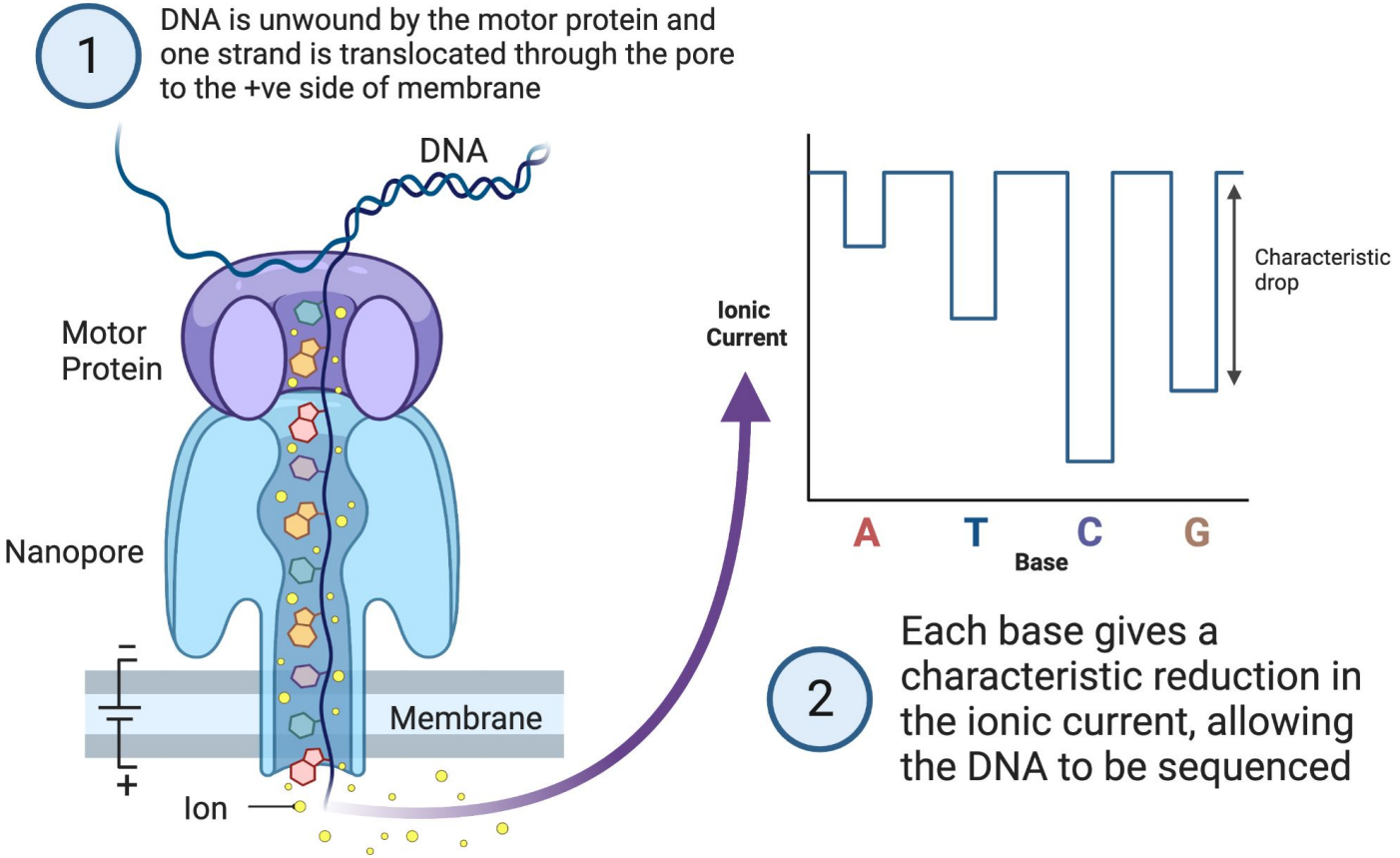
# How can we determine gene sequences?

What is nanopore sequencing?



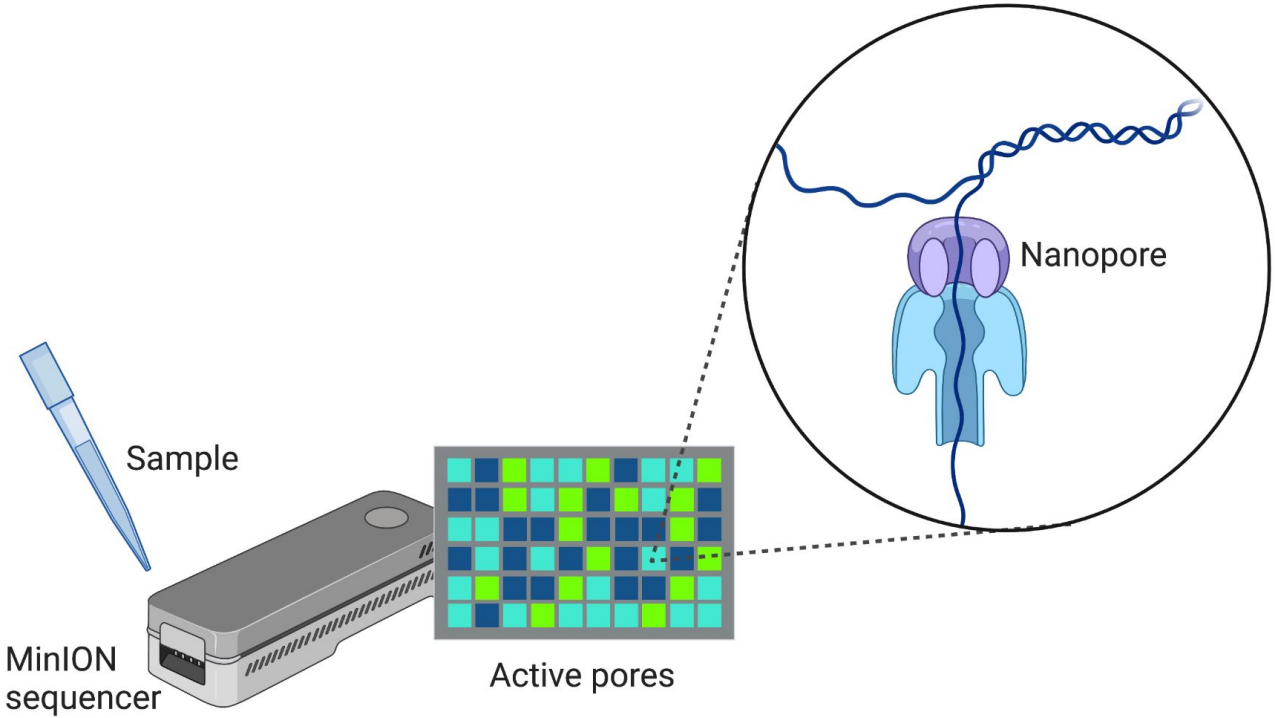
Video link: How nanopore sequencing works (2 min)

# Overview of Nanopore Sequencing




Created with BioRender.com

# Applications




Created with BioRender.com

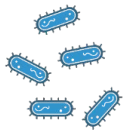
AGTCCCTGAATCGA

# Oxford Nanopore Tech (ONT) Applications

- DNA sequencing
  - Genomes
  - Metagenomes
  - Exomes
  - cDNA
- Direct RNA sequencing
- Methylation analyses
- Multiplexing by PCR or native barcodes

What could **you** use it for?

