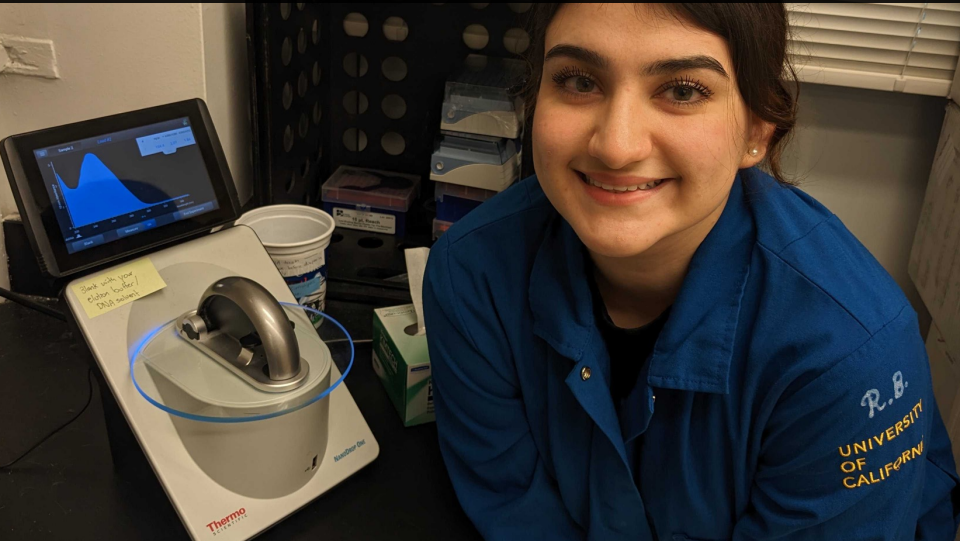




Integrating Undergraduate Researchers into Genome Analysis of Bacterial Culture Collections

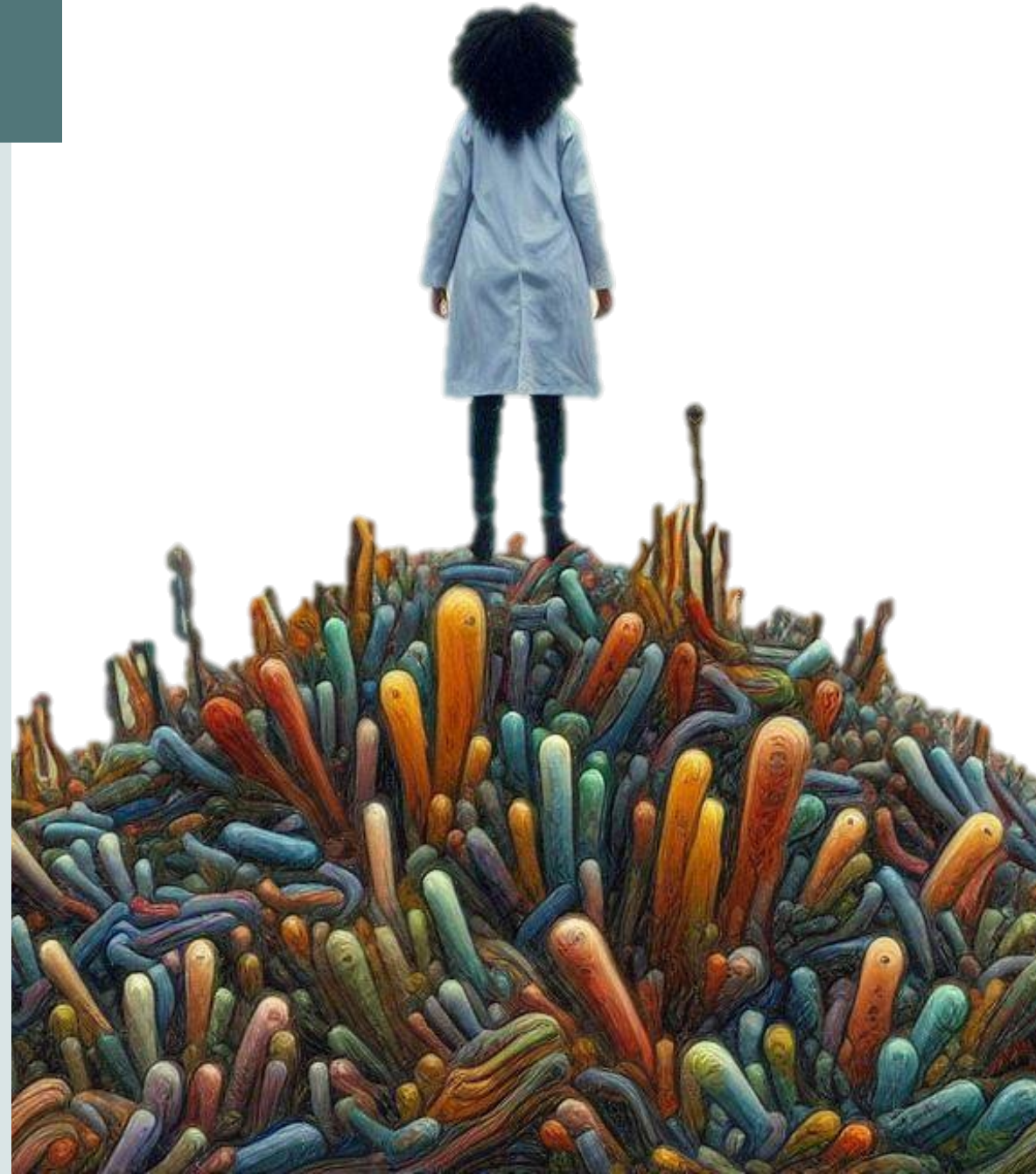
Tiffany Lowe-Power
Asst Prof, UC Davis Plant Pathology



Take a Moment to Reflect:

How did you become a researcher?

- Role Models?
- Mentors?
- Research Opportunities?
- Identity as a Scientist?



My Story of Privilege and Opportunity:

- **Mentor 1** (Sam Parks): High School Bio Teacher who joined a PhD program
 - Connected me to a senior year research opportunity in an environmental micro lab - **identity as a researcher**
- **Mentor 2** (Dr. Leavey): Undergrad Bio Lecturer encouraged me to apply to NSF research experiences for undergrads (REUs).
 - Participated in 2 cohort opportunities (**network!**) with programming to **demystify hidden curriculum** of academia



My Story of Privilege and Opportunity:

- **Mentor 1** (Sam Parks): High School Bio Teacher who joined a PhD program
 - Connected me to a senior year research opportunity in an environmental micro lab - **identity as a researcher**
- **Mentor 2** (Dr. Leavey): Undergrad Bio Lecturer encouraged me to apply to NSF research experiences for undergrads (REUs).
 - Participated in 2 cohort opportunities (**network!**) with programming to **demystify hidden curriculum** of academia
- **My aspirational goal:**
"I want become a professor in order to train the next generation of Scientists"

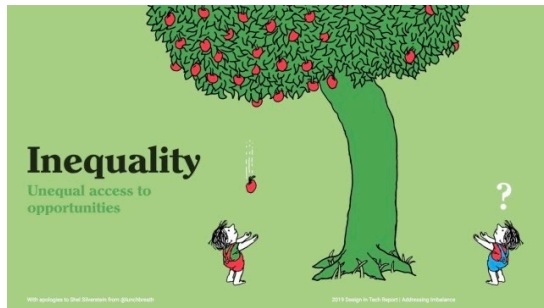


Outline:

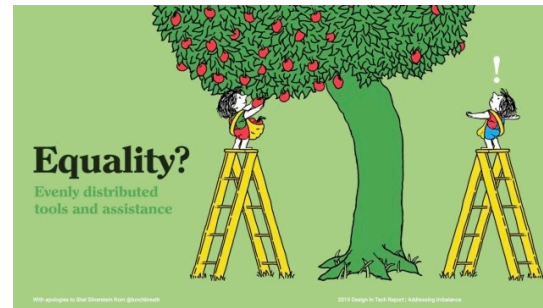
- **Cohort-based undergraduate research experiences (CUREs):**
A semi-scalable approach to training the next generation of scientists
- Development of the *Ralstonia* genomics CURE
- KBase as a platform for bacterial genomics/metagenomics CUREs
- Contributions of *Ralstonia* genomics CURE to scientific knowledge



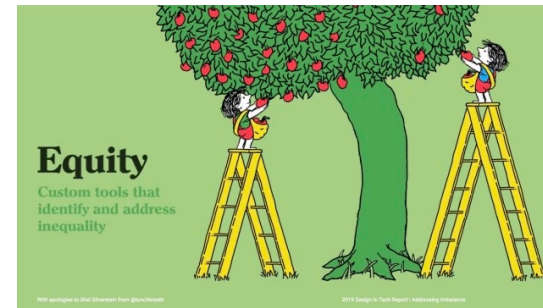
How to equitably train the next generation of scientists?



