Teaching genomics and command-line basics at a primarily undergraduate institution using browser-based activities

Bárbara D. Bitarello

Resources and Programs for Undergraduate Education in Genomics, PAG, 2024

Slides!

About me

- 2021-Present: Assistant professor at Bryn Mawr College (BMC), a small women's liberal arts college
- Research: evolutionary & statistical genomics (humans and other primates)
- Bitarello (dry) Lab: currently 7 undergraduate researchers working on diverse projects in evolutionary & statistical genetics & phylogenetics
- Teaching:
  - 100-level: Intro Bio
  - 200-level: **Genomics (6h/week, 1/2 lab)**, Biostatistics with R
  - 300-level: Evolutionary Genetics & Genomics

Outline

Outline

1. Why browser-based?

Outline

1. Why browser-based?
2. Two projects/experiences from B216 (Genomics) that only require a browser

Outline

1. Why browser-based?
2. Two projects/experiences from B216 (Genomics) that only require a browser
   A. A soft-introduction to the command line and FASTQ files

Outline

1. Why browser-based?
2. Two projects/experiences from B216 (Genomics) that only require a browser
   A. A soft-introduction to the command line and FASTQ files
   B. The Genomics Education Partnership (GEP) and how I've adapted and contributed materials

Outline

1. Why browser-based?
2. Two projects/experiences from B216 (Genomics) that only require a browser
   A. A soft-introduction to the command line and FASTQ files
   B. The Genomics Education Partnership (GEP) and how I've adapted and contributed materials

Bonus: A quick mention about a third project involving R programming!

Why browser-based?

Challenges for teaching genomics/bioinformatics:

Why browser-based?

Challenges for teaching genomics/bioinformatics:

1. getting all tools installed in a variety of OS and versions: often **frustrating** and **time-consuming**

Why browser-based?

Challenges for teaching genomics/bioinformatics:

1. getting all tools installed in a variety of OS and versions: often **frustrating** and **time-consuming**
2. campus computers: often **lack permissions** to get all the **required updates** and **installations** in a timely manner

Why browser-based?

Challenges for teaching genomics/bioinformatics:

1. getting all tools installed in a variety of OS and versions: often **frustrating** and **time-consuming**
2. campus computers: often **lack permissions** to get all the **required updates** and **installations** in a timely manner
3. some students use machines that **lack space or capability** for local installations (e.g. Chromebook)

Why browser-based?

Challenges for teaching genomics/bioinformatics:

1. getting all tools installed in a variety of OS and versions: often **frustrating** and **time-consuming**
2. campus computers: often **lack permissions** to get all the **required updates** and **installations** in a timely manner
3. some students use machines that **lack space or capability** for local installations (e.g. Chromebook)
4. technical challenges intimidate students even more; the browser **keeps it familiar/simple**

Why browser-based?

Challenges for teaching genomics/bioinformatics:

1. getting all tools installed in a variety of OS and versions: often **frustrating** and **time-consuming**
2. campus computers: often **lack permissions** to get all the **required updates** and **installations** in a timely manner
3. some students use machines that **lack space or capability** for local installations (e.g. Chromebook)
4. technical challenges intimidate students even more; the browser **keeps it familiar/simple**

Browser-only activities bypass all of these hurdles!

Examples

- Biostatistics with R: Posit (Studio) Cloud ✓
- Genomics: UCSC Genome Browser, Galaxy, etc ✓

Examples

- Biostatistics with R: Posit (Studio) Cloud ✓
- Genomics: UCSC Genome Browser, Galaxy, etc ✓

But what about running software and learning about the command-line?

Project 1: A soft introduction to the command-line and FASTQ files

Motivation

Motivation

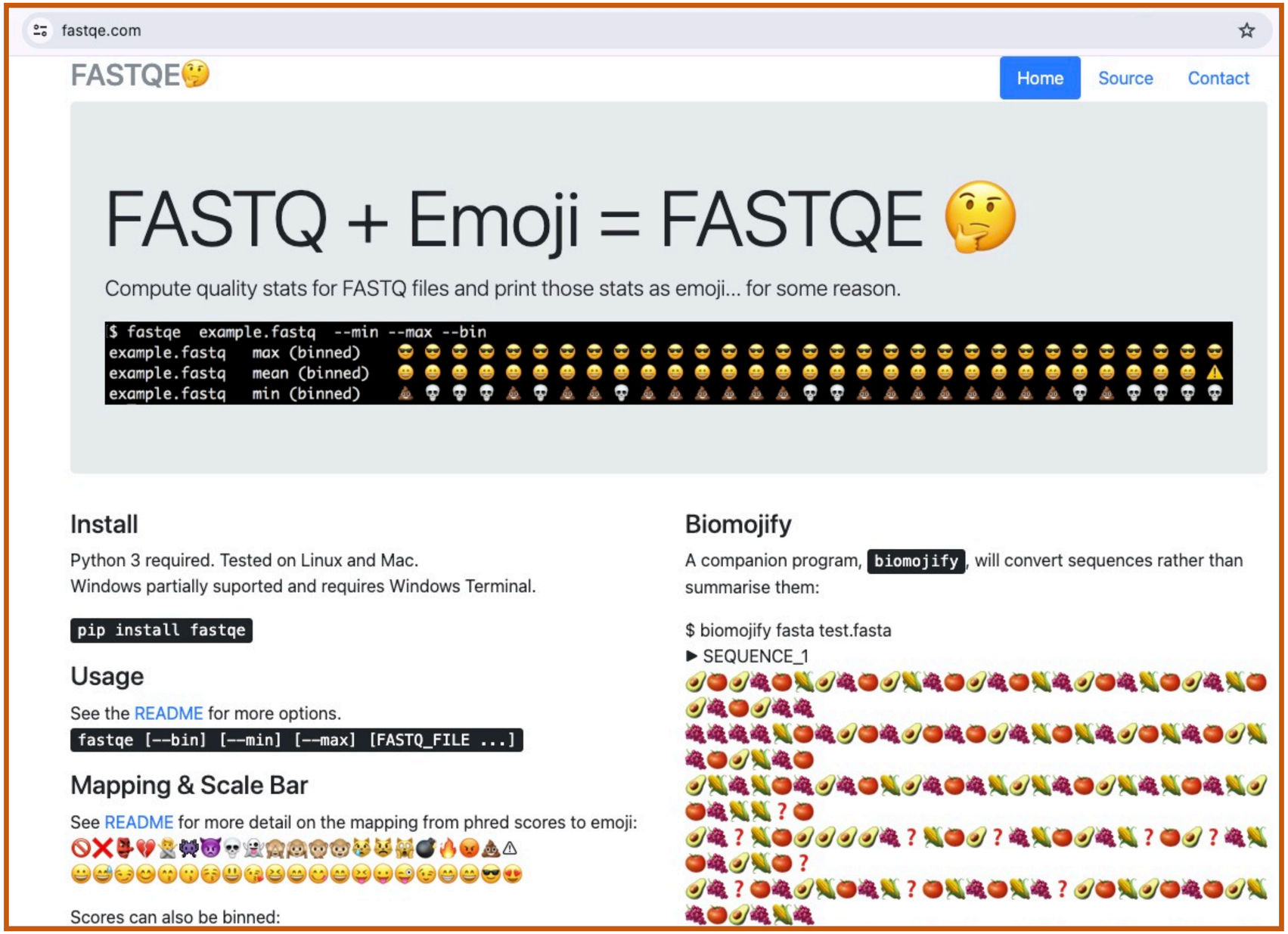
- Excellent materials from St. Jaquest et al. (2021), published in *CourseSource*

