Genetics Society of America
PREP
Peer-Reviewed Education Portal

A Problem Based Learning Exercise on Food Security: Understanding the Role of Genomic Variation and Plant Breeding

Jolie Wax[1], Zhu Zhuo[2], Anna Bower[1], Jessica Cooper[1], Susan Gachara[1], Isaac Kamweru[1], Terrance Mhora[1], Sabari Nath Neerukonda[2], Danielle Novick[1], Julia Winkeler[1], Todd Yoder[1], and Randall J. Wisser[1,*]

[1]Department of Plant and Soil Sciences, University of Delaware, Newark, DE 19716
[2]Department of Animal and Food Sciences, University of Delaware, Newark, DE 19716

[*]**Corresponding Author:** Randall J. Wisser, 531 S. College Ave, 152 Townsend Hall, Newark, DE 19716, 302-831-1356, rjw@udel.edu

SYNOPSIS

Genome Science is concerned with the function and diversity of genomes within and across species as well as with applications policies and ethics surrounding genetic and genomic data. This resource aims to introduce key concepts and current technologies and techniques in genome science to undergraduates majoring in biology or sub-disciplines of biology. This is accomplished by engaging student groups in an activity that explores how genomic variation can benefit food security. This three-stage Problem Based Learning (PBL) exercise, dubbed "*Food Security: Understanding the Role of Genomic Variation and Plant Breeding*," guides students from understanding what food security entails, particularly environmental stress on crop production, to discovering how high-throughput sequencing (HTS) data on genomic variation can aid in solving such problems. The PBL resource is intended to be given over three separate class periods. Each stage includes a reading with new terms defined in the margins and a set of guiding questions. Students work in groups of three or four to complete each stage of the problem. Each stage takes about an hour including time for class-wide discussion of select questions. Students are ultimately challenged to think about how HTS data or genomic variation can help solve issues with food crop production. Following this exercise, students should be able to discuss approaches used to preserve and characterize genetic diversity and how this impacts plant breeding and food security. This resource was developed under a new model for graduate student pedagogical training, where students in a graduate course on genome science participate in the development of PBL exercises for undergraduate coursework.

INTRODUCTION

Genomic diversity of crops is vital to restoring losses due to domestication bottlenecks, controlling diseases, and developing improved varieties. Environmental crop stress such as disease and drought impacts communities reliant upon agricultural products Global progress on an integrated system of germplasm reserves is being made in order to conserve the genetic diversity of our major food crops (Khoury *et al*. 2013) as there is a projected doubling in food demand in the next 50 years (Tilman *et al*. 2002) Additionally, future crop species will need the variety, flexibility and genetic variation to adapt to a changing climate. Initial requirements for the development of improved cultivars are collection, conservation and characterization of diverse germplasms from crop wild relatives and new crops, such as *Apios americana* introduced in stage 2 of this PBL excise. In order to improve existing crops and new crops from semi-domesticated or wild species, it is helpful to understand diversity within the genome and capitalize on molecular genomic approaches that can improve the efficiency of using germplasm resources (Khoury *et al*. 2013).

This resource is intended to introduce undergraduate students to HTS platforms along with some broad concepts in quantitative genetics and their application to solving problems related to food security. The activity opens by introducing a problem, i.e. frequent and prolonged flooding of Nekkanti Subba Raos' rice fields, adopted from a story in Express News Service (2015). After completing the first stage students are asked to read a piece of primary literature, *Submergence Tolerant Tice: SUB1's Journey from Landrace to Modern Cultivar* (Bailey-Serres *et al*. 2010). This paper describes the science that ultimately restores Mr. Subba Raos' crop yields. The second stage of the activity then reiterates key points that the students read about in the Bailey-Serres paper, describes the importance of conserving genetic diversity and lists metrics used by geneticists to characterize diversity within and between populations. These metrics are determined using HTS technologies that are discussed in the final stage of the activity where students transition to addressing questions about the application of this information and contemplate the role of genetic diversity in addressing challenges of food security.

The *Sub1* story exemplifies how fine mapping and positional cloning of casual variant for submergence tolerance in rice was successfully integrated into breeding programs that have contributed significantly to global food security. *Apios americana*, a semi-domesticated crop is introduced in stage 2 as a second example of how genomics tools can be applied to assist efforts in plant breeding. *Apios americana* is locally adapted to the central and eastern U.S., there is abundant germplasm, it has the ability to fix

atmospheric nitrogen and its tubers are rich in protein, carbohydrates, dietary fiber and iron. *A. americana* was a food of the Native American peoples and can be found in abundance around ancient Indian campsites. We chose *A. americana* because it is foreseeable that the PBL exercise could be implemented to include collection, DNA extraction and generation of authentic data for students to analyze in stage three. For the time being, we have included fake data (Table 2 and Figure 1 in Stage 3) generated using ETE Toolkit v 3.0 (Huerta-Cepas *et al*., 2016). However, it would be straightforward to substitute *A. americana* for another crop of interest (e.g. one could consider those listed in Khoury *et al*. 2013).