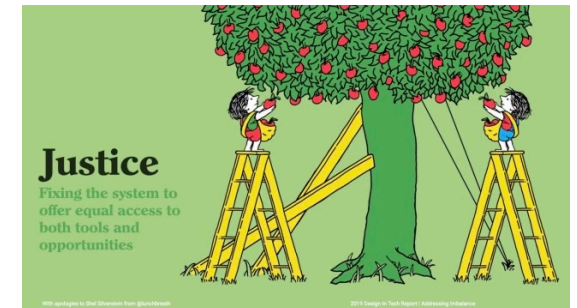
Rely on undergraduates to contact labs



Research requirement for all students



-Paid research
-Peer networks



How to equitably train the next generation of scientists?

Cohort-based Undergrad Research Experiences (**CUREs**)*

Elements of a CURE:

- Learn a variety of research skills
- Uncover the unknown through real research
- Learn, Apply, REPEAT
- Communication & collaboration



* CURE is usually "Class-based", but I am avoiding the paperwork of formalizing my CURE...

Examples of CUREs



Students **isolate bacteriophages** against BSL-1 pathogens. **Sequence bacteriophage genomes** and deposit on NCBI



Students isolate bacteria that produce **antibiotics** that inhibit BLS-1 pathogens. Identify strains by 16S rRNA sequencing and identify the antibiotic (chemistry CURE)



Students **assign function to proteins** by investigating protein structure data and purifying/assaying proteins



Student (competitors) **genetically engineer** microbes for synthetic biology goals

Ingredients for a fruitful and sustainable CURE

Friendly to beginners who are learning by making mistakes

- Safe
- Not too expensive
- Doesn't require specialized technical skill

Scalable Projects

- Standardized methods on a parallel research objective/question
- Students get ownership over their work, but they can learn from each other.
- Doesn't stretch the mentors too thin

Useful to research community

- Connection to widely used databases (NCBI, PDB, etc)
data generation or data analysis
- Published papers... if time/energy/funding permits

Outline:

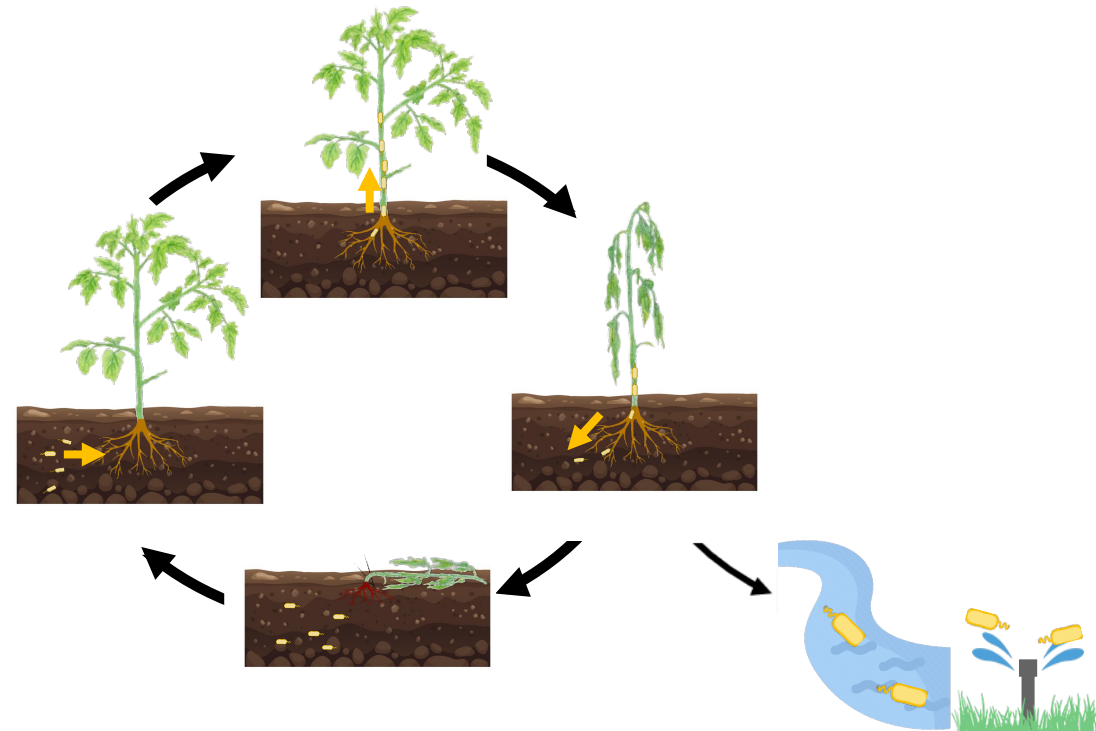
- Cohort-based undergraduate research experiences (CUREs):
A semi-scalable approach to training the next generation of scientists
- **Development of the *Ralstonia* genomics CURE**
- KBase as a platform for bacterial genomics/metagenomics CUREs
- Contributions of *Ralstonia* genomics CURE to scientific knowledge



How can I "CURE" my research interests?

I like understanding **bacterial behavior** in **complex environments** and how **evolution** influences these behaviors.

My focal bacteria: ***Ralstonia solanacearum* species complex**



What gaps in knowledge about *Ralstonia* are CURE-able??

Ivan Buddenhagen:

The Conceptual Problem

‘Bacterial wilt’ is two words but one conceptual entity. *Pseudomonas solanacearum*, likewise, is a single epithet. Although the literature is replete with ‘biotypes’ and ‘strains,’ it generally treats ‘wilt’ as if there is **one** disease and **one** pathogen, created, I presume, by God, all at once, everywhere. We still live in a creationist, Linnean, descriptive, recording of ‘static’ phenomena which occur around us in our approach to our disease research—no matter whether we are generalists, biochemists, or molecular geneticists.

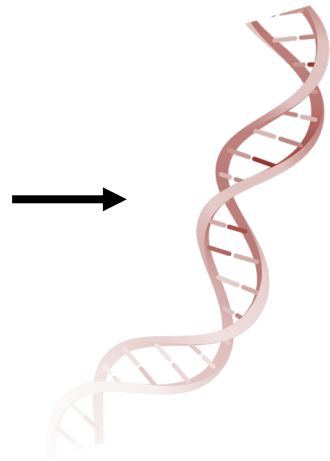
The reality, of course, is very different. There are many bacterial wilts and there are many ‘*Pseudomonas solanacearums*’. They have originated and evolved in widely different places and they have different capabilities with both native flora and introduced hosts, and presumably with different soils and environmental conditions.

