

Compgenomics Wiki and Github

Aroon Chande

Lecture 2

Thursday, January 11, 2018

Agenda for today

- Quick intro to Comp Genomics
- Introduce the class wiki page and Github
- Choose your teams
- Create your wiki profiles

This is a real world problem, in need of computational solutions



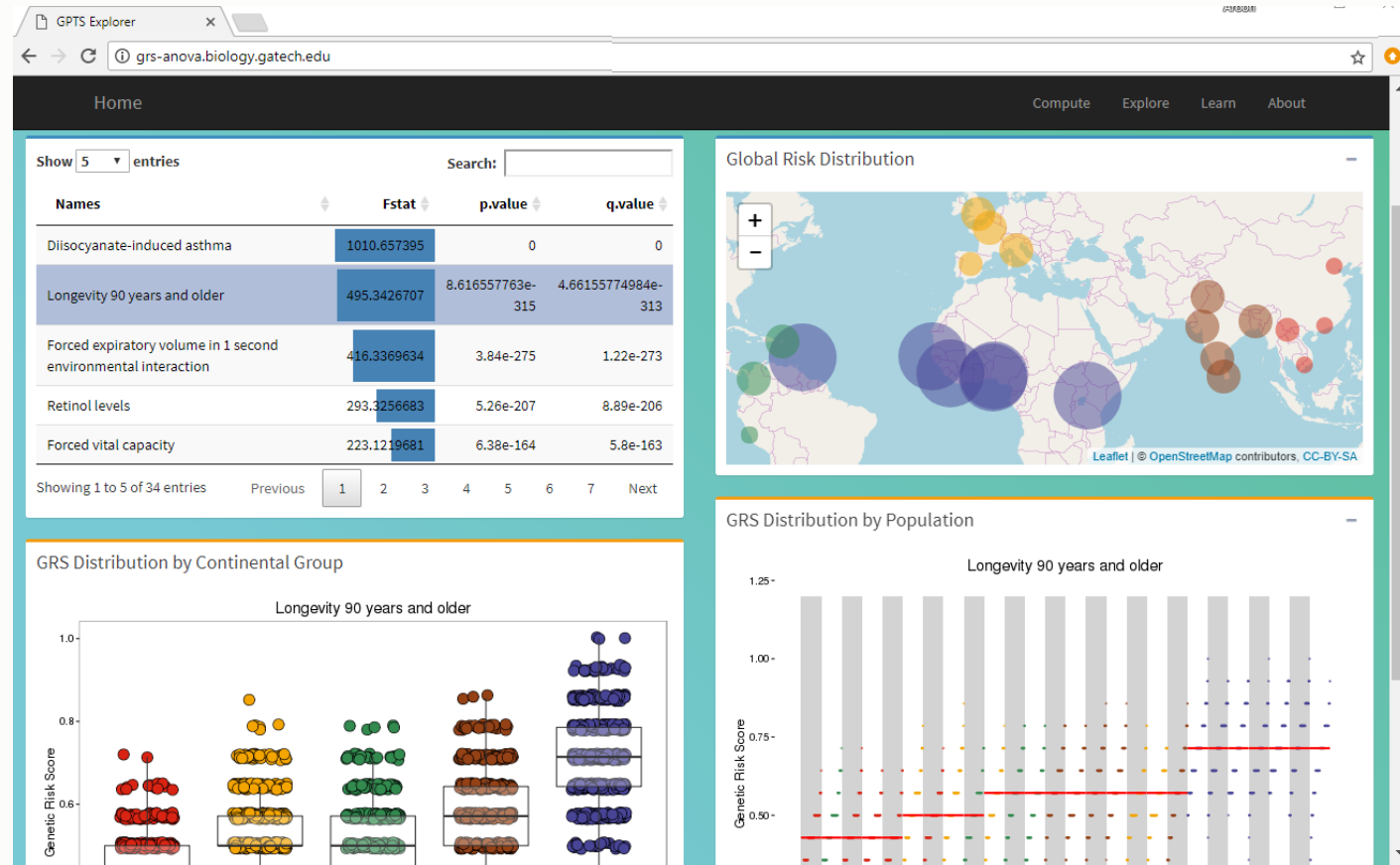
Translating genomics into practice for real-time surveillance and response to carbapenemase-producing Enterobacteriaceae: evidence from a complex multi-institutional KPC outbreak

