BioDIGS a GDSCN project

BioDIGS

BioDiversity and Informatics for Genomics Scholars

Michael Schatz
February 8, 2024
AGBT

GENOMIC DATA SCIENCE
COMMUNITY NETWORK

Disclosures

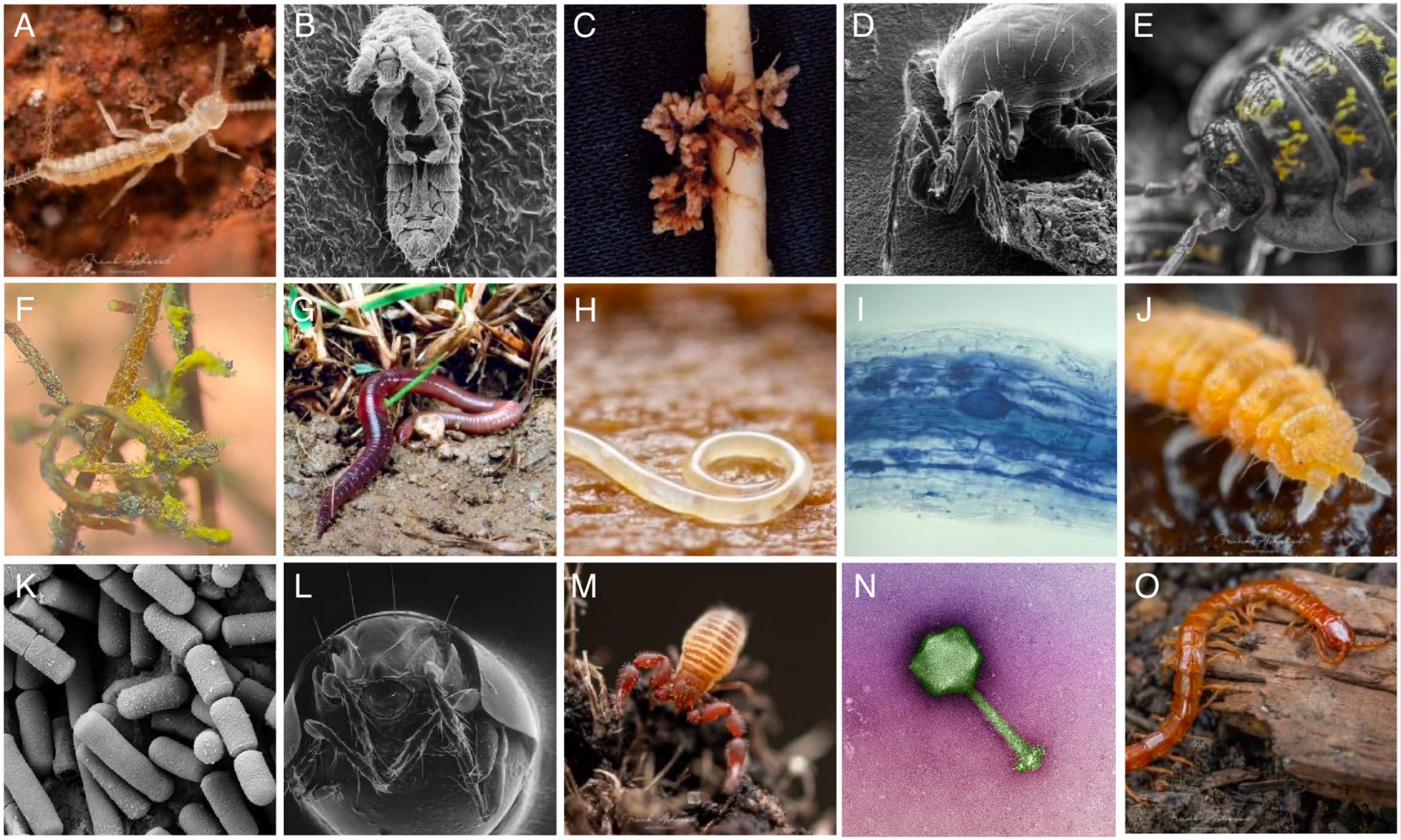
I (Michael Schatz) am a Scientific Advisory Board Member for CosmosID

Rita Colwell is the founder of CosmosID and
an Advocacy Board Member
for the Genomic Data Science Community Network (GDSCN)

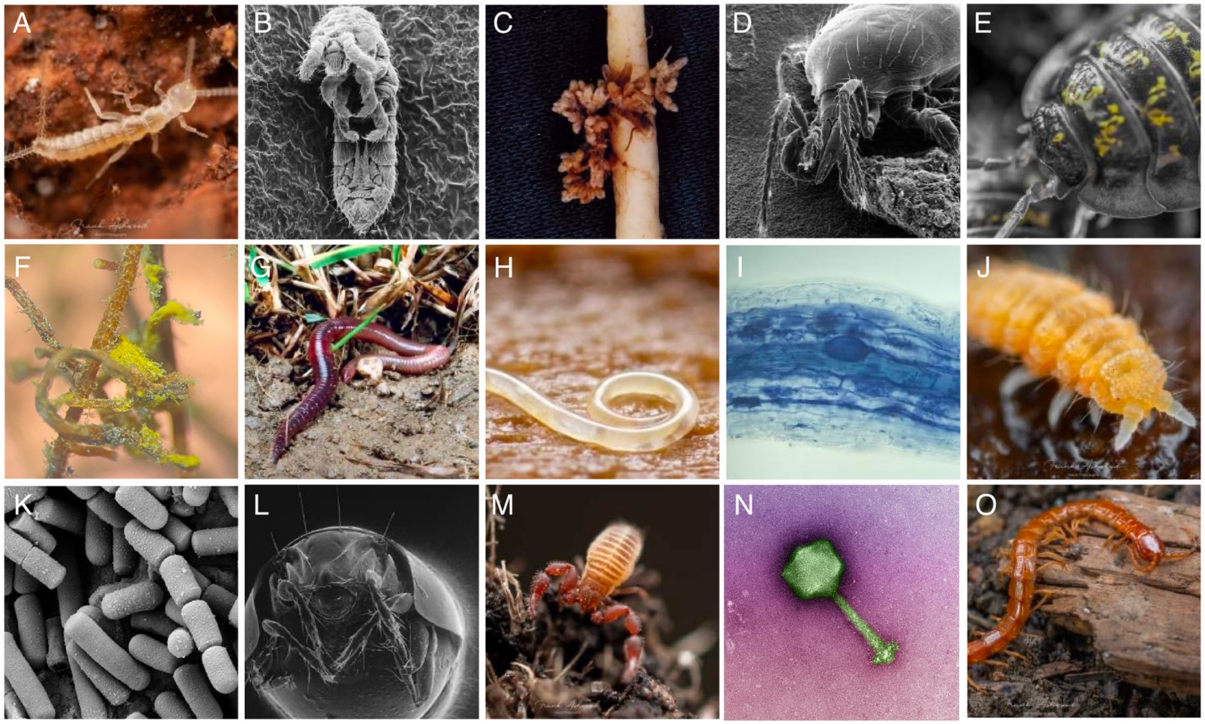
COSMOSID®
Unlocking The Microbiome

GENOMIC DATA SCIENCE
COMMUNITY NETWORK

Genomic Diversity



Genomic Diversity

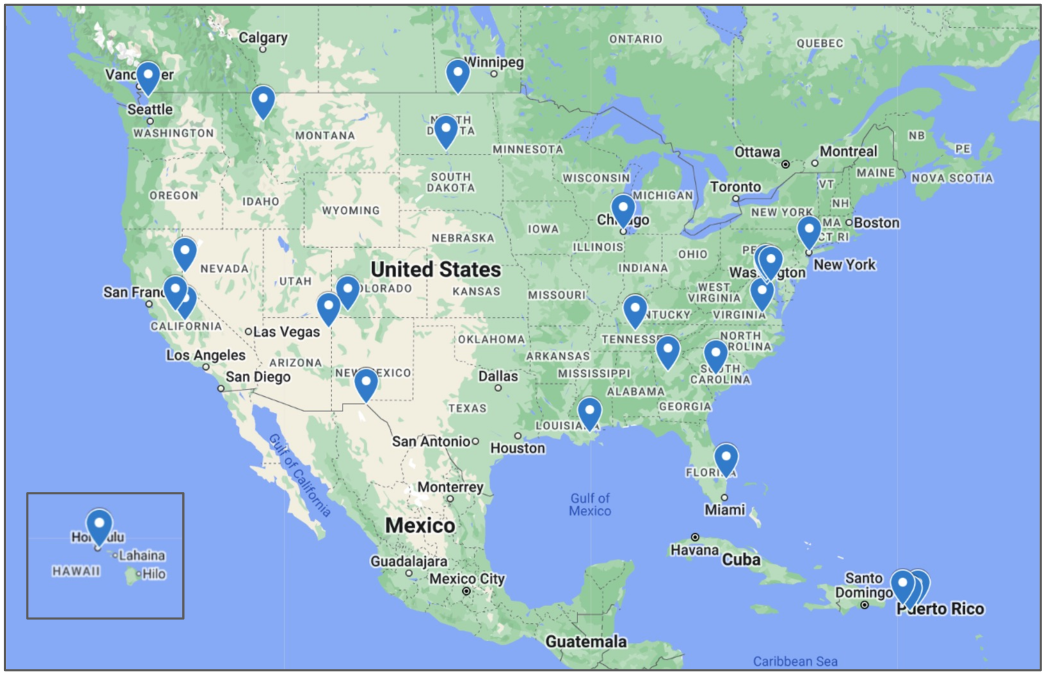


"… soil is likely home to 59% of life … making it the singular most biodiverse habitat on Earth."

Enumerating soil biodiversity
Anthony et al (2023) PNAS.
doi:10.1073/pnas.2304663120

Genomic Data Science Community Network

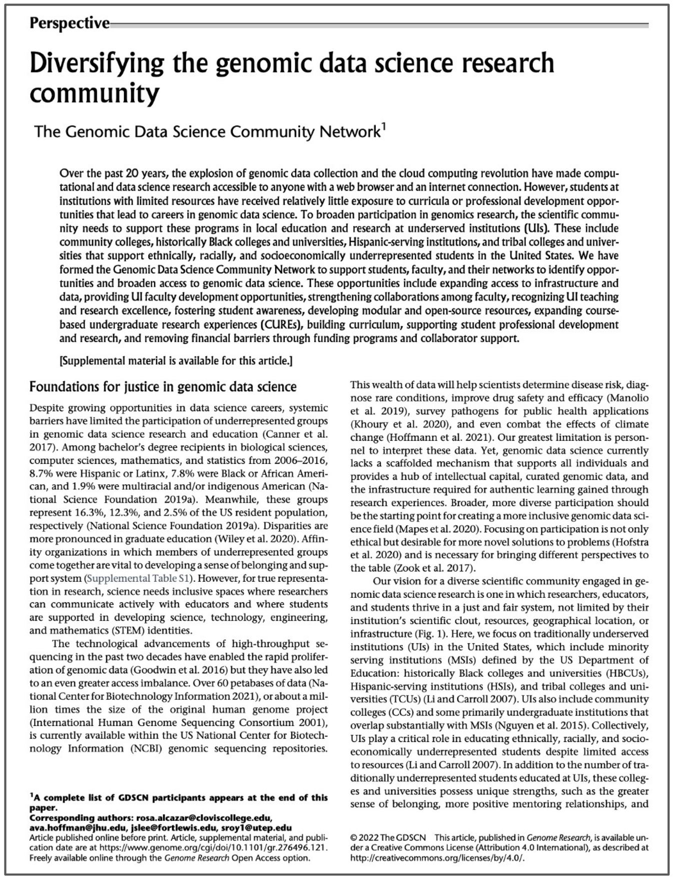
Promoting education and research in genomic data science at HBCUs, HSIs, TCUs, CCs, and other underserved



https://www.gdscn.org/

GDSCN Needs & Resources

GENOMIC DATA SCIENCE
COMMUNITY NETWORK



Perspective

Diversifying the genomic data science research community

The Genomic Data Science Community Network[1]

Over the past 20 years, the explosion of genomic data collection and the cloud computing revolution have made computational and data science research accessible to anyone with a web browser and an internet connection. However, students at institutions with limited resources have received relatively little exposure to curricula or professional development opportunities that lead to careers in genomic data science. To broaden participation in genomics research, the scientific community needs to support these programs in local education and research at underserved institutions (UIs). These include community colleges, historically Black colleges and universities, Hispanic-serving institutions, and tribal colleges and universities that support ethnically, racially, and socioeconomically underrepresented students in the United States. We have formed the Genomic Data Science Community Network to support students, faculty, and their networks to identify opportunities and broaden access to genomic data science. These opportunities include expanding access to infrastructure and data, providing UI faculty development opportunities, strengthening collaborations among faculty, recognizing UI teaching and research excellence, fostering student awareness, developing modular and open-source resources, expanding course-based undergraduate research experiences (CUREs), building curriculum, supporting student professional development and research, and removing financial barriers through funding programs and collaborator support.

[Supplemental material is available for this article.]

Foundations for justice in genomic data science

Despite growing opportunities in data science careers, systemic barriers have limited the participation of underrepresented groups in genomic data science research and education (Canner et al. 2017). Among bachelor's degree recipients in biological sciences, computer sciences, mathematics, and statistics from 2006–2016, 8.7% were Hispanic or Latinx, 7.8% were Black or African American, and 1.9% were multiracial and/or indigenous American (National Science Foundation 2019a). Meanwhile, these groups represent 16.3%, 12.3%, and 2.5% of the US resident population, respectively (National Science Foundation 2019a). Disparities are more pronounced in graduate education (Wiley et al. 2020). Affinity organizations in which members of underrepresented groups come together are vital to developing a sense of belonging and support system (Supplemental Table S1). However, for true representation in research, science needs inclusive spaces where researchers can communicate actively with educators and where students are supported in developing science, technology, engineering, and mathematics (STEM) identities.

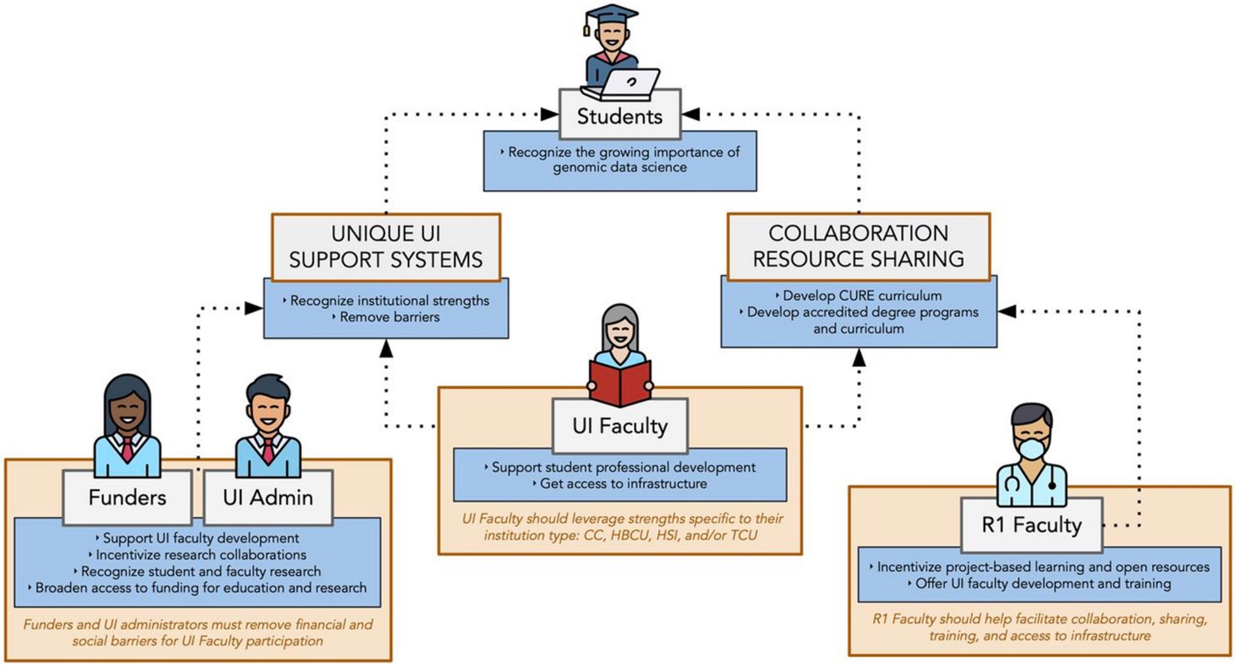
The technological advancements of high-throughput sequencing in the past two decades have enabled the rapid proliferation of genomic data (Goodwin et al. 2016) but they have also led to an even greater access imbalance. Over 60 petabases of data (National Center for Biotechnology Information 2021), or about a million times the size of the original human genome project (International Human Genome Sequencing Consortium 2001), is currently available within the US National Center for Biotechnology Information (NCBI) genomic sequencing repositories. This wealth of data will help scientists determine disease risk, diagnose rare conditions, improve drug safety and efficacy (Manolio et al. 2019), survey pathogens for public health applications (Khoury et al. 2020), and even combat the effects of climate change (Hoffmann et al. 2021). Our greatest limitation is personnel to interpret these data. Yet, genomic data science currently lacks a scaffolded mechanism that supports all individuals and provides a hub of intellectual capital, curated genomic data, and the infrastructure required for authentic learning gained through research experiences. Broader, more diverse participation should be the starting point for creating a more inclusive genomic data science field (Mapes et al. 2020). Focusing on participation is not only ethical but desirable for more novel solutions to problems (Hofstra et al. 2020) and is necessary for bringing different perspectives to the table (Zook et al. 2017).