Bacterial Wilt in the South Pacific (1985 conference proceedings)

Understanding the many faces of *Ralstonia*:

Training junior scientists

Ralstonia genomics CURE



Whole
Genome
Sequencing



Genome Assembly,
Phylogenomics,
Exploration



Graphical User Interface
(Jupyter notebook)
+ Cloud server

My recipe for teaching the *Ralstonia* genomics CURE:

Diverse *Ralstonia* isolates



A Network

Near-peer role models (Grad Students)

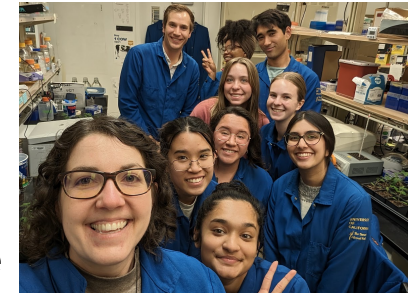


Matt Cope-Arguello



Vienna Elmgreen

Peers



Me

Cost-effective whole genome sequencing for bacteria



Short read sequencing, Illumina

Price

\$77.00

\$101.00

Short reads; small 200 Mbp/1.3 Million reads

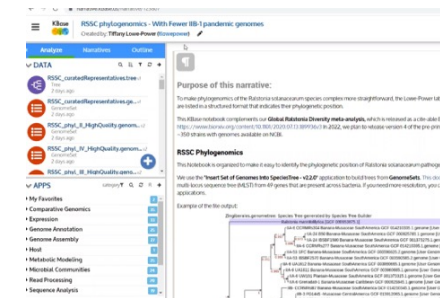
Short reads; medium 400 Mbp/2.7 Million reads

"Co-Working Sessions"

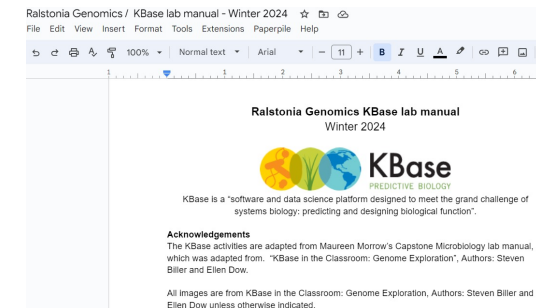
A cozy conference room



KBase



Bioinformatic Lab Manual on Google Docs



Adapted from Maureen Morrow, Steven Biller and Ellen Dow

The *Ralstonia* genomics CURE Experience:

A “Flipped” Research experience?

Dive into **data analysis *before* mastery** of any technical skills

Not everyone wants a career as a researcher, and that is okay!

The *Ralstonia* genomics CURE Experience:

Week 1:

Run FastQC to determine quality of the Illumina sequencing data

“... What does this graph mean?”

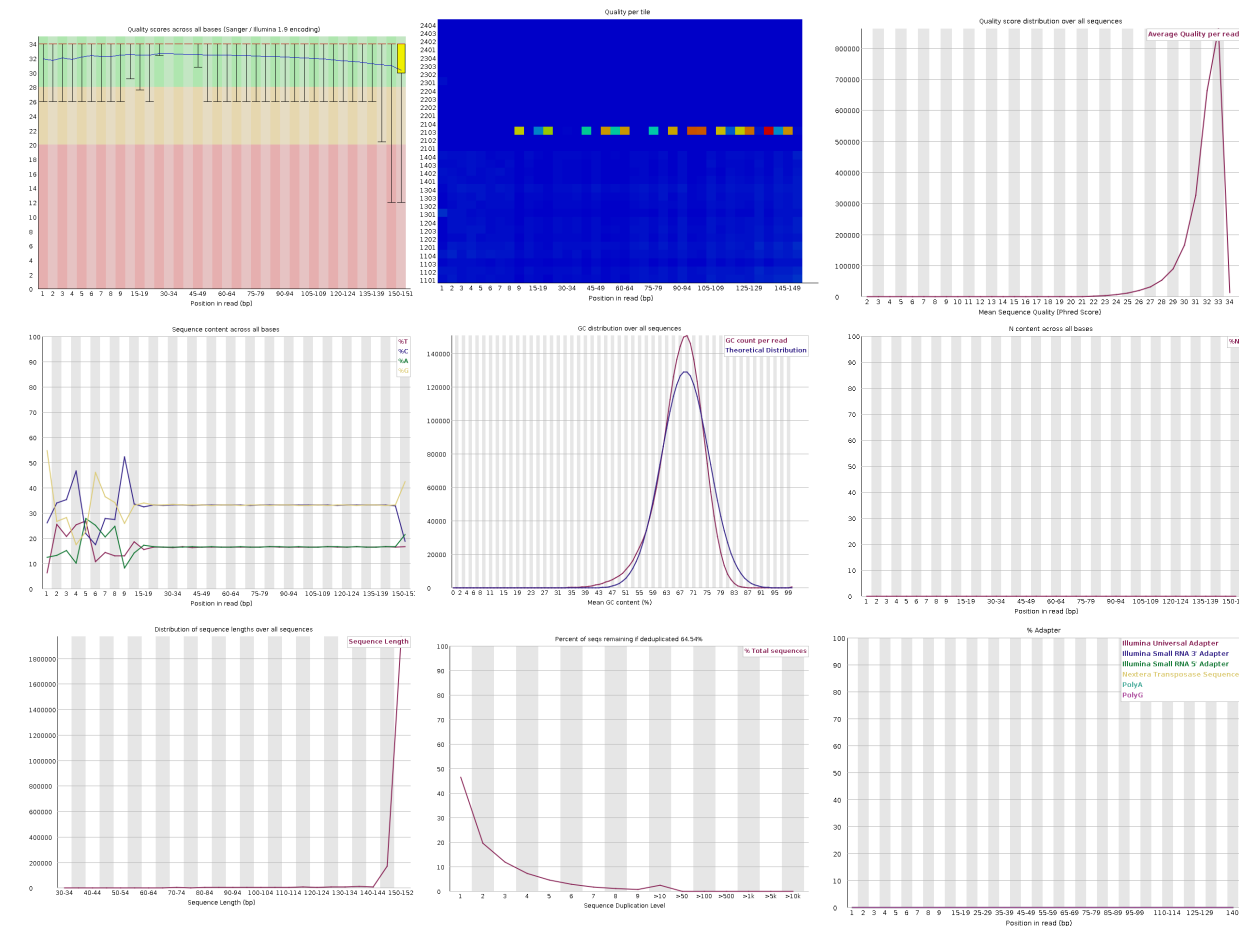
“What are the **axes’** labels?”

“What does the **documentation** say?”

“**Ask your cohort peer** to teach you what we just taught them”

“What does **Google** say?”

“What do **you think** it means?”



Outline:

- Cohort-based undergraduate research experiences (CUREs): A semi-scalable approach to training the next generation of scientists
- Development of the *Ralstonia* genomics CURE
- **KBase as a platform for bacterial genomics/metagenomics CUREs**
- Contributions of *Ralstonia* genomics CURE to scientific knowledge



What is KBase?

A good platform for students to learn how to lab notebook by analyzing genomes. No coding required, but “markdown language” to format notebook entries is a gateway to coding

The screenshot displays a KBase narrative titled "RSSC phylogenomics - With Fewer IIB-1 pandemic genomes" created by Tiffany Lowe-Power. The interface includes a left sidebar with "DATA" and "APPS" sections. The "DATA" section lists several genome sets and trees. The "APPS" section lists various analysis tools with counts. The main content area shows the "Purpose of this narrative:" section, followed by a paragraph about the Ralstonia solanacearum species complex, a link to a meta-analysis, and a section titled "RSSC Phylogenomics" which describes the notebook's organization and the application used to build the species tree. Below this is an example of the file output, a phylogenetic tree titled "Zingiberales.genometree: Species Tree generated by Species Tree Builder".