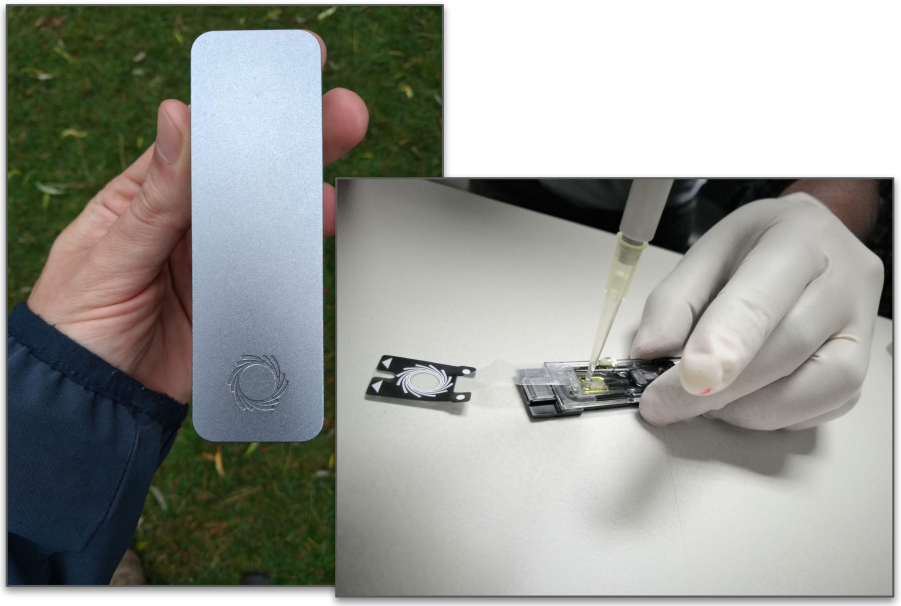
**Share**



AGTCCCTGAATCGA

# You don’t need to purchase much equipment…

- MinION
  - Mk1B - USB powered
  - Mk1C - Screen & computer
- Flongle Adapter
- Flow cells (~50 Gb)
  - R9.4 - variety of kits
  - R10.4 - “Q20+” chemistry
  - R10.**x**…
  - Flongles - ~$90 (2 Gb)
- Voltrax and Miro (Miroculus)



**Left:** MinION Mk1B from Cirosantilli2
**Right:** Loading a rapid barcoded samples onto a Flongle flow cell from WorldFish

AGTCCCTGAATCGA

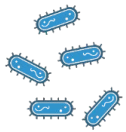
# You can do so much with the devices

- **Genomic DNA** needs to be purified…
  - Enzymatic and mechanical lysis
  - Isolation of high-molecular weight DNA
  - Zymo and New England BioLabs kits and offer educational discounts
- DNA needs to be quantified accurately
  - $A_{260}/A_{280}$ ratios
  - If possible, fluorometric quantification (*e.g.*, Qubit)
  - Sizing (electrophoresis)

