



Welcome to the Workshop!

Workshop Website

<https://go.wisc.edu/3183gc>

Etherpad (for collaborative notes and questions)

<https://go.wisc.edu/f778p7>

Make sure you can connect to your Cyverse account!



Welcome to Data Carpentry! August 5-6 & 12, 2019

Instructors:

Trisha Adamus, Steve Goldstein,
Christina Koch, Tobin Magle, Elizabeth
McDaniel, Sarah Stevens,
Sailendharan Sudakaran

Host:





Elizabeth McDaniel

PhD student in Microbiology Doctoral Training Program

emcdaniel@wisc.edu

Meta-'omics/comparative genomics of engineered wastewater systems



Computational Biology, Ecology, and Evolution Group (ComBEE)

- Peer-led study group of researchers interested in applying computational biology techniques to topics in ecology and evolution
- Bi-weekly study groups in R/python
- Monthly meetings/informal seminars on topics in ecology & evolution
- Fall Social scheduled for **September 6th @ 4PM – MSB 6201**
- Join the listserv by sending an email to join-combee@lists.wisc.edu
- combee-uw-madison@github.io

What is Data Carpentry?

- Workshop by The Carpentries
 - Non-profit, international organization
- Workshops to help researchers adopt reproducible computational practices
- Instructors and helpers are all volunteers
- Materials developed by an open science community
- Code-along learning model

Code of Conduct

- Use welcoming and inclusive language
- Be respectful of different viewpoints and experiences
- Gracefully accept constructive criticism
- Focus on what is best for the community
- Show courtesy and respect towards other community members
- Contact Sarah Stevens if you experience conduct violations

<http://software-carpentry.org/conduct/>

Workshop Logistics

Restrooms:

- Across the hall (or on the first floor)

Food:

- Coffee/tea in morning
- No lunch provided (Aldo's café, Saffron downstairs, Union South across street)
 - Lunch social – instructors will be around
 - Microwave and fridge available

Schedule

Morning:

Introduction to R

Lunch

Afternoon:

Data wrangling and visualization in R

Workshop Logistics

Workshop Website

<https://go.wisc.edu/3183gc>

Etherpad (for collaborative notes and questions)

<https://go.wisc.edu/f778p7>

Lessons

<https://datacarpentry.org/genomics-workshop/>

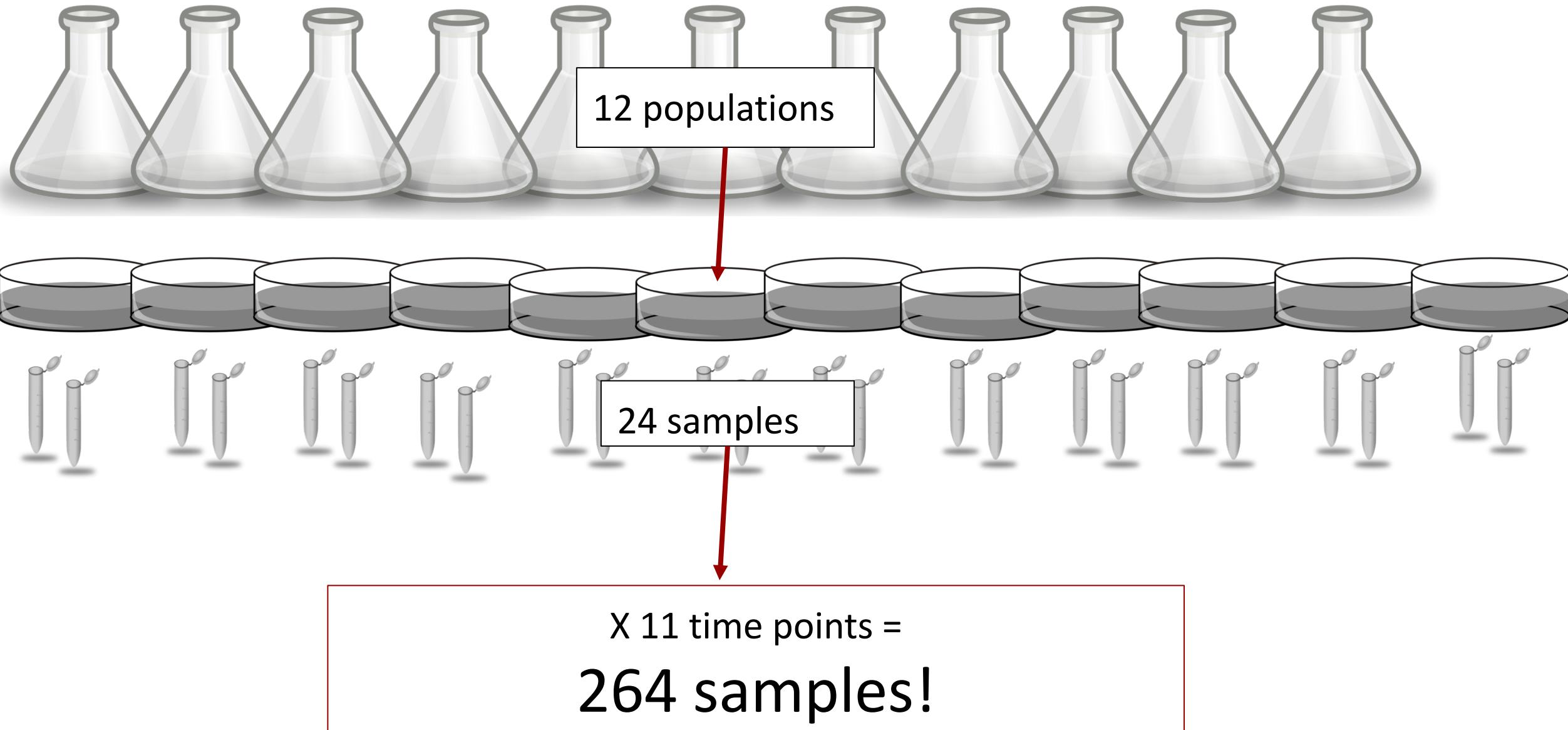
Make sure you can connect to your Cyverse account!