Jason C. Kwong^{1,2,3,4}, Courtney R. Lane^{4,5}, Finn Romanes⁵, Anders Gonçalves da Silva^{1,4}, Marion Easton⁵, Katie Cronin⁶, Mary Jo Waters⁶, Takehiro Tomita⁴, Kerrie Stevens⁴, Mark B. Schultz^{1,2,4}, Sarah L. Baines^{1,2}, Norelle L. Sherry^{1,2,3}, Glen P. Carter^{1,2}, Andre Mu^{1,2}, Michelle Sait⁴, Susan A. Ballard^{1,4}, Torsten Seemann^{1,7}, Timothy P. Stinear^{1,2} and Benjamin P. Howden^{1,2,3,4}

DOI 10.7717/peerj.4210

Very quick overview of the class

- This is the **tenth** year that this course is being offered.
- In the end you will produce something like this →



Some Tips

- **Start Early** – Most of these processes will take time *i.e.*, don't start immediately after your background presentation, you need to present some preliminary results right after it.
 - **Server people** – you have to start even earlier. Setting up the server could be easy or extremely difficult depending on how much you've read and understood about the tool (and your skill level).

Some Tips

- **Document continuously** – It is tempting to leave documentation of your work and results to the end; don't! Document your code, analysis, and results as you go. This results in better documentation and won't leave you awake at 3am the day before your group final presentation
 - Wiki writing is just as important as analysis documentation.