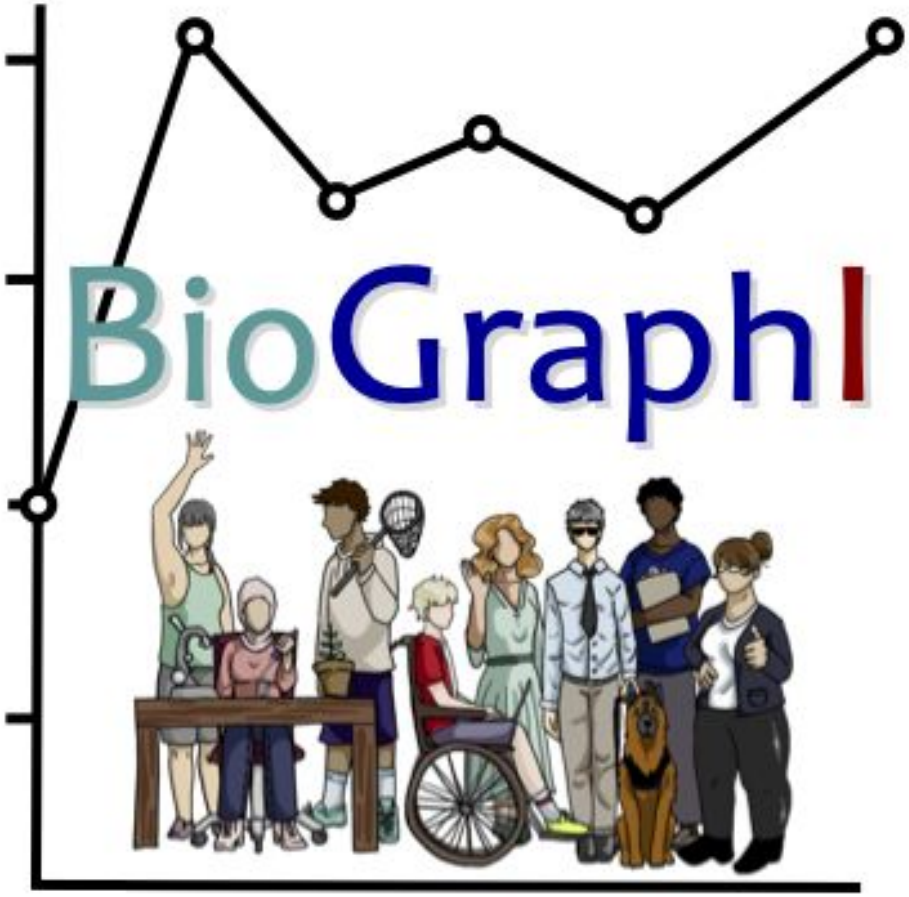
➔ **You will sequence your samples!**

# What is the output?

- Files
- Graphs
- More files



# Interview with Jason Williams



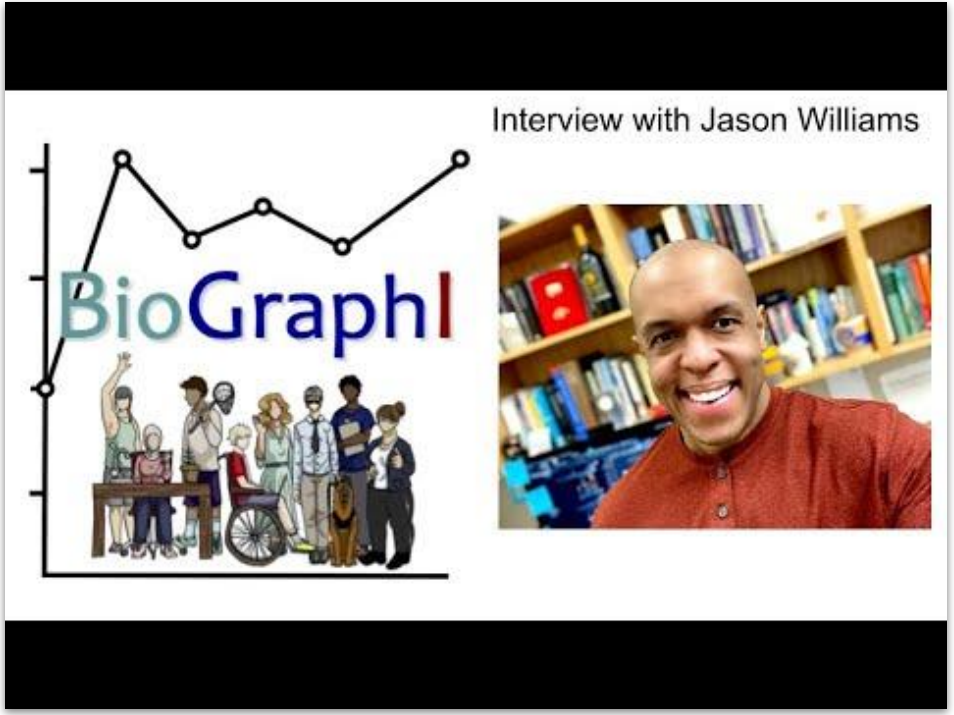
# Meet the scientist who makes these graphs: Jason Williams

**Jason Williams**

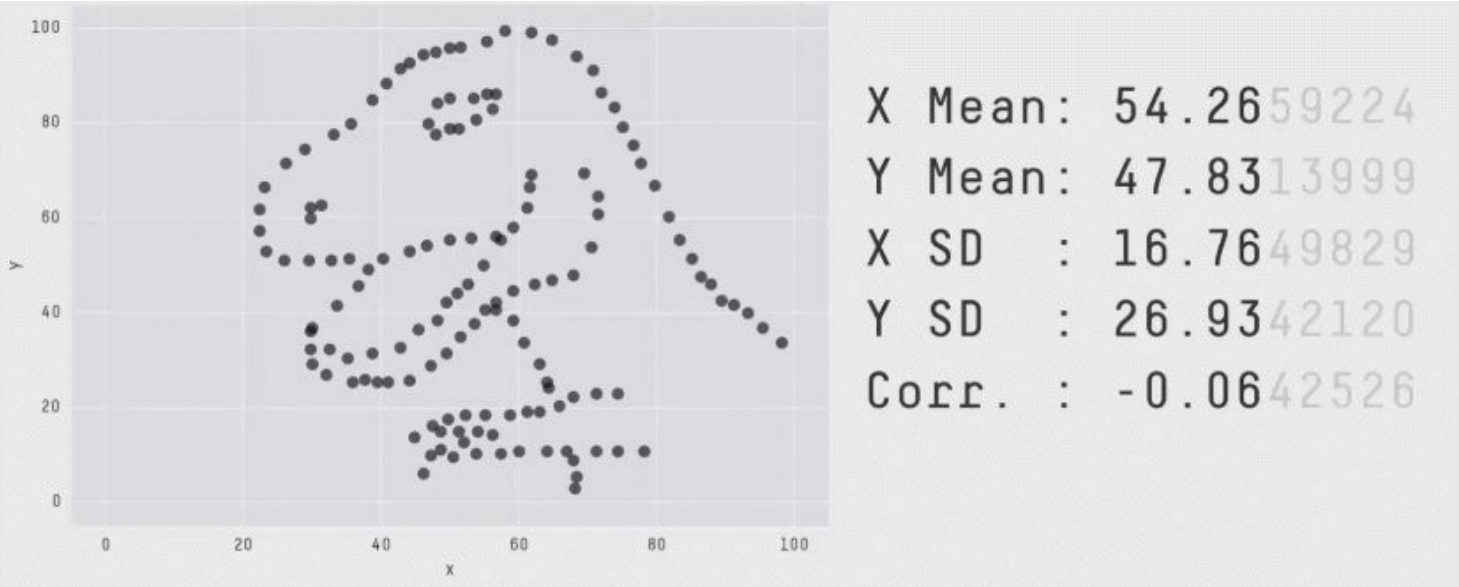
Assistant Director, Diversity and Research Readiness

DNA Learning Center,

Cold Spring Harbor Laboratory

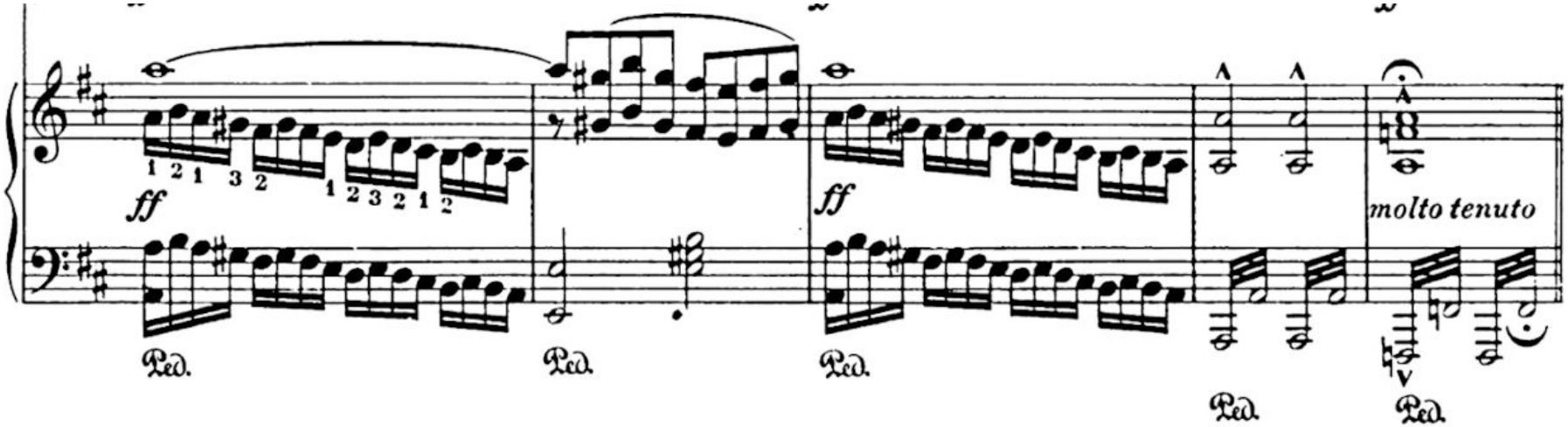


# The power of “visualization” (communicating patterns)




Gif credit: [Download the Datasaurus: Never trust summary statistics alone; always visualize your data]

# The power of “visualization” (communicating patterns) #2



# Quality control of sequence data

## Basic Statistics

| Measure | Value |
|---|---|
| Filename | good_sequence_short.txt |
| File type | Conventional base calls |
| Encoding | Illumina 1.5 |
| Total Sequences | 250000 |
| Sequences flagged as poor quality | 0 |
| Sequence length | 40 |
| %GC | 45 |