Our vision for a diverse scientific community engaged in genomic data science research is one in which researchers, educators, and students thrive in a just and fair system, not limited by their institution's scientific clout, resources, geographical location, or infrastructure (Fig. 1). Here, we focus on traditionally underserved institutions (UIs) in the United States, which include minority serving institutions (MSIs) defined by the US Department of Education: historically Black colleges and universities (HBCUs), Hispanic-serving institutions (HSIs), and tribal colleges and universities (TCUs) (Li and Carroll 2007). UIs also include community colleges (CCs) and some primarily undergraduate institutions that overlap substantially with MSIs (Nguyen et al. 2015). Collectively, UIs play a critical role in educating ethnically, racially, and socioeconomically underrepresented students despite limited access to resources (Li and Carroll 2007). In addition to the number of traditionally underrepresented students educated at UIs, these colleges and universities possess unique strengths, such as the greater sense of belonging, more positive mentoring relationships, and

[1] **A complete list of GDSCN participants appears at the end of this paper.**

Corresponding authors: rosa.alcazar@cloviscollege.edu, ava.hoffman@jhu.edu, jslee@fortlewis.edu, sroy1@utep.edu

Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.276496.121. Freely available online through the *Genome Research* Open Access option.



Students
- Recognize the growing importance of genomic data science

UNIQUE UI SUPPORT SYSTEMS
- Recognize institutional strengths
- Remove barriers

COLLABORATION RESOURCE SHARING
- Develop CURE curriculum
- Develop accredited degree programs and curriculum

Funders | UI Admin
- Support UI faculty development
- Incentivize research collaborations
- Recognize student and faculty research
- Broaden access to funding for education and research

Funders and UI administrators must remove financial and social barriers for UI Faculty participation

UI Faculty
- Support student professional development
- Get access to infrastructure

UI Faculty should leverage strengths specific to their institution type: CC, HBCU, HSI, and/or TCU

R1 Faculty
- Incentivize project-based learning and open resources
- Offer UI faculty development and training

R1 Faculty should help facilitate collaboration, sharing, training, and access to infrastructure

"Diversifying the Genomic Data Science Research Community." Genome Research (2022). doi:10.1101/gr.276496.121

North Carolina Agricultural
and Technical State University

Kristen
Rhinehardt

Foundations of GRADS-4C

Knowledge Transfer and Sustainability

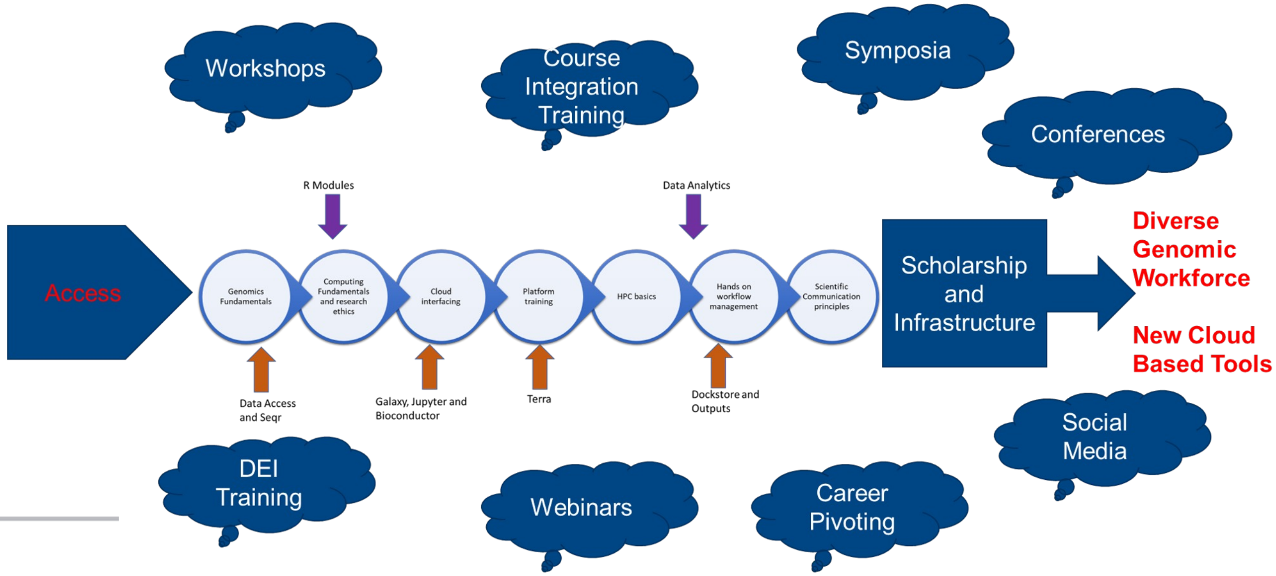
Partnership

Growth Model Infrastructure Development

Trust

GRADS-4C: Genomic Research and Data Science Center for Computation and Cloud Computing

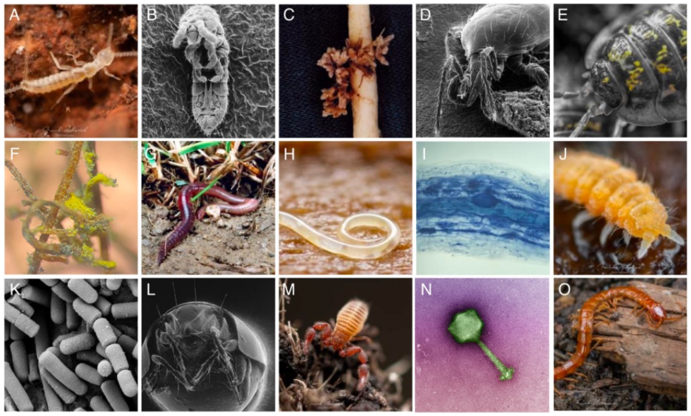
Our Mission is to develop an educational and hands-on research training network with resources in computational genomics, data sciences, and cloud computing (CGDS) to investigate and improve human health, as well as develop the future CGDS workforce, particularly among underserved populations



Website: GRADS4C.ncat.edu

AGGIES DO

BioDIGS: What can we learn from the soil?

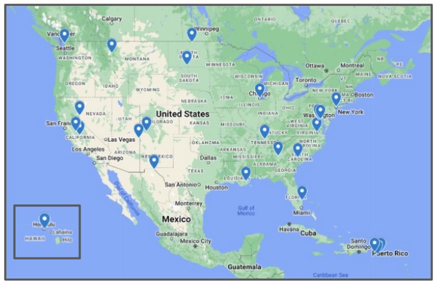


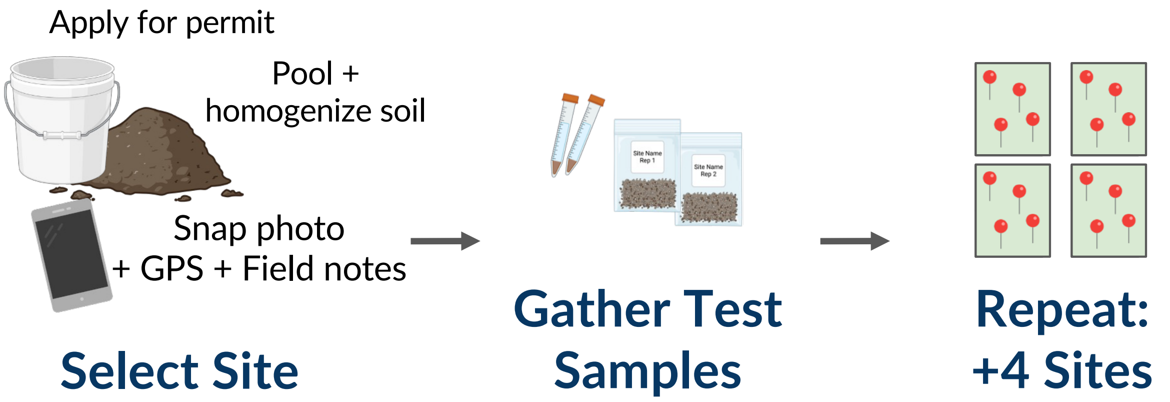
BioDIGS
a GDSCN project

- *Microbiology & Metagenomics*
  - What's there? How does the genomic composition change in time and space?
- *Genomics & Bioinformatics*
  - Optimal approaches for metagenome assembly and classification? Merits of short- vs long-read sequencing?
- *Agriculture & Environment*
  - How do characteristics of the soil & soil microbiome modulate plant & animal development?
- *Public Health*
  - How does the soil microbiome influence the human microbiome & human health outcomes?

BioDIGS Sampling Protocol

GENOMIC DATA SCIENCE COMMUNITY NETWORK





Apply for permit

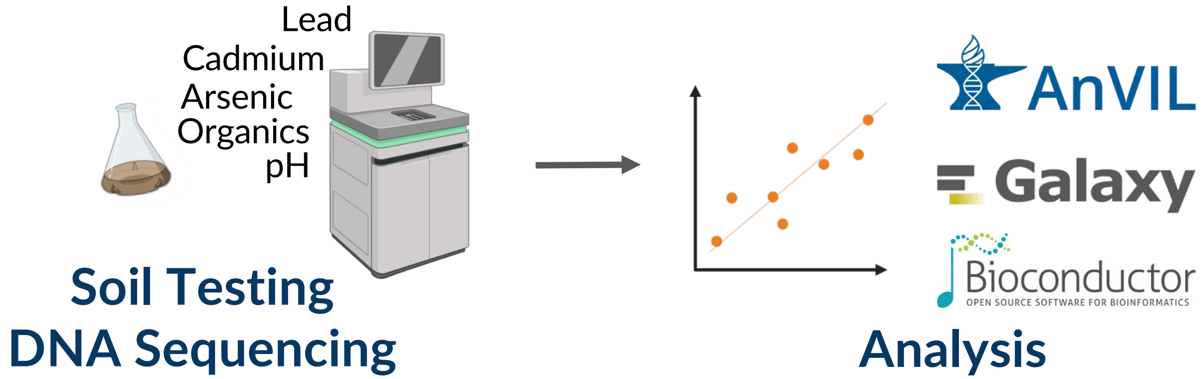
Pool + homogenize soil

Snap photo + GPS + Field notes

Select Site

Gather Test Samples

Repeat: +4 Sites



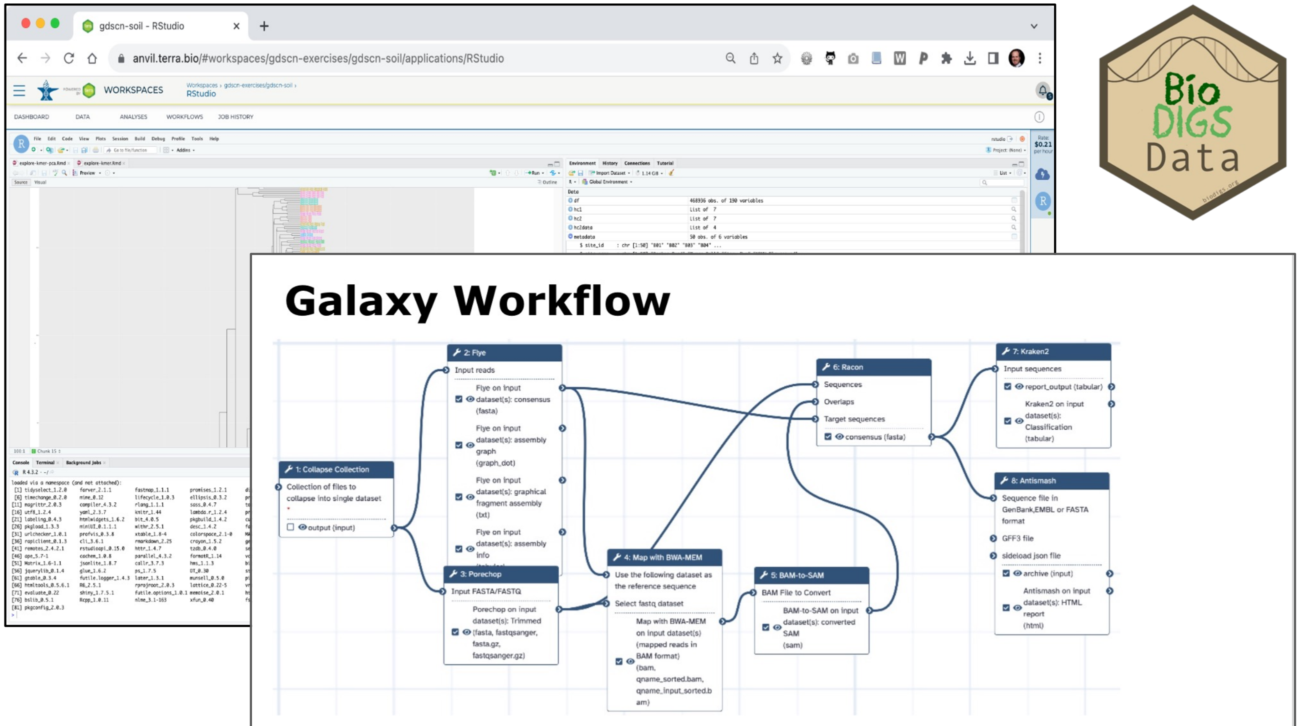
Lead
Cadmium
Arsenic
Organics
pH

Soil Testing DNA Sequencing

Analysis

BioDIGS Analysis

GENOMIC DATA SCIENCE
COMMUNITY NETWORK


gdscn-soil - RStudio
anvil.terra.bio/#workspaces/gdscn-exercises/gdscn-soil/applications/RStudio
WORKSPACES
DASHBOARD
DATA
ANALYSES
WORKFLOWS
JOB HISTORY
1: Collapse Collection
Collection of files to collapse into single dataset
output (input)
2: Flye
Input reads
Flye on input dataset(s): consensus (fasta)
Flye on input dataset(s): assembly graph (graph_dot)
Flye on input dataset(s): graphical fragment assembly (txt)
Flye on input dataset(s): assembly info
3: Porechop
Input FASTA/FASTQ
Porechop on input dataset(s): Trimmed (fasta, fastqsanger, fasta.gz, fastqsanger.gz)
4: Map with BWA-MEM
Use the following dataset as the reference sequence
Select fastq dataset
Map with BWA-MEM on input dataset(s) (mapped reads in BAM format) (bam, qname_sorted.bam, qname_input_sorted.bam)
5: BAM-to-SAM
BAM File to Convert
BAM-to-SAM on input dataset(s): converted SAM (sam)
6: Racon
Sequences
Overlaps
Target sequences
consensus (fasta)
7: Kraken2
Input sequences
report_output (tabular)
Kraken2 on input dataset(s): Classification (tabular)
8: Antismash
Sequence file in GenBank,EMBL or FASTA format
GFF3 file
sideload json file
archive (input)
Antismash on input dataset(s): HTML report (html)

BioDIGS Data

Galaxy Workflow

1. Flye (assembly)
2. Porechop (trim & QC)
3. BWA-MEM (align reads)
4. Racon (polish contigs)
5. Kraken2 (classify contigs)
6. AntiSMASH (BGC annotation)
7. Kallisto (BGC abundance)

Galaxy PROJECT

g themselves and a better
n helping to protect plant
our soil/roots. Identified
may be involved in
Galaxy PROJECT
NATIONAL PLANT & ANIMAL

