

182

Washington State University
MAJOR CURRICULAR CHANGE FORM -- COURSE
(Submit original signed form and **ten** copies to the Registrar's Office, zip 1035.)

RECEIVED
FEB 04 2014
WSU REGISTRAR

Future Effective Date: 01/01/2015 New course Temporary course Drop service course
(effective date cannot be retroactive) There is a course fee associated with this course (see instructions)

- Variable credit _____
 - Increase credit (former credit _____)
 - Number (former number _____)
 - Crosslisting (between WSU departments)
(Must have both departmental signatures)
 - Conjoint listing (400/500)
 - Request to meet Writing in the Major [M] requirement (Must have All-University Writing Committee Approval)
 - Request to meet GER in _____ (Must have GenEd Committee Approval) Fulfills GER lab (L) requirement
 - Professional course (Pharmacy & Vet Med only) Graduate credit (professional programs only)
 - Other (please list request) _____
- Repeat credit (cumulative maximum _____ hours)
 - Lecture-lab ratio (former ratio _____)
 - Prefix (former prefix _____)
 - Cooperative listing (UI prefix and number _____)
taught by: WSU UI jointly taught
 - S, F grading

BIOLOGY 534 Modern methods in Population Genomics
course prefix course no. title

<u>3</u>	<u>3</u>	<u>0</u>	<u>0</u>	<u>Graduate Standing; BIOLOGY 519</u>
credit	lecture hrs per week	lab hrs per week	studio hrs per week	prerequisite

Description (**20 words or less**) Problems and prospects of designing a study with genomic data: from raw data to demography and selection inferences

Instructor: Omar Cornejo Phone number: 335-0179 Email: omar.cornejo@wsu.edu
Contact: Justine Rupp Phone number: 335-8649 Email: ruppj@wsu.edu
Campus Zip Code: 4236

- Please attach rationale for your request, a current and complete syllabus, and explain how this impacts other units in Pullman and other branches (if applicable).
- Secure all required signatures and provide 10 copies to the Registrar's Office.

[Signature] 28 January 2014 Chair/date
[Signature] 2/13/14 Dean/date
 _____ General Education Com/date

_____ Chair (if crosslisted/interdisciplinary)*
 _____ Dean (if crosslisted/interdisciplinary)*
 _____ Graduate Studies Com/date

_____ All-University Writing Com/date
 _____ Academic Affairs Com/date
 _____ Senate/date

***If the proposed change impacts or involves collaboration with other units, use the additional signature lines provided for each impacted unit and college.**

Rationale for the creation of the course: Modern Population Genomic analysis

Prof. Omar E Cornejo

The increased ability to produce next generation sequence data (either DNA or RNA) from plants, animals, and microorganisms has opened the possibility to address important questions in evolutionary and functional biology. The decreasing costs of next generation sequence technologies permit the study of genome-wide genetic data to characterize the variation from natural as well as experimental populations and understand questions like: what has been the demographic history of organisms? what are loci that have contributed to the adaptation of organisms to the environment they live in? what are effective barriers to gene flow in populations that are separated geographically? what are the loci that contribute to phenotypic differences in organisms? It is understandable that there is an increasing need for preparing biologists, plant scientists, immunologists, geneticists among other professionals with the necessary technical skills to make front to a new era in which big data has permeated almost all field of biology.

Currently there are no courses taught at Washington State University that teaches graduate students how to analyze raw sequence data (obtained with next generation sequencing technologies) in a population genetics framework. This proposed course will offer graduate students the opportunity to explore the prospects and problems of designing a study with genomic data, and have hands on experience on its analysis from beginning (raw fastq data) to end (Principal component studies, model inference, etc). The students will be required to synthesize this information and analyses performed and will have to present a manuscript presenting their results in a similar format to that required to that required for the submission to a scientific journal.

Including this course in the WSU catalog is particularly important for the growing number of students in the School of Biological Sciences and other department in other schools across campus that are attempting to engage in an increased use of next generation sequence data in their research. I believe that graduate students will certainly benefit from enrolling in this course because it will expose them to the increasing needs of learning the computational skills necessary for the analysis of genomic data and the use of resources like High Performance Computing. Previous enrollment (as BIOL 589-Advanced Topics in Biology) included zoology, botany, anthropology, and environmental science graduate students.