## Sequence Duplication Levels

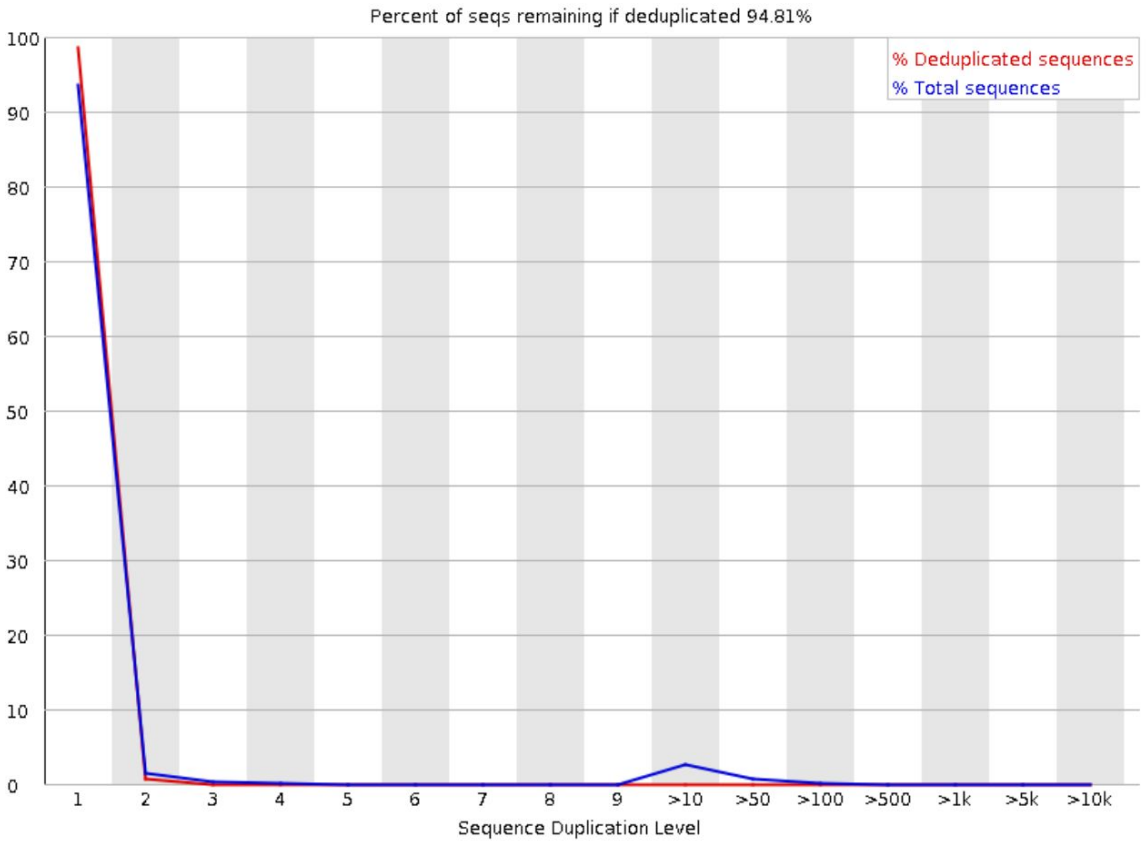



Photo credit:
https://www.bioinformatics.babraham.ac.uk/projects/fastqc/bad_sequ
ml#M8

# Quality control of sequence data 2

## FastQC Report

### Summary

- Basic Statistics
- Per base sequence quality
- Per tile sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content

### Basic Statistics

| Measure | Value |
|---|---|
| Filename | bad_sequence.txt |
| File type | Conventional base calls |
| Encoding | Illumina 1.5 |
| Total Sequences | 395288 |
| Sequences flagged as poor quality | 0 |
| Sequence length | 40 |
| %GC | 47 |

### Sequence Duplication Levels

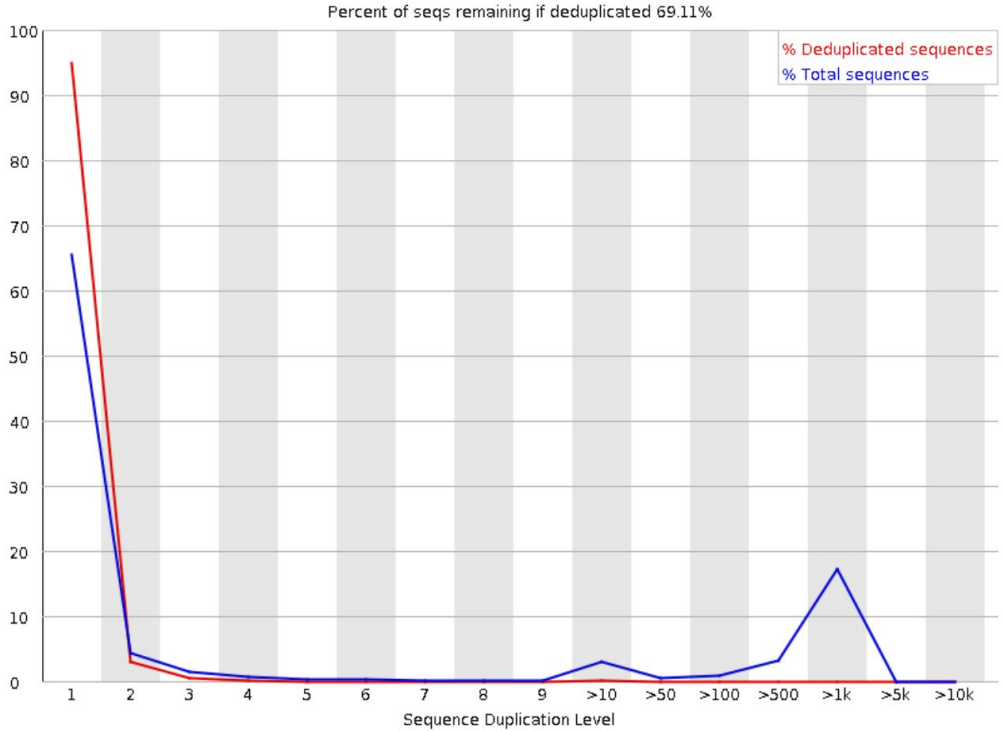
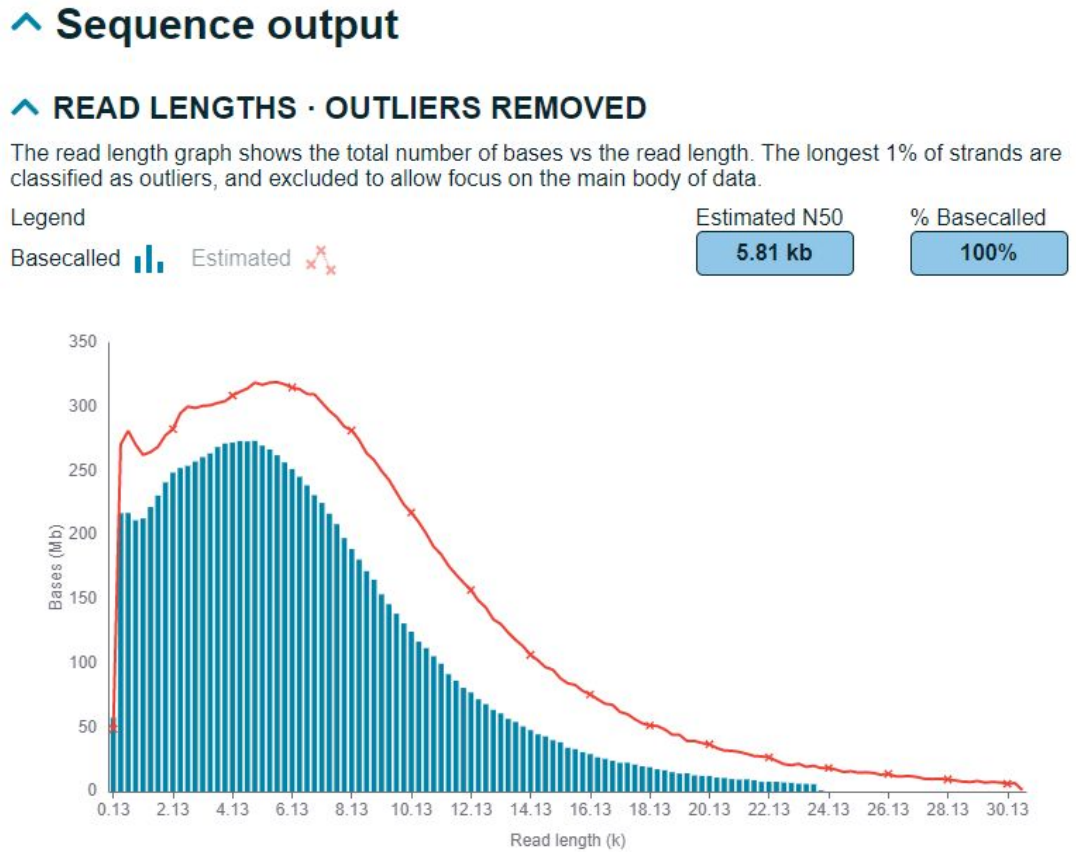



