

# Section I-1:

## Introduction to Course structure, Computing, Bioinformatics, Sequencing, and Genomics

Introduction to Bioinformatics  
Analysis

Phillip Richmond

# Foreword

- This is not a comprehensive detailing of any of the subjects
- This is simply to orient you and provide a useful base of knowledge

- Screencasts and Slides for lectures are pre-recorded and should be watched before coming to class
- I recommend trying to follow along the screencast with the tasks I'm performing
- In class, there will be a short presentation, followed by working on a problem-set together



# Course Website

- To use the course website, you need to sign up for a GitHub account (it's free). This was a prerequisite for the course

<https://github.com/>

- GitHub and repositories are topics that we may go over later in the course, but if you decide to pursue computational biology then it's quite a nice tool to have
- Then type into the search tab:

Bioinformatics-Introductory-Analysis-Course

# Course Web page

- The address is:



[https://github.com/Phillip-a-richmond/  
Bioinformatics-Introductory-Analysis-Course](https://github.com/Phillip-a-richmond/Bioinformatics-Introductory-Analysis-Course)

- By scrolling down, you will see the README, which has details about the course

## Bioinformatics-Introductory-Analysis-Course

This repository contains course material for Phillip Richmond's Introduction to Bioinformatics Analysis Course

### Course Purpose

The purpose of this course is to provide an introduction to Bioinformatics and Genomics as it pertains to short-read sequencing analysis, with a specific focus on brewing and fermentation yeast strains. Taught in an inverted classroom format, there will be screencasts and lecture notes for each section that can be gone through independently, and in class we will simply work on example datasets and problem sets.

### General Course Information & Prerequisites

- Prerequisites
  - ◊ WestGrid access is required, and is different for a PI vs. a student
    - Westgrid Account information [here](#)
    - For a Westgrid Account, register by following these [instructions](#). Realize that the process can take several days!
  - ◊ GitHub Account
    - It's free, and you can sign up [here](#)
  - ◊ For mac/linux users, only need native terminal which comes with the operating system
  - ◊ For PC users, download and install [MobaXterm](#)
- General Course Info
  - ◊ Videos are hosted via YouTube at Phil Richmond's YouTube Channel [here](#)
  - ◊ Slides, Assignments, and Quizzes are hosted via Google Drive [here](#)
  - ◊ Extra resources:
    - [Linux/Unix Cheatsheet](#)
    - [Linux/Unix online tutorial](#)

### Course Outline

#### Section I: Introductions, Linux/Unix, WestGrid

1. Introduction to Next Generation Sequencing, Bioinformatics, and Computing
  - ◊ [Slides](#)
  - ◊ [Video](#)
2. Getting set-up on WestGrid and using terminal
  - ◊ Open up a terminal
  - ◊ Customize terminal
  - ◊ Login to WestGrid
  - ◊ [Specs of Orcinus](#)
  - ◊ [Slides](#)
  - ◊ [Video](#)

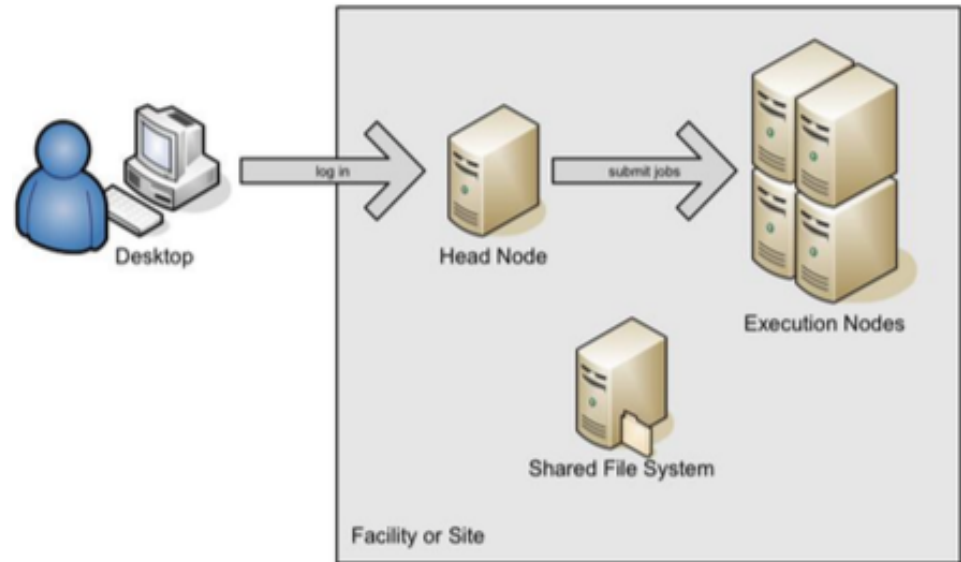
You should  
have already  
done these!!

