American Museum of Natural History

RGGS Comparative Genomics 2 – Computational Methods (Session 6)

Jose Barba

Gerstner Scholar in Bioinformatics & Computational Biology

AMNH Arachnology Lab
Institute for Comparative Genomics
✉ jbarba@amnh.org
🌐 josebarbamontoya.github.io

October 10, 2024
ICG conference room

Session 5 outline

- **Quiz 1 (oral recapitulation of previous topics)**

- **Paper presentations**

Additional matters

- **Session 9 — October 31**
  - Unfortunately, Jessica Goodhart will not be available to teach on October 31
  - However, she has kindly offered to teach on the morning of November 1
  - Who is available to attend the class on that day?
    * Hopefully, most of you

`Oral recapitulation quiz`

- **Reproducibility and organization**
  - Why is reproducibility important in computational biology, and how do you ensure your work is organized for reproducibility?

- **Introduction to Unix/Unix-like OS**
  - What are the advantages of using Unix or Unix-like operating systems for computational biology?

- **Basic navigation in the terminal**
  - Can you walk me through basic terminal navigation, including commands for viewing, listing, moving between directories, and editing text files?

Oral recapitulation quiz

- **<u>Command line computing basics 1</u>**
  - Mention useful commands to navigate directories, view/edit/create files, among other tasks

- **<u>Command line computing basics 2</u>**
  - What is a batch script?
  - What are shell pipelines, for- and while-loops?

- **<u>Connecting to remote servers & package managers</u>**
  - What steps would you take to connect to a remote server, and how can a package manager assist in managing the software on that server?

`Oral recapitulation quiz`

- **Introduction to GitHub & version control**
  - What are the key benefits of using GitHub for version control, and how does it help in collaborative research?

- **Working on remote servers (HPC architecture, scheduler, jobs)**
  - Could you explain what a HPC environment is and how to submit a job using a scheduler?
  - Can all computational process be parallelized?

- **Scripting:**
  - How can scripting automate tasks in computational biology, and what are some typical tasks that can be scripted?
  - Mention some examples of scripting languages

`Oral recapitulation quiz`

- **<u>Introduction to R and Python for phylogenomics</u>**
  - Mention some examples of tasks in Python or R that involve data manipulation, statistical computing and graphics
  - Can you describe a scenario where you would use either R or Python for phylogenomics, and why you would choose one over the other?

Oral recapitulation quiz

- **Summary of tasks that can be scripted**
  - Data retrieval and management
  - Data preprocessing
  - Analysis and modeling
  - Visualization
  - Batch processing
  - Result compilation and reporting
  - Integrating tools and pipelines

Paper presentations

- **Each student has chosen an -omics paper that they find innovative, exciting, relevant to their work, or particularly interesting**

- **Today (October 10, 2024), they will deliver a 10-minute presentation that provides a concise overview of the research question addressed and a thorough explanation of the computational methods employed**