Photo credit:
https://www.bioinformatics.babraham.ac.uk/projects/fastqc/bad_sequence_fastqc.html#M8

# Discussion Question 1

With your neighbors, discuss this mystery graph!

- Describe the trend of the data.
- What do you think these data represent?
- Be ready to share your ideas in **3 minutes**.



# Mystery continued: how reliable are these sequences?

## QUALITY SCORE

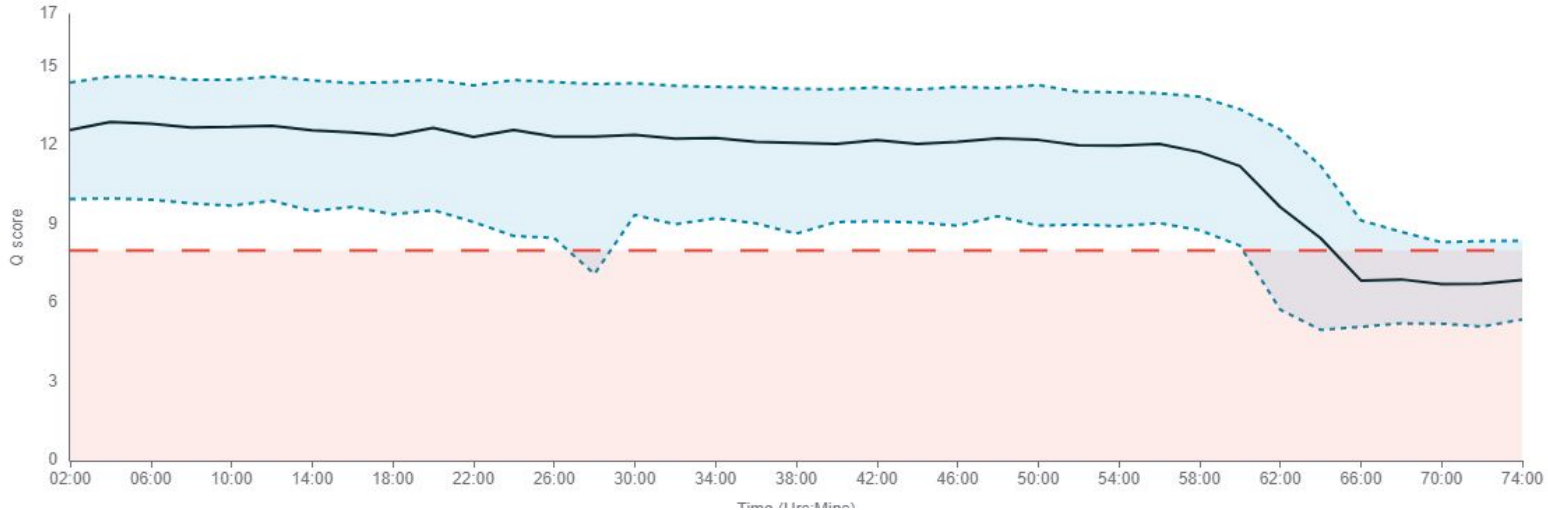
The quality score is calculated as basecalling is performed on your device. Quality score is measured on the logarithmic PHRED scale. Reads that fall below the minimum value of 8 will be classified as failed reads. You can alter the accepted minimum quality score in MinKNOW.

Legend

**Mode** — The most frequent quality score of reads in the run.

**Spread** — The spread of quality scores, found by calculating full width half maximum.

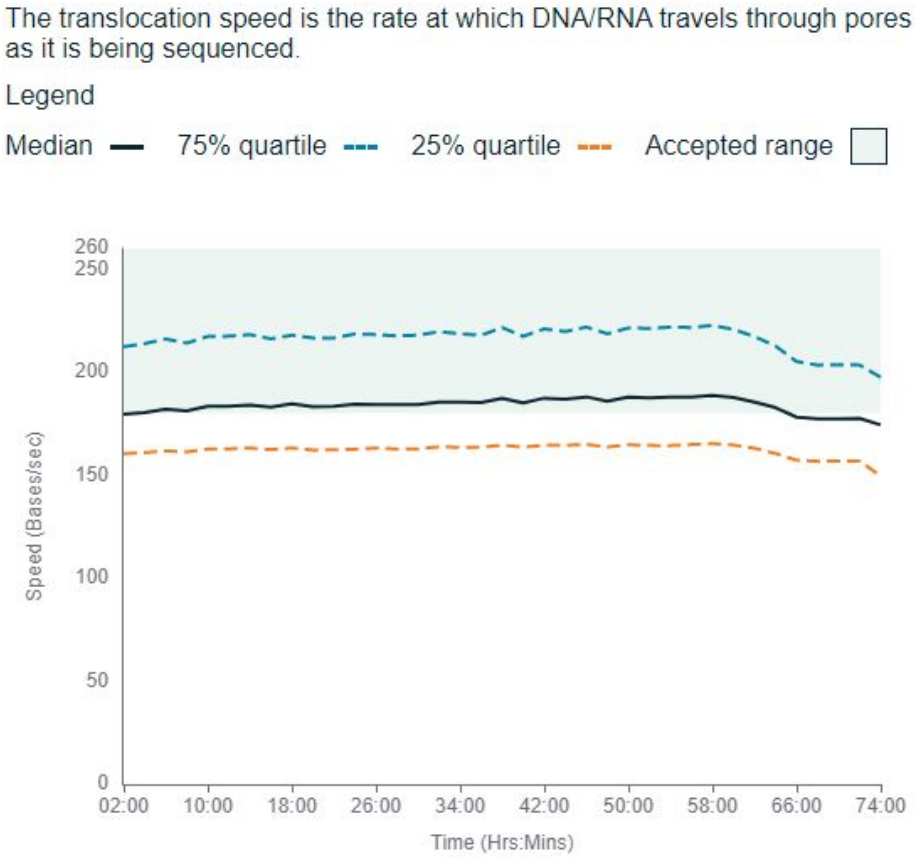
**Min. quality score** --- Minimum quality score to be accepted as a passed read.



