Event Schedule (all times EDT)

Monday, September 27th

1:00–2:00 pm Whole group meeting: Welcome and introduction

- Logistical information (including GitHub, Slack, Zoom, Cloud Access, Support)
- Introduction to team projects (Team Leaders, ~1 min each)
- Discussion of team roles
- Introduction to prepared datasets
  Dr. Kjiersten Fagnan, DOE/LBNL-JGI; Dr. Migun Shakya, DOE/LANL
- Presentation of an in-development metagenomics index
  Dr. Richa Agarwala, NIH/NLM/NCBI

2:00–5:00 Team breakout meetings: Initial planning

- Introduce team members
- Assign team roles
- Decide on initial approach;
  Outline readme introduction;
  Create workflow diagram
- Choose cloud platform(s) and prepare data

Tuesday, September 28th

1:00–1:30 Whole group meeting: Teams present proposed workflows
  (5 minutes per team with slides)

1:30–5:00 Team meetings and free work time

4:30–5:00 Team Leaders' meeting | Push work to GitHub

Wednesday, September 29th

1:00–1:30 Whole group meeting: Lightning stand up

1:30–5:00 Team meetings and free work time

4:30–5:00 Writers' meeting | Push work to GitHub

Thursday, September 30th

1:00–1:30 Whole group meeting: Lightning stand-up

1:30–5:00 Team meetings and free work time

4:30–5:00 Push work to GitHub

Friday, October 1st

1:00–3:00 Team meetings and free work time

3:00–4:00 Push work to GitHub | Prepare final presentations

4:00–5:00 Final team presentations (8 minutes per team with slides)

Schedule Details

Team roles

Team Leader conveys scientific objectives to team, coordinates work.

Tech Lead coordinates software installation and data acquisition,
manages version control and the team's GitHub repository,
troubleshoots technical issues with tech support.

Writer ensures that all work is documented,
manages GitHub README and Team Report

Flex fills various roles and assumes responsibilities for tasks as needed.

Proposed workflow presentations (5 min per team)

Each team will share their plans and get feedback from other teams.
Suggested topics to cover briefly:

- What specific sequence search task are you benchmarking?
- What are the biological applications?
- What software currently exists?
- What metrics will you use to measure performance?
- What datasets will you use?
- What will your team build (e.g., snakemake/nextflow pipelines, docker containers)?
- What is your planned workflow for the codeathon?

Lightning stand-ups (5 min per team)

Each team will report on the previous day's progress and any roadblocks they've encountered.
This is an opportunity to troubleshoot and share expertise across teams.

Final team presentations (8 min per team, plus discussion)

Each team will present their codeathon product and suggest opportunities for follow-up work.

Team Leaders' meeting

All Team Leaders and the event organizers meet to discuss team management issues,
opportunities for inter-team collaboration, and questions about the codeathon.

Writers' meeting

All Writers and the event organizers meet to discuss documentation, including the GitHub
READMEs and Team Reports.

Push work to GitHub

We encourage you to continuously commit your progress to git and push to the team GitHub
repository often. To ensure that every participant's work is committed and shared at least once
a day, we will reserve the last half hour of the end of each working day for committing,
merging, and pushing any unfinished work.

Codeathon products

GitHub Repositories

All code generated by your team should be included in your team's repository on GitHub. The repository must include a README that documents your team's objectives and products. The contents of the repository should be sufficient to reproduce and build on your work.

Team 1: https://github.com/NCBI-Codeathons/bothie
Team 2: https://github.com/NCBI-Codeathons/psss-team2
Team 3: https://github.com/NCBI-Codeathons/psss-team3-hard-annotation
Team 4: https://github.com/NCBI-Codeathons/psss-team4

Examples of previous NCBI codeathon repositories:

- https://github.com/NCBI-Hackathons/PubRunner
- https://github.com/NCBI-Hackathons/MetagenomicAntibioticResistance
- https://github.com/NCBI-Codeathons/Using-PubMed-MEDLINE-Metadata-to-Generate-Knowledge

Team Reports

Throughout the week, please document your project in a Team Report. We've provided a basic structure, but you may modify it to suit your project if necessary. These reports will be integrated into a manuscript along with the output of June's *Emerging Solutions in Petabyte-Scale Sequence Search* workshop.

Team 1
Team 2
Team 3
Team 4

Petabyte-Scale Sequence Search:
Metagenomics Benchmarking Codeathon

Support and tutorials

For general support, please use the **#help-desk** channel in Slack or send the organizers a direct message (**@Dan Rice** and **@Alexa Salsbury**). We will have experts from NIH, DOE, Amazon Web Services (AWS), and Google Cloud Platform (GCP) available to answer your questions.

Cloud tutorials from Titus Brown's group

- **AWS** tutorial:
  https://ngs-docs.github.io/2021-august-remote-computing/making-use-of-on-demand-cloud-computers-from-amazon-web-services.html
  Recording:
  https://video.ucdavis.edu/media/Workshop+11A+Making+Use+of+On-Demand+%22Cloud%22+Computers+from+Amazon+Web+Services/1_l9wfn0lj
- **GCP** draft tutorial:
  https://training.nih-cfde.org/en/latest/Cloud-Platforms/Introduction-to-GCP/

Cloud tutorials from AWS

- **Basic AWS** tutorials can be browsed here. Some highlights: Deploying a Kubernetes Application with AWS EKS and Run a serverless "Hello, world" with AWS Lambda.
- In addition, **AWS Research and Education** hosts 10 minute tutorials, including RStudio on AWS, spinning up Jupyter notebooks using AWS SageMaker Studio, and moving your storage to AWS.
- Public Sector **webinars**:
  - Launching an EC2 Instance in an AWS Educate account (20m)
  - Serverless Jupyter on AWS (60m)

Cloud tutorials from GCP

See following pages.