Links in  
blue for  
important  
websites

Links in blue  
for videos on  
YouTube, or  
slides on  
Google Drive

# Computing via servers

- User interacts with their own desktop
- Through a terminal, they can communicate with the head node
- The head node communicates with the execution nodes through the job scheduler



# Compute Canada and WestGrid

- Compute Canada
  - Organization that works to bring computational resources to researchers in Canada
- WestGrid
  - One branch of Compute Canada, composed of 15 institutions across BC, Alberta, Saskatchewan, and Manitoba
  - <https://www.westgrid.ca/>



# WestGrid Orcinus (hosted at UBC)

[orcinus.westgrid.ca](http://orcinus.westgrid.ca)

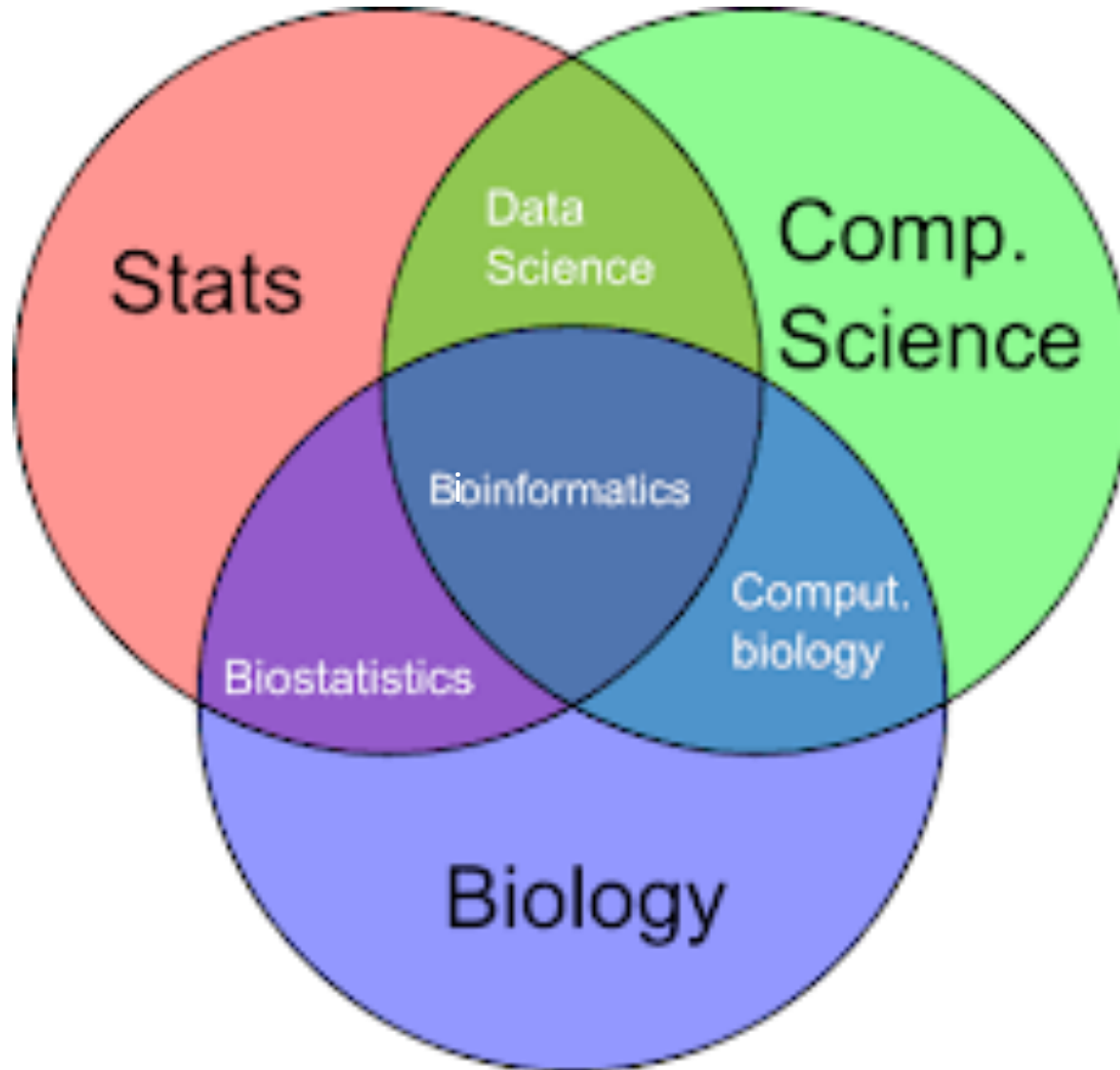
- 384 Nodes
  - 8-processors
  - 12Gb RAM
- 544 Nodes
  - 12-processor
  - 24 Gb Ram
- Total of 9600 “cores” or “processors”
  - That’s a LOT



# So when we start using the server

- We will need:
  - A way to interact with the head node
    - Terminal (Mac)
    - MobaXterm (Windows)
  - A way to interact with the compute nodes
    - SGE Job scheduler, by submitting job scripts

# Bioinformatics



# Next Generation Sequencing

- There are a lot of online resources describing next generation sequencing
- In this course, we will be using data from the Illumina sequencing platform
- <https://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course/what-next-generation-dna-sequencing/illumina->
- [Illumina:](#)
  - <https://www.youtube.com/watch?v=womKfikWlxM>

# Genomics

- “The branch of molecular biology concerned with the structure, function, evolution, and mapping of genomes.” – Merriam Webster Dictionary



# Genomics and Wine Yeast

- Through this course I hope to expose you to the tools and platforms on which you will analyze yeast genomics data
- The problem sets, and examples, will all be using publicly available yeast data
- The main genomics topics in this course will include:
  - Resequencing and Variant Calling
  - Genome Assembly and Comparative Genomics
  - RNA-seq Differential Expression Analysis