Through inquiry-based exercises such as this PBL, students build partnerships that support open discussion, feedback and community learning. These are important components of the scientific process and learning environment (Tanner 2013). In an interview with Bill Wood (2016) he notes an approximately 30% increase in learning is achieved through active learning strategies. Wood also acknowledges that an important change necessary in biology education is to make students more comfortable with using mathematics and quantitative thinking. This why we have embedded metrics used to quantify genomic variation in the content of the PBL, and students are required to interpret this quantitative information. Working in small groups and creating culturally relevant materials is an effective means of engaging students and supporting inclusiveness in learning (Tanner 2013). This PBL includes elements that we feel will be relevant to diverse students. Addressing real world issues, food and nutrition, cutting-edge technology, and natural variation are complex and we hope this will allow space for the exploration of different ideas. Breeding highly productive crops that are tolerant to disease, flood, drought, and salinity is a major challenge being addressed by plant geneticists and breeders. Our PBL explores how genetic and genomic technologies and techniques can support the efforts to overcome this challenge.

APPROACH

Overview

We suggest students be placed in groups of three or four individuals and work together throughout the exercise. Typical of group learning, it is also recommended to maximize the diversity of individuals per group based on skills, background etc. If students are asked to complete each stage in class, the PBL can be done in three 60-minute periods, but some reading outside of class is also required. Each of the three stages of the PBL is to be completed in each of three successive classes. After the students have finished answering the questions, open discussion is encouraged. There are two opportunities for students to work independently. First, a key piece of literature "*Submergence Tolerant Rice: SUB1's Journey from Landrace to Modern Cultivar*" is cited in the first stage of the PBL and is assigned reading before moving on to stage two (Bailey-Serres *et al*. 2010). Below is a summary of this publication. This can be used to guide discussion prior to embarking on stage two or as a document in and of itself to accompany the primary literature. Second, after completing stage three, a reflection and answer to the central problem – How analysis of genomic diversity can help in addressing challenges associated with food security? – could be assigned as individual work. Examples of student answers to the *Guiding Questions* can be found below the summary of the *SUB1* story.

Summary of Submergence Tolerant Rice: SUB1's Journey from Landrace to Modern Cultivar

Two quantitative trait loci (QTL) are discussed in the Bailey-Serres paper. Both contain genes that respond to the phytohormone ethylene and are called ethylene response factors (*ERF*). These loci control different response strategies in submergence tolerant rice. Deepwater landraces contain the *SNORKEL* QTL and a biochemical pathway that leads to cell division and elongation, allowing the plants to keep enough aerial tissue for photosynthesis. Unfortunately, this takes a toll by decreasing overall plant yields. The *SUBMERGENCE1* (*SUB1*) QTL discovered in lowland landraces was found to contribute to the majority of phenotypic variation of tolerance. All rice in these studies were found to have *SUB1B* and *SUB1C*, but the presence of *SUB1A* was unique to submergence tolerant landraces. The *SUB1A* locus was discovered in a particular landrace called FR13A that has the ability to limit underwater elongation by a submergence dependent dormancy. Two alleles of *SUB1A* were also found that differ by only a single amino acid but only *SUB1A-1* confers submergence tolerance. The *SUB1A-1* allele produces more transcripts in young and submerged rice plants. The effects of these transcripts limit responsiveness to gibberllin which regulates growth. Using lines differing only by

SUB1A-1 it was found that submerged tolerant plants maintain a regular growth rate rather than consuming leaf starch and soluble sugars to enhance elongation and escape. These examples illustrate the balance between utilization and conservation, escape versus quiescence. When flood waters rise slowly elongation growth is feasible, however, when flooding is deep and prolonged the protection of energy reserves and growth meristems is advantageous.