Motivation

- Excellent materials from St. Jaquest et al. (2021), published in *CourseSource*
- Introduces the command-line and FASTQ files by using the FASTQE software

FASTQE: FASTQ + EMOJI

- Official Page: fastqe.com
- Github: github.com/fastqe/fastqe
- St. Jacques et al. (2021). CourseSource. doi.org/10.24918/cs.2021.17



Challenges in implementing the lesson

Challenges in implementing the lesson

- FASTQE is a python package that needs to be installed (as well as its dependencies) — and python installations are ~~always~~ often a nightmare!

Challenges in implementing the lesson

- FASTQE is a python package that needs to be installed (as well as its dependencies) — and python installations are ~~always~~ often a nightmare!
- Proposed implementation in publication: either a) local installation or b) CyVerse

Challenges in implementing the lesson

- FASTQE is a python package that needs to be installed (as well as its dependencies) — and python installations are ~~always~~ often a nightmare!
- Proposed implementation in publication: either a) local installation or b) CyVerse
- 2022: lost one entire class installing locally for each students and one student still could not get it to work; Cyverse did not work for any of them despite many efforts over several days

Challenges in implementing the lesson

- FASTQE is a python package that needs to be installed (as well as its dependencies) — and python installations are ~~always~~ often a nightmare!
- Proposed implementation in publication: either a) local installation or b) CyVerse
- 2022: lost one entire class installing locally for each students and one student still could not get it to work; Cyverse did not work for any of them despite many efforts over several days
- FASTQE on galaxy: did not have time to test; removes the command-line experience

Updated implementation

Solution: using mybinder.org

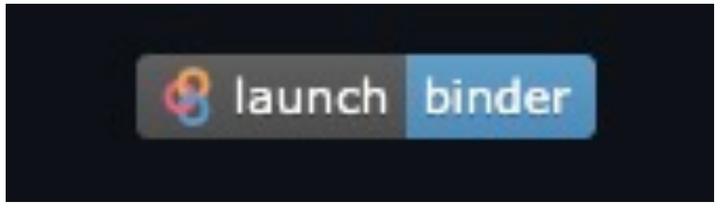
- Binder allows you to create **custom computing environments** that can be shared and used by many remote users.
- A Binder service is powered by BinderHub, an open-source tool.
- One such deployment lives at mybinder.org, and is free to use.

Freely available

- The Github repo https://github.com/bitarellolab/Genomics_Teaching contains:
  - The code in R markdown and html formats
  - A slide deck for students
  - A direct link to launch the url for the activity

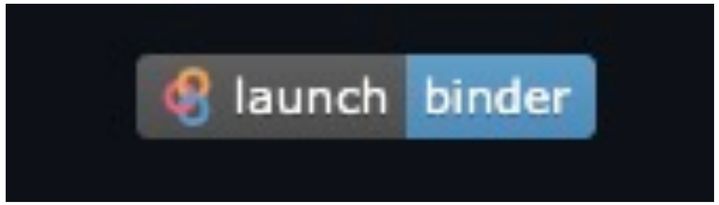
Freely available

- The Github repo https://github.com/bitarellolab/Genomics_Teaching contains:
  - The code in R markdown and html formats
  - A slide deck for students
  - A direct link to launch the url for the activity



Freely available

- The Github repo https://github.com/bitarellolab/Genomics_Teaching contains:
  - The code in R markdown and html formats
  - A slide deck for students
  - A direct link to launch the url for the activity



- Shortened link I made for this presentation (try it!): http://tinyurl.com/33wkwjwt

Learning goals

Learning goals

1) Have a soft introduction to the command line

Learning goals

1) Have a soft introduction to the command line

- `ls`
- `mkdir`
- `cd`
- `pwd`
- `more`
- `less`
- `wc`
- `pip`
- `conda`

Learning goals

1) Have a soft introduction to the command line

- `ls`
- `mkdir`
- `cd`
- `pwd`
- `more`
- `less`
- `wc`
- `pip`
- `conda`

2) Have a soft introduction to FASTA and FASTQ files

Learning goals

1) Have a soft introduction to the command line

- `ls`
- `mkdir`
- `cd`
- `pwd`
- `more`
- `less`
- `wc`
- `pip`
- `conda`

2) Have a soft introduction to FASTA and FASTQ files

3) Build an intuition around next-generation sequencing quality scores based on emojis

