Statistical Genomics
CROPS 545
3 credit hours- Spring 2017

Professor: Zhiwu Zhang
Office 105 Johnson Hall.
Phone 509-335-2899
Email Zhiwu.Zhang@wsu.edu

Course Objective: Develop concepts and analytical skills for modern breeding by using Genome-Wide Association Study and genomic prediction in framework of mixed linear models and Bayesian approaches.

Lecture room: Johnson Hall 204 with AMS.

Class schedule: W/F 3:10-4:25 PM.

Office hours: M 3:00-4:00 PM (130 Johnson Hall)

Course Description: Concepts and applications of statistical methods in genomics. The course includes three sections: Fundamental, Genome Wide Association Study (GWAS) and Genomic Prediction/Selection (GS). The fundamental section covers the essential knowledge and skills of statistics, computer programming (R) and genomics. GWAS and GS sections cover the mechanisms, methods, and computing tools in GWAS and GS, respectively. We start from genotypes and pick up some of them as genes to simulate phenotypes. Then we examine how well we can map the genes and predict the phenotypes starting with very intuitive methods such as correlation and regression. Then we vary relevant factors to evaluate their strength and pitfall. We also evolve statistical methods and computing tools all the way to their state of art, including mixed model and Bayesian methods. The course is beneficial for experimental design, data analyses to map genes controlling complex traits, and predict their underlying genetic potential among individuals. Analytical skills, critical thinking and hand-on operations are emphasized throughout the teaching.

Text Book: There is no required textbook. Each lecture will be accompanied by a handout that covers all of the in class material and more in-depth material that is beyond this course. For students who would like to have a general reference book, I recommend a free book (academia):
Genome-Wide Association Studies and Genomic Prediction
http://link.springer.com/book/10.1007%2F978-1-62703-447-0

Prerequisites: BIOLOGY 474 or MBIS478 is suggested, but not required. Student backgrounds in statistics, linear algebra, computer programming, and genetics will be helpful towards success in this course.

Assessments: Attendance (10%), participation (10%), midterm exam (20%), final exam (30%) and Homework (30%). Late homework receives penalties of half off per day except unexpected excused university absences.

Grade: A (93%-100%); A- (90%-93%); B+ (87%-90%); B (83%-87%) B- (80%-83%); C+ (77%-80%); C (73%-77%); C- (70%-73%) D+ (66%-70%); D (60%-66%); F(0%-60%). Note: The upper grade will be assigned to a score at a cutting point without rounding. For examples, score of 93.00% receives "A" and score 92.99% receives "A-".

Exams: Midterm (two hours) and final (three hours)

Lab: An assignment will be given, but not graded, for each lab class to enhance the understanding of the theory and help the completion of homework.

Attendance and participation: Attendance in each lecture and lab is expected. Students are encouraged to actively participate through asking of questions and contributing ideas in response to questions from the instructor or students. Among the total points, 10% is assigned to attendance/present and 10% to participation. In accordance with Academic Regulation 73, absences impede a student's academic progress and should be avoided. Those students who must miss a lecture for university-sponsored activities such as field trips, judging teams, sports, conferences, etc. should obtain an official Class Absence Request form from the faculty or staff member supervising the off-campus activities. Scheduling conflicts with employment and non-university activities will be considered unexcused absences.

Student Learning Outcomes: Upon completion of the course, students will be able to:

1) Apply quantitative and scientific reasoning to solve problems in statistical genomics;
2) Understand the development of the statistical methods for gene mapping, molecular breeding and health management;
3) Integrate concepts, principles, methods, and skills in statistics, genetics and computer programming to conduct in a variety of genomic research;
4) Communicate effectively using emerging graphic and graphic media.

All the outcomes will be evaluated by the last four assessments (participant, midterm exam, final exam and Homework).

WSU Work statement: For each hour of lecture equivalent, students should expect to have a minimum of two hours of work outside class.

WSU Safety Statement: Classroom and campus safety are of paramount importance at Washington State University, and are the shared responsibility of the entire campus population. WSU urges students to follow the “Alert, Assess, Act” protocol for all types of emergencies and the “Run, Hide, Fight” response for an active shooter incident. Remain ALERT (through direct observation or emergency notification), ASSESS your specific situation, and ACT in the most appropriate way to assure your own safety (and the safety of others if you are able).

WSU Disability Statement: Reasonable accommodations are available for students with a documented disability. If a student has a disability and may need accommodations to fully participate in this class, the student should 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.

WSU Academic Honesty Statement: As an institution of higher education, Washington State University is committed to principles of truth and academic honesty. All members of the University community share the responsibility for maintaining and supporting these principles. When a student enrolls in Washington State University, the student assumes an obligation to pursue academic endeavors in a manner consistent with the standards of academic integrity adopted by the University. To maintain the academic integrity of the community, the University cannot tolerate acts of academic dishonesty including any forms of cheating, plagiarism, or fabrication. Academic integrity is the cornerstone of the university and will be strongly enforced in this course. Any student caught cheating on any assignment or exam will be given an F grade for the course, will not have the option to withdraw from the course, and will be reported to the Office of 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: http://apps.leg.wa.gov/WAC/default.aspx?cite=504-26-010.