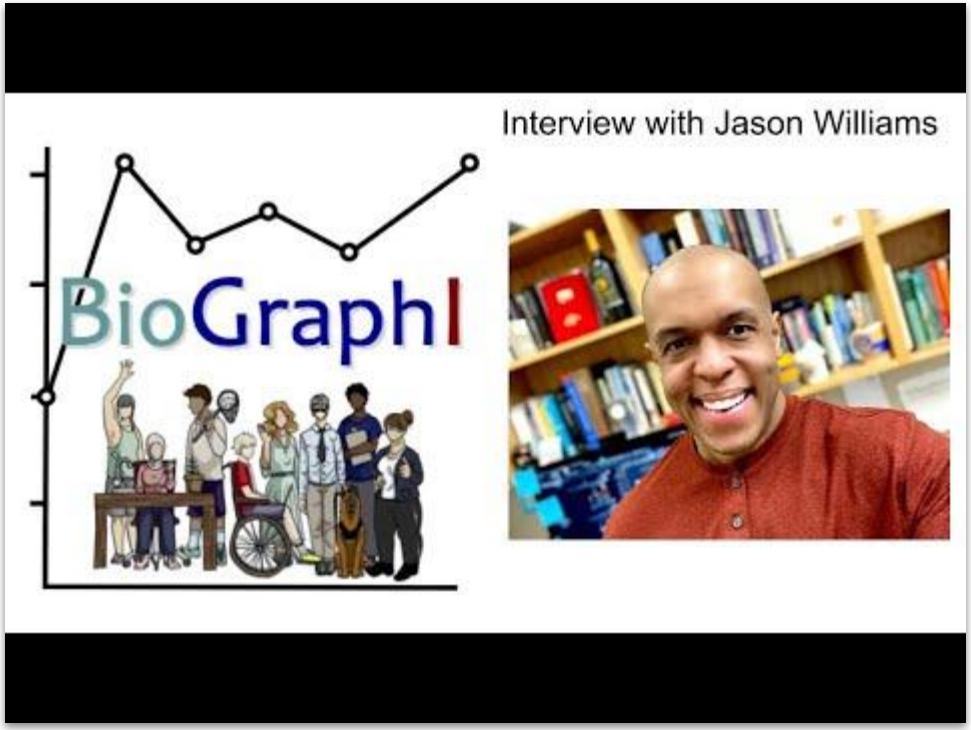
# Discussion Question 2

With your neighbors, answer these questions:

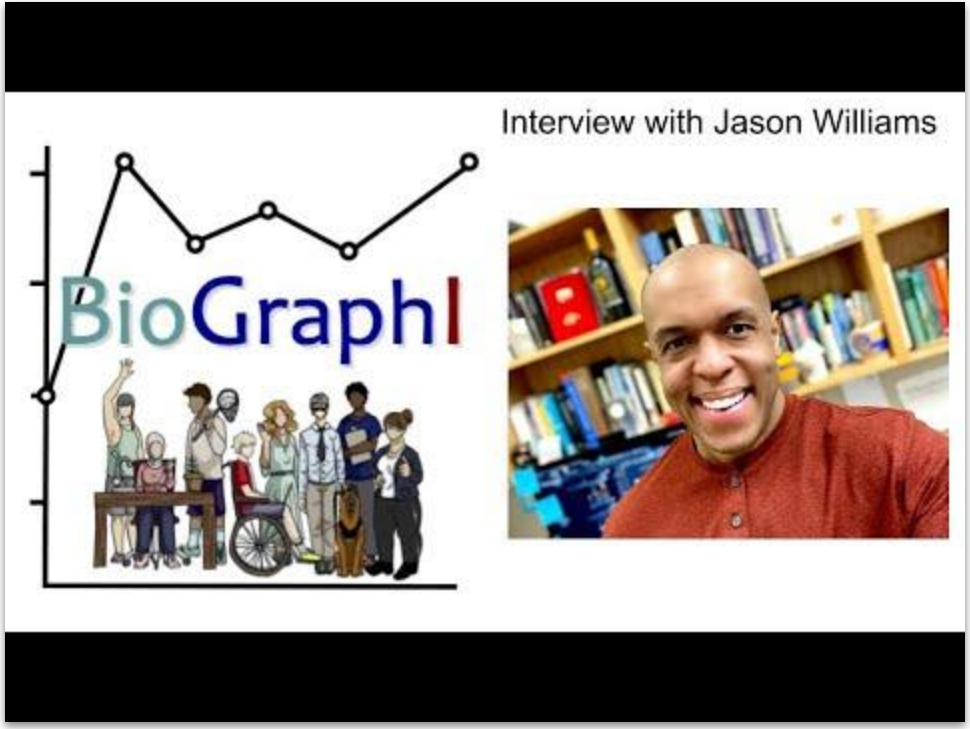
- What do you think the hypothesis was for the experiment?
- Was this hypothesis supported (or not) by these data?