Tempo and mode of genome evolution in a 50,000-generation experiment

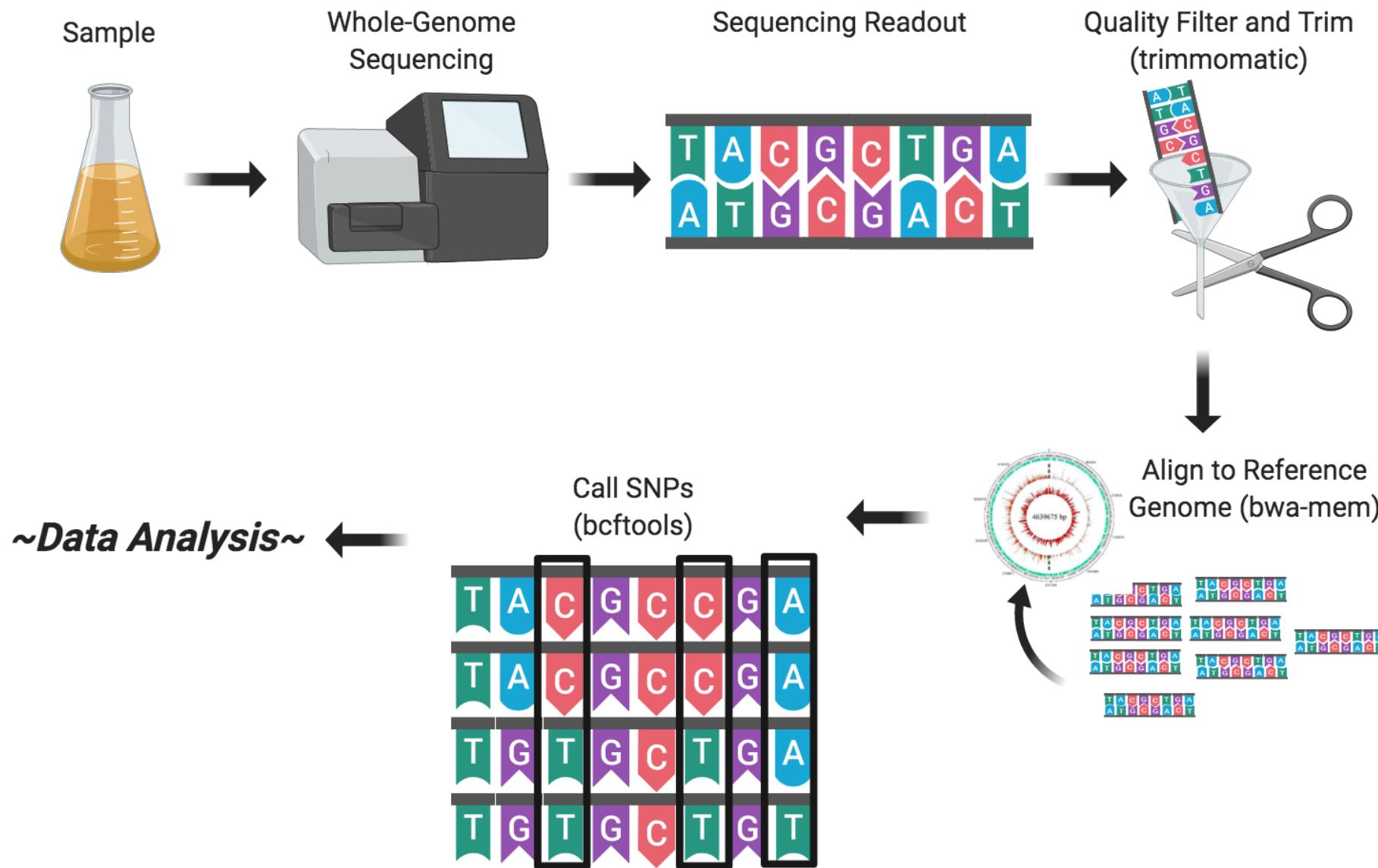
Olivier Tenaillon^{1*}, Jeffrey E. Barrick^{2,3*}, Noah Ribeck^{3,4}, Daniel E. Deatherage², Jeffrey L. Blanchard⁵, Aurko Dasgupta^{2†}, Gabriel C. Wu², Sébastien Wielgoss^{6,7}, Stéphane Cruveiller⁸, Claudine Médigue⁸, Dominique Schneider^{7,9} & Richard E. Lenski^{3,4*}