Campus Resources

- Writing Center, https://writingprogram.wsu.edu/writing-center-peer-tutoring/
- Library Services, http://www.wsulibs.wsu.edu/
- CACD, Center for Advising and Career Development, https://ascc.wsu.edu
- Office of Student Conduct, http://conduct.wsu.edu
- Counseling and Testing Services, http://counsel.wsu.edu/
- Academic Integrity, http://academicintegrity.wsu.edu

Statistical Genomics

CROPS 545, Spring 2017

Lecture	Lecture	Section	Title	HW Due
1	1/11/17	Fundamental	Syllabus/course overview and introduction	
2	1/13/17		Computer programming in R	
3	1/18/17		Random variables and distribution	
4	1/20/17		Statistical inference	
5	1/25/17		Linear algebra[1]	
6	1/27/17		Genotyping By Sequencing (GBS)[2]	
7	2/1/17		Missing genotype imputation[3]	HW1
8	2/3/17		Genetic architecture and simulation of phenotype	
9	2/8/17		Linkage disequilibrium	
10	2/10/17	GWAS	GWAS by correlation	
11	2/15/17		Power, type I error and False Discovery Rate	HW2
12	2/17/17		Population structure and principal component analysis	
13	2/22/17		General Linear Model (GLM)	
14	2/24/17		Kinship[4]	Midterm
15	3/1/17		Mixed Linear Model (MLM)[5]	HW3
16	3/3/17		Compressed MLM[6]	
17	3/8/17		Efficient Mixed Model Association (EMMA)[7]	
18	3/10/17		Population Parameter Previously Determined (P3D)[6,8]	
19	3/22/17		SUPER GWAS method[9,10]	HW4
20	3/24/17		Multiple Loci Mixed Model (MLMM)[11]	
21	3/29/17		FarmCPU[12]	
22	3/31/17	Genomic Prediction	Marker Assisted Selection (MAS)	
23	4/5/17		Model fit and cross validation accuracy[13]	
24	4/7/17		genomic Best Linear Unbiased Prediction(gBLUP)[4,14,15]	
25	4/12/17		Ridge regression (rrBLUP)[16]	HW5
26	4/14/17		Kernel and machine learning[16]	
27	4/19/17		Bayesian theory	
28	4/21/17		Bayesian methods[17]	
29	4/26/17		Bayesian implementation[18]	
30	4/28/17		BLUP alphabet	HW6

Reference

1. Lynch, M. & Walsh, B. *Genetics and analysis of quantitative traits*. *Genetics and analysis of quantitative traits.* (1998).
2. Elshire, R. J. *et al.* A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* **6,** e19379 (2011).
3. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* **11,** 499–511 (2010).
4. VanRaden, P. M. Efficient methods to compute genomic predictions. *J Dairy Sci* **91,** 4414–4423 (2008).
5. Yu, J. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38,** 203–208 (2006).
6. Zhang, Z. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* **42,** 355–360 (2010).
7. Kang, H. M. *et al.* Efficient control of population structure in model organism association mapping. *Genetics* **178,** 1709–1723 (2008).
8. Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* **42,** 348–354 (2010).
9. Wang, Q., Tian, F., Pan, Y., Buckler, E. S. & Zhang, Z. A SUPER Powerful Method for Genome Wide Association Study. *PLoS One* **9,** e107684 (2014).
10. Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nature Methods* **8,** 833–835 (2011).
11. Segura, V. *et al.* An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics* **44,** 825–830 (2012).
12. Liu, X., Huang, M., Fan, B., Buckler, E. S. & Zhang, Z. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLoS Genet.* **12,** e1005767 (2016).
13. Zhou, Y., Isabel Vales, M., Wang, A. & Zhang, Z. Systematic bias of correlation coefficient may explain negative accuracy of genomic prediction. *Briefings Bioinforma.* (2016).
14. Zhang, Z., Todhunter, R. J., Buckler, E. S. & Van Vleck, L. D. Technical note: Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. *J. Anim. Sci.* **85,** 881–885 (2007).
15. Bernardo, R. Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop Sci.* **34,** 20–25 (1994).
16. Endelman, J. Ridge regression and other kernels for genomic selection in the R package rrBLUP. *Plant Genome* **4,** 250–255 (2011).
17. Meuwissen, T. H., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157,** 1819–1829 (2001).
18. Pérez, P., de los Campos, G., Crossa, J. & Gianola, D. Genomic-Enabled Prediction Based on Molecular Markers and Pedigree Using the Bayesian Linear Regression Package in R. *Plant Genome J.* **3,** 106 (2010).