This proposed course meets the vision and mission of the College of Arts and Sciences and allows me to contribute by bringing my expertise to the educational plan to the college. The study of genomic data has become a highly interdisciplinary field and the continues advances introduce the need for teaching biology students from all areas the necessary computational tools and new ways of thinking that escape the traditional approaches within each field. The ability to understand and synthesize information will be emphasized all over the course and specific tasks within the class will be designed to reinforce systematic thinking and design. The tools and "way of thinking" taught in this class will help students beyond the immediate skills learned during the course to tackle complex problems and design studies to address problems that require the analysis of big data sets.

Creation of this class will require access to the HPC facilities on campus. It is my understanding that students have right of access to those facilities already and no additional expenses are required. Access to a computer lab would be ideal, but if the

university does not count on computer labs, then students would be required to work on their laptops.

**Biology 534: Modern methods in Population Genomics 3 credits
Spring 2015**

Instructor: Prof. Omar E. Cornejo
Office: Eastlick 395
email: omar.cornejo@wsu.edu
Phone: 335-0179

Office hours: Please email to make an appointment.

This course does not have a Teaching Assistant; all grading will be done by the instructor

Catalog Description: Problems and prospects of designing a study with genomic data: from raw data to demography and selection inferences

Prerequisites: graduate standing, Population Genetics Biology 519

Lectures: W 2:10 – 5:00. Course location TBD.

Required course materials. This course integrates across biological and computational sciences disciplines thus there is not a single text that encompasses course content. All required materials for the course will be provided on the course webpage (lms.wsu.edu). No text is required for the course. Citations and handouts will be provided on the course website. The lectures will be presented using PowerPoint. Slides will be placed on the course web site before each lecture. Handouts with basic command line material or access to tutorials will be provided.

Course objectives

Population Genetics and Evolutionary Biology are fields within biology that are being revolutionized by the possibility of generating genomic sequence data. Interestingly, much of theoretical population genetics is based on simple Mendelian inheritance played out at a population-scale. While many long-lasting insights in population genetics were derived mathematically before it was even known that DNA was known to be the molecular basis of heredity, modern population genetics nowadays is highly influenced by our ability to generate and analyze genetic data for whole genomes in multiple samples from a population.

This course offers graduate students the opportunity to have a first hand experience in the analysis of genomic data and go from raw sequence data obtained with new technology to finalized population genetic analyses. Some of these final results will include principal component analysis to assess population genetic structure, fitting site frequency spectrum (distribution of genetic variation genome wide) to demographic models and simple scans of selection.

The course presents the following general structure:

- Introduction to High Performance Computing (HPC) and linux (Week 1 - 2)
- Development of sequence analyses pipelines for quality control of raw sequence data, mapping, processing and generation of SNP calls in multiple individuals (Weeks 2 - 9)
- Modern tools in population genetic analysis (Weeks 9 - 16)

Specific objectives of the class include:

1. Develop a conceptual understanding of assumptions underlying basic population genetic/genomic analysis
2. Critically explore and evaluate current relevant literature and scientific hypotheses in population genomics and the tools designed for the analysis of big data
3. Foster interdisciplinary thinking and development to address complex evolutionary questions

I expect that students will:

1. Develop a fundamental set of skills to work in linux and to work on an HPC
2. Develop basic understanding of the error structure of the DNA sequence data generated with new technologies
3. Use critical thinking to design and set analysis pipelines for the mapping and post-processing of raw sequence data, and understand the underlying assumptions of the methods employed.
4. Develop a minimal set of skills to design simple analytical pipelines and identify single nucleotide variants genome wide
5. Develop skills to address major biological concepts in evolution, ecology and organismal biology as it relates to genomes using population genetic analytical tools.
6. Effectively interpret and communicate the results of population genetic analyses in full oral presentations as well as written reports

Student learning outcomes

	At the end of this course, students should be able to:	Course assignment that advances the learning goal	This objective will be evaluated primarily by:
Depth, Breadth and Integration of Learning	Integrate knowledge from diverse and broad fields and apply their knowledge to contemporary issues in population genetics/genomics. Understand the underlying assumptions of the methods employed.	Written and oral evaluation of scientific literature from diverse fields and in-depth, integrative oral and written analysis of data developed throughout the course. Depth and breadth will be test Final oral and written reports.	<ul style="list-style-type: none"> • Written manual of developed analysis pipeline • Oral presentation of results and analytical revision of the subject in light of current literature • Final written report in the form of a manuscript
Critical and Creative Thinking	Critically evaluate information and data from diverse sources	Written final report that will have to integrate the	<ul style="list-style-type: none"> • Written manual of developed analysis