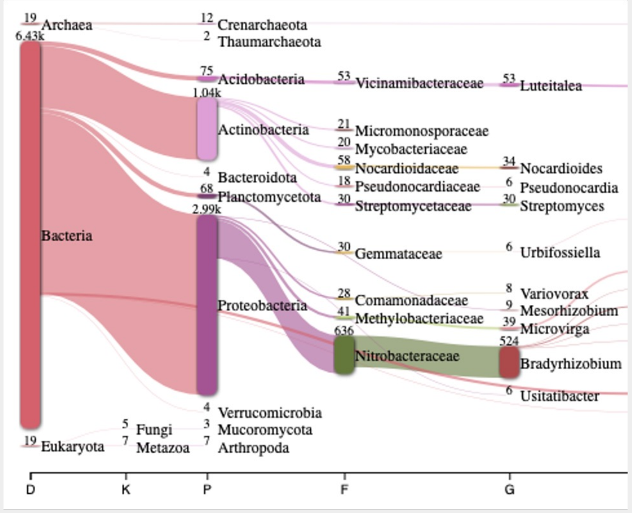
Katherine Ulbricht

Nia Davis

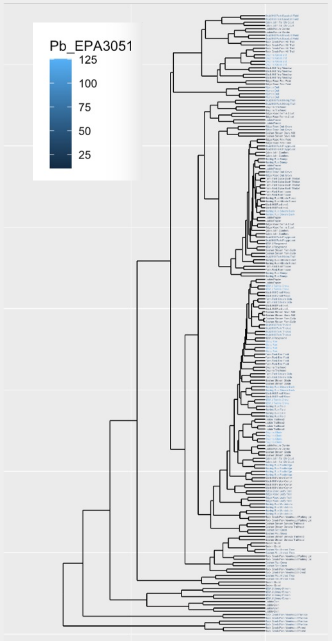
Ayalew lab
Spelman College

BioDIGS Analysis

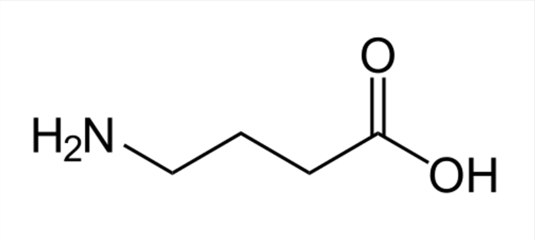
Genomic Diversity



Environmental Associations

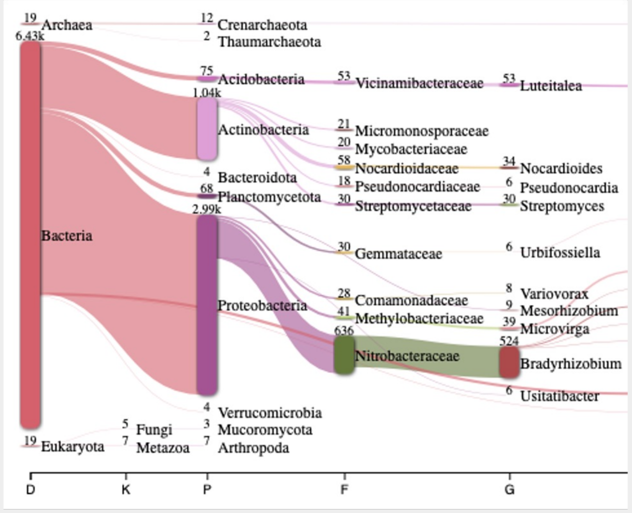


Human Health & Disease

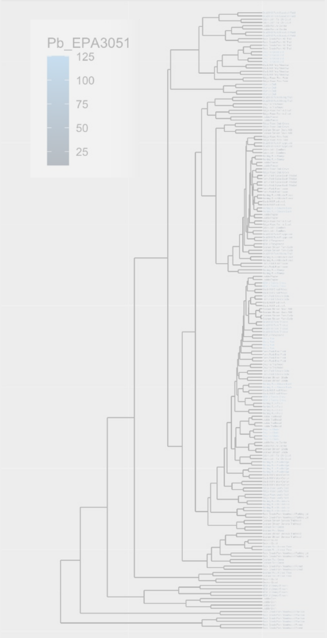


BioDIGS Analysis

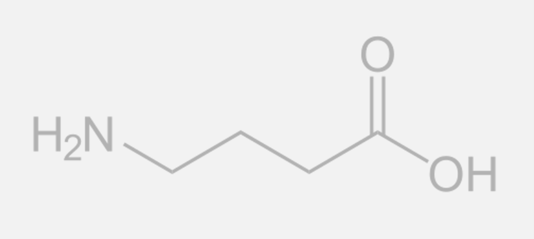
Genomic Diversity



Environmental Associations

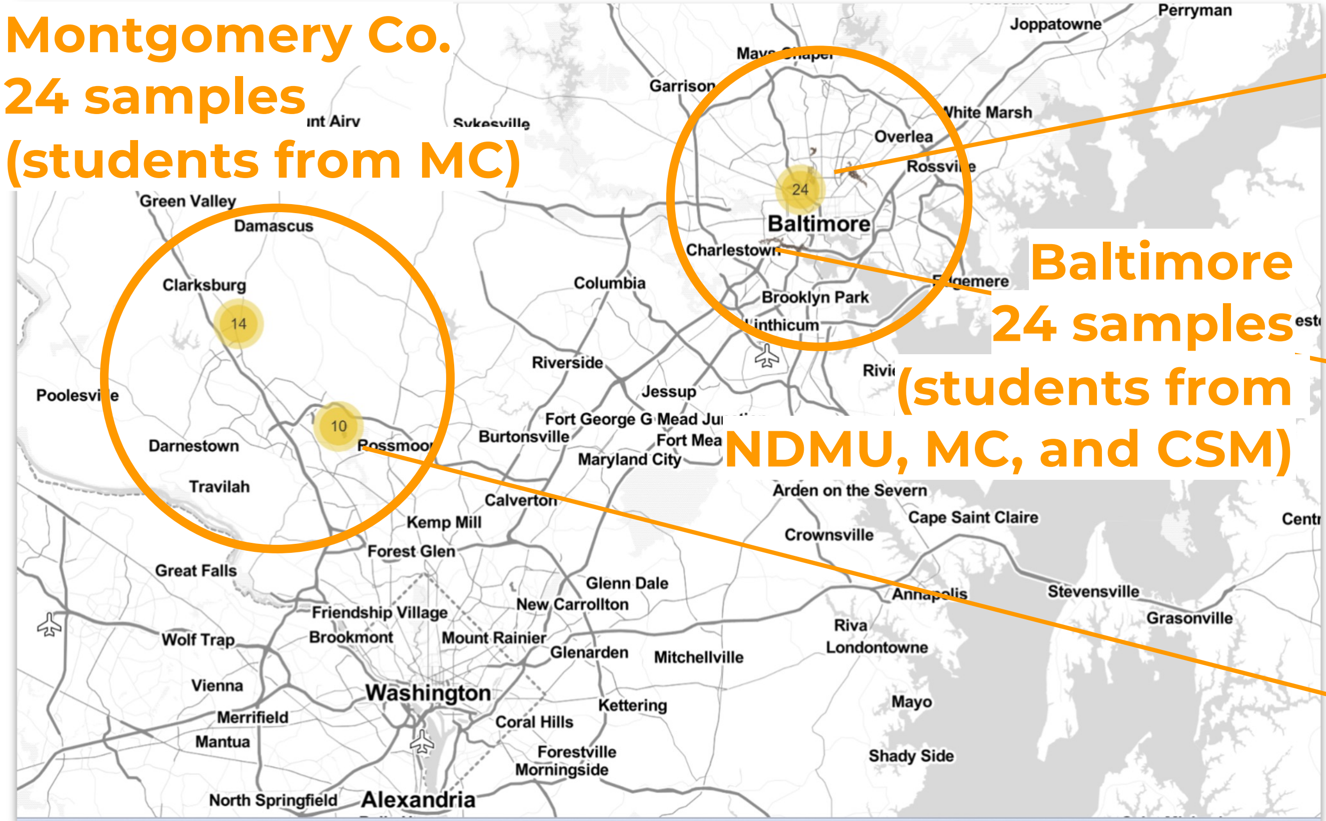


Human Health & Disease



DC+Baltimore pilot study sampling

GENOMIC DATA SCIENCE
COMMUNITY NETWORK



Montgomery Co.
24 samples
(students from MC)

Baltimore
24 samples
(students from
NDMU, MC, and CSM)

Interactive map at: http://biodigs.org

Stony Run (near JHU)

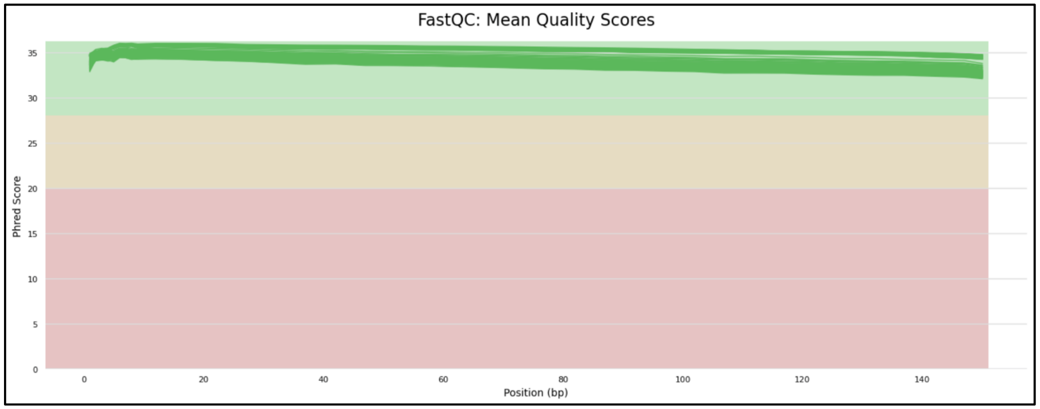
Gwynns Falls Trailhead

Lake Needwood

DC+Baltimore Data

Sampling & Sequencing

- Zymo DNA/RNA shield
- Qiagen DNeasy PowerSoil Pro kit
- 2x150bp @ NovaSeq SP
- ~20M reads / site



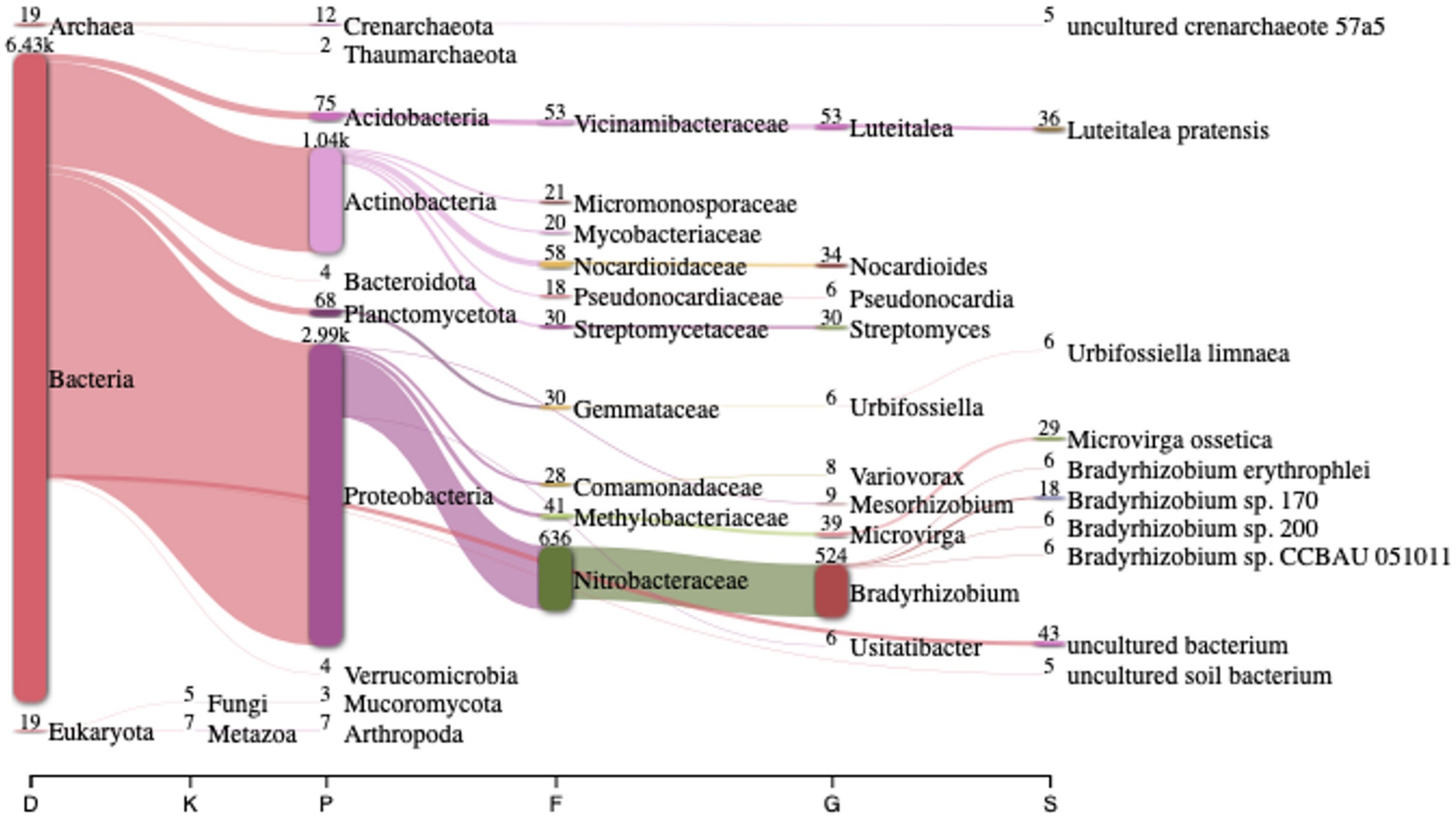
grcf
grcf.jhmi.edu

UNIVERSITY OF DELAWARE
COOPERATIVE
EXTENSION

Soil Testing

- Routine soil testing: Soil pH, Organic matter, Mehlich 3 extractable nutrients (P, K, Ca, Mg, Mn, Zn, Cu, Fe, B, S, Na and Al), Phosphorus Saturation Ratio
- Heavy metals: Arsenic, Cadmium, Chromium, Copper, Nickel, Lead, Zinc

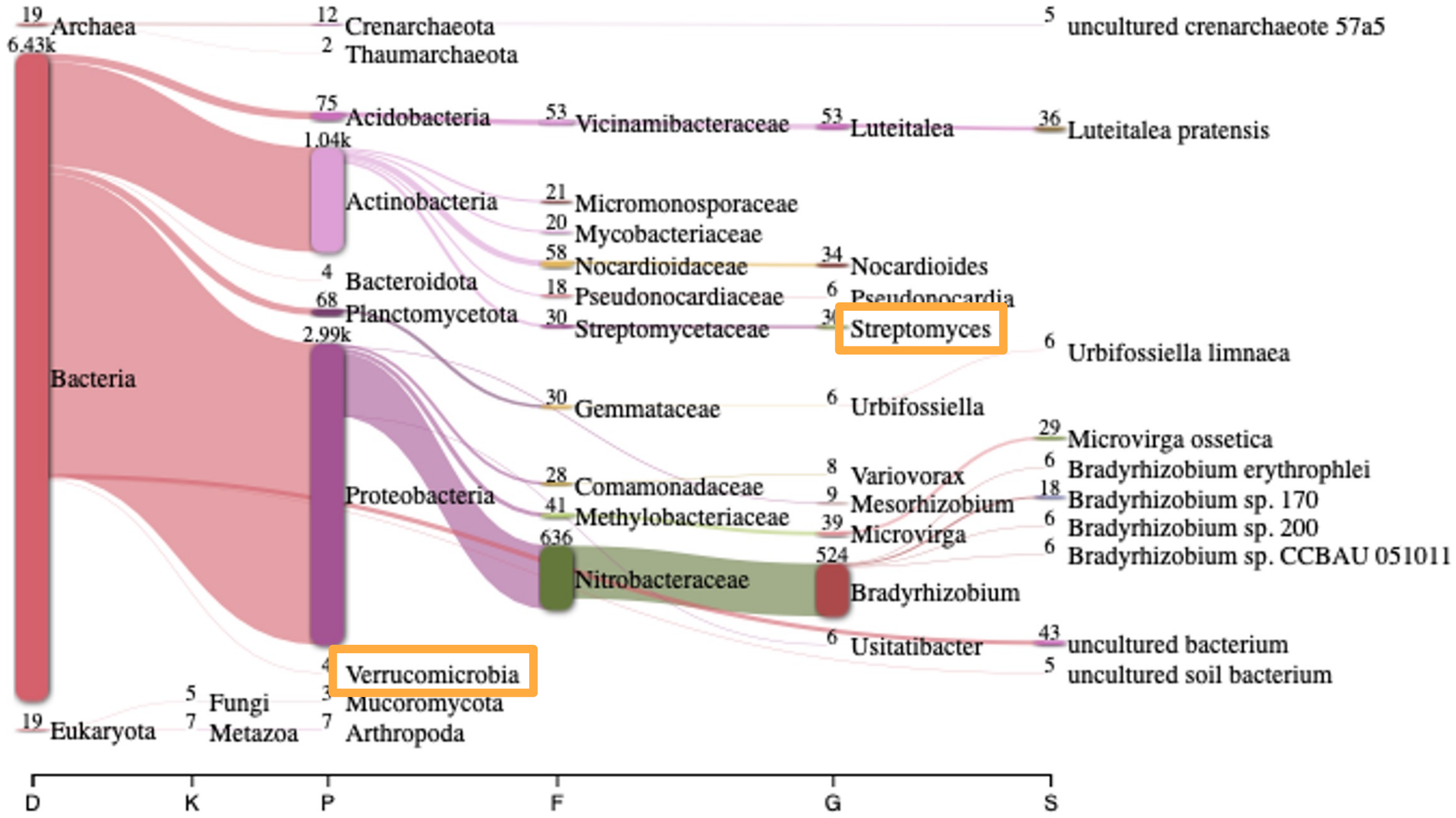
Taxonomic classification



- Mostly common soil microbes; Several species associated with nitrogen fixation

Improved metagenomic analysis with Kraken 2
Wood, Lu, Langmead (2019) Genome Biology do:10.1186/s13059-019-1891-0