Structure

Structure

- Students follow the two tutorials through the mybinder.org link on their browser in order:

Structure

- Students follow the two tutorials through the mybinder.org link on their browser in order:
  - Bash Basics tutorial

Structure

- Students follow the two tutorials through the mybinder.org link on their browser in order:
  - Bash Basics tutorial
  - FASTQE tutorial

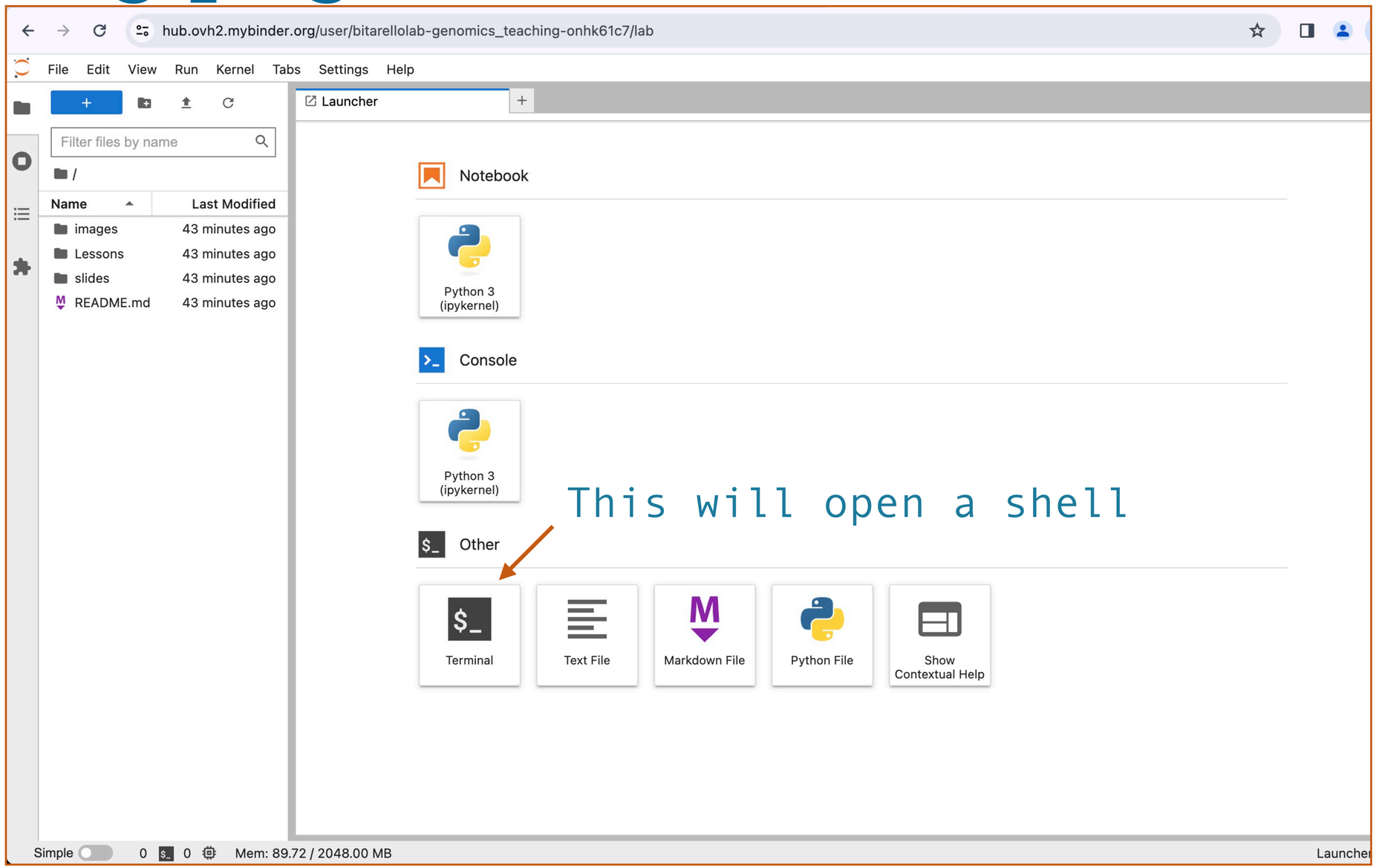
Structure

- Students follow the two tutorials through the mybinder.org link on their browser in order:
  - Bash Basics tutorial
  - FASTQE tutorial
- Students hand in answer sheet at the end of class