	and apply this to their own professional goals	analyses performed during the class with current literature and put their results in a much general framework. Students will be required to justify their choice of analytical techniques in the light of the questions addressed.	<p>pipeline</p> <ul style="list-style-type: none"> • Oral presentation of results and analytical revision of the subject in light of current literature • Evaluation of continuous discussions about potential analysis based on current literature
Communication	Use oral communication and presentation skills to discuss current global change topics to both scientists and non-scientists	Participation in continuous group discussions of scientific literature. Preparation and presentation of a 30 min PowerPoint talk to peers, and evaluation of peers' scientific presentations	<ul style="list-style-type: none"> • Oral presentation of final results • Evaluation of peers' presentations
Information literacy	Use appropriate internet and library databases, library search tools, and library delivery services to find scientific literature to effectively define the scope of a research question or topic and obtain the necessary sources of scientific information to evaluate current topics in HPC usage and genomic analysis	Library and internet research of scientific literature for student review papers and presentations. Completeness of literature searches using library and internet tools.	<ul style="list-style-type: none"> • Written report on analysis pipelines and Final reports • Oral presentation on final results

Course schedule

Introduction to HPC usage and next generation sequence data

Week 1

Topic: Introduction to HPC and Linux

Week 2

Topics: Shell scripting and preparation of job submission to HPC (Torque or Sun Grid engine)

Quality control of high throughput sequence data

Mapping of genomic DNA data and Identification of Variants

Week 3

Topic: Mapping of raw sequence data to reference genomes (bwa)

Week 4

Topic: Post-processing of aligned reads (Samtools, Bamtools and Picard)

Week 5

Topic: Post-processing of aligned reads: local insertion and deletion realignment and base quality recalibration (Genome Analysis Tool Kit -GATK-)

Week 6

Topic: Visualization of mapped genomes and Summary Statistics (IGV and others)

Week 7

Topic: Single Nucleotide identification -calling- and insertion and deletion identification (GATK)

Week 8

Topic: Hard versus soft filters after SNP identification (GATK)
Individual and site based statistics

Introduction to Population Genomic Analysis

Week 9

Topic: Summary statistics and Population Structure: Hardy Weinberg, Principal component analysis

Week 10

Topic: Inference of admixture patterns among individuals and local ancestry deconvolution

Week 11

Topic: Introduction to forward and backward simulators (MS, SFS_code and $\delta a \delta i$)

Week 12

Topic: Analysis of historical demographic history of organisms in a maximum likelihood framework

Week 13

Topic: Introduction to genomic methods for the identification of regions under selection (SweepFinder and Composite Likelihood Ratio methods). Interpretation of selection analyses.

Oral Presentations of Results

Weeks 14-15

Required assignments and grading

Grades are based on homework assignments, oral report and written reports as shown below. Working with genomic data in your homework assignments and terms paper exploring a specific genome topic are important to your grade and understanding of the material.

Table 1: Assignments and associated grades

Homework assignments (weeks 2 – 9)	(7 X 20 pts each)
Report on analytical pipeline (week 10)	100
First draft of analyses (week 13)	100
Final report of analyses (end week 15)	130
Oral and written report (weeks 14 - 15)	130
<hr/>	
Grand total	600

Continuous work evaluation: will be based on Homework. The culmination of analysis and assignments each week is necessary for the continuous development of the material. Each section is necessarily dependent on the previous and it will be expected that students would have culminated their assignments every week.

Major assignments: The evaluation will be based on reports of progress aimed to have two main outcomes: A document describing in detail the analysis pipeline and a report in the form of a manuscript where results, analysis and interpretation will be presented as if it were a report for a scientific journal.

Oral Presentations: The oral presentations will be performed in teams. Members of the team will present part of their research and progress and evaluation will be done as specified below.

Detailed directions for homework will be posted on Angel. You may work with fellow students on homework, but the actual writing must be your own. Homework may either be submitted on e-mail and I will be revising the advance of the work as we progress.

Attendance policy. You will be responsible for all materials presented in lecture. You are also responsible for all announcements made in lecture even if they do not appear in the syllabus. We will not take attendance, but expect you to attend every class. We will give demonstrations of use of genetic databases that you will need to use for many of the homeworks and if you miss such a class, do not expect us to give you one-to-one tutorial.

