Special Topics:

Techniques in Genotyping-by-Sequencing (GBS)

Course Number: MCB 599 CRN 58928 006: 1 graduate credit P/N

**Course Meeting Times/Days:**

T/R, 1:30 p.m. - 2:50 p.m., ALS 4000. Spring 2016, April 12 to May 5.

**Course Credits and Grading:** This course combines approximately 20 hours of instruction and assignments for 1 credit. This course will be graded on a Pass/No Pass basis.  
  
**Course Prerequisites:** Graduate standing and/or approval by the instructor. Familiarity with the technology and biological use cases of high throughput sequencing and some experience with command line is required, preferably attendance in MCB 599 Intro to Unix/Command Line. Some basic knowledge or R and command line as well as a basic statistics or previous data analysis course is required in addition to instructor permission.   
  
**Instructor: Adelaide Rhodes**

* Office: 3133 ALS; (541)737-1628
* Adelaide.Rhodes@cgrb.oregonstate.edu
* Scheduled office hours: Tuesday and Thursday, 11:30 a.m. to 12:30 p.m.
* Office visits also can be arranged via email.

**Course Webpage:** <http://teaching.cgrb.oregonstate.edu/MCB/Rhodes/GBS_Spring_2016/index.html>

**Course Content:** This is a lecture/computer lab practicum course to provide an overview of and practical experience in some of the most commonly used methods for GBS analysis. Guest lectures and guided computer exercises will enable participants to apply a variety of methods to the generation and analysis of GBS data. We will also cover variant detection and quality control for SNP calling. Guest lecturers and guided computer exercises will expose participants to practical data management, data quality assessment, visualization tools, and the statistics underlying genotyping-by-sequencing. This class will not cover population genetics in any detail, but will focus more on the preparation of data for downstream analysis from laboratory and field studies.

**Course Specific Measureable Student Learning Outcomes:**

At the end of this course, you should be able to:

1. Interact more efficiently with core lab personnel on designing and implementing GBS experiments.
2. Utilize the cloud computing environment at the Center for Genome Research and Biocomputing at Oregon State University.
3. Run scripts interactively and in batch using command line for GBS Analysis.
4. Utilize R-scripts for experimental design before GBS and data visualization afterwards.

**Learning Resources:**  
There is no textbook. Course material will be made available as needed on the website or through Blackboard/Canvas throughout the semester.   
  
**Evaluation of Student Performance:**  
Students will be evaluated through computational assignments, quizzes and attendance. Attendance for enrolled students is mandatory. Absences due to reasons other than illness must be approved with the instructor at least 48 hours before the class session.   
**Module I Evaluation Components:**

1. Attendance 80 points (25%)
2. Homeworks and Quizzes 160 points (50%)
3. In-class Tutorials 80 points (25%)

**Statement Regarding Students With Disabilities:**  
Accommodations are collaborative efforts between students, faculty and Disability Access Services (DAS). Students with accommodations approved through DAS are responsible for contacting the faculty member in charge of the course prior to or during the first week of the term to discuss accommodations. Students who believe they are eligible for accommodations but who have not yet obtained approval through DAS should contact DAS immediately at 737-4098. DAS website:[http://ds.oregonstate.edu](http://ds.oregonstate.edu/)   
  
**Academic Honesty**   
The students are expected to be honest and ethical in their academic work. Any incident of academic dishonesty will be handled according to the University's Academic Regulations <http://catalog.oregonstate.edu/ChapterDetail.aspx?key=75#Section2883>   
  
**Link to Statement of Expectations for Student Conduct:**  
<http://studentlife.oregonstate.edu/studentconduct/>  
<http://oregonstate.edu/main/about/copyright>  
<http://oregonstate.edu/main/about/disclaimer>