Structure

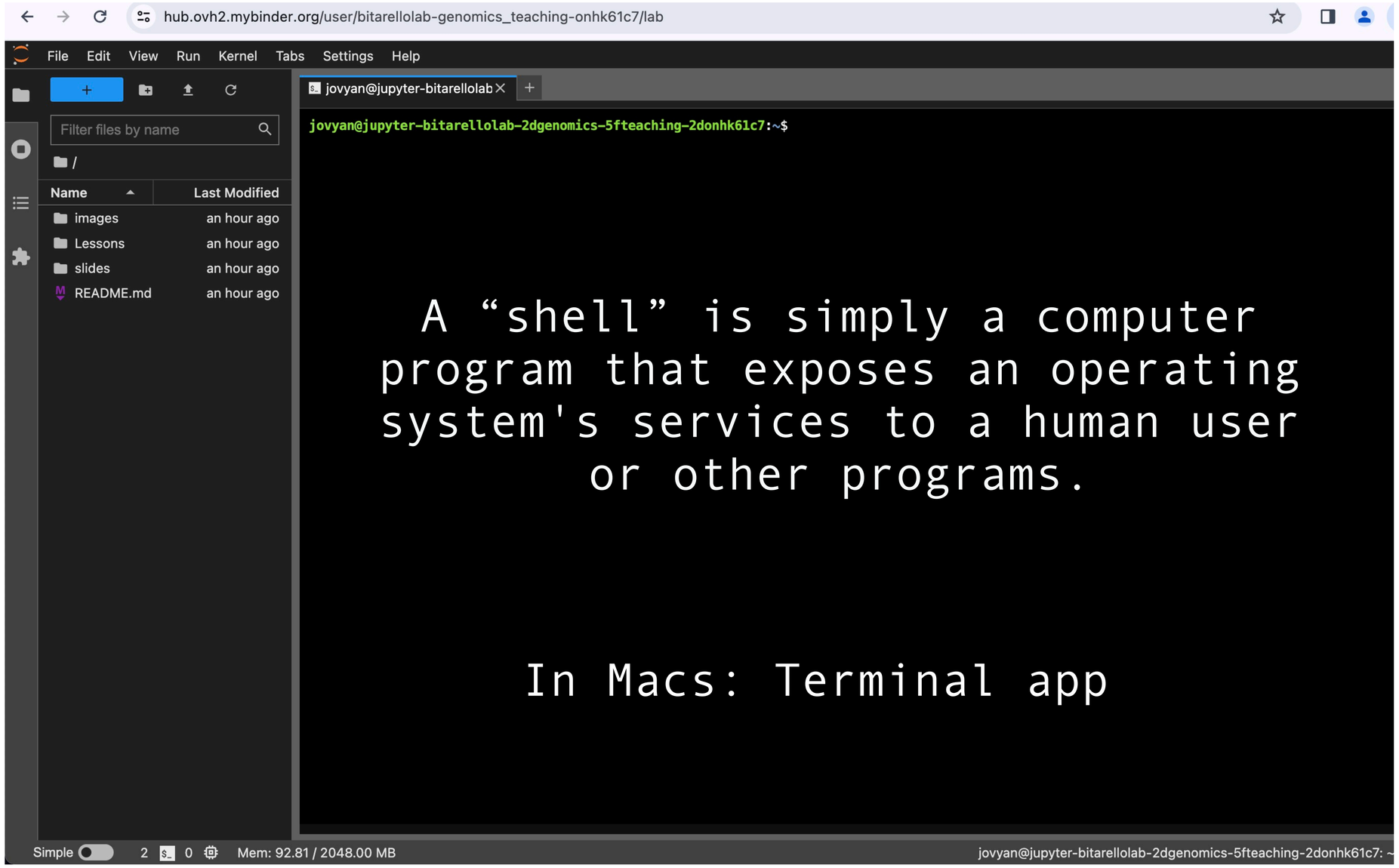
- Students follow the two tutorials through the mybinder.org link on their browser in order:
  - Bash Basics tutorial
  - FASTQE tutorial
- Students hand in answer sheet at the end of class
- Later in the semester: problem set question where they had to go back to this and analyze different sequence files