Identification of the *SUB1A-1* allele was accomplished by first fine-mapping the *SUB1* QTL using a large F_2 population, followed by positional cloning and transgenic validation. *SUB1A* is thought to result from a duplication of *SUB1B* following domestication of *indica* rice. Using single nucleotide polymorphisms discovered within *SUB1A* and *SUB1C* breeders were able to precisely integrate the FR13A-derived *SUB1A-1* allele into high yielding, farmer preferred, landraces. This was done by continued backcrossing to the recurrent/preferred parent and selecting genotypes with the *SUB1A-1* allele in a recurrent parent background. Because these varieties retain the characteristics farmers like and have submergence-dependent tolerance they have become widely adopted in flood prone areas of South and Southeast Asia. With the varieties created using this marker-assisted selection method, farmers of low-lying land are now protected from flash floods.

Submergence imposes complex abiotic stressors such as temperature, turbidity and light penetration, with reduction in photosynthesis and respiration being the major consequences. Even so, populations that differed only by the presence or absence of *SUB1A-1* showed a nearly 10-fold difference in viability following 16 days of submergence.

Quantitative Genetics and Quantitative Traits

As a matter of background for discussion on quantitative traits and quantitative genetics, if students are already familiar with Mendelian genetics, it may be useful to describe how quantitative genetics may be considered as an extension of Mendelian genetics. Instead of one gene underlying the variation many genes do, yet they are still transmitted according to Mendelian laws of inheritance. Phenotypes vary in different ways, and the way in which phenotypes vary is often predictive of complexity in the genetic basis for that variation. At one extreme are dichotomous Mendelian traits with two distinct phenotypes (like the classical wrinkled and round peas) which is typically caused by a single gene. At the other end are traits, such as height, which vary continuously, giving rise to the notion of a quantitative trait for which the distribution of phenotypic variation is often portrayed (and observed) as a normal distribution or bell-shaped curve. Quantitative traits are typically caused by multiple genes where each

contributes a small amount of the total variation; the recombination and Mendelian segregation among multiple alleles at multiple loci is what gives rise to a continuum of phenotypic variation. Many traits that are of interest to breeders are quantitative traits regulated by more than one locus, influenced by environmental factors, and show several intermediate phenotypes. The *SUB1* story is a great example for introducing students to concepts in quantitative genetics, but it is not exactly the typical example of a quantitative trait, since such a large proportion of the variation is attributable to a single gene.

Learning Objectives and Examples of Answers to Guiding Questions

This following portion of this document contains the learning objectives and examples of student answers to the guiding questions that conclude each of the three sections of the PBL Student Handout. The learning objectives can be assessed by student responses to the guiding questions. We recommend that these be answered in groups of three or four students and that specific questions are opened for classroom discussion that lead in to the next stage of the PBL.

Stage 1. Food Security: Learning Objectives

1. Discuss the ways in which the environment affects food security in terms of crop production.
2. Describe strategies or mechanisms plants use to adapt to an environment.
3. Find ways that phenotypic diversity within a species could help overcome environmental effects on food security.

Possible Answers to Guiding Questions

1. Based on the information provided, what could be done to help Nekkanti Subba Rao and other farmers in this region to mitigate crop loss due to flood events?

Based on the available information, the most obvious solution would be to switch to growing the landrace accessions that are tolerant to inundation. However, it was indicated that these landraces lack other important agronomic characteristics, so we need a way to combine the flood tolerance with these other important agronomic characteristics into a single variety.

2. List and discuss some of the factors that affect food security.

There are many facets that affect food security, including human population growth, agricultural productivity and sustainability, climate change and extreme weather events, economics, infrastructure, politics, and more. (more specific answers might include things mentioned in the PBL, such as drought or disease). Students can discuss amongst themselves about how the factors they list affect food security.

3. What is phenotypic variation, and how can it be useful for breeding and for food security?

Phenotypic variation is an observable difference for a set of phenes between individuals of a population. In general, selection for a phenotypic state that is desirable could help address an environmental stressor like flood tolerance, which in turn could help provide greater food security. More specifically, if favorable phenotypes are conceived and individuals that show those phenotypic states are identified, a breeder can select and mate only those individuals — discarding the individuals that have unfavorable phenotypic states. This tends to increase the frequency of the desired phenotypic state in the next generation, and over time the desired phenotypic state, which may initially be rare, can become the norm. If these phenotypes provide an environmental advantage like flood tolerance or disease resistance then they will contribute to food security.

4. What are the factors that influence phenotypic variation, and which do you think is of greatest value to plant breeders? Why?

The three factors that affect phenotypic variation are 1) genotype, 2) environment, and 3) genotype-by-environment interaction. The genotype can be controlled by the breeder, i.e. a breeder can select favorable genotypes, while the environment (at least the climate and weather, as opposed to agronomic practices) cannot. Therefore, the genotype is of most value. However, this depends on whether the phenotype is controlled genetically, and not only influenced by the environment. Genotype-by-environment interaction complicates matters but is of similar value to a breeder as is genotype per se, since breeders may select specific genotypes that perform best in specific environments.