Purpose of this narrative:

To make phylogenomics of the *Ralstonia solanacearum* species complex more straightforward, the Lowe-Power lab is creating a repository (this narrative) where *Ralstonia* genomes are listed in a structured format that indicates their phylogenetic position.

This KBase notebook complements our **Global *Ralstonia* Diversity meta-analysis**, which is released as a cite-able BioRxiv pre-print with a DOI. <https://www.biorxiv.org/content/10.1101/2020.07.13.189936v3> In 2022, we plan to release version 4 of the pre-print, which will have meta-data on over 8,000 strains including ~350 strains with genomes available on NCBI.

RSSC Phylogenomics

This Notebook is organized to make it easy to identify the phylogenetic position of *Ralstonia solanacearum* pathogens within the context of other sequenced strains.

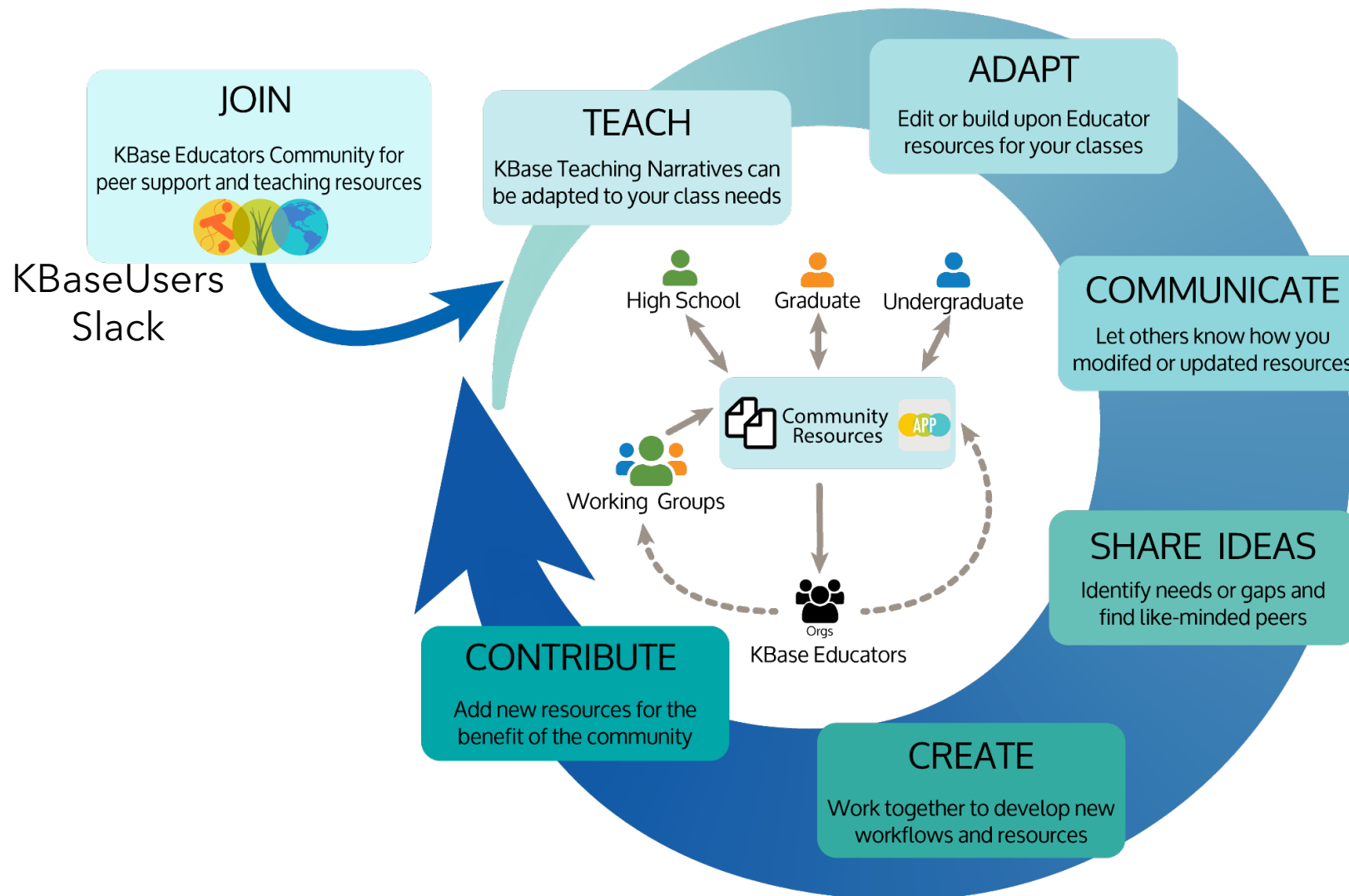
We use the "Insert Set of Genomes Into SpeciesTree - v2.2.0" application to build trees from **GenomeSets**. This [documentation page](#) explains how the app works. In brief, it builds a multi-locus sequence tree (MLST) from 49 genes that are present across bacteria. If you need more resolution, you can download the genomes and run more customizable applications.

Example of the file output:

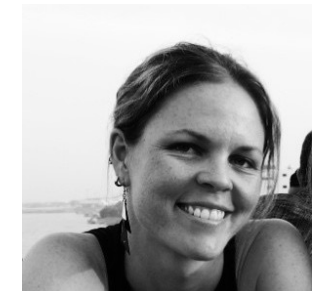
```
Zingiberales.genometree: Species Tree generated by Species Tree Builder
  Ralstonia mannitolilytica [GCF 000953875.1]
  IIA-6 CCRMRs304 Banana-Musaceae SouthAmerica GCF 014210335.1.genome [User Genome 123807/149/1]
  IIA-24 B50 Banana-Musaceae SouthAmerica GCF 000825785.1.genome [User Genome 123807/272/1]
  IIA-24 IBSBF1900 Banana-Musaceae SouthAmerica GCF 001373275.1.genome [User Genome 123807/78/1]
  IIA-6 CCRMRs277 Banana-Musaceae SouthAmerica GCF 014210395.1.genome [User Genome 123807/79/1]
  IIA-53 SFC Banana-Musaceae SouthAmerica GCF 003590625.2.genome [User Genome 123807/150/1]
  IIA-53 IBSBF2570 Banana-Musaceae SouthAmerica GCF 003590585.2.genome [User Genome 123807/151/1]
  IIA-6 UA1612 Banana-Musaceae SouthAmerica GCF 003860665.1.genome [User Genome 123807/147/1]
  IIA-6 UA1611 Banana-Musaceae SouthAmerica GCF 003860685.1.genome [User Genome 123807/148/1]
  IIA-6 UW181 Plantain-Musaceae SouthAmerica GCF 001373315.1.genome [User Genome 123807/146/1]
  IIA-6 Grenada9-1 Banana-Musaceae Caribbean GCF 000825845.1.genome [User Genome 123807/273/1]
  IIB- CCRMRsB7 Banana-Musaceae SouthAmerica GCF 014210345.1.genome [User Genome 123807/132/1]
  IIB-3 PD1445 -Musaceae CentralAmerica GCF 015912065.1.genome [User Genome 123807/139/1]
  IIB-3 CFBP1416 Plantain-Musaceae CentralAmerica GCF 000825925.1.genome [User Genome 123807/274/1]
```