Taxonomic classification

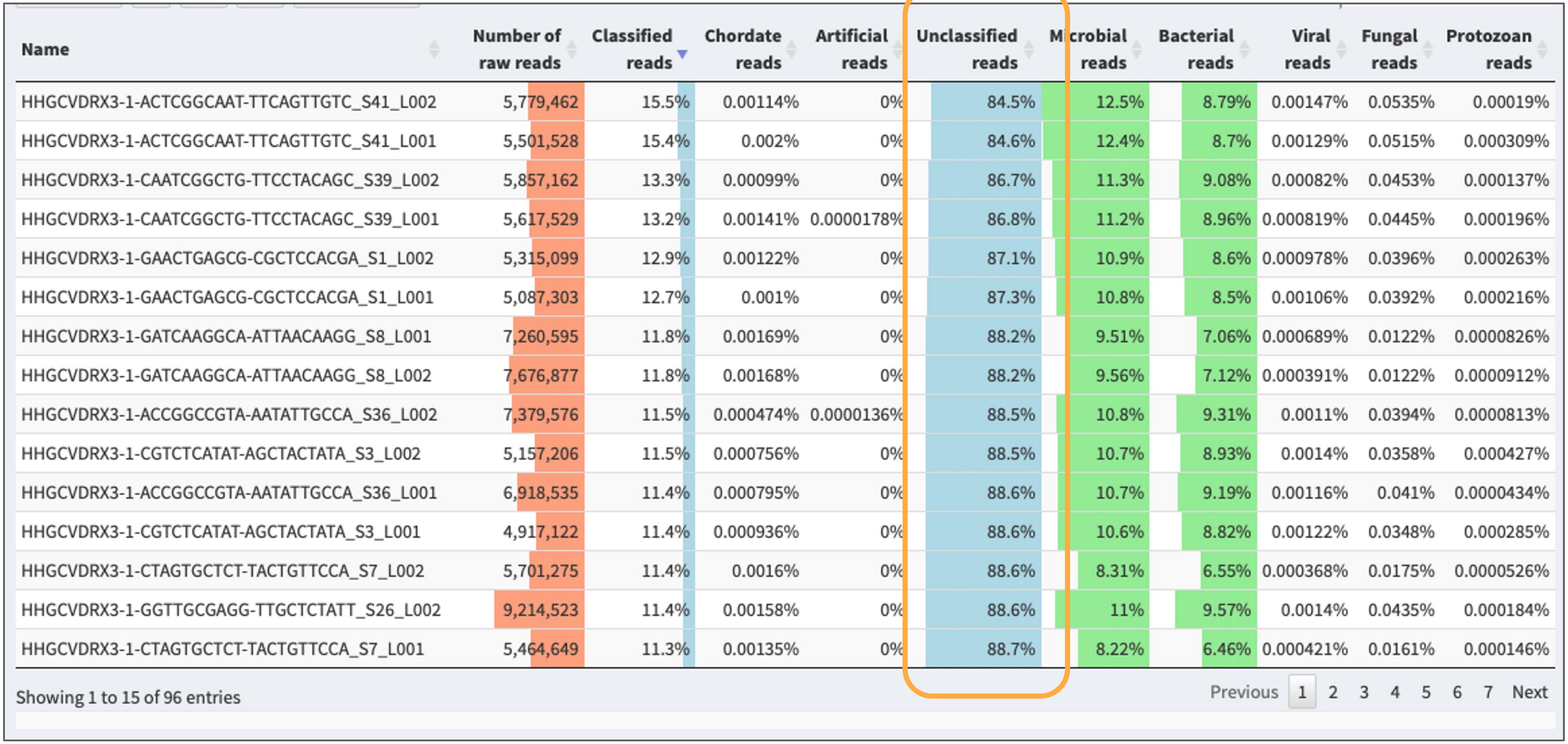


- Mostly common soil microbes; Several species associated with nitrogen fixation
- Several hits to the genus Streptomyces, which produce many naturally occurring antimicrobials
- Certain species of Verrucomicrobia are important human probiotics for GI metabolism (Akkermansia)

Improved metagenomic analysis with Kraken 2
Wood, Lu, Langmead (2019) Genome Biology do:10.1186/s13059-019-1891-0

Species classification

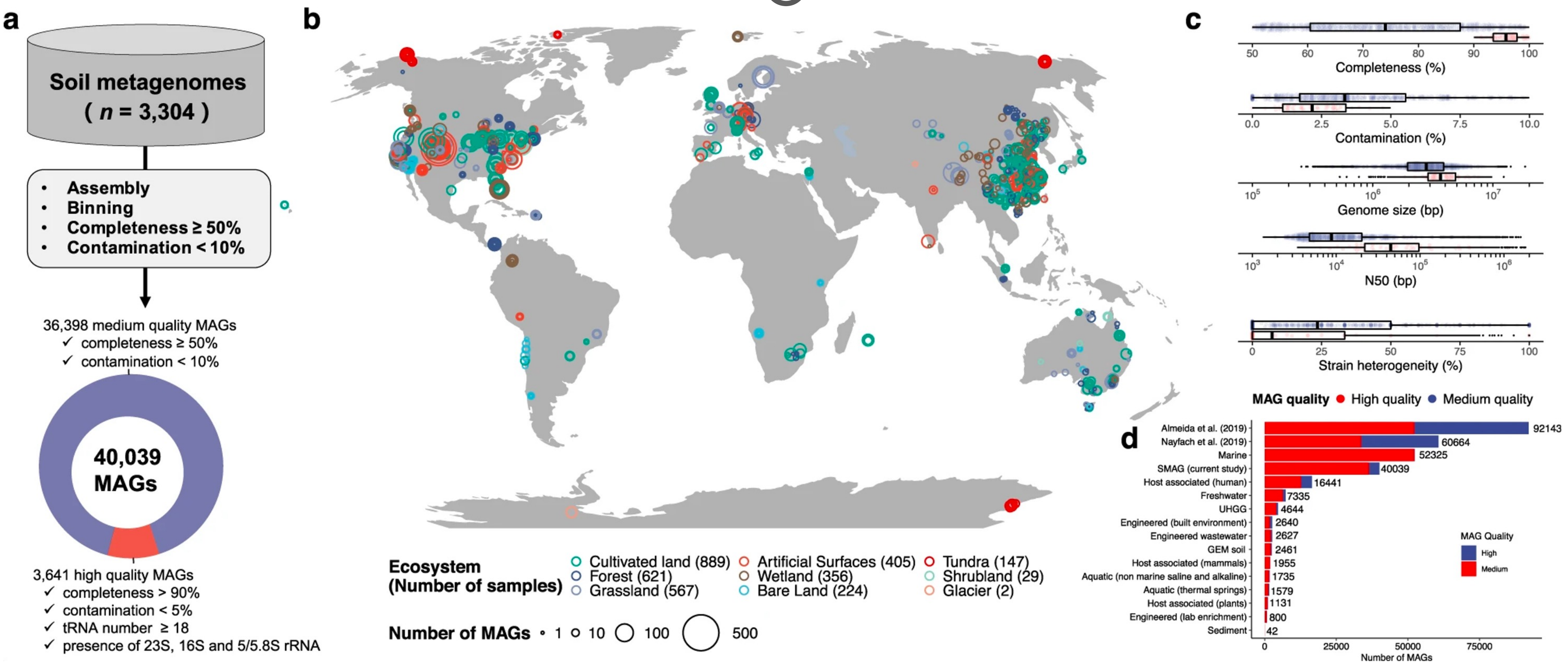
GENOMIC DATA SCIENCE
COMMUNITY NETWORK



Name	Number of raw reads	Classified reads	Chordate reads	Artificial reads	Unclassified reads	Microbial reads	Bacterial reads	Viral reads	Fungal reads	Protozoan reads
HHGCVDRX3-1-ACTCGGCAAT-TTCAGTTGTC_S41_L002	5,779,462	15.5%	0.00114%	0%	84.5%	12.5%	8.79%	0.00147%	0.0535%	0.00019%
HHGCVDRX3-1-ACTCGGCAAT-TTCAGTTGTC_S41_L001	5,501,528	15.4%	0.002%	0%	84.6%	12.4%	8.7%	0.00129%	0.0515%	0.000309%
HHGCVDRX3-1-CAATCGGCTG-TTCCTACAGC_S39_L002	5,857,162	13.3%	0.00099%	0%	86.7%	11.3%	9.08%	0.00082%	0.0453%	0.000137%
HHGCVDRX3-1-CAATCGGCTG-TTCCTACAGC_S39_L001	5,617,529	13.2%	0.00141%	0.0000178%	86.8%	11.2%	8.96%	0.000819%	0.0445%	0.000196%
HHGCVDRX3-1-GAACTGAGCG-CGCTCCACGA_S1_L002	5,315,099	12.9%	0.00122%	0%	87.1%	10.9%	8.6%	0.000978%	0.0396%	0.000263%
HHGCVDRX3-1-GAACTGAGCG-CGCTCCACGA_S1_L001	5,087,303	12.7%	0.001%	0%	87.3%	10.8%	8.5%	0.00106%	0.0392%	0.000216%
HHGCVDRX3-1-GATCAAGGCA-ATTAACAAGG_S8_L001	7,260,595	11.8%	0.00169%	0%	88.2%	9.51%	7.06%	0.000689%	0.0122%	0.0000826%
HHGCVDRX3-1-GATCAAGGCA-ATTAACAAGG_S8_L002	7,676,877	11.8%	0.00168%	0%	88.2%	9.56%	7.12%	0.000391%	0.0122%	0.0000912%
HHGCVDRX3-1-ACCGGCCGTA-AATATTGCCA_S36_L002	7,379,576	11.5%	0.000474%	0.0000136%	88.5%	10.8%	9.31%	0.0011%	0.0394%	0.0000813%
HHGCVDRX3-1-CGTCTCATAT-AGCTACTATA_S3_L002	5,157,206	11.5%	0.000756%	0%	88.5%	10.7%	8.93%	0.0014%	0.0358%	0.000427%
HHGCVDRX3-1-ACCGGCCGTA-AATATTGCCA_S36_L001	6,918,535	11.4%	0.000795%	0%	88.6%	10.7%	9.19%	0.00116%	0.041%	0.0000434%
HHGCVDRX3-1-CGTCTCATAT-AGCTACTATA_S3_L001	4,917,122	11.4%	0.000936%	0%	88.6%	10.6%	8.82%	0.00122%	0.0348%	0.000285%
HHGCVDRX3-1-CTAGTGCTCT-TACTGTTCCA_S7_L002	5,701,275	11.4%	0.0016%	0%	88.6%	8.31%	6.55%	0.000368%	0.0175%	0.0000526%
HHGCVDRX3-1-GGTTGCGAGG-TTGCTCTATT_S26_L002	9,214,523	11.4%	0.00158%	0%	88.6%	11%	9.57%	0.0014%	0.0435%	0.000184%
HHGCVDRX3-1-CTAGTGCTCT-TACTGTTCCA_S7_L001	5,464,649	11.3%	0.00135%	0%	88.7%	8.22%	6.46%	0.000421%	0.0161%	0.000146%

Showing 1 to 15 of 96 entries

SMAG catalogue: 40,039 soil MAGs



Soil metagenomes (*n* = 3,304)

- Assembly
- Binning
- Completeness ≥ 50%
- Contamination < 10%

36,398 medium quality MAGs
✓ completeness ≥ 50%
✓ contamination < 10%

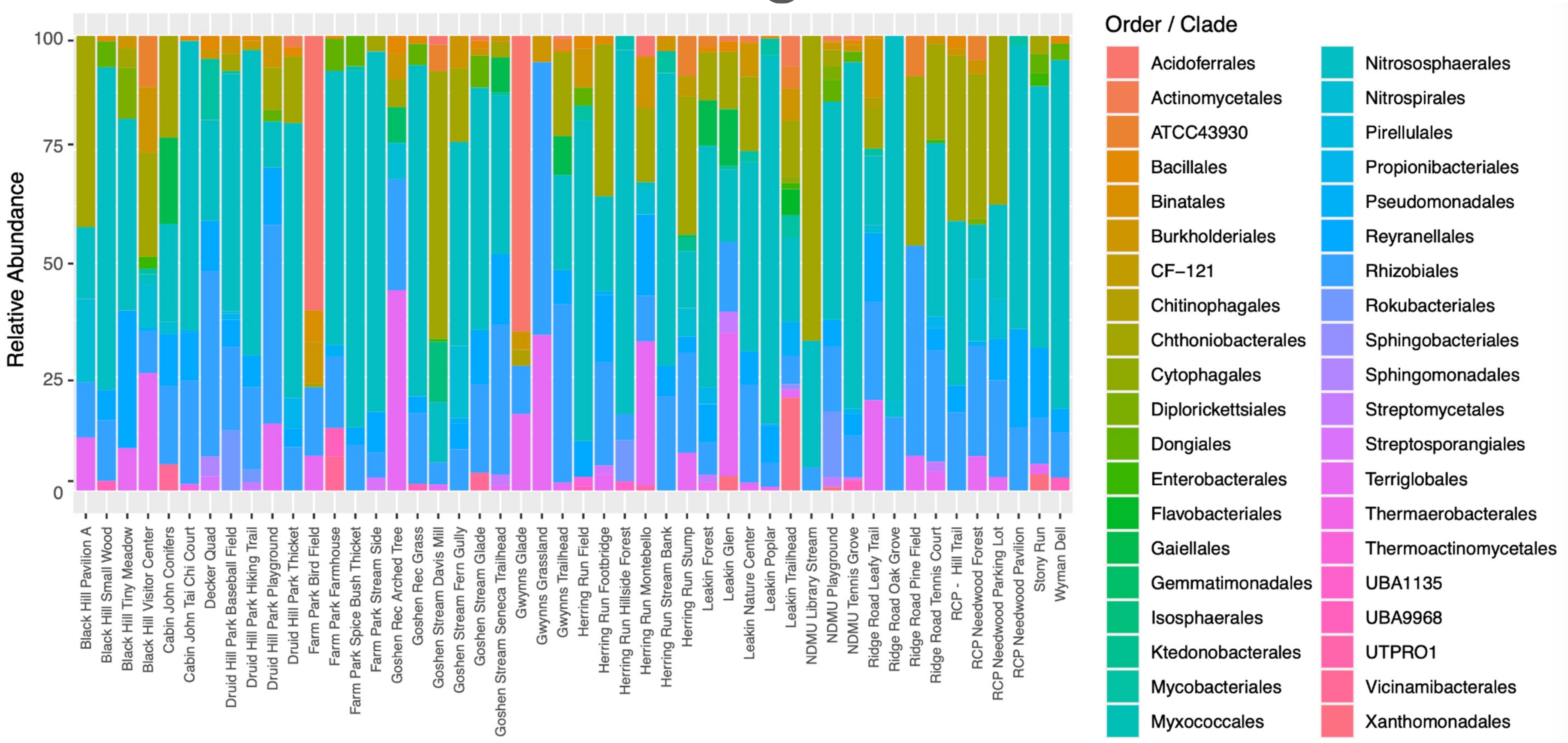
40,039 MAGs

3,641 high quality MAGs
✓ completeness > 90%
✓ contamination < 5%
✓ tRNA number ≥ 18
✓ presence of 23S, 16S and 5/5.8S rRNA

A genomic catalogue of soil microbiomes boosts mining of biodiversity and genetic resources

Ma *et al* (2023) Nat Communication https://doi.org/10.1038/s41467-023-43000-z

Genomic Diversity



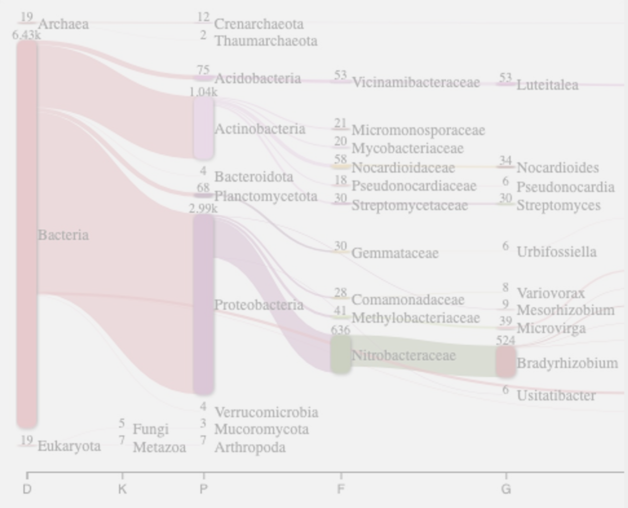
Metagenome profiling and containment estimation through abundance-corrected k-mer sketching with sylph