Landing page



This will open a shell

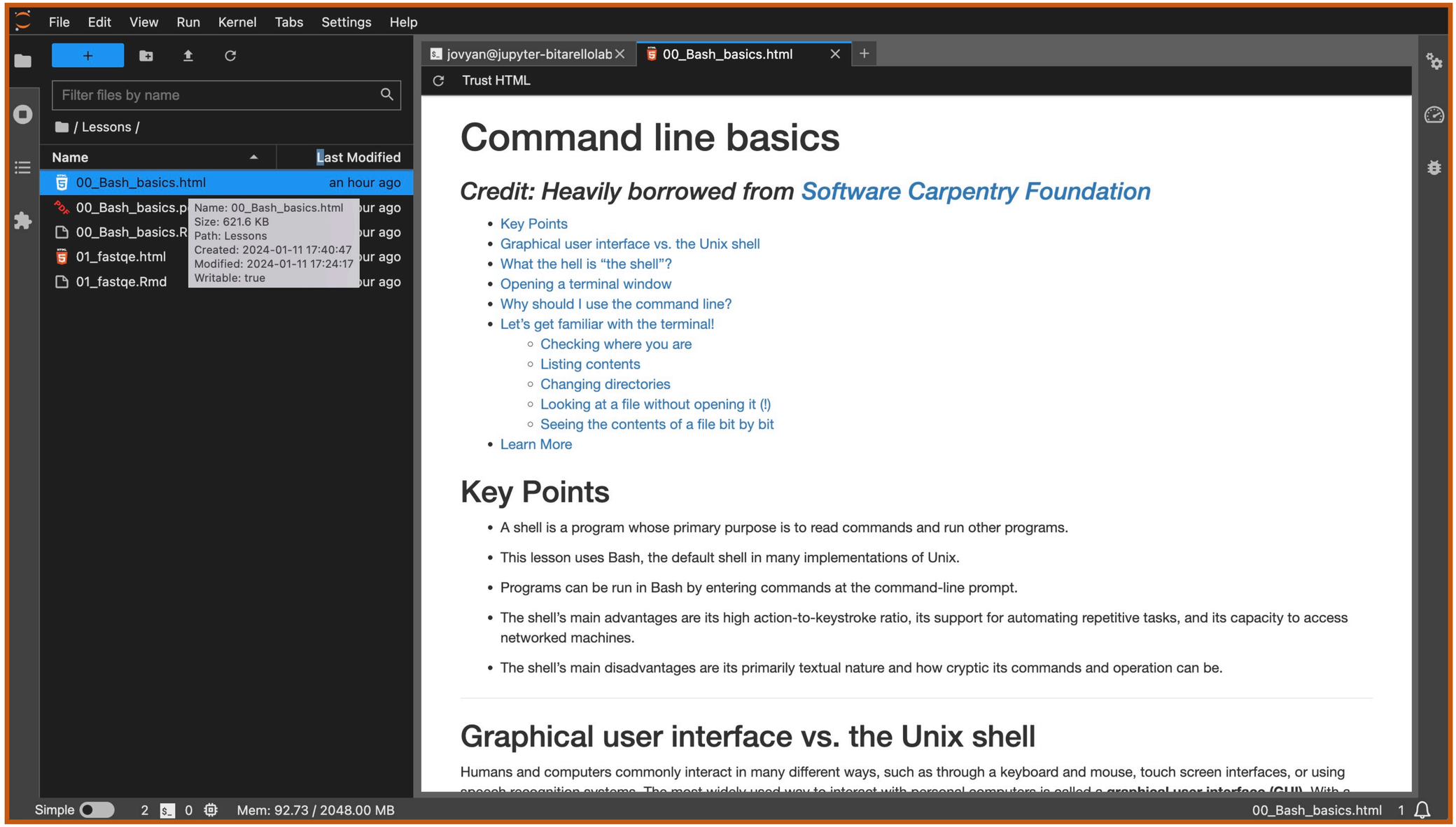
A shell is now open



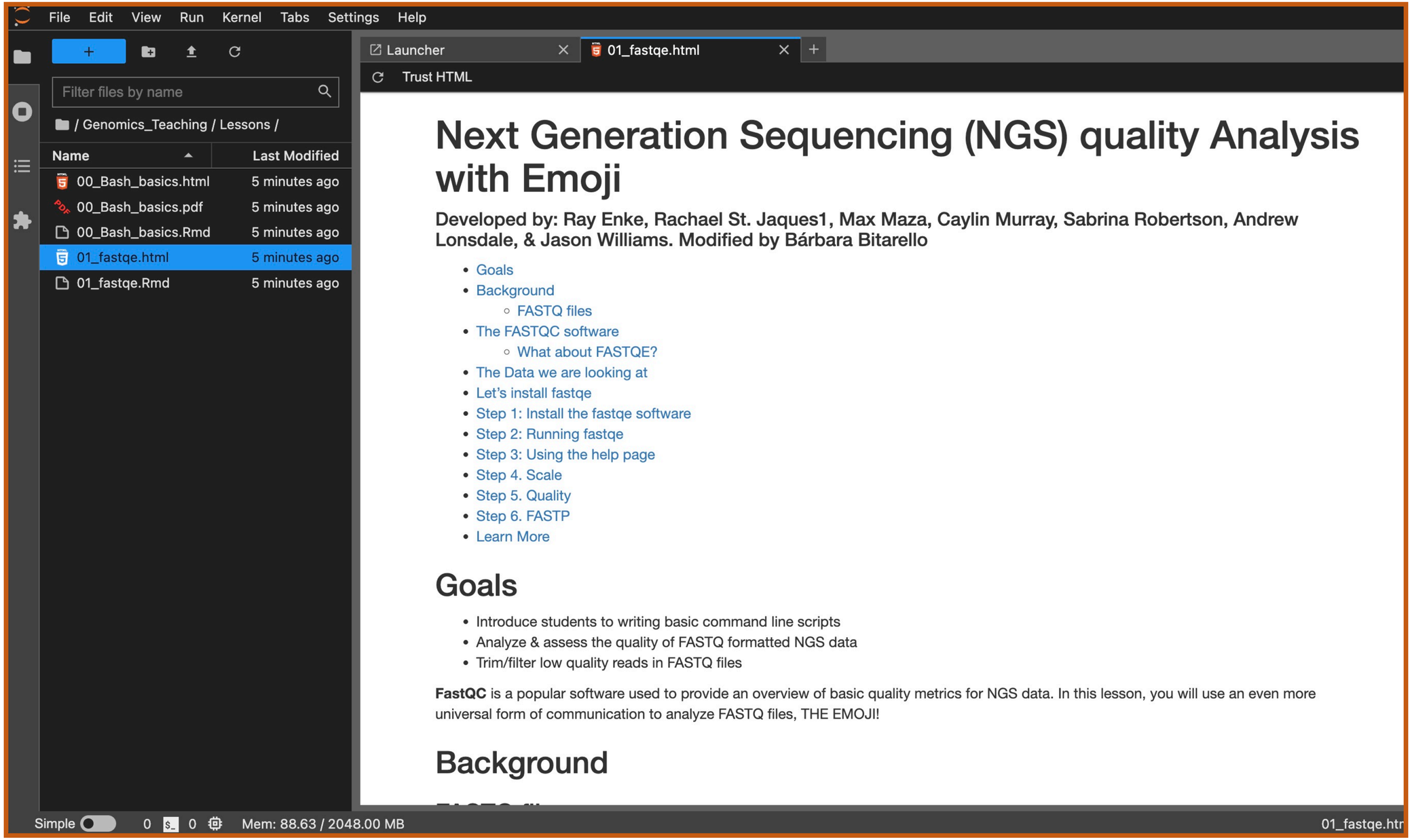
A “shell” is simply a computer program that exposes an operating system's services to a human user or other programs.

In Macs: Terminal app

Soft introduction to the command line



Learning about FASTQ files with FASTQE



Spread the love!

Spread the love!

- In its current implementation I have had zero issues with folks accessing it
- Occasionally can take a few minutes to load; gets faster the more it gets used

Spread the love!

- In its current implementation I have had zero issues with folks accessing it
- Occasionally can take a few minutes to load; gets faster the more it gets used
- Instructors with coding experience/teaching goals: can use this as a template for other activities
- Instructors without coding experience/teaching goals: can access and follow the activity seamlessly through my github

Spread the love!