There is **support** for bringing KBase into the classroom

KBase Educators community



Dr. Ellen Dow
KBase Educators
Program Lead



Dr. Elisha
Wood-Charlson
KBase User Working
Group Lead

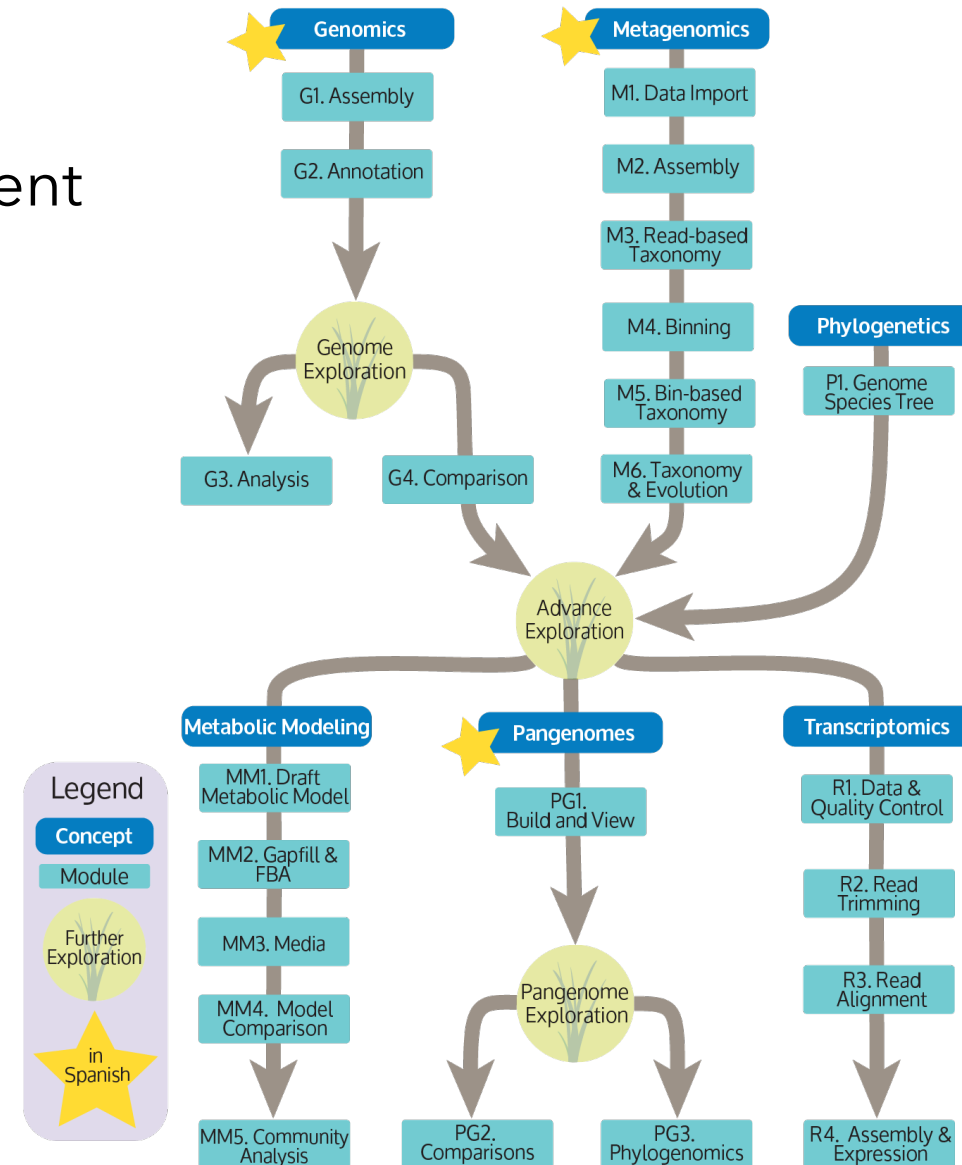
There are **resources** for bringing KBase into the classroom

Teaching modules

Adaptable - customizable, flexible content

Sharable - many existing resources!

English and Spanish

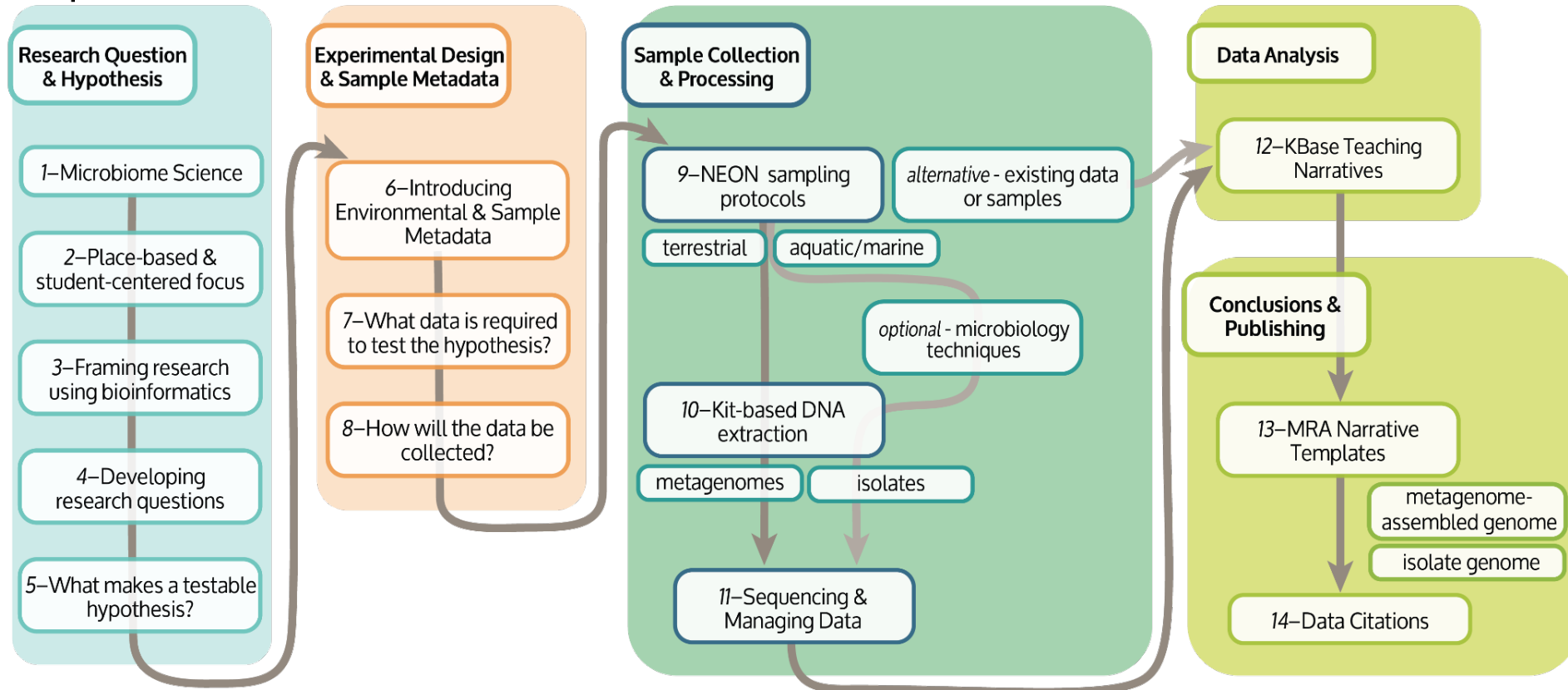


There are **resources** for bringing KBase into the classroom

MICROnet:

Microbiomes in **C**omputational **R**esearch **O**pportunities Network

Goal: Develop modular curriculum



Outline:

- Cohort-based undergraduate research experiences (CUREs): A semi-scalable approach to training the next generation of scientists
- Development of the *Ralstonia* genomics CURE
- KBase as a platform for bacterial genomics/metagenomics CUREs
- **Contributions of *Ralstonia* genomics CURE to scientific knowledge**