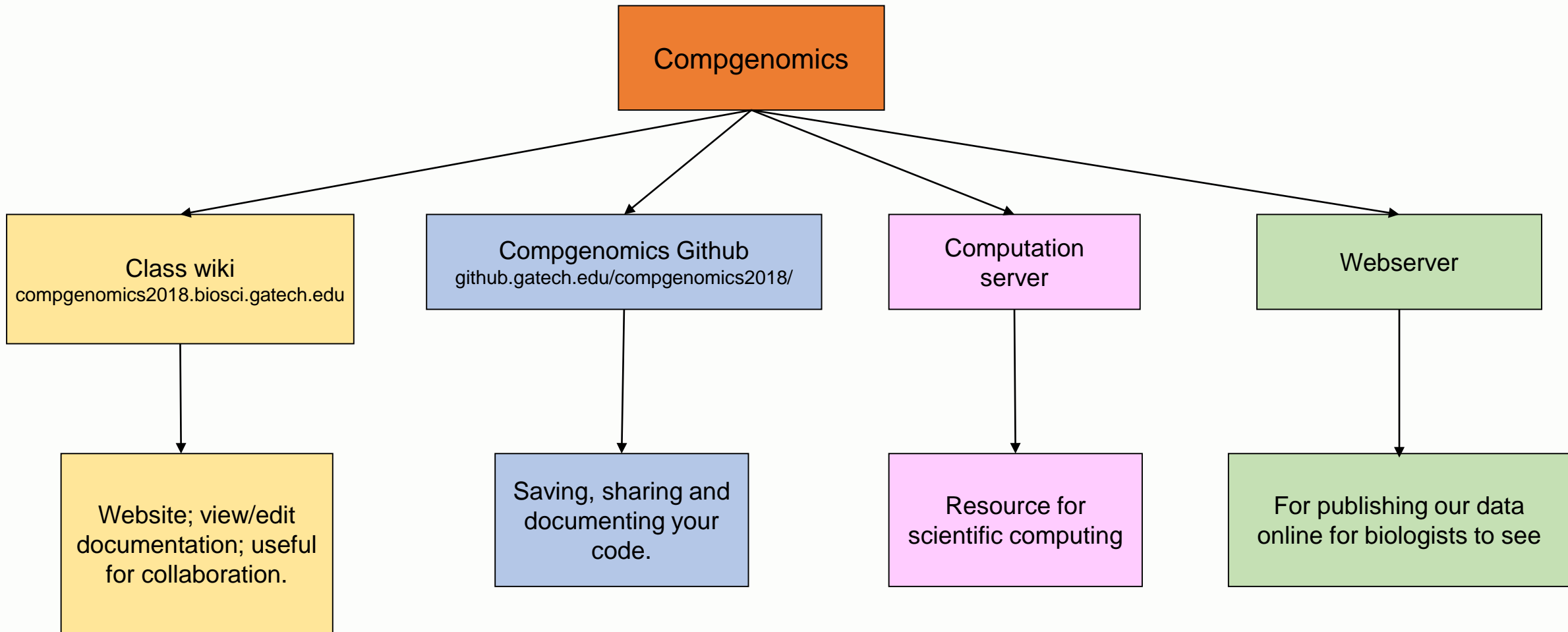
Some Tips

- **Stick to an accession convention across the groups** when processing genes/proteins – this will avoid confusion, unnecessary mapping and will make sure you do not wind up with duplicates
- **Have proper channels for exchange of data** – again, this avoids confusions

Some Tips

- **Everyone likes the new stuff** – Try to search for what is recent in the literature and attempt to use it along with the classical tools
- **Understand what you are doing** – don't blindly follow what other classes have done, have a proper understanding and basis of what you are doing

CompGenomics Resources



Course Wiki page

<http://compgenomics2018.biosci.gatech.edu/>

- The course wiki page is a public display of your work
- Lectures, readings, and all other course material can be found here
- We ***will not*** be using T-square, other than to send email announcements to the class

Previous wiki pages

- We are using different hosting infrastructure this year, past CompGenomics wiki pages can be found at:
`compgenomics[2009-2016].biology.gatech.edu`
- Use previous classes as a reference for what kind of documentation is expected

Don't reinvent the wheel

- Or at least, don't reinvent it every time
- Scientific research is building and expanding on existing knowledge
- For some of the tasks you'll be given, there may only be one “great” way to do them
 - We won't penalize doing the same thing as previous years *if* you provide your own, data-driven reasoning

Wiki accounts

- All students are required to create a wiki account and contribute to the wiki
- In order to register you must use your “user@gatech.edu” email address
- You have until Monday, January 15th to register an account

Wiki Accounts

- The wiki is public and where your collaborators, other scientists, and maybe future employers will see your work
- Once we have assigned groups, add yourself on the Profiles page. Be sure to include a picture.