- In its current implementation I have had zero issues with folks accessing it
- Occasionally can take a few minutes to load; gets faster the more it gets used
- Instructors with coding experience/teaching goals: can use this as a template for other activities
- Instructors without coding experience/teaching goals: can access and follow the activity seamlessly through my github

https://github.com/bitarellolab/Genomics_Teaching

Tl;dr

Tl;dr

- Adapts St. Jacques et al. (2021) materials so that everything can be installed and run from a browser

Tl;dr

- Adapts St. Jacques et al. (2021) materials so that everything can be installed and run from a browser
- Preserves the learning process of installing the packages while providing a uniform environment for all students

Tl;dr

- Adapts St. Jacques et al. (2021) materials so that everything can be installed and run from a browser
- Preserves the learning process of installing the packages while providing a uniform environment for all students
- Hopefully useful to instructors with different degrees of programming experience

Tl;dr

- Adapts St. Jacques et al. (2021) materials so that everything can be installed and run from a browser
- Preserves the learning process of installing the packages while providing a uniform environment for all students
- Hopefully useful to instructors with different degrees of programming experience
- Upcoming: expanding/modifying the intro to command-line portion

Tl;dr

- Adapts St. Jacques et al. (2021) materials so that everything can be installed and run from a browser
- Preserves the learning process of installing the packages while providing a uniform environment for all students
- Hopefully useful to instructors with different degrees of programming experience
- Upcoming: expanding/modifying the intro to command-line portion
- Possibility: publish on QUBES/CourseSource to increase visibility

Project 2: The Genomics Education partnership (GEP)

The Genomics Education Partnership (GEP)

The Genomics Education Partnership (GEP)

- Active member since July 2021

The Genomics Education Partnership (GEP)

- Active member since July 2021
- Integrates **active learning** into the undergraduate curriculum through CURES centered in **bioinformatics** and **genomics**

The Genomics Education Partnership (GEP)

- Active member since July 2021
- Integrates **active learning** into the undergraduate curriculum through CURES centered in **bioinformatics** and **genomics**
- Use of GEP materials depends on course structure and student background experience

The Genomics Education Partnership (GEP)

- Active member since July 2021
- Integrates **active learning** into the undergraduate curriculum through CURES centered in **bioinformatics** and **genomics**
- Use of GEP materials depends on course structure and student background experience
  - incorporating short lessons into an existing course (e.g., using a genome browser to investigate eukaryotic gene structure)

The Genomics Education Partnership (GEP)

- Active member since July 2021
- Integrates **active learning** into the undergraduate curriculum through CURES centered in **bioinformatics** and **genomics**
- Use of GEP materials depends on course structure and student background experience
  - incorporating short lessons into an existing course (e.g., using a genome browser to investigate eukaryotic gene structure)
  - participating in a genomics CURE centered around comparative gene annotation

Basic Curriculum: Understanding Eukaryotic Genomes (UEG)

All use the UCSC genome browser:

- Module 1: Intro to the UCSC Browser
- Module 2: Transcription
- Module 3: Post-transcription processing
- Module 4: Removal of introns
- Module 5: Translation
- Module 6: Alternative Splicing

- Pre and post-course, students fill out a survey that shows how much they know the content.
- Helps keep funding and assess learning gains

My experience with UEG materials

- Very customizable - materials are already great but all is easily available and free for anyone to edit
- Questions and keys are provided to instructors
- Each module fits well into a 3 hour lab
- Very positive feedback from students

Beyond UEG materials

- There are other highly curated materials + faculty-generated curriculum
- Particularly awesome for junior faculty (my opinion)
- Very helpful when designing new activities without having to start from scratch
- Lots of space to offer suggestions and contributions
- Conferences

Example 1: Mixing and adding

Genomics Education Partnership
BIOL B216 Genomics
Spring 2023
Bitarello

A hands-on Introduction to sequence homology with BLAST[1]

- Adapted by Bárbara Bitarello from two GEP lessons:
  - *An introduction to NCBI* (By: Wilson Leung)
  - *Introduction to Blast using human leptin* (By: Justin DiAngelo & Alexis Nagengast)
- Optional: submit this for evaluation by GEP to become part of curated curriculum

Example 2: New but GEP-inspired

BIOL B216 Genomics
Spring 2023
Bitarello

Using UNIPROT to explore protein sequences and build a phylogenetic tree

Example 2: New but GEP-inspired

- Activity in a google doc with links to:
  - Uniprot: https://www.uniprot.org/
  - Clustal Omega, Blast (inside Uniprot)
  - iTol Interactive Tree of Life page: https://itol.embl.de/

Example 2: New but GEP-inspired

- Activity in a google doc with links to:
  - Uniprot: https://www.uniprot.org/
  - Clustal Omega, Blast (inside Uniprot)
  - iTol Interactive Tree of Life page: https://itol.embl.de/