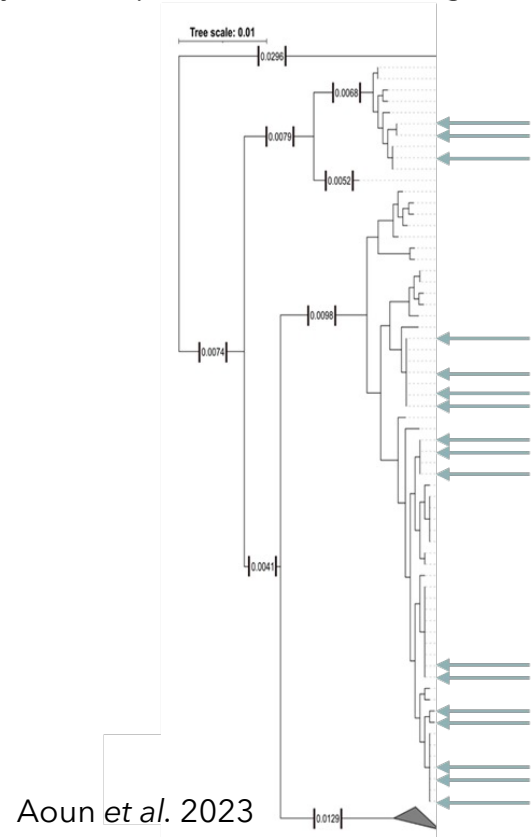


Generating public *Ralstonia* genomic data

Spring 2022

17 new genomes

Collected in **Central/South America** in 1960s
by Luis Sequeira, Ivan Buddenhagen, and Ed French

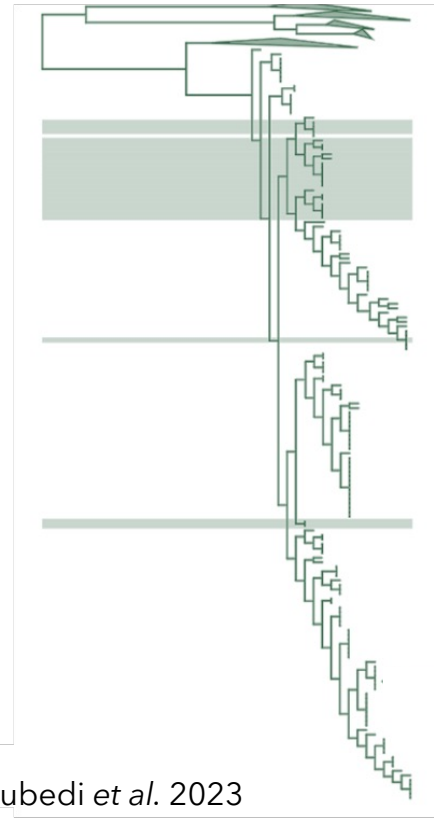


Aoun et al. 2023

Winter 2023

24 new genomes

Collected in **Bangladesh/Nepal** in 2019
by Nagendra Subedi and Sally Miller (Ohio State)

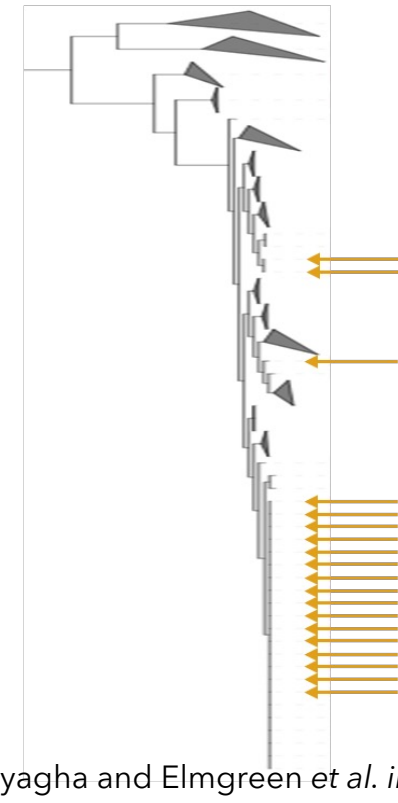


Subedi et al. 2023

Winter 2024

19 new genomes

Collected in **Tanzania** in 2019
by Helen Kanyagha and Sally Miller (Ohio State)



Kanyagha and Elmgreen et al. in prep



Release genomes to
NCBI Assembly/SRA

Ralstonia genomics CURE – A playground for scientific discovery

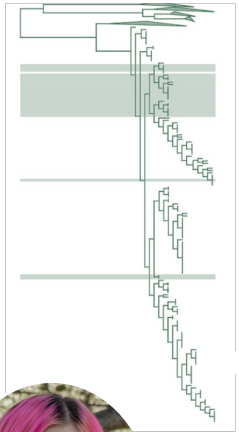
Tabitha Cowell:
Undergrad from CURE 2023



Ralstonia genomics CURE – A playground for scientific discovery

Tabitha Cowell's exploration:

Genomes of
Bangladesh/ Nepal
Ralstonia



Pangenome
Analysis

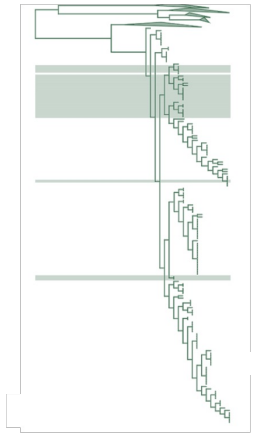
Genomes had differential
abundance of genes



Ralstonia genomics CURE – A playground for scientific discovery

Tabitha Cowell's exploration:

Genomes of
Bangladesh/ Nepal
Ralstonia



Pangenome
Analysis

Genomes had differential
abundance of genes

"Type IV secretion TraX protein"
"Dot-Icm conjugal transfer protein"
"VirD4 conjugal coupling protein"



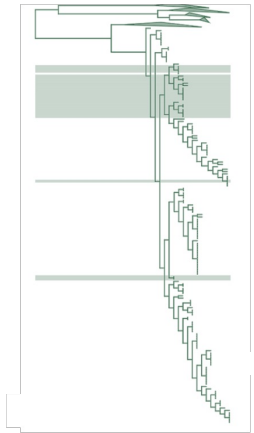
Tabitha noticed
commonalities in
gene annotations



Ralstonia genomics CURE – A playground for scientific discovery

Tabitha Cowell's exploration:

Genomes of
Bangladesh/ Nepal
Ralstonia



Pangenome
Analysis

Genomes had differential
abundance of genes

"Type IV secretion TraX protein"
"Dot-Icm conjugal transfer protein"
"VirD4 conjugal coupling protein"



Tabitha noticed
commonalities in
gene annotations



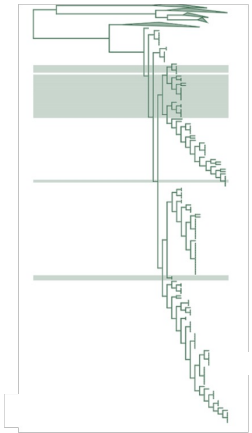
Background knowledge

- Conjugation/T4SS relates to horizontal gene transfer
- Conjugation has not been studied in *Ralstonia*.

Ralstonia genomics CURE – A playground for scientific discovery

Tabitha Cowell's exploration:

Genomes of
Bangladesh/ Nepal
Ralstonia



Pangenome
Analysis

Genomes had differential
abundance of genes

"Type IV secretion TraX protein"
"Dot-Icm conjugal transfer protein"
"VirD4 conjugal coupling protein"

Tabitha noticed
commonalities in
gene annotations

Background knowledge

- Conjugation/T4SS relates to horizontal gene transfer
- Conjugation has not been studied in *Ralstonia*.

I suggested a scientific question:

What is the phylogenetic
distribution of conjugal /
T4SS genes in the *Ralstonia*
pangenome?