The grade scale will be:

90%+ = A	77-79 = B-	64-66 = D+
87-89 = A-	74-76 = C+	60-63 = D
84-86 = B+	70-73 = C	<60 = F
80-83 = B	67-69 = C-	

Incompletes. University policy (Acad. Reg. #90) states that Incompletes may only be awarded if: "the student is unable to complete their work on time due to circumstances

beyond their control". Granting of Incompletes is rare and must be justified following these guidelines.

Students with Disabilities: Reasonable accommodations are available for students with a documented disability. If you have a disability and need accommodations to fully participate in this class, please either visit or call the Access Center (Washington Building 217; 509-335-3417) to schedule an appointment with an Access Advisor. All accommodations MUST be approved through the Access Center. For more information contact a Disability Specialist on your home campus:

Pullman or WSU Online: 509-335-3417. <http://accesscenter.wsu.edu>,
Access.Center@wsu.edu

Academic Integrity. Academic integrity will be strongly enforced in this course. Any student caught cheating on any assignment will be given an F grade for the course and will be reported to the Office Student Standards and Accountability. Cheating is defined in the Standards for Student Conduct WAC 504-26-010 (3). It is strongly suggested that you read and understand these definitions.

Cheating on an exam or a writing assignment (including plagiarism) will result in a final grade of F for the entire course, will be reported to the Office of Student Affairs, and will result in additional disciplinary action by the University.

Campus Safety and Emergencies. Washington State University is committed to enhancing the safety of the students, faculty, staff, and visitors. It is highly recommended that you review the Campus Safety Plan (<http://safetyplan.wsu.edu/>) and visit the Office of Emergency Management web site (<http://oem.wsu.edu/>) for a comprehensive listing of university policies, procedures, statistics, and information related to campus safety, emergency management, and the health and welfare of the campus community.

Report on Analysis Pipeline Grading Sheet

Topic Coverage (Logical and Complete?)	(40)	_____
Organization	(20)	_____
Quality of Tables and Figures	(10)	_____
Overall Writing (Clarity, Readability, Spelling)	(20)	_____
Overall Synthesis	(10)	_____
Total	(100)	_____

First draft analysis report Grading Sheet

Topic Coverage (Logical and Complete?)	(40)	_____
Organization	(20)	_____
Quality of Tables and Figures	(10)	_____
Overall Writing (Clarity, Readability, Spelling)	(20)	_____
Overall Synthesis	(10)	_____

Total (100) _____

Final Report Grading Sheet

Topic Coverage (Logical and Complete?) (40) _____
Organization (20) _____
Quality of Tables and Figures (10) _____

Overall Writing (Clarity, Readability, Spelling) (20) _____
Overall Synthesis (10) _____
Analysis and success in addressing revisions (30) _____

Total (130) _____

Team Oral Presentations

Each team will deliver a presentation to the class based on their team review paper. You are expected to use PowerPoint (or something similar), and to divide the presentation equally among your team members. Your presentation should be 35 to 40 minutes to ensure enough time is left for questions. More details will be provided in class. The grade scale will be:

Oral Presentation

1. Organization and Clarity (30 Points) _____
2. Effective use of time (30 Points) _____
3. Thorough coverage of material (30 Points) _____
4. Integration among team members (30 Points) _____

Total Points for Presentation (120 Points) _____

Team Self Evaluations

It is important that each team divide the tasks equally among members. Unfortunately, it is the case for some teams that one or two members complete a majority of the tasks due to lack of effort by others. You will be asked to evaluate the contribution of each team member to the final written review paper and the presentation. Your evaluations will be kept private and will not be distributed to other team members. *You will not receive a grade for your paper and presentation until you have turned in the team evaluation form for each assignment.*

Due dates: Paper assessment: Due with final paper
Presentation assessment: Due with presentation

Team Project _____

Please assess your team's performance on the written review and oral presentation. If the whole team effort represents 100%, how would you assess individual contributions of you and your teammates to that 100%?

If you believe that everyone contributed equally, divide 100 by the number of team members and report those equal percentages (e.g., an equal effort by 4 team members represents 25% from each member).

Name

1. _____
2. _____
3. _____
4. _____
5. _____ ME _____

Comments:

Evaluation of Team Presentations

A goal of this course is to critically evaluate scientific results and conclusions. Please take notes during each presentation using the grading criteria above. You will turn in your notes at the end of each class, and each review is worth 10 points.