- My own, but using the GEP "format" as inspiration saved me a lot of time

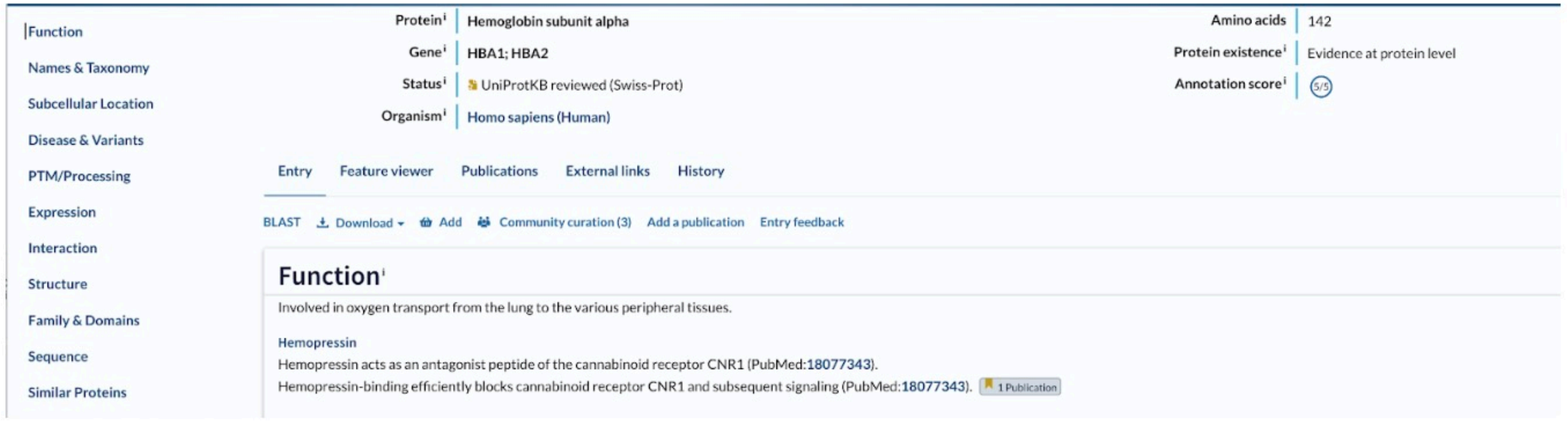
Example 2: New but GEP-inspired

- Activity in a google doc with links to:
  - Uniprot: https://www.uniprot.org/
  - Clustal Omega, Blast (inside Uniprot)
  - iTol Interactive Tree of Life page: https://itol.embl.de/

- My own, but using the GEP "format" as inspiration saved me a lot of time
- Optional: submit this for evaluation by GEP to become part of curated curriculum

Activity 1: Learning your way around UniProt

- Go to https://www.uniprot.org/
- Search for "hemoglobin subunit alpha" (in quotes)
- You will get something like 2,991 results (as of April 9th, 2023).
- Scroll down and find the human entry.
- Click on the human entry (it should be the first one): P69905
- This takes you to a page with a lot of information about this specific protein.



- Take some time to explore the data in this page.

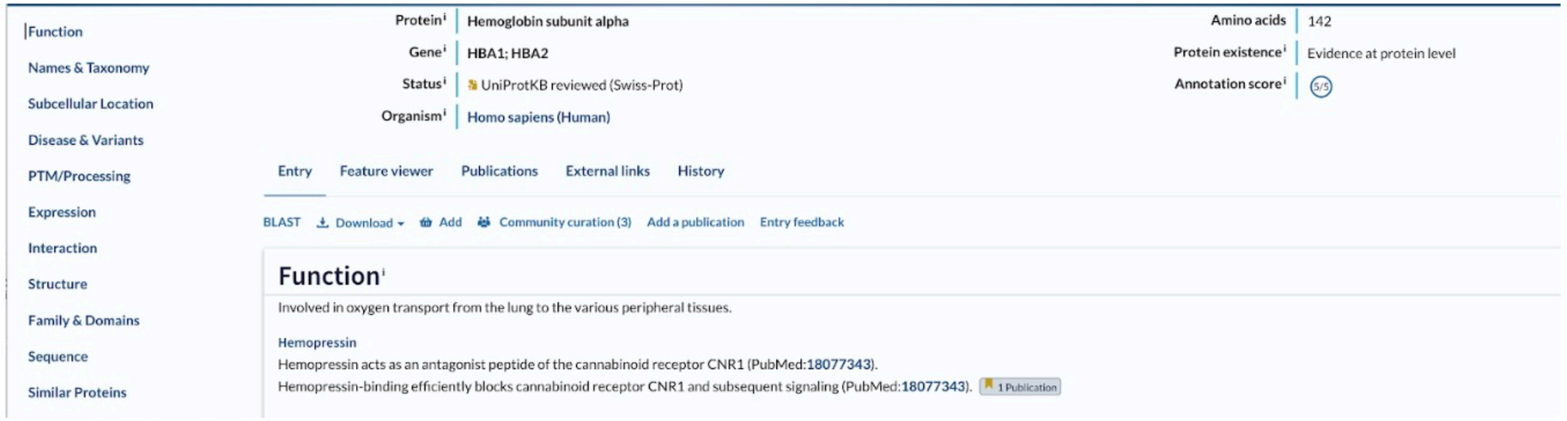
Questions:

1) How many aminoacids long is this protein?
2) How many isoforms of this protein are there?
3) What is the main function of this protein (according to the description in the page?)

Step-by-step

Activity 1: Learning your way around UniProt

- Go to https://www.uniprot.org/
- Search for "hemoglobin subunit alpha" (in quotes)
- You will get something like 2,991 results (as of April 9th, 2023).
- Scroll down and find the human entry.
- Click on the human entry (it should be the first one): P69905
- This takes you to a page with a lot of information about this specific protein.



- Take some time to explore the data in this page.

Questions:

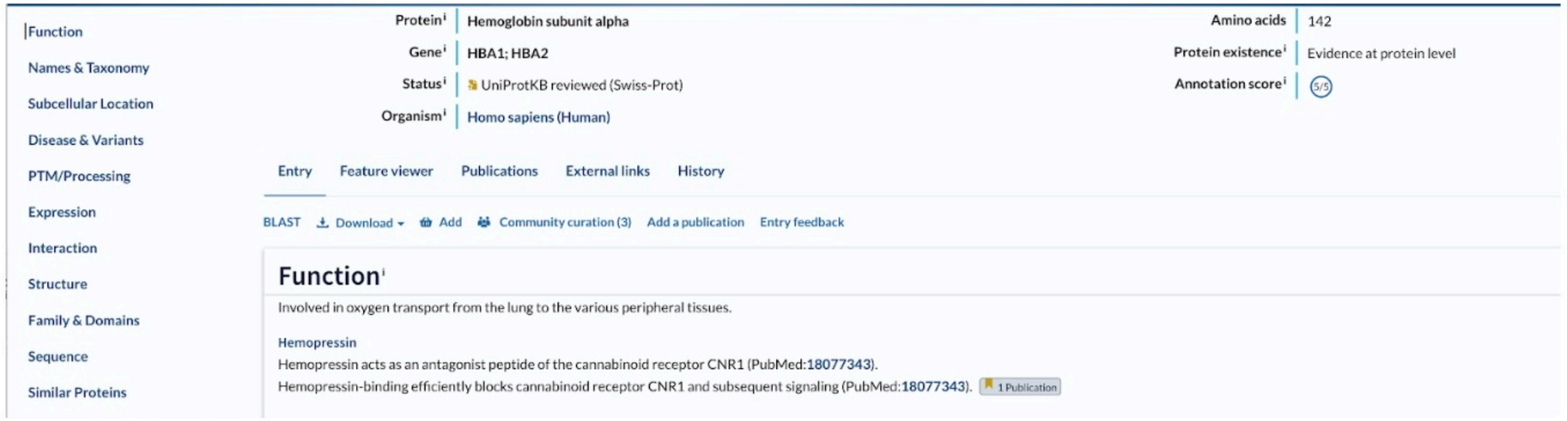
1) How many aminoacids long is this protein?
2) How many isoforms of this protein are there?
3) What is the main function of this protein (according to the description in the page?)

Step-by-step

Lots of screenshots

Activity 1: Learning your way around UniProt

- Go to https://www.uniprot.org/
- Search for "hemoglobin subunit alpha" (in quotes)
- You will get something like 2,991 results (as of April 9th, 2023).
- Scroll down and find the human entry.
- Click on the human entry (it should be the first one): P69905
- This takes you to a page with a lot of information about this specific protein.



- Take some time to explore the data in this page.

Questions:

1) How many aminoacids long is this protein?
2) How many isoforms of this protein are there?
3) What is the main function of this protein (according to the description in the page?)

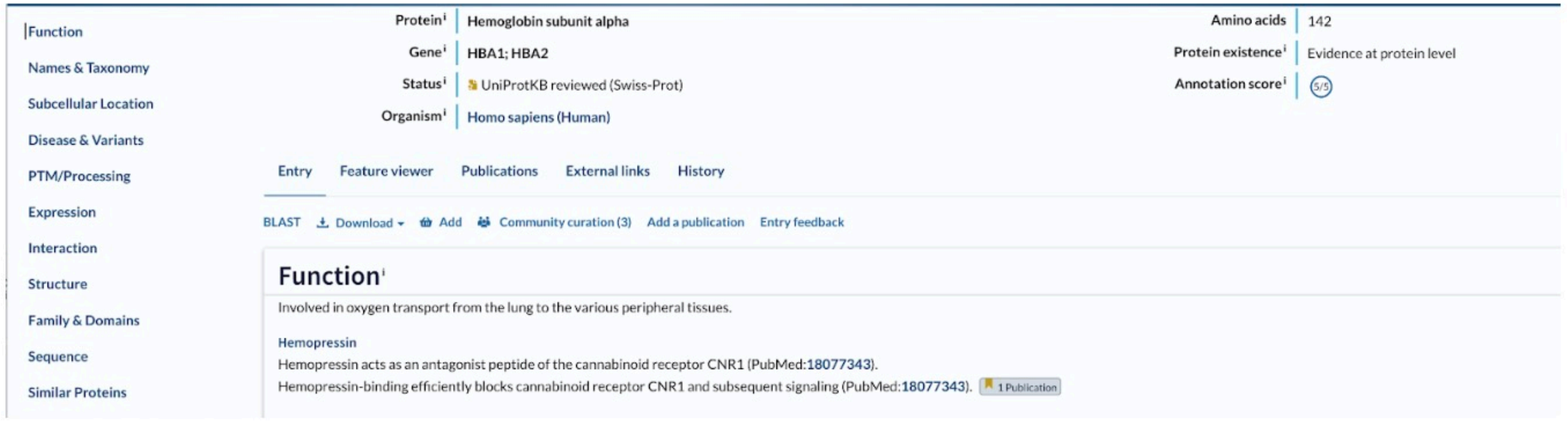
Step-by-step

Lots of screenshots

Questions handed in at the end of lab

Activity 1: Learning your way around UniProt

- Go to https://www.uniprot.org/
- Search for "hemoglobin subunit alpha" (in quotes)
- You will get something like 2,991 results (as of April 9th, 2023).
- Scroll down and find the human entry.
- Click on the human entry (it should be the first one): P69905
- This takes you to a page with a lot of information about this specific protein.



- Take some time to explore the data in this page.

Questions:

1) How many aminoacids long is this protein?
2) How many isoforms of this protein are there?
3) What is the main function of this protein (according to the description in the page?)

Bonus: Other projects

- Digital Scholarship Grant from BMC: Developing an R package with materials for the *B215: Biostatistics with* R course.
- Many great resources out there, but not very specific to biology
- $5,000$ to pay two UG students to help me with this (2023-2024)
- Currently not a package but materials were used this fall
- Considering making this a mybinder.org repo instead

Acknowledgments

For the command-line activity portion, I took heavy inspiration from:

- The Software Carpentry. https://swcarpentry.github.io/shell-novice/01-intro.html (Accessed March 22, 2023)

- https://thegep.org/

Thank you!

Questions?

Email: bbitarello@brynmawr.edu
Website: https://bitarellolab.digital.brynmawr.edu/
GitHub: https://github.com/bitarellolab
Twitter (X): @dudutchy

Thank you!

Questions?

Email: bbitarello@brynmawr.edu (happy to share materials!)
Website: https://bitarellolab.digital.brynmawr.edu/
GitHub: https://github.com/bitarellolab
Twitter (X): @dudutchy