Shaw and Yu (2023) bioRxiv. doi:10.1101/2023.11.20.567879

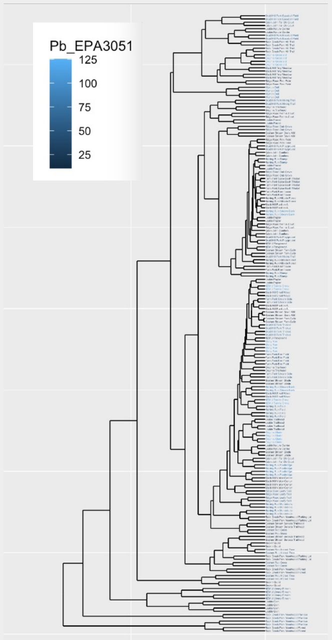
BioDIGS Analysis

GENOMIC DATASCIENCE
COMMUNITY NETWORK

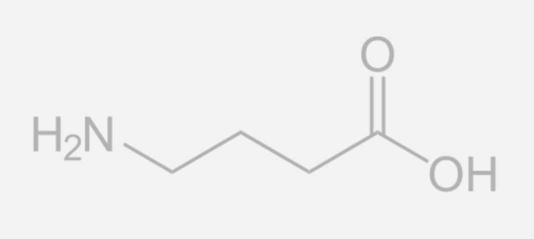
Genomic Diversity



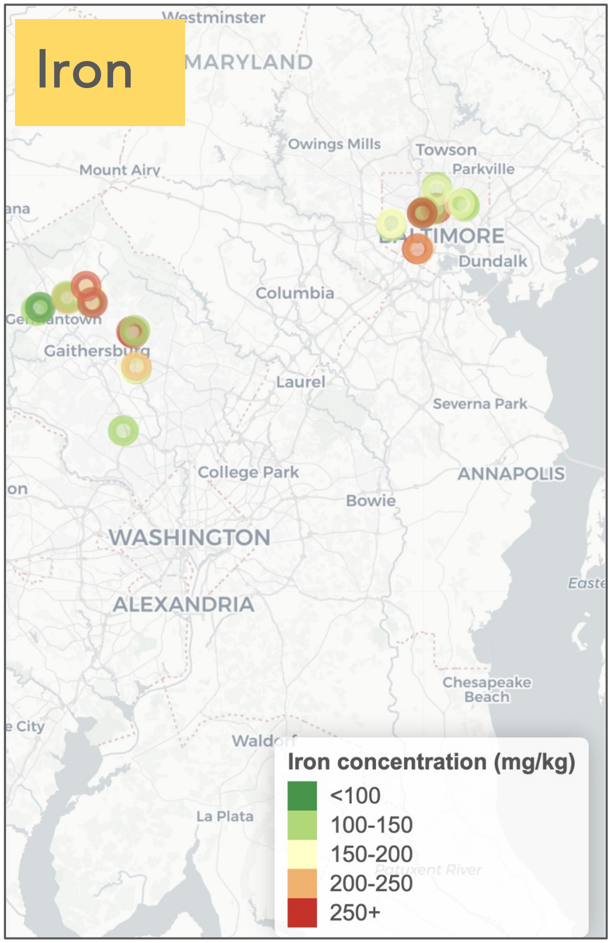
Environmental Associations

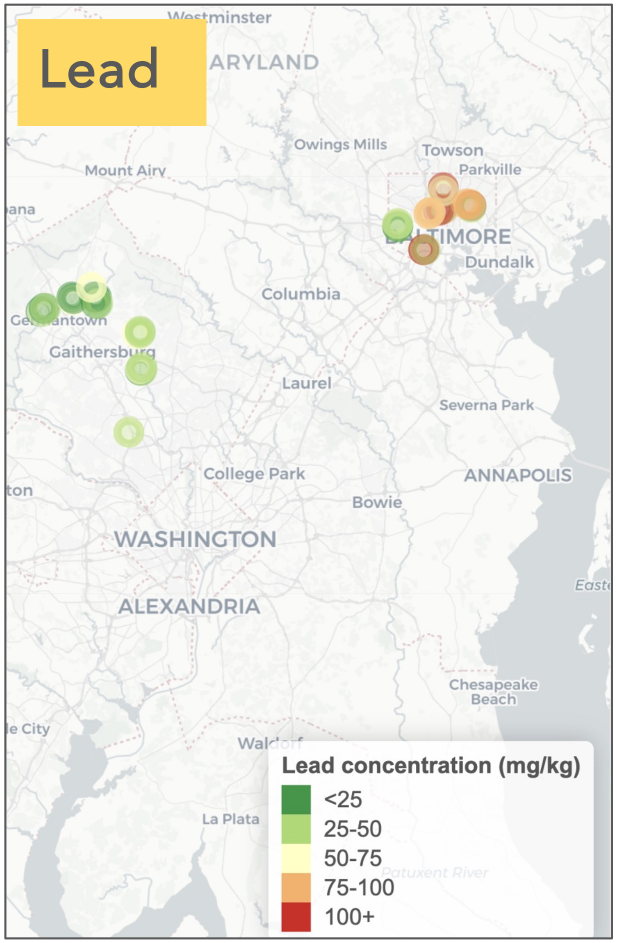


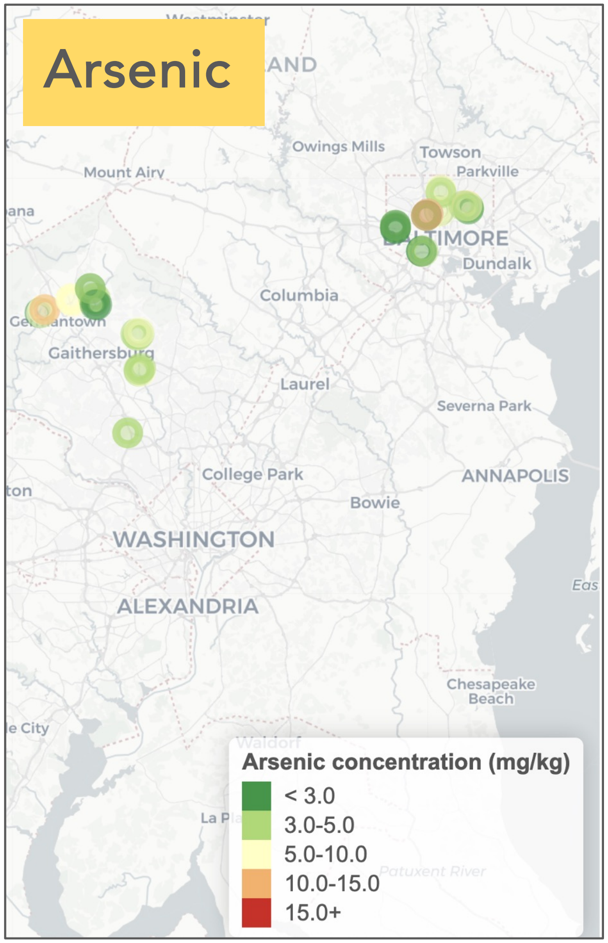
Human Health & Disease



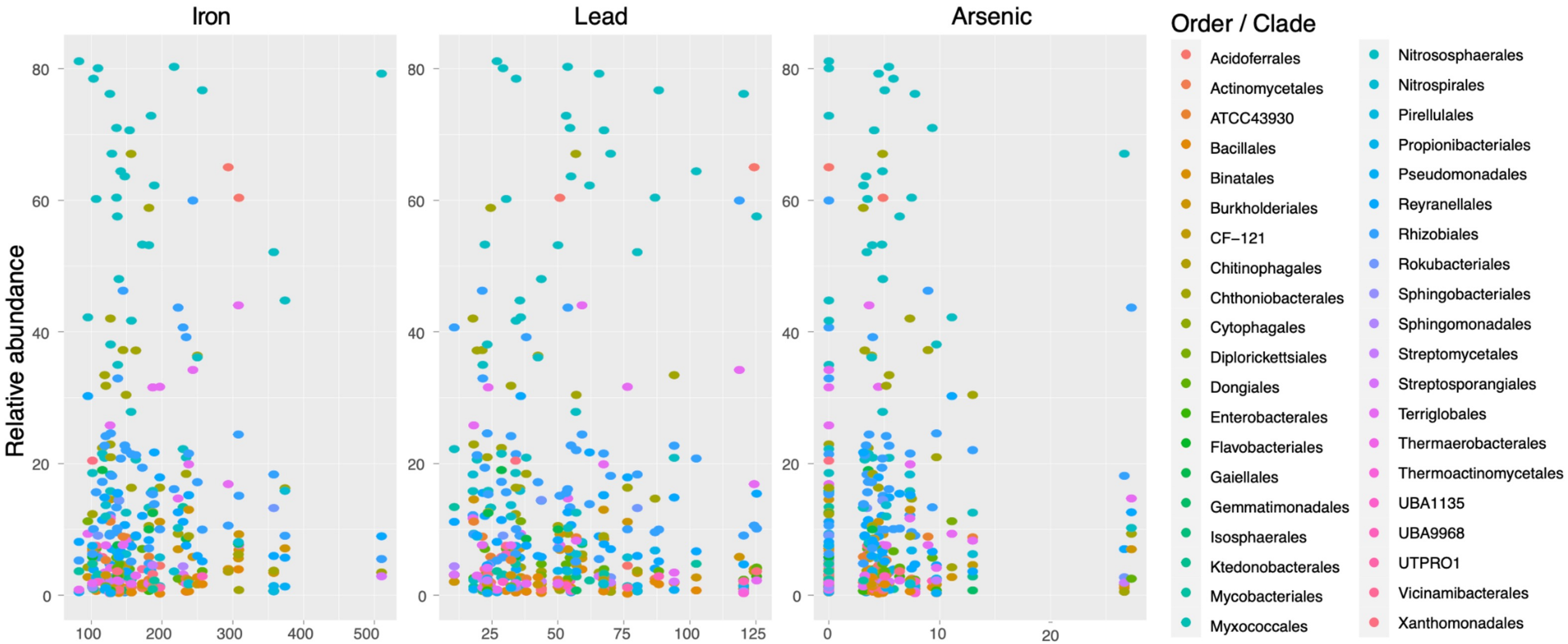
Heavy metal concentration



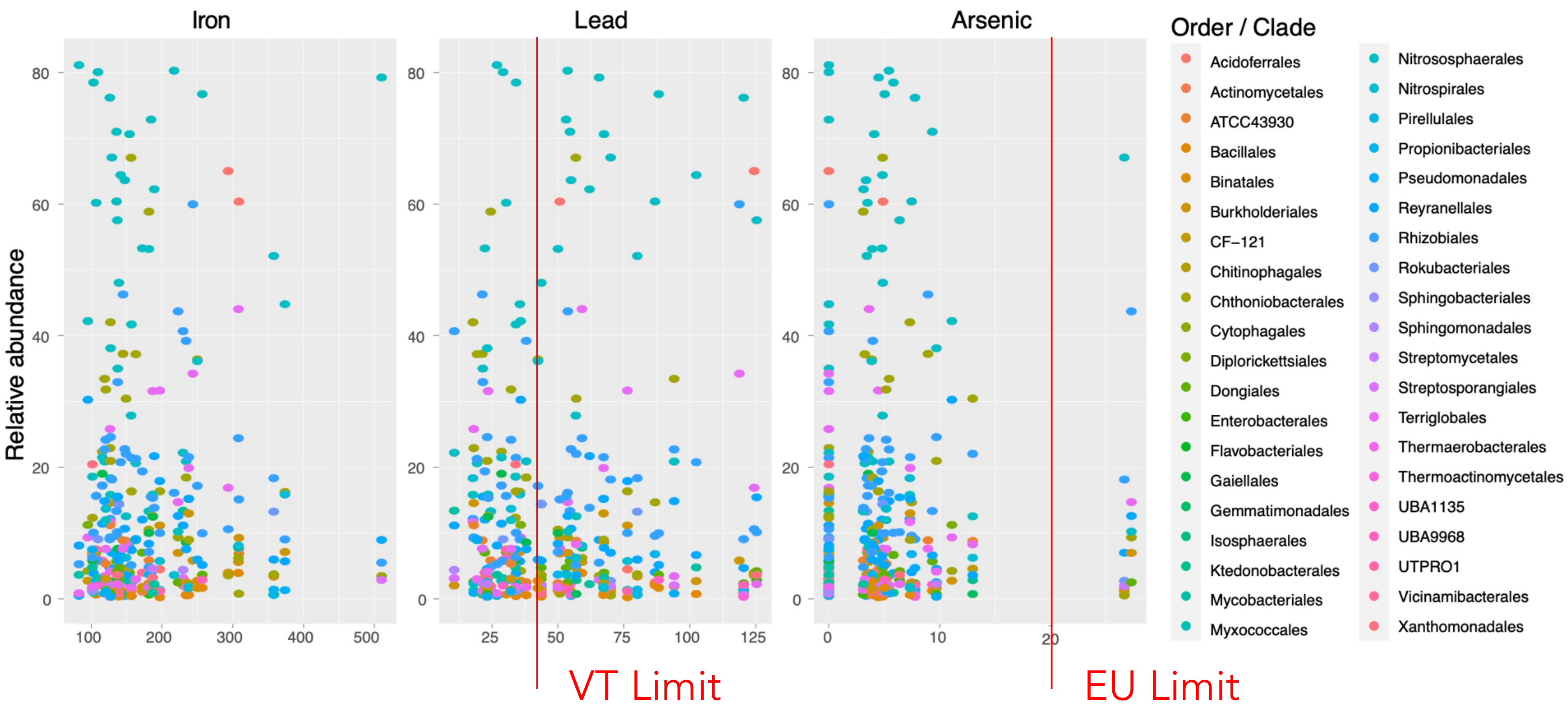




Heavy metal associations



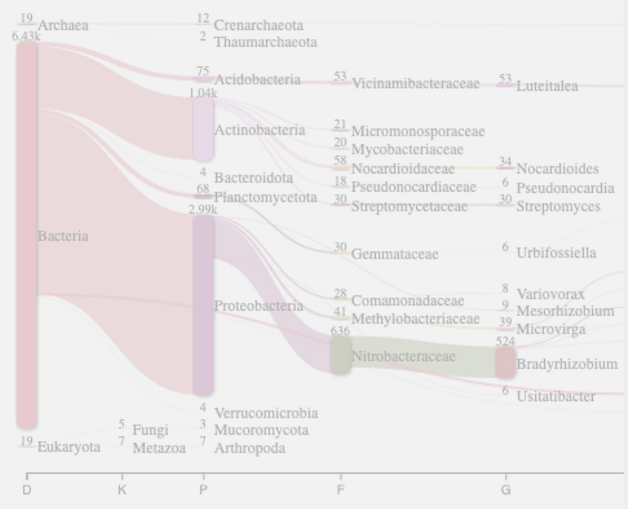
Heavy metal associations



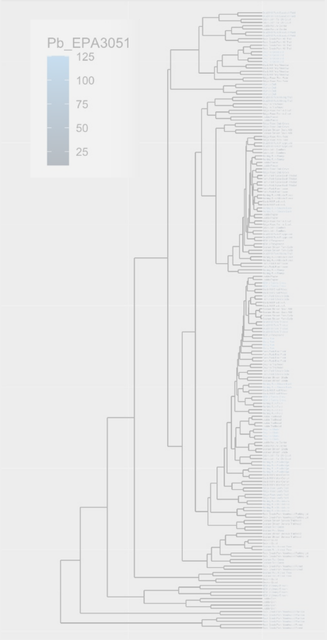
BioDIGS Analysis

GENOMIC DATA SCIENCE
COMMUNITY NETWORK

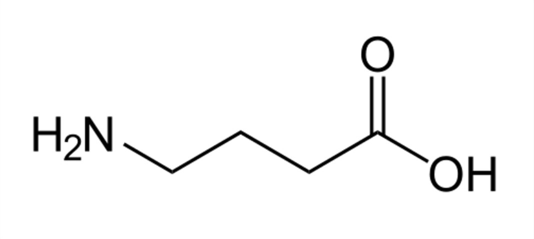
Genomic Diversity



Environmental Associations



Human Health & Disease

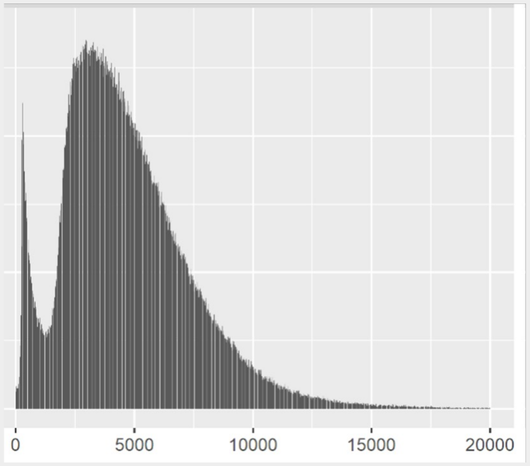


Soil metagenomes and human health

GENOMIC DATA SCIENCE
COMMUNITY NETWORK

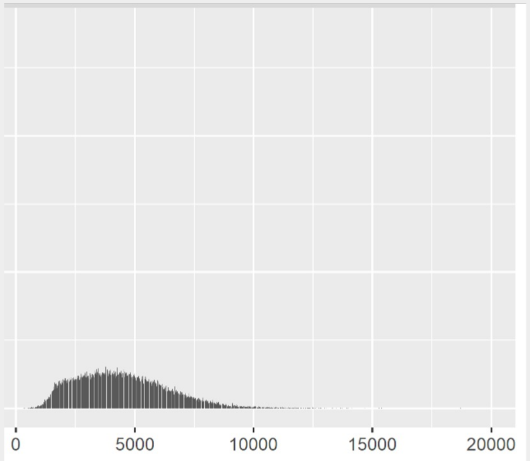
Enter long reads…

Oxford Nanopore PromethION



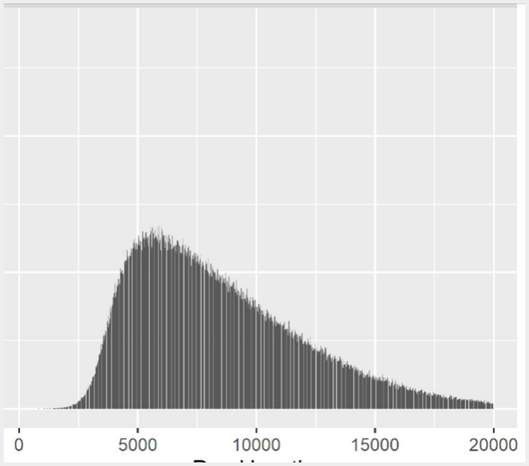
N: 14.82M
Total Yield: 70.2Gb
10kbp+: 8.3Gb
Read N50: 5,788bp

