Statement of Research

Nicholas Santantonio

In the face of climate change, farms will have less access to agronomic inputs while they strive to maintain and update sustainable agricultural practices. This leaves genetics as the primary target for improvement of our food systems. Plant response to stress is largely quantitative in nature, driven by the complex interaction of genes and their environment throughout growth and development. Integration of quantitative genetics, genomics, and digital agriculture will be essential for development and deployment of the 21st century breeding technologies that will expedite genetic improvement to unique environments.

Increasing selection intensity and reducing cycle times

The most easily exploited terms in the breeder's equation are cycle time and selection intensity. With affordable marker platforms that can allow information sharing across relatives, resources can be reallocated toward increasing the number of lines that are genotyped and evaluated while reducing replication. Increasing the flow of genotype-phenotype information from a breeding program will allow for genomics-informed decision making to advance lines faster and recycle materials earlier. Through simulation and collaboration with the new wheat breeder, I intend to investigate strategies to effectively deploy these new technologies.

Aerial phenotypes can be used in multi-trait models to substantially increase trial sizes without increasing the number of harvested plots, thus increasing selection intensity. Optimal contributions can leverage genome-wide information to recycle materials at earlier stages while minimizing the loss of genetic diversity, while maximizing genetic improvement and product development. Using a genetic loss function to burn-in on genetic predictors, I intend to train convolutional neural networks (CNNs) on plot images to find genotype specific growth patterns associated with yield performance. I also aim to investigate the potential to harness three dimensional deep learning models to extract genetic information on growth, modeling time as the third dimension.

The infrastructure to support regular genotyping and proximal sensing must be developed before a breeding program is inundated with incoming data. I intend to help establish databases, implement quality control and develop standard operating procedures within SCS. Genotypic information will be used to link otherwise unrelated trials and traits, building the data resources necessary for model training. As informatics infrastructure is developed, the impact of pushing generation times toward biological and logistical limits through rapid-cycling can be evaluated.

Integration of genomics and digital agriculture to shed light on G×E

Valued in 2018 at \$539 million (USDA, 2019), alfalfa is an important perennial legume forage for Colorado farmers. I intend to build a lab with a focus on the quantitative genetics of plant growth and development using alfalfa as a model, with the potential to release locally adapted alfalfa varieties. Forages are an almost ideal model organism for integrating quantitative genomics and proximal sensing to understand plant growth and response to differential stress. Multiple cuts through several years allow for repeated measurements of growth through time, while the harvestable product can be imaged directly during each regrowth cycle.

I recently obtained funding through a US Alfalfa Farmer Research Initiative (USAFRI) grant to develop a population-level genomic prediction framework to help model growth through time. Using bulked DNA from many individuals, the genomic prediction framework operates on allele frequencies in a population as opposed to allele counts in an individual. Genomic relationships

between populations can then be used in longitudinal, random regression models to build genotype specific growth curves. Moving forward, I want to evaluate how plasticity in growth and response to stress allows for stability or sensitivity to varying environments. This framework can also be used to track allele frequency changes through time, allowing for identification of loci under natural selection for persistence to differential stress.

New populations will be formed using mathematical optimization to determine optimal contributions of parent varieties. Populations will then be planted in optimal proportions and randomly inter-mated to maximize beneficial population allele frequencies in the resulting seed lot, while minimizing inbreeding. Once shown to be effective, this strategy could revolutionize the way alfalfa seed is produced, with new varieties being defined and created by their *parents*. Farmer seed would be produced by planting parental populations in optimal proportions, making new varieties rapidly available. The reduced cost of seed production could make alfalfa an affordable one or two year rotation crop for field rejuvenation.

Cropping systems integration

Crops are a single component of a larger agro-ecological system, which includes soil microorganisms, endophytes, the animals that feed on those crops and the microbiome of those animals. Other than host-pathogen interactions, little attention has been paid to genomic interactions between these organisms (i.e. G×G), despite an overall notion that they are important. These interactions can be thought of as a special case of the G×E problem, where the genetic covariance of the "environment" (e.g. soil microorganisms) can be determined by genotyping that "environment".