Ralstonia genomics CURE – A playground for scientific discovery

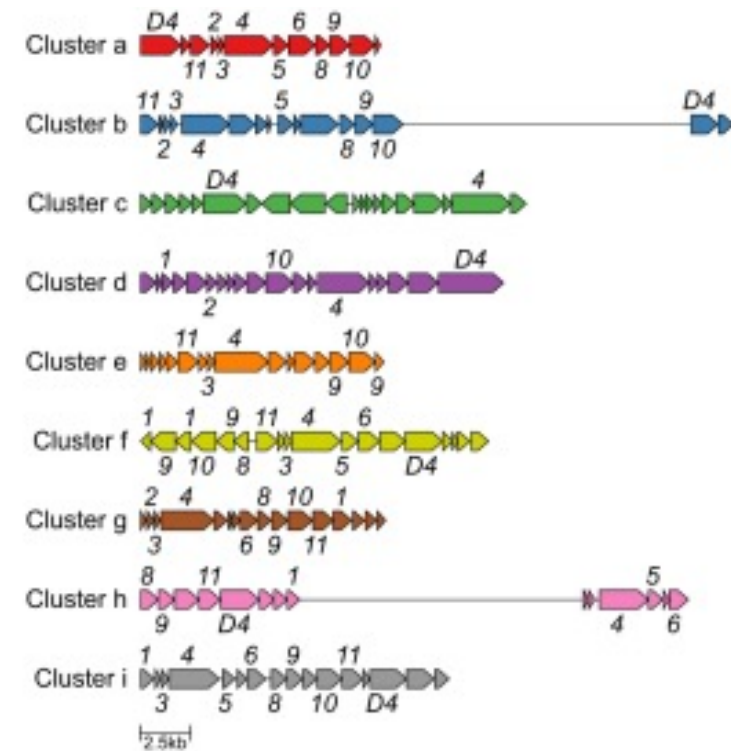
What is the phylogenetic distribution of conjugative / T4SS genes in the *Ralstonia* pangenome?

Tabitha has classified 9 types of conjugative gene clusters

BLAST searches
in KBase

**Investigate gene organization/
synteny**
Clinker

Visualize phylogenetic distribution on iToL (Interactive Tree of Life)



Ralstonia genomics CURE – A playground for scientific discovery

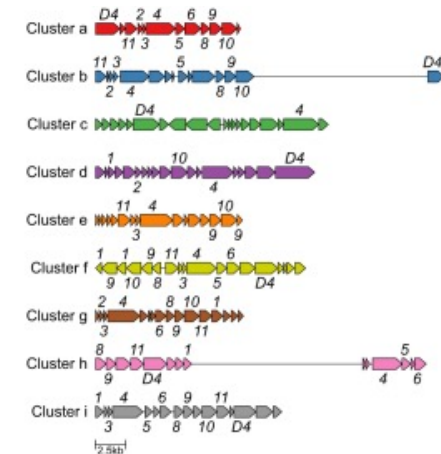
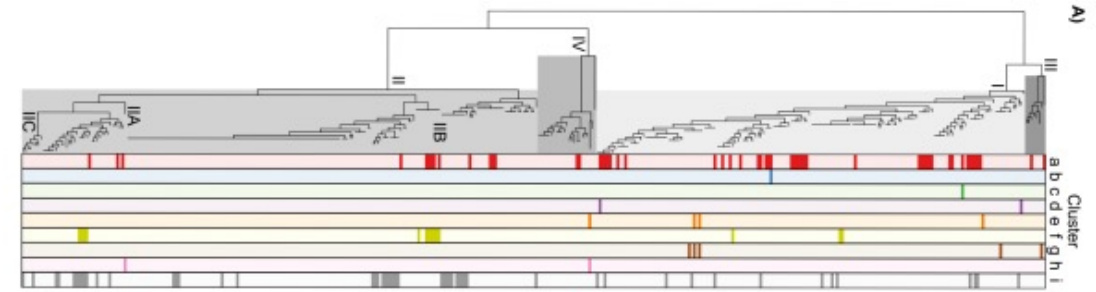
What is the phylogenetic distribution of conjugative / T4SS genes in the *Ralstonia* pangenome?

BLAST searches
in KBase

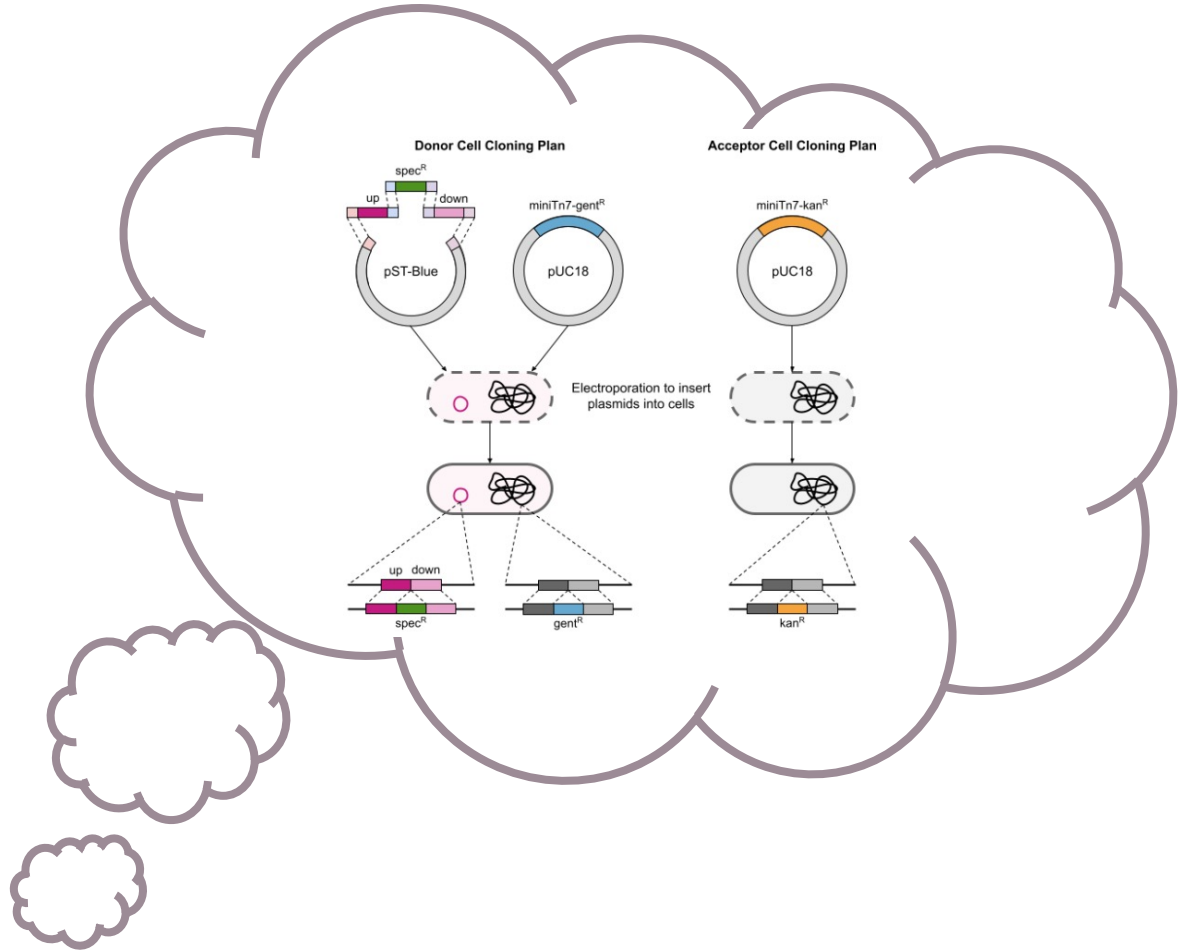
**Investigate gene organization/
synteny**
Clinker

Visualize phylogenetic distribution on
iToL (Interactive
Tree of Life)

Tabitha discovered that conjugative/T4SS gene clusters have sporadic phylogenetic distribution, consistent with the hypothesis that these are mobile genetic elements



Ralstonia genomics CURE – A playground for scientific discovery



I am an **accidental plant pathologist**.

The *Ralstonia* CURE is based on my knowledge base, resources at my disposal, and the student population I work with.