PacBio HiFi Sequel IIe



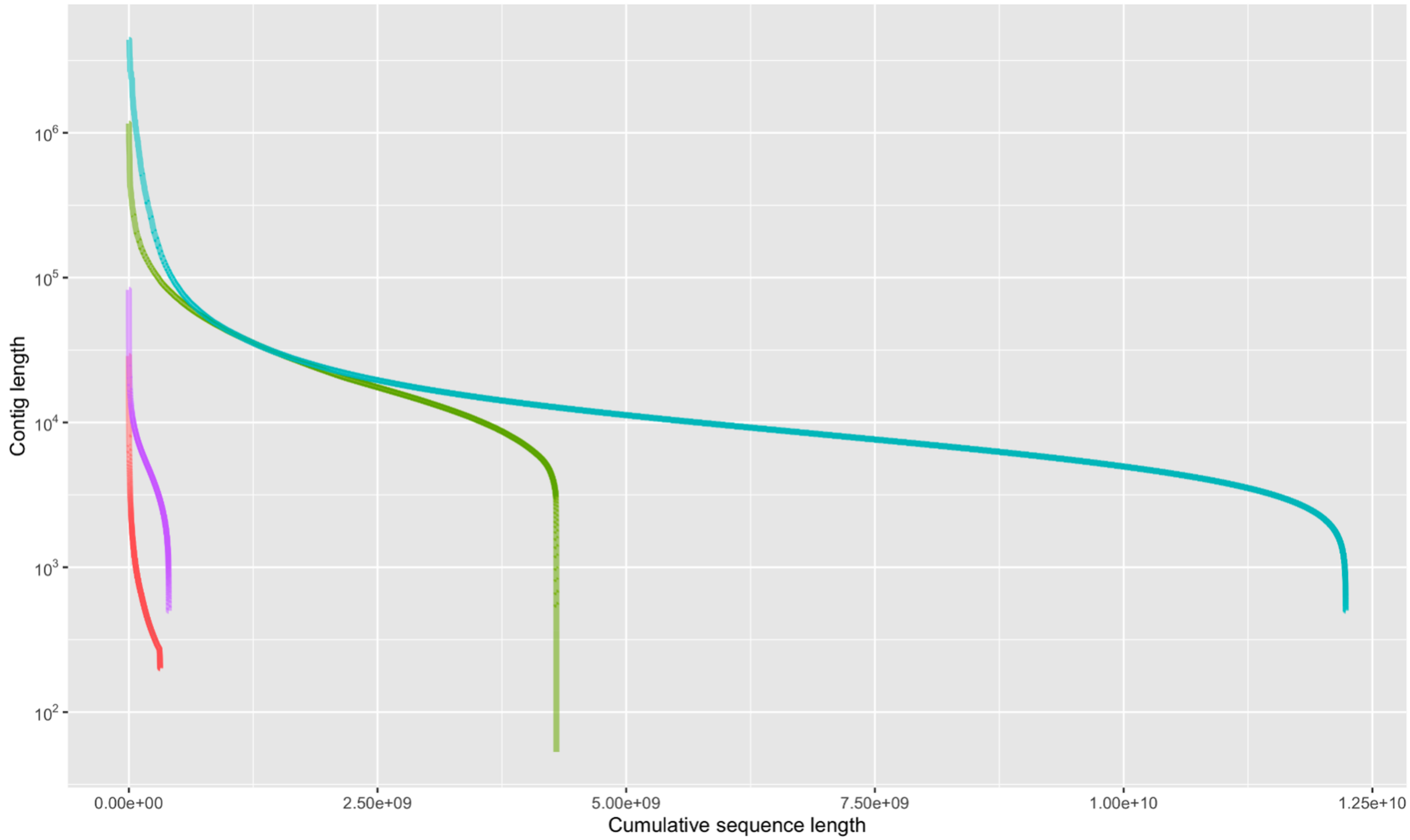
N: 1.31M
Total Yield: 5.9Gb
10kbp+: 0.1Gb
Read N50: 5,180bp