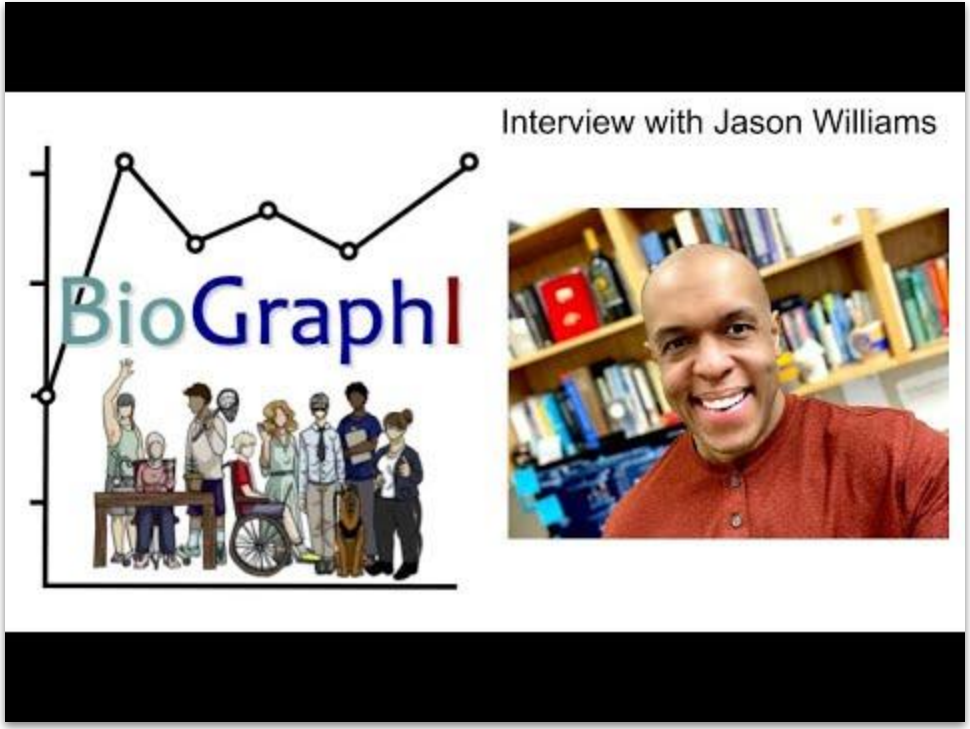
# What’s the significance of these findings?



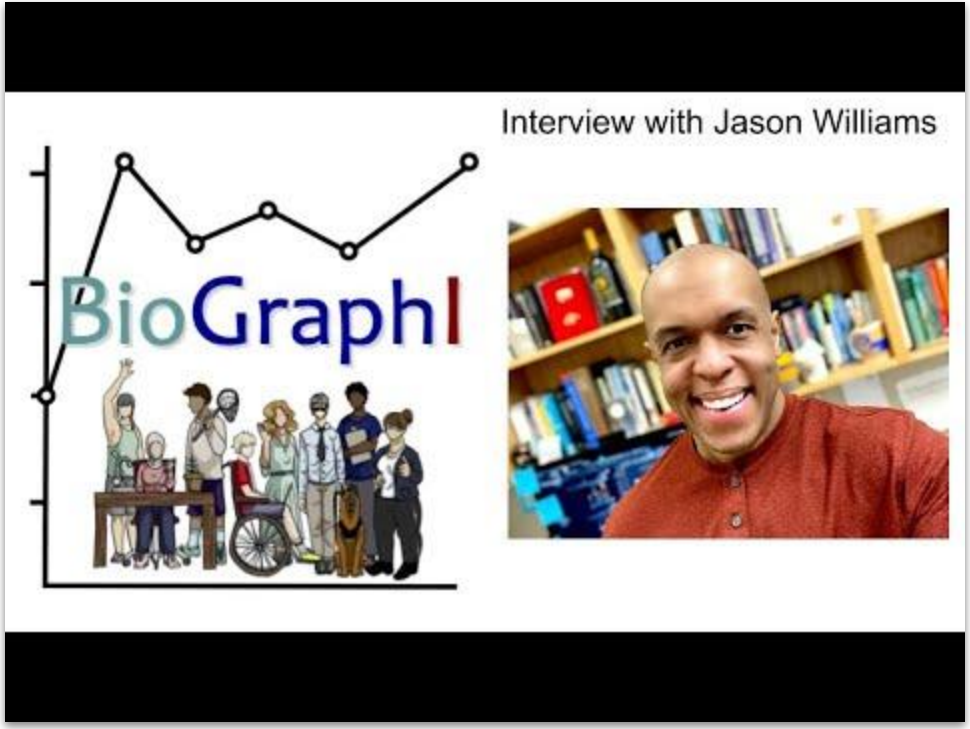
# What kinds of scientific questions interest you the most?



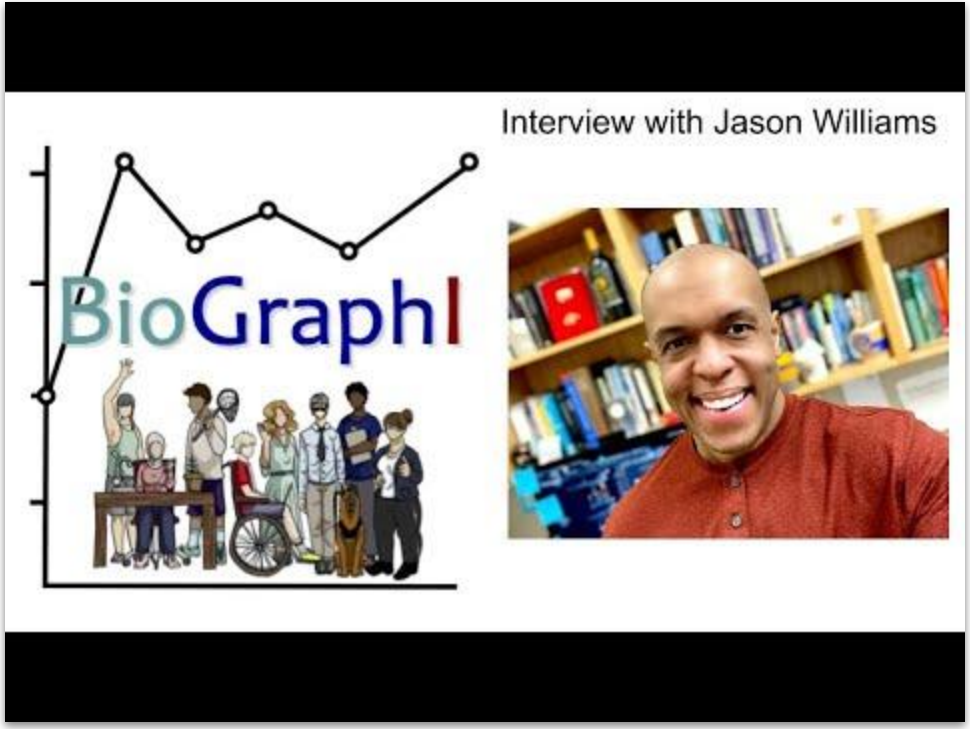
# When and how did you know you wanted to be a scientist?