- Long-term evolution experiment (LTEE) in *E. coli*
- Wanted to understand the genomic and phenotypic changes of evolution
- Founded 12 replicate populations under identical conditions in February 1988
- Sequenced whole genomes from 2 clones from each population at 11 time points



Last Week: Creating a Genomics Pipeline



What biological questions pertaining to the LTEE could you ask with our SNP data?

Discuss amongst your table and write your thoughts in the Etherpad

What is R?

- “R” refers to both the programming language itself, and the software platform to interpret script written in it

Why should I learn R?

- Less pointing and clicking
- Writing scripts makes analyses more reproducible than clicking a series of buttons > remember what you did and can share with others
- Interdisciplinary (field-specific extensions or packages) for genomics, statistical simulations, etc.
- Takes data of all shapes and sizes
- Produce high-quality graphics without \$\$\$
- Large, welcoming, open source community always developing new packages

What is RStudio?

- Rstudio is an Integrated Development Environment (IDE) for working with R
- Can write code, navigate files, view variables you created, and inspect plots within Rstudio
- Lots of other IDEs that can suit a wide variety of programming languages

If you do not:

- A) Have a Cyverse instance running and can access the IP address**
- B) Have RStudio and the tidyverse package installed on your local computer**

Raise your tent to the red side

Before we start:

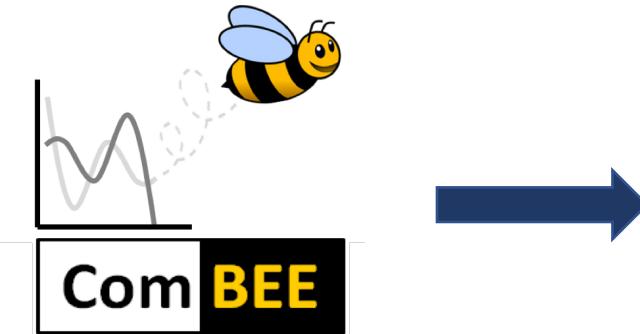
- **Resources to get help**
- **Resources on campus**
- **RStudio universe/packages**

How do I find help?



- stack overflow
- R-bloggers
- R Studio Blog
- ...and more!

Resources on campus



Elizabeth McDaniel
emcdaniel@wisc.edu

Kirsten Gotting
gotting@wisc.edu

Social Sept. 6th @ 4 PM in MSB 6201
- Study groups
- Monthly seminars



Sarah Stevens
facilitator@datascience.wisc.edu

Office hours to get help on all coding needs!



Tobin Magle
tobin.magle@wisc.edu

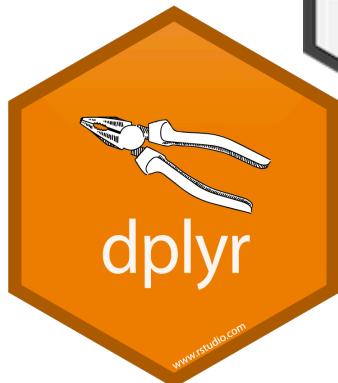
R workshops at Steenbock library



chtc@cs.wisc.edu
Christina Koch
Lauren Michael

Office hours to help with your coding needs on CHTC cluster

The world of R/Rstudio



<https://www.bioconductor.org/>

- Over 1700 packages for analyzing and visualizing genomic data
- Extensive documentation and vignettes (example demos) for packages
- Workflows for common analyses: differential gene expression, annotation, etc.

Accessing the RStudio server

- Cyverse account, find your IP address in Projects > enter project that your instance was created > click on project > find IP address such as 128.196.142.78 (do not use this address!)
- Enter the IP address into your web browser, followed by :8787, such as 128.196.142.78:8787
- Username: dcuser
- Password: data4Carp
- Green side of tent when you are logged in, or already have RStudio/tidyverse packages installed locally