Stage 2. Genetic Diversity: Learning Objectives

1. Talk about where crop scientists might find diverse germplasm.
2. Describe the relationship between phenotypic and genotypic variation.
3. List metrics used to quantify genotypic variation within individuals and populations.

Answers to Guiding Questions

1. What can plant scientists do to avoid or minimize the loss of natural diversity in plant species?

Beyond in situ conservation, plant scientists can create ex situ gene banks and living collections to conserve germplasm. It is already feasible to genetically transform species with specific genes, and it may one day be possible to re-create an individual from its nuclear DNA. Therefore, preserving tissue and DNA from plant species may also provide a means of preserving diversity.

2. How can scientists quantify genetic diversity among individuals?

Scientist can quantify genetic diversity by recording the phenotypes of individuals. Because phenotypes are influenced by the environment, this measure of variation is context dependent and replicated experimental designs are needed to collect robust information on phenotypic variation. Due to advances in sequencing technology, it has become increasingly common to survey genome-wide variation in the DNA sequence, and use measures such as nucleotide diversity, inbreeding coefficient and F-statistics. It is also possible to measure variation in other cellular processes, such as transcription and translation. However, like phenotypes, these too can exhibit context dependency.

3. How might estimates of inbreeding for *A*. *americana* individuals be useful to breeders?

If the genotypes of new individuals are determined, and inbreeding is estimated, this could be used to predict their phenotypic outcome. For instance, individuals with low levels of inbreeding would be expected to produce a limited number of relatively large tubers.

4. Without phenotypic data, how can scientists use measures of F_{ST} for managing germplasm collections?

F_{ST} can be used for guiding the establishment, curation and reduction or expansion of germplasm collections. Because F_{ST} provides a measure of differentiation between groups of individuals, where higher levels of F_{ST} are indicative of more differentiated groups, F_{ST} can be used to characterize natural populations located in different ecological zones. This information could help guide how to sample germplasm from the wild. Moreover, comparing F_{ST} of an existing collection to new candidate germplasm from the wild would inform which populations might be best to sample from. In addition, multiple collections may

exist for a species at different places in the world. F_{ST} can be used to assess how different the collections are from one another. In general, F_{ST} can facilitate strategic thinking in how best to manage germplasm collections.

Stage 3. High Throughput Sequencing: Learning Objectives

1. Compare/contrast different types of HTS platforms.
2. Deduce which HTS platforms are best suited for different metrics of genetic diversity.
3. Evaluate how HTS data can be leveraged to solve climate related crop issues.

Stage 3: Answers to Guiding Questions

1. Belamkar *et al*. (2016) estimated the genome size of *A*. *americana* to be approximately 1.65 Gbp. Your team needed to assay SNPs across the genome to examine genetic diversity in your new collection of *A. americana*. Which HTS platform would you have used for this project and why?

In order to identify and survey SNPs across a genome size of 1.6 Gb in a large population, the most cost-effective sequencing platform that is able to produce large amounts of data with relatively low error rate should be chosen. Illumina HiSeq 2500 is idea for this application.

2. Summarize the findings from your team's project on HTS-based characterization of genetic diversity in *A*. *americana*.

A. americana accessions sampled in nearby locations are clustered and less differentiated than samples from more distant locations For instance, NJ and DE, NC and SC, and GA and FL each have relatively low pairwise F_{ST}. This suggests accessions in those regions share genomic variation and are likely to contain alleles that may be useful at similar frequencies, such that sampling across those locations could optimize the discovery of favorable variants.

3. Given the context of the situation and summary of your findings, what would be the top three recommendations your team would make to the breeding community?

1) It would be helpful to also genotype Blackmon and Reynolds' breeding material and investigate how it is related to the germplasm characterized in the project. The geographical trend in the results suggests they would most closely related to germplasm from FL and GA.

2) Because the results show an overall latitudinal trend in differentiation with some pairs of locations being more similar, samples from extremes of the latitudinal range would be expected to differ the most in terms of the potentially unique DNA sequence variants. However, the gradual trend in differentiation suggests that novel or favorable alleles may be spread across all locations.

3) It is important to recall that genotype information alone is not fully predictive of phenotypes. However, the genotype information is a useful guide to sampling diversity. Since A. americana is being considered as an alternative crop to potato, which is grown outside the eastern U.S., it would be useful to perform phenotypic screens with samples from locations that show at least moderate differentiation (e.g. >0.1). For instance, initial screens may be performed for germplasm from FL, NC, DE, and ME. This could be followed by deeper sampling of germplasm from the locations that show promising phenotypes (as well as nearby locations). That is, the information can be used recursively to advance breeding objectives.