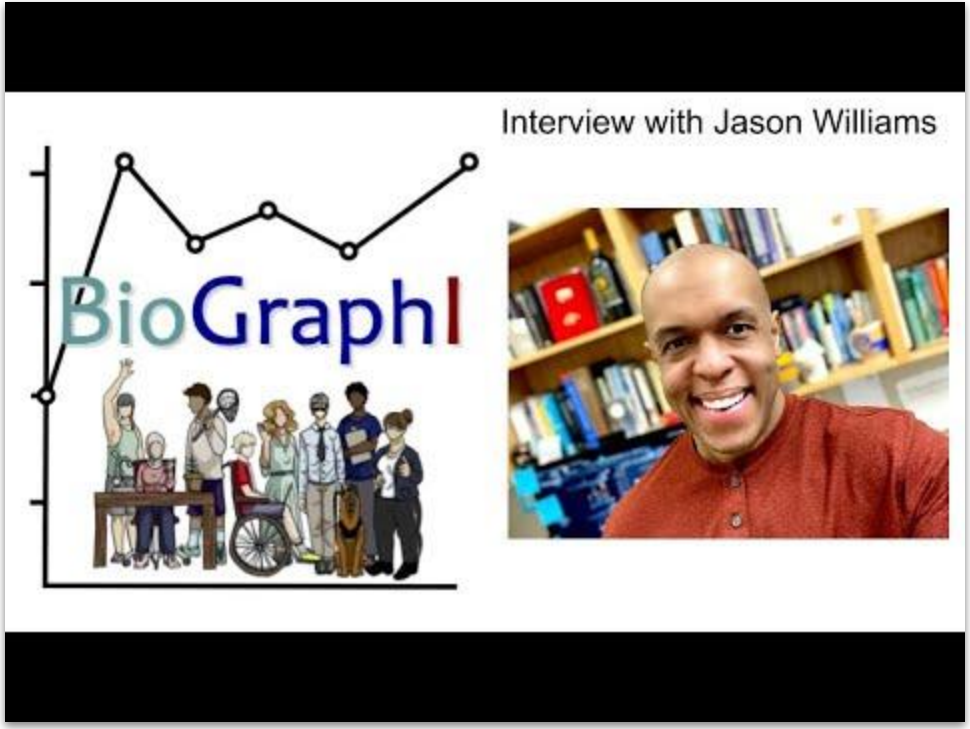
# Tell us about the path that led you to your current job.



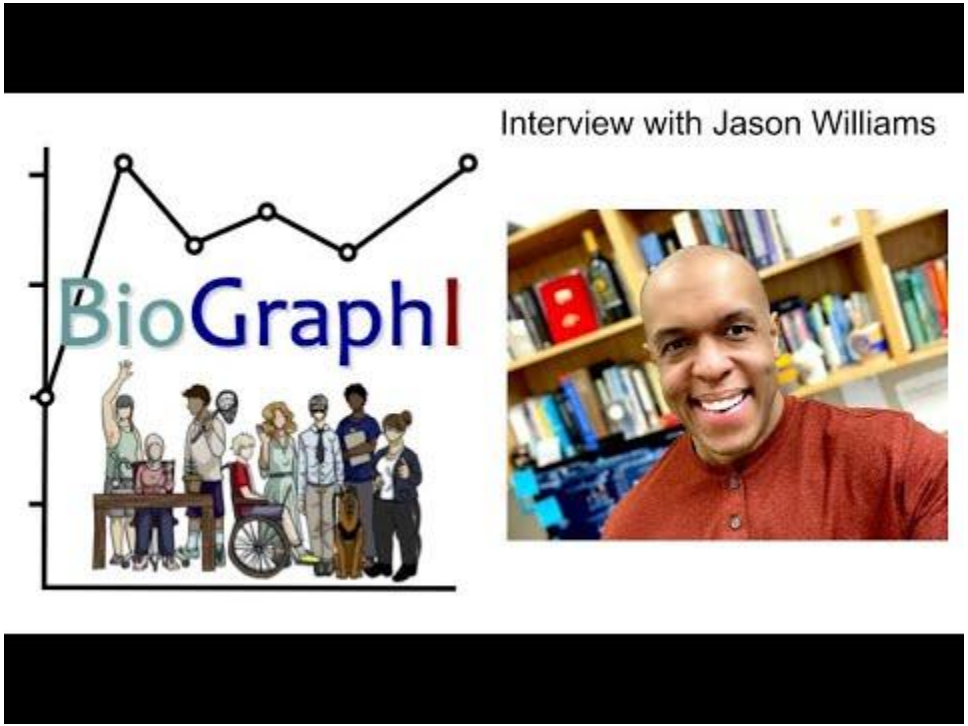
# What is it like being a scientist who identifies as…?



# What is a typical day like for you?



Tell me about a moment when you felt like you really **belonged** in the field of science.



# Discussion Question 6

With your peers, brainstorm ideas:

- What questions do you have about Jason’s path to this work?

# Homework assignment

Submit one more question for Jason Williams that wasn't answered in this interview.

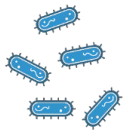
Complete the post-lesson survey.

AGTCCCTGAATCGA

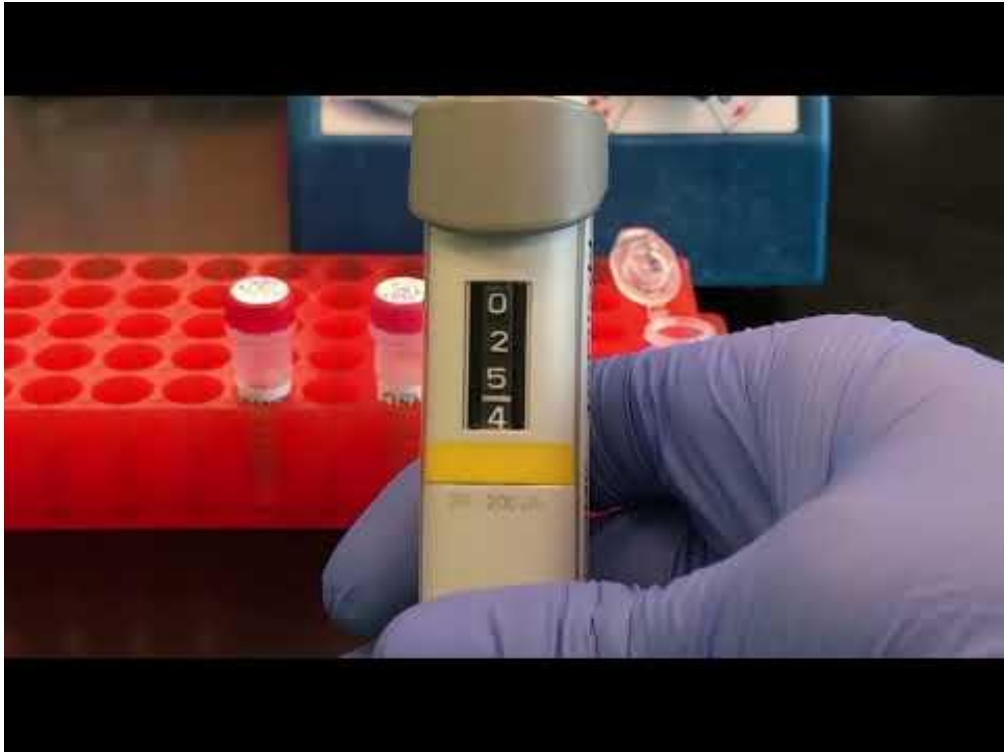
# Brainstorming

Share your ideas about REEs, electronics waste, and Nanopore uses in the **Discussion Forum**!

**Contact us:**
ccgoller@ncsu.edu
@CarlosGoller



# FYI/Optional: Microfluidics and Library Prep for Nanopore

Example of students preparing DNA for sequencing using digital microfluidics device. Caleb Carter and Thomas Morgan worked in the summer of 2022 to learn how to use the Miroculus Miro device and prepare DNA for sequencing. They then recorded this tutorial.

Video by Caleb Carter and Thomas Morgan, 11-min