PacBio HiFi Revio



N: 9.97M
Total Yield: 85.2Gb
10kbp+: 39.5Gb
Read N50: 9,600bp

Assembly Results



🟦 Revio HiFi

Total: 10.9 Gb
10kb+: 5.4 Gb
metaMDBG

🟩 Nanopore

Total: 4.3 Gb
10kb+: 3.5 Gb
metaFlye

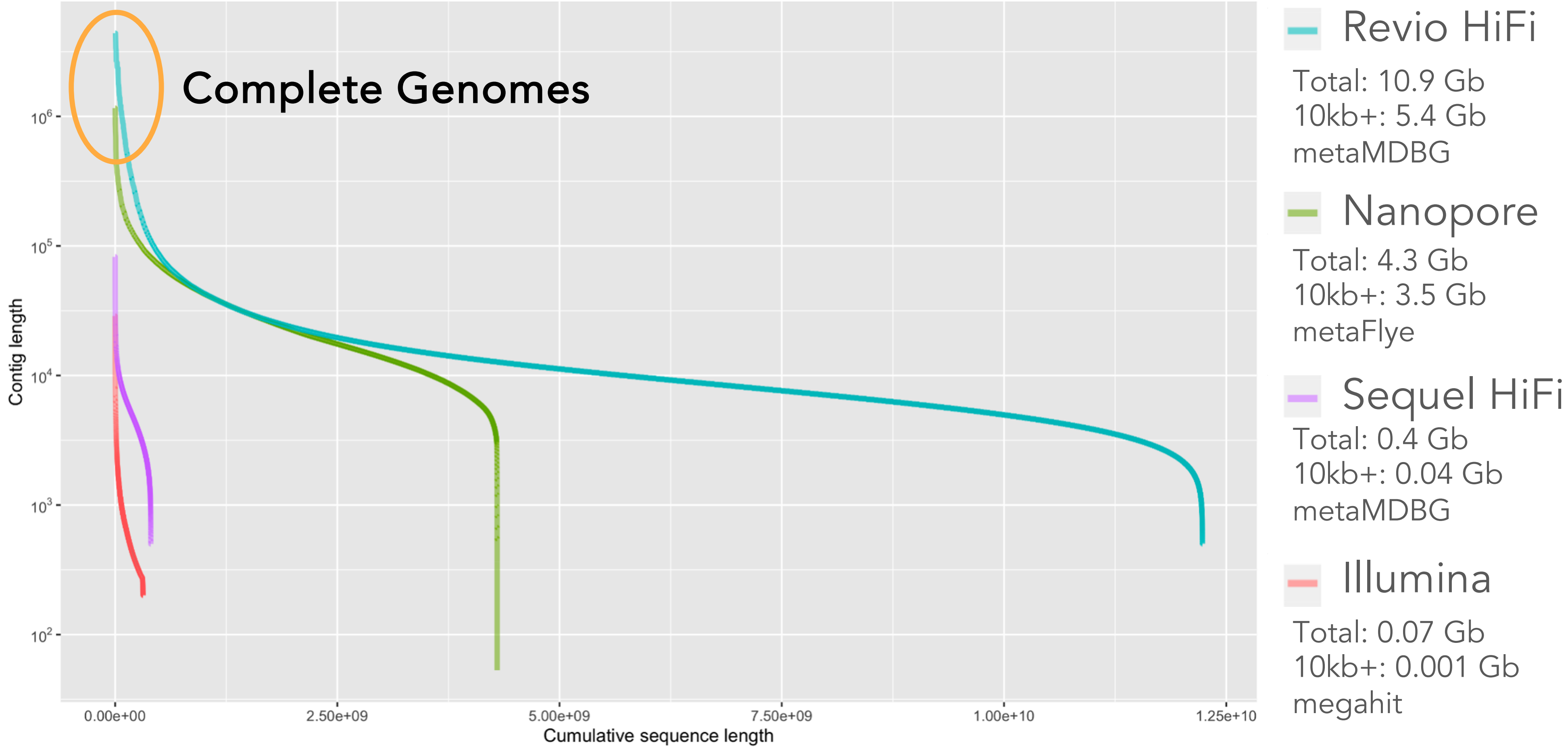
🟪 Sequel HiFi

Total: 0.4 Gb
10kb+: 0.04 Gb
metaMDBG

🟥 Illumina

Total: 0.07 Gb
10kb+: 0.001 Gb
megahit

Assembly Results



Complete Genomes

🟦 Revio HiFi

Total: 10.9 Gb
10kb+: 5.4 Gb
metaMDBG

🟩 Nanopore

Total: 4.3 Gb
10kb+: 3.5 Gb
metaFlye

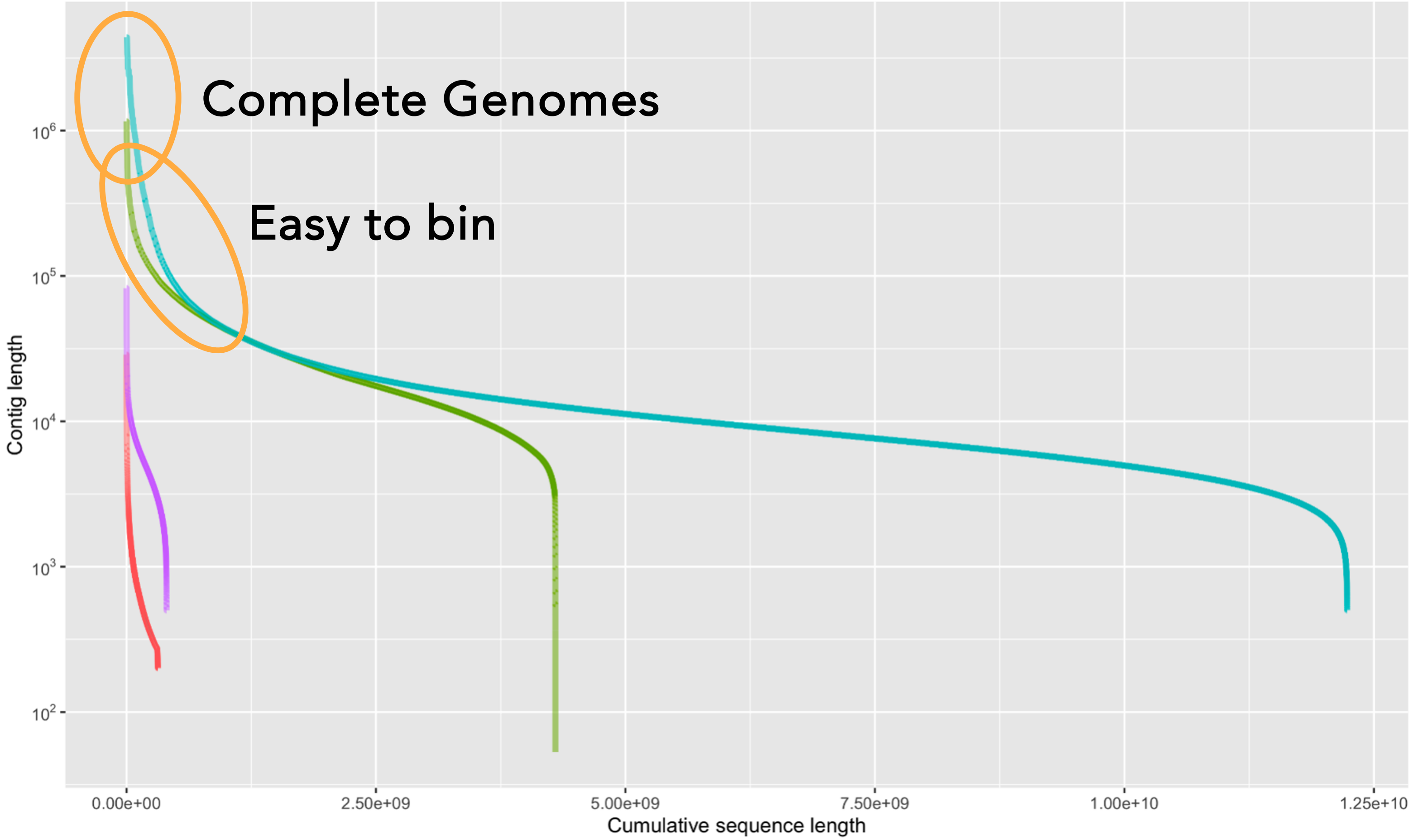
🟪 Sequel HiFi

Total: 0.4 Gb
10kb+: 0.04 Gb
metaMDBG

🟥 Illumina

Total: 0.07 Gb
10kb+: 0.001 Gb
megahit

Assembly Results



Complete Genomes

Easy to bin

🟦 Revio HiFi

Total: 10.9 Gb
10kb+: 5.4 Gb
metaMDBG

🟩 Nanopore

Total: 4.3 Gb
10kb+: 3.5 Gb
metaFlye

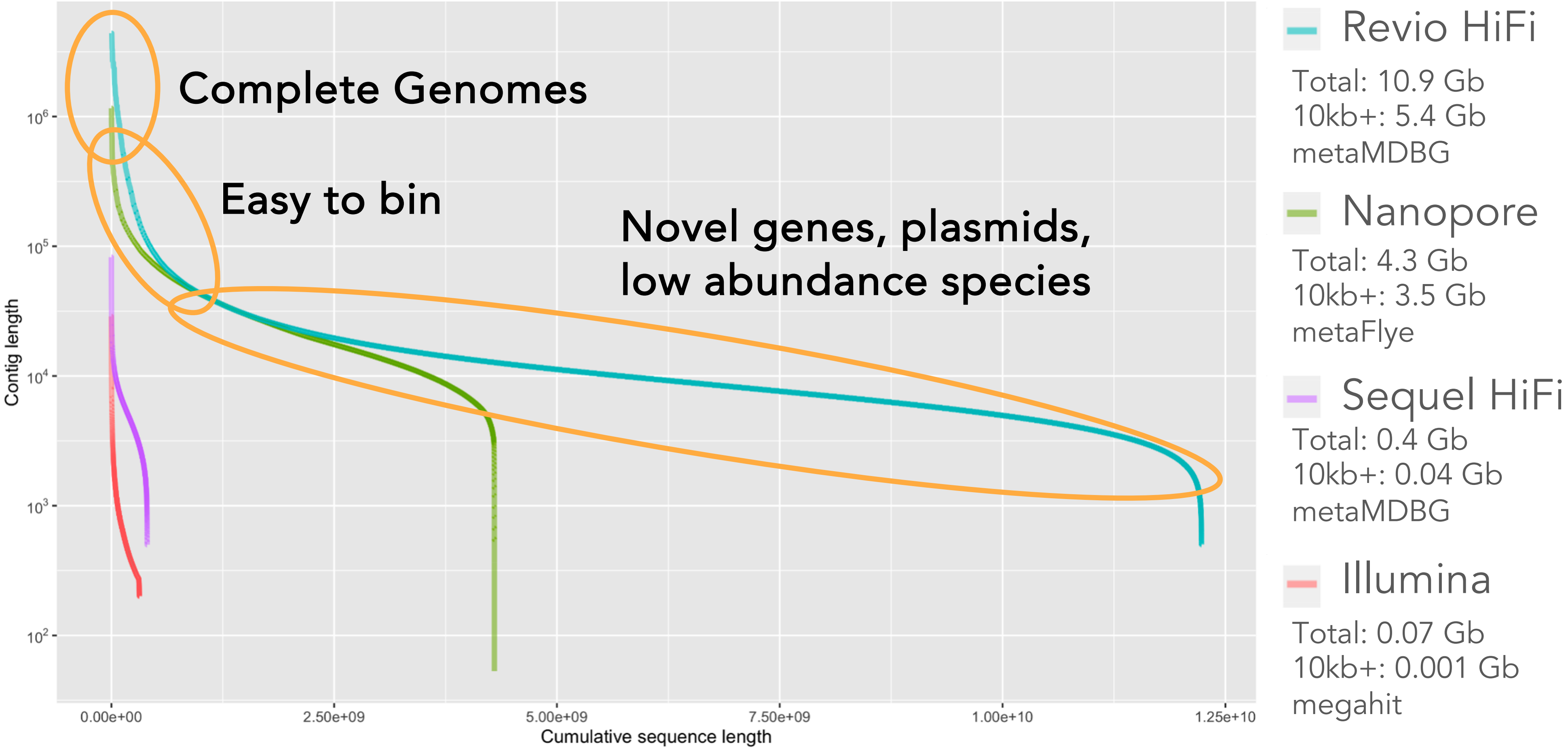
🟪 Sequel HiFi

Total: 0.4 Gb
10kb+: 0.04 Gb
metaMDBG

🟥 Illumina

Total: 0.07 Gb
10kb+: 0.001 Gb
megahit

Assembly Results



Complete Genomes

Easy to bin

Novel genes, plasmids, low abundance species

Revio HiFi

Total: 10.9 Gb
10kb+: 5.4 Gb
metaMDBG

Nanopore

Total: 4.3 Gb
10kb+: 3.5 Gb
metaFlye

Sequel HiFi

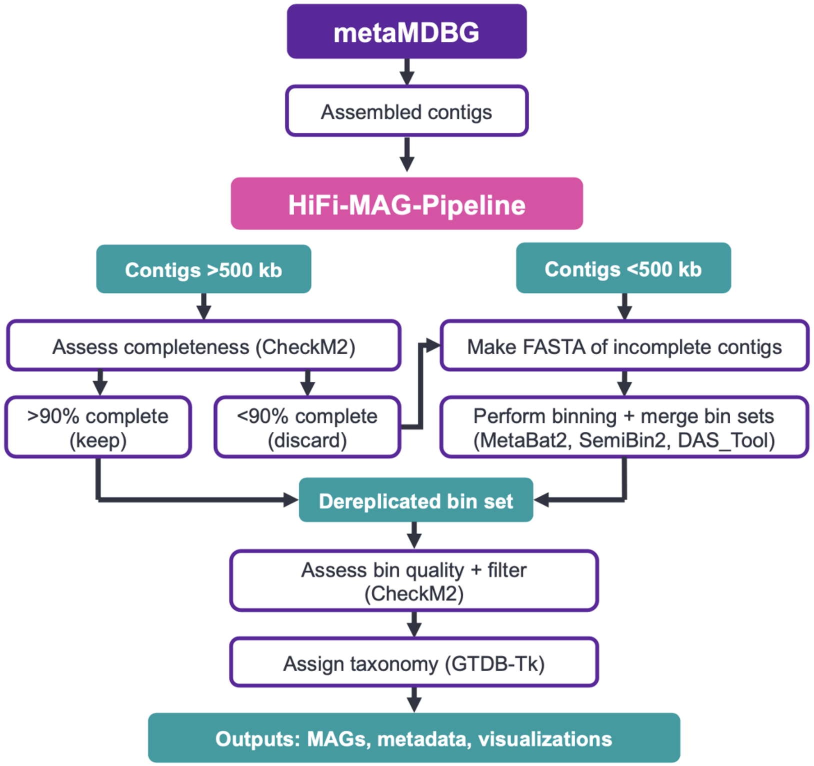
Total: 0.4 Gb
10kb+: 0.04 Gb
metaMDBG

Illumina

Total: 0.07 Gb
10kb+: 0.001 Gb
megahit

PacBio HiFi - MAG Analysis

GENOMIC DATASCIENCE
COMMUNITY NETWORK



metaMDBG

Assembled contigs

HiFi-MAG-Pipeline

Contigs >500 kb

Contigs <500 kb

Assess completeness (CheckM2)

>90% complete (keep)

<90% complete (discard)

Make FASTA of incomplete contigs

Perform binning + merge bin sets (MetaBat2, SemiBin2, DAS_Tool)

Dereplicated bin set

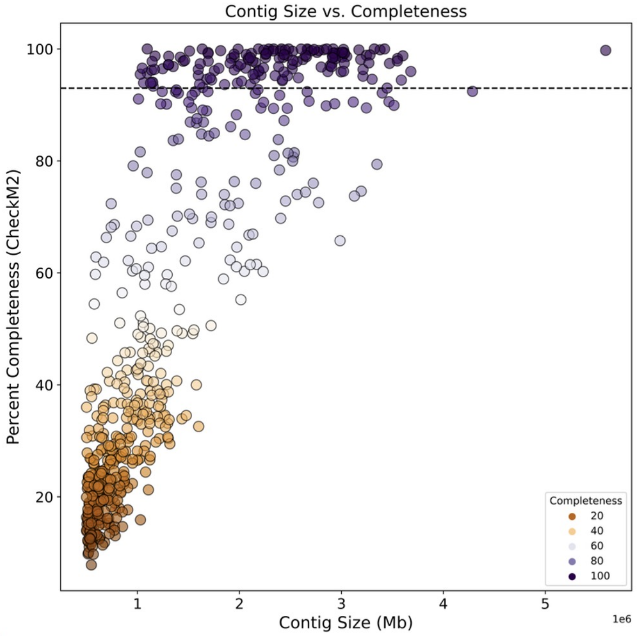
Assess bin quality + filter (CheckM2)

Assign taxonomy (GTDB-Tk)

Outputs: MAGs, metadata, visualizations

Analysis: Dan Portik, PacBio

https://github.com/PacificBiosciences/pb-metagenomics-tools

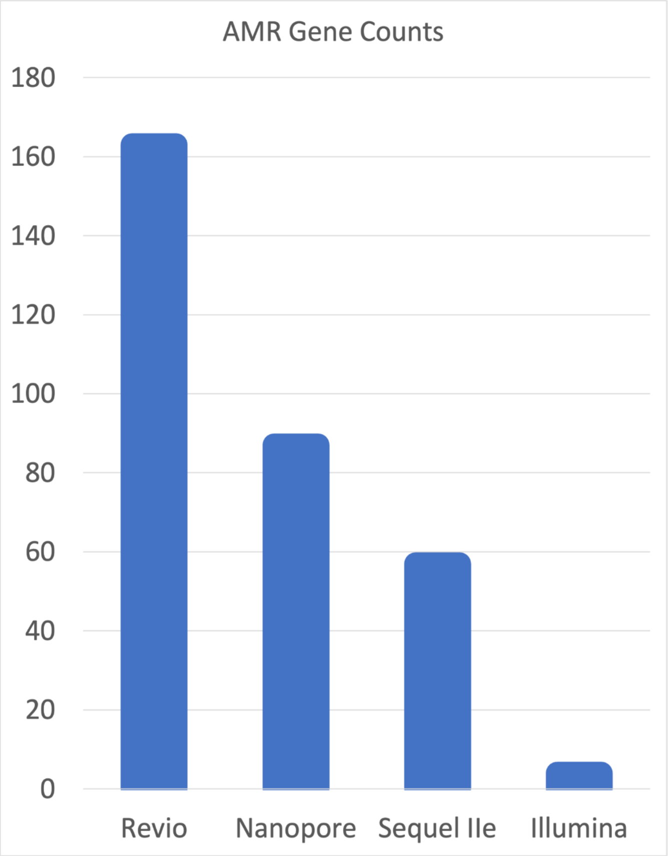


Out of 158 MAGs (55 HQ), only 1 could be assigned to a known species

- 27 single contig genomes!
- ~25x more HQ MAGs / site than short reads
- Archaea are highly represented

Antimicrobial Resistance

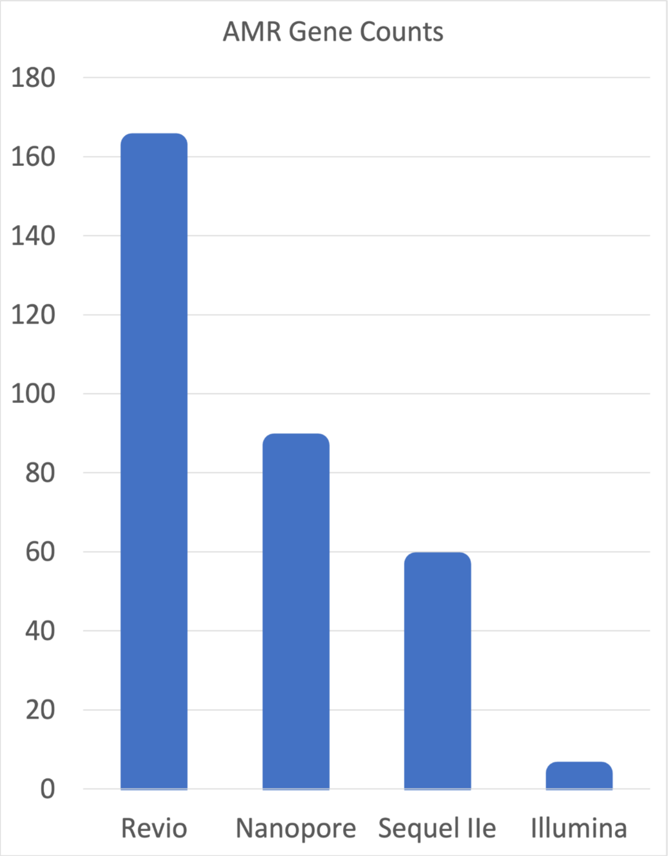
GENOMIC DATASCIENCE
COMMUNITY NETWORK

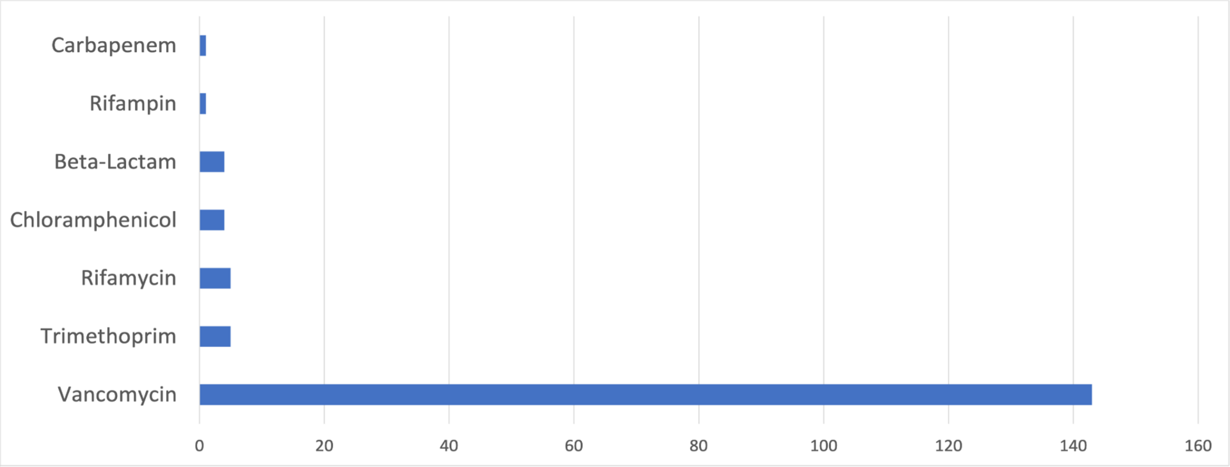


https://github.com/tseemann/abricate

Antimicrobial Resistance

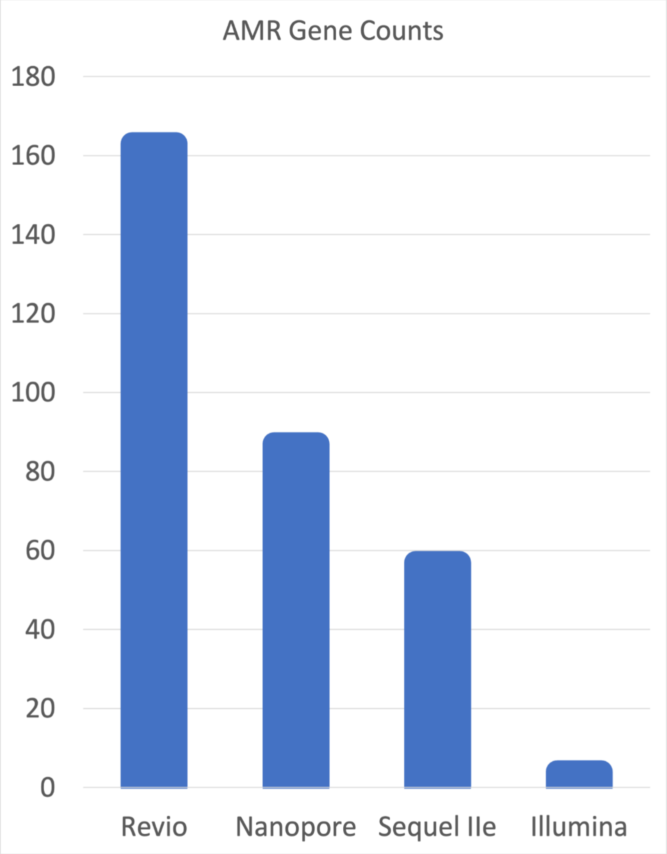
GENOMIC DATASCIENCE
COMMUNITY NETWORK



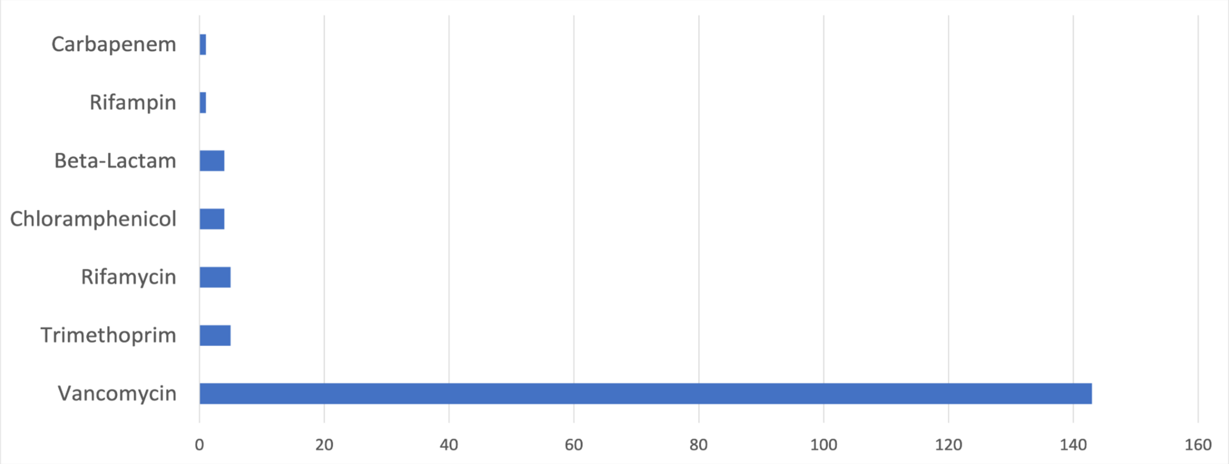


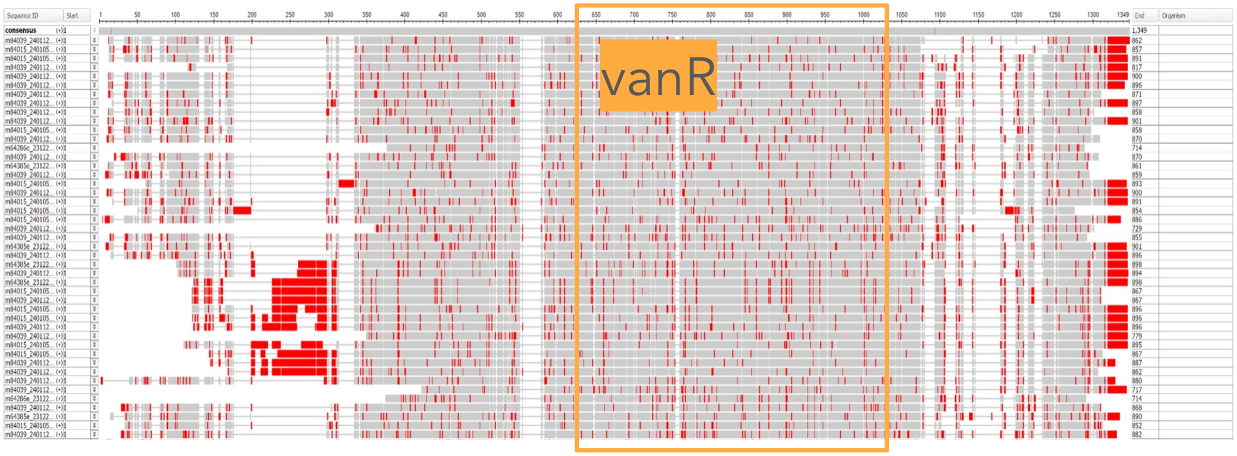
https://github.com/tseemann/abricate

Antimicrobial Resistance



https://github.com/tseemann/abricate

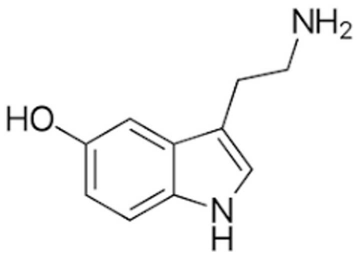




Metabolites in Humans & Microbes

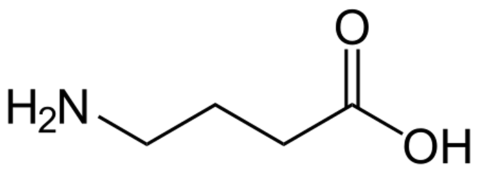
The neuroactive potential of the human gut microbiota in quality of life and depression
Valles-Colomer et al. (2019) Nature Microbiology. doi:10.1038/s41564-018-0337-x