4. How can HTS analysis of genomic diversity help in addressing agricultural challenges associated with food security?

The characterization of genomic diversity facilitates conservation and breeding efforts, which enables effective utilization of germplasm for the improvement of crop adaptation, stress tolerance and productivity. HTS provides a rapid, context independent means of measuring genetic diversity in very large samples. This information helps geneticists and breeder wield DNA sequence variation and germplasm resources when tackling the challenges of food security related to varietal development of crop species (existing or new). For instance, HTS facilitates the introgression of unique sequence variants in the genome (e.g. controlling submergence tolerance) from one population into another that is more agronomically relevant, and HTS can enable strategic use of genetic diversity in the development of new crops. HTS-facilitated diversification of existing crops and development of new crop species are expected to mitigate instability in cropping systems due to climate change, generate new economic opportunities and health benefits that are each critical to a stable and safe food supply.

JUSTIFICATION

This problem based learning (PBL) exercise provides students with a basis for understanding the importance of genomic variation and plant breeding for agriculture, with an emphasis on modern high-throughput sequencing technologies. This serves as an introductory educational resource for undergraduate students to help prepare them for advanced study and research using genomics tools. The concepts and guiding questions of each stage gradually introduce new topics, requiring comprehension and reflection of previous stages, to facilitate learning. Through this excise, students will improve their learning skills, such as searching reading and comprehending literature, thinking critically, interpreting scientific data and figures, etc. Moreover, this PBL excise engages students to work both individually and with peers. The experience helps students to become adapted to a collaborative environment and also emphasizes independent learning.

REFERENCES

Bailey-Serres, J., T. Fukao, P. Ronald, A. Ismail, S. Heuer, and D. Mackill (2010) Submergence tolerant rice: *SUB1*'s journey from landrace to modern cultivar. *Rice* 3:138-147.

Belamkar, V., A. Wenger, S. R. Kalberer, V. G. Bhattacharya, W. J. Blackmon, and S. B. Cannon (2015) Evaluation of phenotypic variation in a collection of *Apios americana*: An edible tuberous legume. *Crop Science* 55:712-726.

Belamkar, V., A. D. Farmer, N. T. Weeks, S. R. Kalberer, W. J. Blackmon, and S. B. Cannon (2016) Genomics-assisted characterization of a breeding collection of *Apios americana*, an edible tuberous legume. *Scientific Reports* 6:34908.

DeHaan, L. R., D. L. Van Tassel, J. A. Anderson, S. R. Asselin, R. Barnes, G. J. Baute, D. J. Cattani, S. W. Culman, K. M. Dorn, B S. Hulke, M. Kantar, S. Larson, M. D. Marks, A. J. Miller, J. Poland, D. A. Ravetta, E. Rude, M. R. Ryan, D. Wyse, and X. Zhang (2016) A pipeline strategy for grain crop domestication. *Crop Science* 56(3):917-930.

Express News Service. "Rain batters Nellore, relief operations hit." *The New Indian Express*, 19 Nov. 2015. Web. 01 Aug. 2017.

Fedoroff, N. V., D. S. Battisti, R. N. Beachy, P. J. M. Cooper, D. A. Fischhoff, C. N. Hodges, V. C. Knauf, D. Lobell, B. J. Mazur, D Molden, M. P. Reynolds, P. C. Ronald, M. W. Rosegrant, P. A. Sanchez, A. Vonshak, and J.-K. Zhu (2010) Radically rethinking agriculture for the 21st century. *Science* 327(5967):833-834.

Huerta-Cepas, J., F. Serra, & P. Bork (2016) ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Molecular Biology and Evolution* 33(6):1635-1638.

Khoury, C. K., S. Greene, J. Wiersema, N. Maxted, A. Jarvis, & P. C. Struik (2013) An inventory of crop wild relatives of the United States. *Crop Science* 53(4):1496-1508.

Tanner, K. D. (2013) Structure matters: twenty-one teaching strategies to promote student engagement and cultivate classroom equity. *CBE-Life Sciences Education* 12(3):322-331.

Tilman, D., K. G. Cassman, P. A. Matson, R. Naylor, and S. Polasky (2002) Agricultural sustainability and intensive production practices. *Nature* 418(6898):671-677.

Wood, W. B. (2016) Learning and doing: An interview with Bill Wood. *Genetics 203*(2):631-633.