Account registration

Compgenomics 2018

Not Secure www.compgenomics2018.biosci.gatech.edu/Main_Page

Not logged in Talk Contributions **Create account** Log in

Main page Discussion Read View source View history Search Compgenomics 2018

Main Page

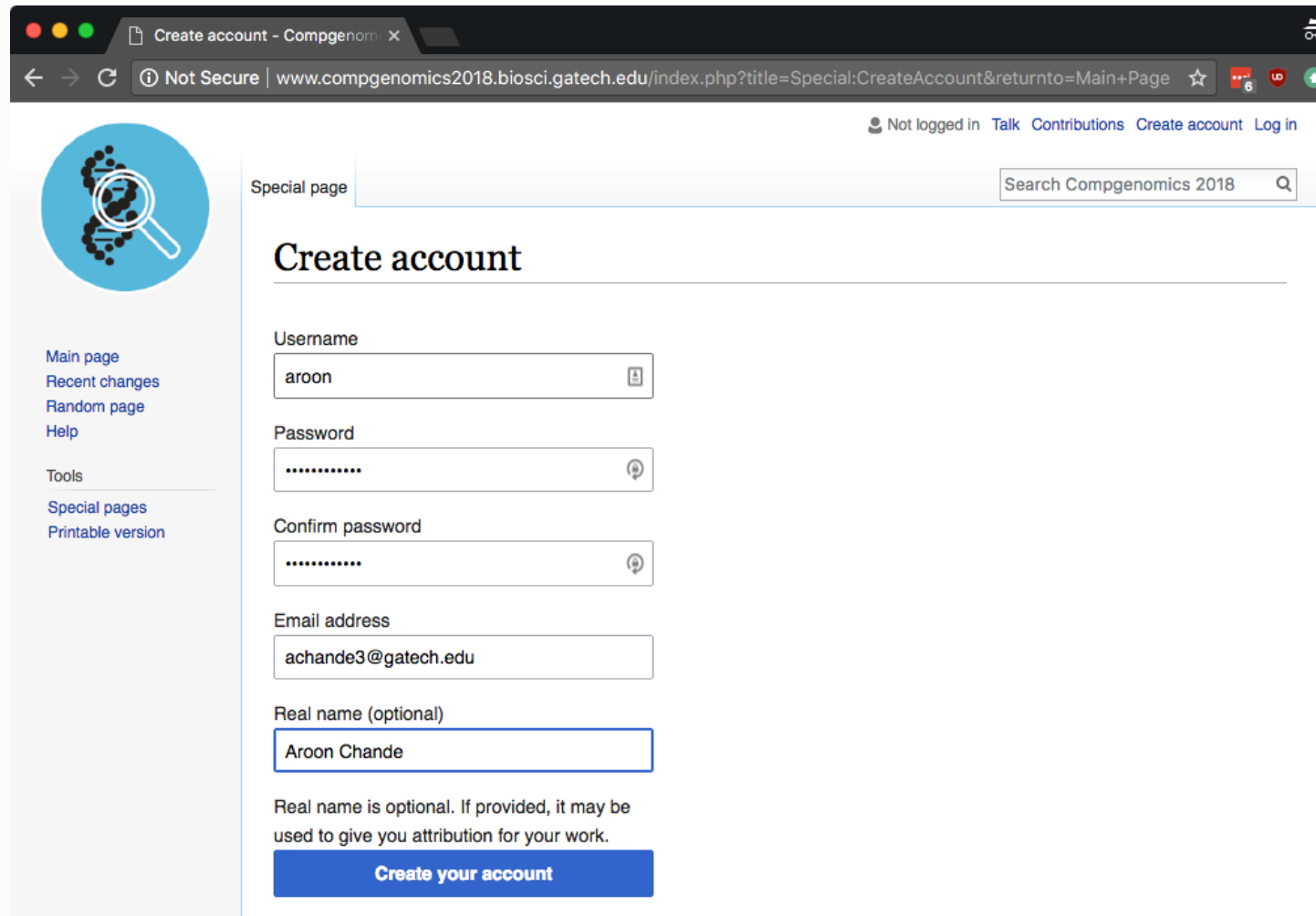
Click “create account”

Contents [hide]

- 1 Computational Genomics
- 2 Course Materials
- 3 Guest Lecturers
- 4 Working Groups
- 4.1 Team I:

Main page
Recent changes
Random page
Help
Tools

Account registration



Special page

Create account

Username

Password

Confirm password

Email address

Real name (optional)

Real name is optional. If provided, it may be used to give you attribution for your work.

[Create your account](#)

[Main page](#)
[Recent changes](#)
[Random page](#)
[Help](#)

[Tools](#)

[Special pages](#)
[Printable version](#)

Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Class Github

<https://github.gatech.edu/compgenomics2018/>

- I'll talk more about what git and source code management are next week on Tuesday
- This site will contain **all** your code, raw analysis and other work product
- During the semester, these data will be private, afterwards they will be open-sourced (on Github.com)

Why use git?

- Git and other source code management (SCM) tools are very useful for storing and tracking the files and scripts you create when working
- It provides an audit trail of *who* made what *changes*, *when*
- Use and proficiency with SCM is quickly becoming a required skill in both industry and the public sector (i.e. CDC)

Why use git?

- Git keeps everyone accountable
- Keep collaborative files in one place
- Provides a platform to organize, document, and share your work

Github student pack

<https://education.github.com/pack>

- Get free and discounted access to AWS (Cloud servers), DigitalOcean (Virtual servers), and other useful software tools
- Great for learning and refining some of the skills you'll learn this semester

Group selection