|  |  |  |
| --- | --- | --- |
| **Week/Date** | **Activity/Location** | **Assignments & Due Dates** |
| April 5, 2016 | Orientation I:  Logging In and File Management  **ALS 4000** | HW0: Basic Command Line (ungraded)  Due: April 6th by 5 p.m., please email directly to Matthew. |
| April 7, 2016 | Orientation II:  More shell commands  **ALS 4000** | **PLEASE READ BEFORE CLASS BEGINS ON TUESDAY, APRIL 12th.**  A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species, Elshire et al. 2011  <http://teaching.cgrb.oregonstate.edu/MCB/Rhodes/GBS_Spring_2016/>  Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Davey et al. 2011  [http://teaching.cgrb.oregonstate.edu/MCB/Rhodes/GBS\_Spring\_2016/Davey\_2011\_NatureGenotypingReview.pdf](http://teaching.cgrb.oregonstate.edu/MCB/Rhodes/GBS_Spring_2016/Davey_2011_NatureGenotypingReview.pdf%20)  GBS Bioinformatics Pipeline ...or, “Where Your Data Go After Sequencing” Powerpoint Slides <http://teaching.cgrb.oregonstate.edu/MCB/Rhodes/GBS_Spring_2016/Buckler_FilterImpTools111028.pdf> |
| April 12, 2016 | Syllabus  Intro Lecture  “Intro to GBS” – Javier Tabima  Guest Lecture –  “Laboratory Methods for GBS” – Aaron Trippe  **ALS 4000** | Please finish the above readings before class.  HW1 (20 pts.): Please read through the handout on *in silico* digest using command line that we will be running as Tutorial 1 on April 14th.  Read this paper: Flexible and scalable genotyping-by-sequencing strategies for population studies. Hefelfinger et al. 2014 <http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-979> and answer the following questions. Please email me your responses before noon on April 14th.   1. (4 pts.) What is the difference between a blunt and staggered end restriction enzymes? 2. (4 pts.) What are the two organisms they used to test the blunt end enzymes? Name two of the eight blunt end restriction enzymes tested. 3. (4 pts.) Why did the Mlyl not retain the appropriate restriction motif as often as the other enzymes chosen? 4. (4 pts.) If we had a limited amount of money for sequencing, what fragment size range might be a good choice to achieve near saturation of sites within a limited size spectrum at a lower depth of coverage? 5. (4 pts.) Figure 3 discusses how the observed sites matched or did not match the predicted sites from *in silico* digest. [We are going to be doing *in silico* digest in class on Thursday]. 6. How many reads from maize aligned to predicted sites, and 7. How many reads from rice aligned to predicted sites? (Hint read down two paragraphs from the figure). |
| April 14, 2016 | Tutorial 1: *In silico* digest on command line (10 pts.)  Tutorial 2: Estimating coverage, using R-script (10 pts.)  **ALS 4000** | HW2 (20 pts.): Repeat Tutorial 2 from In-Class (the R-script) with a larger genome (a link is provided inside the script) and answer the following questions.   1. 5 pts. Discuss whether the individual enzymes and enzyme pair gave a different result than the genome used in class and provide a table comparing the in-class and homework genomes. 2. Did the general pattern stay the same – which enzyme would you choose for this project?   Assigned Reading to be Completed Before April 19:  TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. Glaubitz et al. 2014  <http://teaching.cgrb.oregonstate.edu/MCB/Rhodes/GBS_Spring_2016/Glaubitz_2014_TASSEL.pdf>  TASSEL: software for association mapping of complex traits  in diverse samples. Bradbury et al. 2007. <http://teaching.cgrb.oregonstate.edu/MCB/Rhodes/GBS_Spring_2016/Bradbury_et_al_2007.pdf>  Optional Reading:  Structured Association Mapping using STRUCTURE and  TASSEL. Vinod 2011.  <http://teaching.cgrb.oregonstate.edu/MCB/Rhodes/GBS_Spring_2016/TASSEL_STRUCTURE.pdf> |
| April 19, 2016 | Guest Lecture: Matthew Peterson, CGRB  “Tassel Pipeline”  **ALS 4000** | HW3 (20 pts): Read Tutorial 3 (Tassel Pipeline Command Line) and fill in the pipeline diagram (provided as a handout). Turn in this diagram and short answer worksheet at the beginning of the next class (Thursday, April 21 at 1:30 p.m.)  Make sure that the Tassel GUI program is loaded onto your laptop before class on Thursday.  If you are motivated, you can start running the tutorial before class, at least looking at the commands and making sense of them is the first step to running the pipelines on the infrastructure. Please bring your questions to class. |
| April 21, 2016 | Tutorial 3:  Running the Tassel Pipeline (20 pts.)  Interacting with the data on the Tassel GUI interface  **ALS 4000** | HW 4 (20 pts.): Data analysis on TASSEL program on your computer (20 pts.)  Using the worksheet from the in-class tutorial, answer the questions in the handout. Please turn this in at the beginning of class on April 26th at 1:30 p.m.  Assigned Reading to be Completed Before April 26:  Stacks: Building and Genotyping Loci De Novo  From Short-Read Sequences, Catchen et al. 2011  <http://teaching.cgrb.oregonstate.edu/MCB/Rhodes/GBS_Spring_2016/Catchen_2011_Stacks_de_Novo.pdf>  Stacks: an analysis tool set for population genomics, Catchen et al. 2013  <http://teaching.cgrb.oregonstate.edu/MCB/Rhodes/GBS_Spring_2016/Catchen_2013_Stacks_population_genomics.pdf>  GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data Melo et al. 2016. : <http://www.biomedcentral.com/1471-2105/17/29> |
| April 26, 2016 | Guest Lecture:  Kelly Vining  “Reference-based GBS Analysis”  Javier Tabima  “Non-reference based GBS Analysis”  Discussion  **ALS 4000** | HW 5 (20 pts.): Read through the Stacks pipeline handout and fill in the pipeline diagram (provided as a handout). Turn in this diagram and short answer worksheet at the beginning of the next class (Thursday, April 28 at 1:30 p.m.) |
| April 28, 2016 | Tutorial:  Running the Stacks Pipeline (20 pts.) | Assigned Reading to be Completed Before May 3:  Genome-wide association mapping for wood characteristics  in Populus identifies an array of candidate single nucleotide  polymorphisms. Porth et al. 2013  <http://teaching.cgrb.oregonstate.edu/MCB/Rhodes/GBS_Spring_2016/Porth_et_al_2013_GWAS.pdf>  VcfR: an R package to manipulate and visualize VCF format data. Knaus and Grunwald 2016  <http://biorxiv.org/content/early/2016/02/26/041277>  HW6 (20 pts.): TBD |
| May 3, 2016 | Lecture  “Quallity Control and Population Genetics”  Javier Tabima  **ALS 4000** | HW7 (20 pts.): TBD |
| May 5, 2016 | Tutorial 4: VcfR tool for data visualization in R. (20 pts.) | In class quiz – putting it all together (20 pts. for participation) |