How can **you** have **impact**
...on your **field**?
...on the **next generation of scientists**?



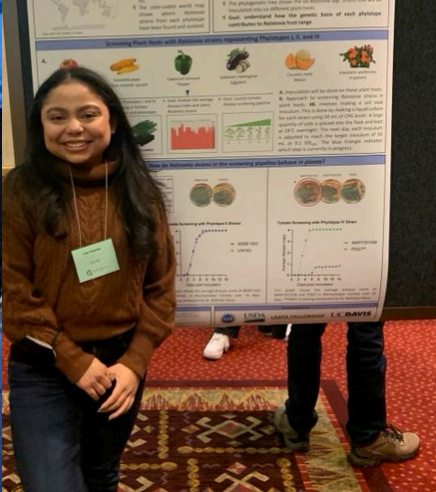
Dr. Nathalie Aoun



Jason Avalos



Lisa Repetto



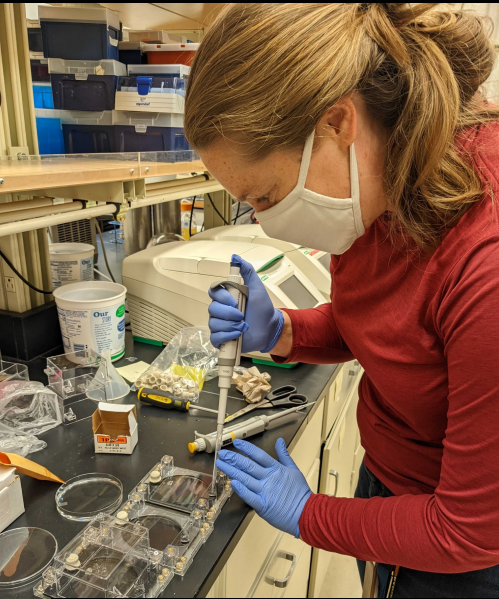
Stratton Georgoulis



Vienna Elmgreen



Acknowledgements



Dr. Becca Schomer



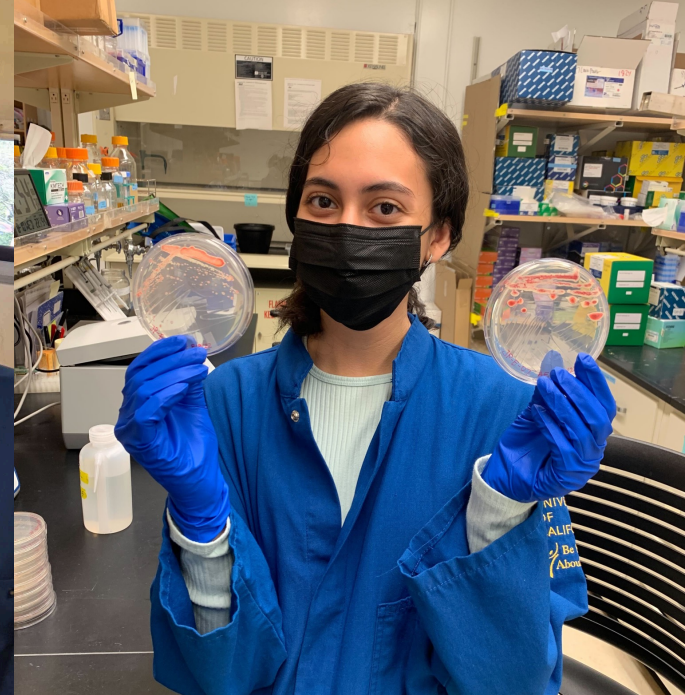
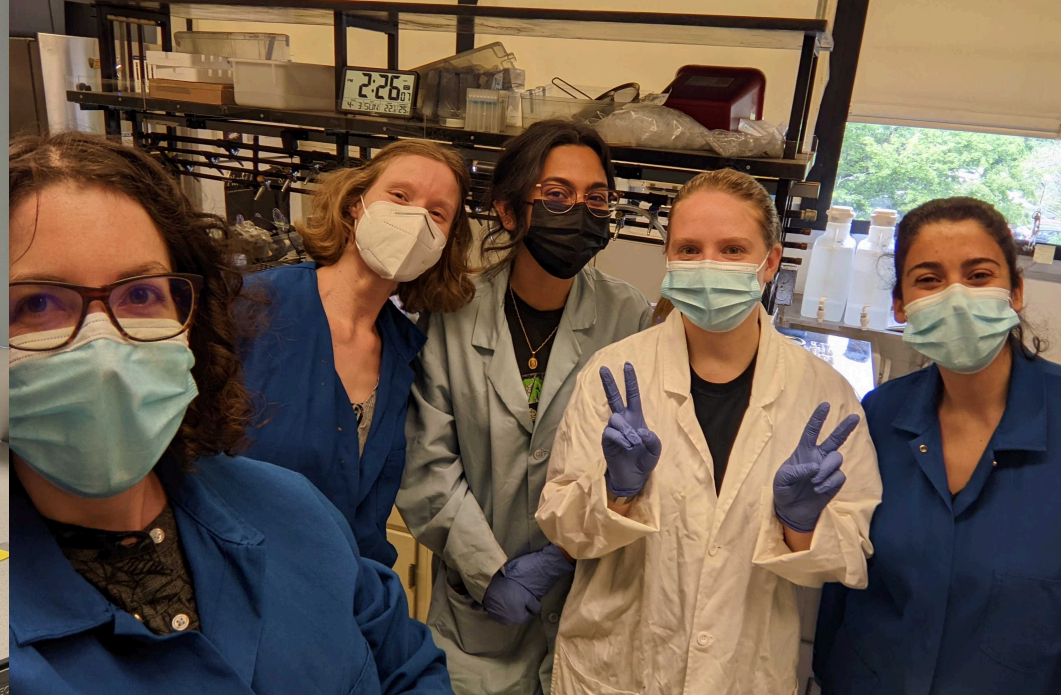
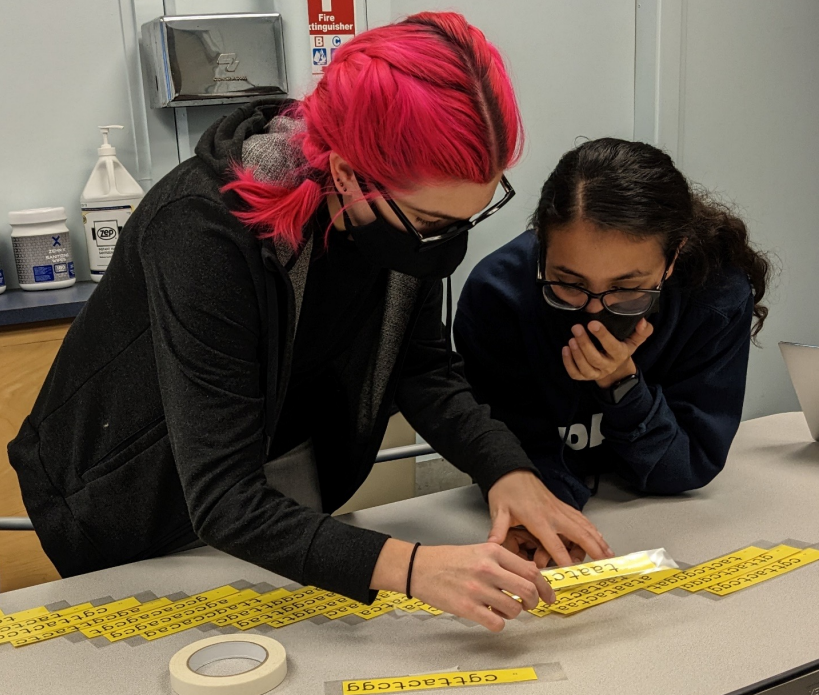
Matt Cope-Arguello



Jonathan Beutler



Agnès Bellon



Questions?