Metabolites in Humans & Microbes



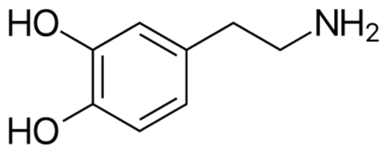
Serotonin

mood, digestion, and sleep



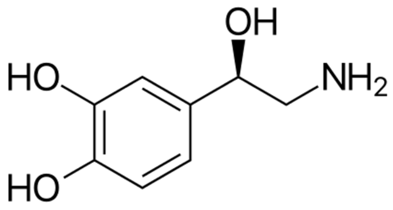
GABA

chief inhibitory neurotransmitter



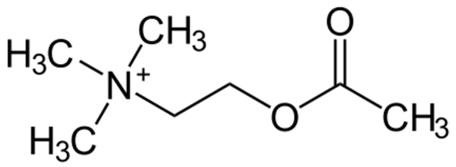
Dopamine

neurotransmitter for “pleasure”



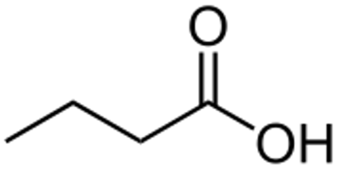
Norepinephrine

Fight or flight response



Acetylcholine

Motor neuron neurotransmitter



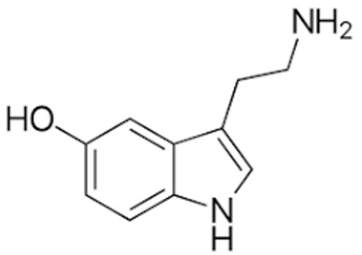
Butyrate

host immune homeostasis

The neuroactive potential of the human gut microbiota in quality of life and depression
Valles-Colomer et al. (2019) Nature Microbiology. doi:10.1038/s41564-018-0337-x

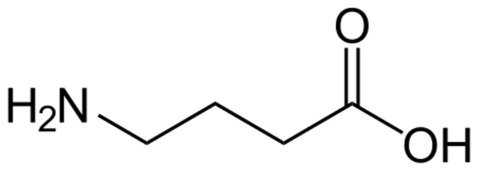
Metabolites in Humans & Microbes

GENOMIC DATASCIENCE COMMUNITY NETWORK



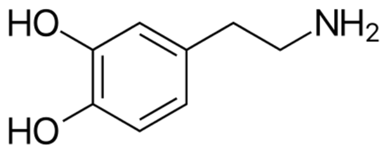
Serotonin

mood, digestion, and sleep



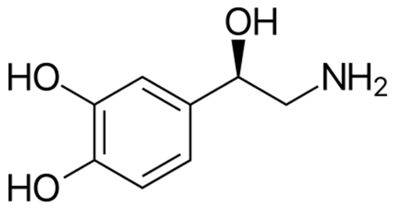
GABA

chief inhibitory neurotransmitter



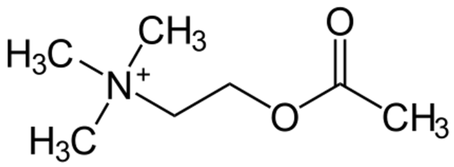
Dopamine

neurotransmitter for “pleasure”



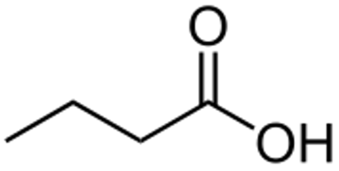
Norepinephrine

Fight or flight response



Acetylcholine

Motor neuron neurotransmitter



Butyrate

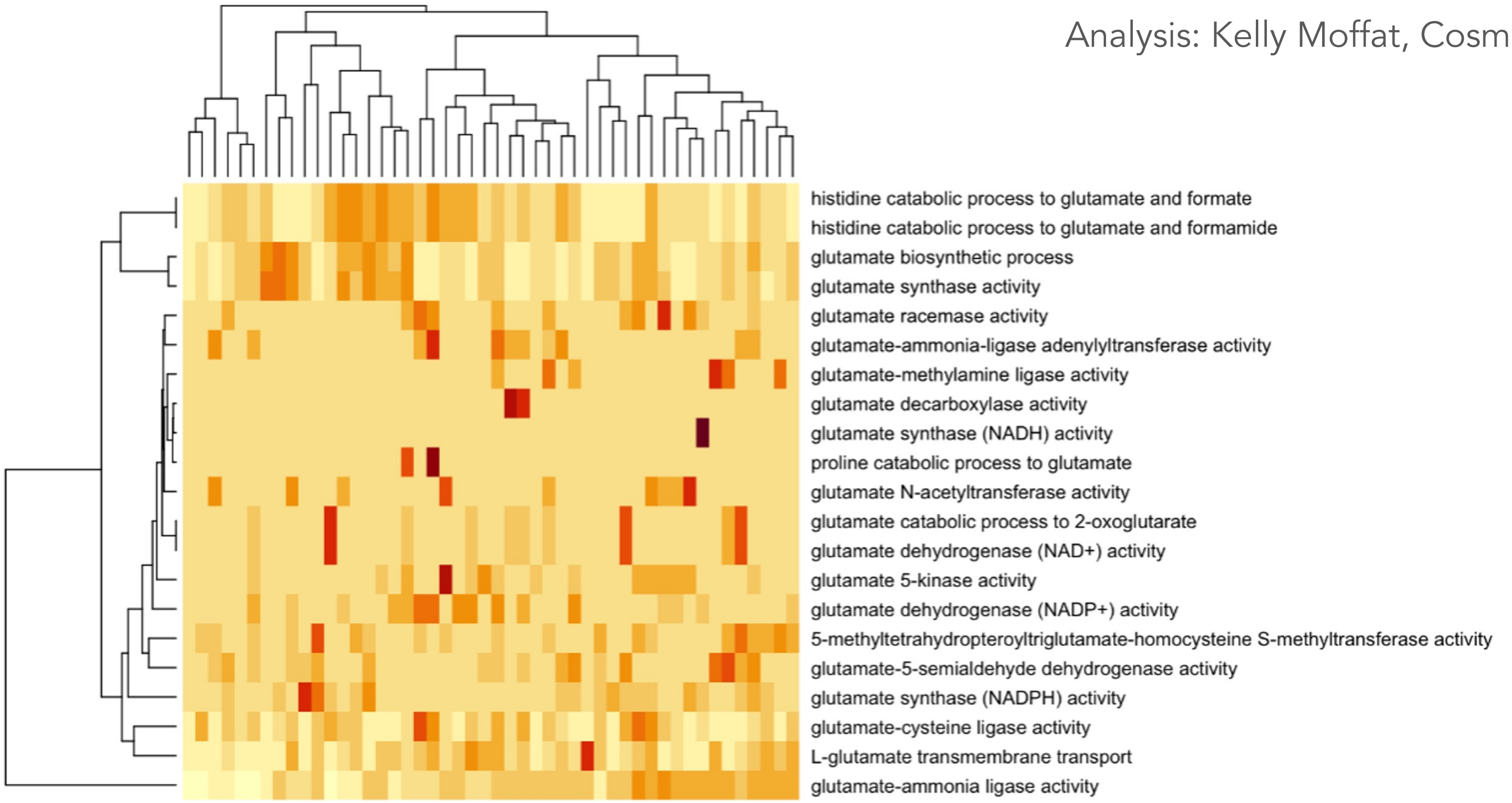
host immune homeostasis

The neuroactive potential of the human gut microbiota in quality of life and depression
Valles-Colomer et al. (2019) Nature Microbiology. doi:10.1038/s41564-018-0337-x

GABA pathway abundances

GENOMIC DATASCIENCE
COMMUNITY NETWORK

Analysis: Kelly Moffat, CosmosID



Species-level functional profiling of metagenomes and metatranscriptomes (HUMAnN 2.0)
Franzosa et al (2018) Nature Methods. doi: doi.org/10.1038/s41592-018-0176-y

Summary

GENOMIC DATA SCIENCE
COMMUNITY NETWORK

- Exploring the interaction between soil, metagenomic diversity, and human health
  - Discover new genomes and genes
  - Discover new antimicrobial resistance genes, neurotransmitters present in the soil
  - Orders of magnitude improvements from long reads
- Empower the next generation of scientists
  - Teach state-of-the-art genomics & data science
  - Provide open access data & compute for community research

Summary

- Exploring the interaction between soil, metagenomic diversity, and human health
  - Discover new genomes and genes
  - Discover new antimicrobial resistance genes, neurotransmitters present in the soil
  - Orders of magnitude improvements from long reads
- Empower the next generation of scientists
  - Teach state-of-the-art genomics & data science
  - Provide open access data & compute for community research
- Next Steps: Dynamics across space & time!
  - More institutions, longitudinal analysis
  - Exposures, Climate, Health data, etc
  - Training materials

We need your help!

linktr.ee/biodigs

Student Acknowledgements

Montgomery County
Madeline Graham, Daniel Chin, David Soussana

Baltimore
Loraye Smith, Tyler Smith, Madeline Graham

Seattle
Matheus Fernandes, Carl Pontino, Randon Serikawa, Joelle Taganna

Clovis Community College
Malachi Whitford, Grace Freymiller, Domonique Advincula, Troy Burgess, Janet Castillo, Jennifer Elziade, Dorthy Esparza, Nicholas Foreman, Ana Hernandez, Glenda Medina, Christina Munoz, Nicole Potter, Quince Quintana, Nickie Ruiz, Ryan Wilder, Orion York

Meharry Medical College
Lincoln Liburd II, Sydney Jamison, Destiney Ball, Claude Albritton, Arjun Pratap

Virginia State University
Michael Marone

Northern Virginia Community College
Rachel Marie Ametin, Joceph Duncan, Noha Elnawam, Sarah-Leila Kaci

University of Texas - El Paso
Frida Delgadillo, Armando Jimenez, Keyan Ozuna

El Paso Community College
Efren Barragan, Faith Chanhuhwa, Tania Da Silveira, Marco Ferrel, Josh Samuel Ikechi-Konkwo, Olivia Kelly, America Pinela, Ryley Stewart

Spelman College
Natajha Graham, Nia Davis, Katherine Ulbricht

University of Hawaii - Mānoa
Sudhir Kumar Rai, Yujia Qin, Ba Thong Nguyen, Mohammadamin Mahmanzar, Yu Chen, Isam Mohd Ibrahim, Donna Lee Kuehu, Asmita Pandey

Clovis Community College

UTEP

VSU VIRGINIA STATE Since 1882

NOTRE DAME OF MARYLAND UNIVERSITY 1895

El Paso Community College

Spelman College®
A Choice to Change the World

MEHARRY® MEDICAL COLLEGE

UNIVERSITY of HAWAIʻI® MĀNOA

NOVA Northern Virginia Community College

Acknowledgements

GDSCN Faculty

Andrew Lee, Northern Virginia Community College
Brandi Kamermans, Northwest Indian College
Edu Suarez Martinez, University of Puerto Rico
Emily Biggane, United Tribes Technical College
James Sniezek, Montgomery College
Joslynn Lee, Fort Lewis College
Kamal Chowdhury, Claflin University
Karla Fuller, Guttman Community College
Loyda B. Méndez, University of Puerto Rico at Aguadilla
Lyle Best, Turtle Mountain Community College
Maria Alvarez, El Paso Community College
Mentewab Ayalew, Spelman College
Michael Campbell, University of Southern California
Michele Nishiguchi, University of California Merced
Miguel P Mendez Gonzalez, University of Puerto Rico
Peter Vos, Montgomery College
Rachel Saidi, Montgomery College
Robert Meller, Morehouse School of Medicine
Rosa Alcazar, Clovis Community College
Shazia Tabassum Hakim, Dine College
Sidd Pratap, Meharry Medical College
Sourav Roy, University of Texas at El Paso
Xianfa Xie, Virginia State University
Youping Deng, University of Hawaii at Manōa

GDSCN Organizers

Johns Hopkins University: Michael Schatz, Lance O'Connor, Matthew Nguyen, Stephen Mosher, Natalie Kucher, Tyler Collins, Alex Ostrovsky

Fred Hutch Cancer Research Center: Jeffrey Leek, Ava Hoffman, Elizabeth Humphries, Carrie Wright

Carnegie Institute for Science: Frederick Tan

National Human Genome Research Institute: Shurjo Sen, Christina Daulton, Valentina Di Francesco

JHU GRCF

David Mohr, Alejandra Gutierrez, Samantha Zimmer, Alan Scott, Kim Doheny

PacBio

Mark Van Oene, Kelvin Liu, Jeffrey Burke, Primo Baybayan, Michelle Kim, Dan Portik, Jeremy Wilkinson, Trang Dahlen, Jonas Korlach

CosmosID

Kelly Moffat, Manoj Dadlani, Rita Colwell

Tiny Earth

Jennifer Kerr, Nichole Broderick, Sarah Miller

National Human Genome Research Institute

COSMOSID
Unlocking The Microbiome

PacBio

Oxford NANOPORE Technologies

tiny earth

Bloomberg Professors

Mentewab (Ment) Ayalew, Ph.D.
Associate Professor of Biology
Spelman College. Atlanta, GA

Emily Biggane, Ph.D.
Research Faculty
Environmental Toxicology
United Tribes Technical College
Bismarck, ND

Rosa Alcazar, Ph.D.
Biology Faculty
Clovis Community College
Fresno, CA

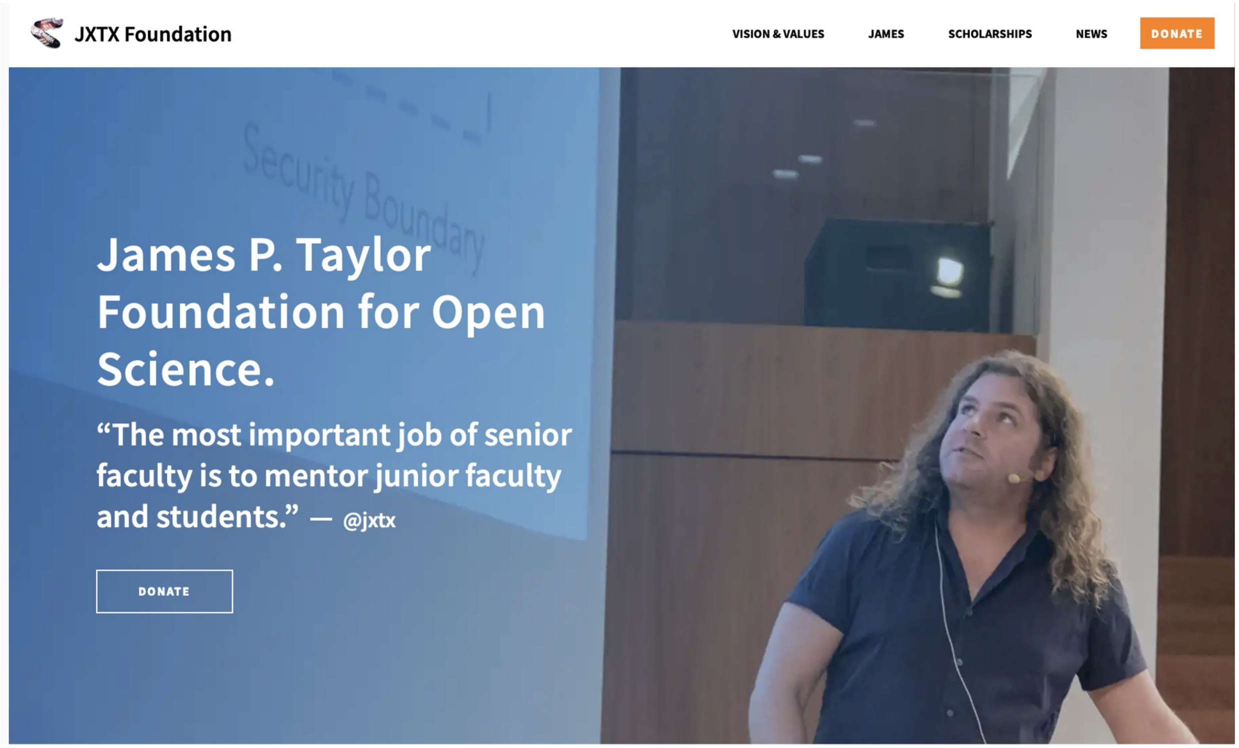
Natalie Kucher
Program Manager
Johns Hopkins University
Baltimore, MD

Kristen L Rhinehardt, Ph.D.
Assistant Professor
Computational Science & Engineering
North Carolina Agricultural and Technical
State University, Winston-Salem, NC

Siddharth (Sidd) Pratap, Ph.D.
Director of Bioinformatics
Meharry Medical College
Nashville, TN

Ava Hoffman, Ph.D.
Senior Staff Scientist
Fred Hutch Cancer Center
Seattle, WA

>BioDIGS
a GDSCN project



JXTX Foundation

VISION & VALUES JAMES SCHOLARSHIPS NEWS DONATE

James P. Taylor Foundation for Open Science.

“The most important job of senior faculty is to mentor junior faculty and students.” — @jxtx

DONATE

Thank you!

http://biodigs.org

Fre
Can
Arnold
Thoma
Visitor
1100 Fairvi
NOBODY CARES
WORK HARDER
DRI
ANVIL