Legumes can form symbiotic relationships with nitrogen fixing bacteria, *Rhizobium*, when available soil nitrogen is low. Genetic increases in signaling, infection and nodulation could allow for the use of less chemical fertilizer, reducing the environmental impact of nitrogen runoff. I aspire to build G×G prediction models for simultaneous selection of host variety and symbiont to increase the rate of colonization and nitrogen fixation, especially under moderate soil nitrogen levels.

These synergistic genomic selection models could also be adapted to other goals, allowing for much wider collaborations between animal breeders, microbiologists and soil scientists. Examples include root/soil microbiome using wheat as a model, forage/animal and forage/rumen-microbiome using alfalfa as a model, and even barley/yeast interactions using malting barley as a model. Products specifically improved for one another could be marketed as a package (e.g. maltsters could offer specific yeasts with certain barley malts), or a service (e.g. genotyping a farmers field to determine optimal varieties for their soil microbiome).

Research philosophy

In the era of big data, a shift away from small designed experiments to large observational studies at the breeding program scale is inevitable. When genotyped and made publicly available with FAIR data principals, the vast amount of data generated in a breeding program becomes a treasure trove for asking questions and informing breeding decisions. Clever experimental design will offset genotyping costs by trading replication at an individual level for replication at the genetic level.

I believe in a collaborative model, where breeding programs do not operate in isolation. They share germplasm, resources, expertise, and most importantly, ideas. I intend to contribute to the collaborative effort at CSU and across the globe to build the foundational capabilities needed to deploy the latest technology for variety development. As climate change progresses, heat, drought, intense storms and hard frosts will be the new norm, and we must work together to help defend our food security through accelerated genetic improvement.

Statement of Teaching

Nicholas Santantonio

Plant breeders have traditionally been generalists, combining genetics with a range of plant sciences to identify farmers' needs and turn out products to meet those needs. However, the range of skills required in the field is rapidly increasing. Proficiency in statistics, programming, bioinformatics and machine learning are now expected in addition to the traditional skills in physiology, pathology, agronomy and genetics. As educators, we must update our curriculum to best prepare students for careers in the new era of genomics and digital agriculture.

The widening breadth of the plant breeding discipline may demand multiple paths of instruction. This should include a more structured series of quantitative courses to benefit students seeking specialization. Students who chose this path will finish graduate school with the comprehension and ability to effectively adapt the latest computational techniques to meet future breeding goals. I plan to work closely with the new wheat breeding professor to develop quantitative courses that complement the plant breeding course material.

Teaching experience

During graduate school, I led several activity-based learning sections as a Teaching Assistant (TA) for an introductory biology course at Cornell, BIOMG1350. As a TA for an introductory plant breeding course, PLBRG2010, I built a pdf based online homework medium, managed the course website, coordinated labs, aided in exam development, and constructed and delivered several lectures. In my third year of graduate school, I was asked by several members of the second year cohort to lead a weekly discussion of a quantitative genetics textbook (Bernardo, 2010). As a postdoc, I continued to hone my teaching experience and philosophy by co-instructing an advanced graduate-level course on the evolution of genetic modeling in plant breeding, PLBRG7420.

Teaching philosophy

Inclusive teaching tactics are key to ensuring equal access to knowledge in the classroom. Expectations must be made clear, and reinforced throughout the semester so that students do not get behind and fail to meet milestones. It is important for students to be exposed to plant breeding and quantitative genetics ideas from multiple perspectives, and they must be given the opportunity to demonstrate critical thinking on different levels. While some students may show analytical thinking and synthesis on exams, others may shine in more hands-on projects. Most mathematical and computational learning occurs through doing, not watching. The lecture is important to present material in a concise, structured manner, but concepts are cemented when the student reconstructs the ideas on their own time.

Longer term projects provide students the opportunity to practice and apply concepts in a more autonomous environment, and are invaluable for assessing comprehension and critical thinking. Regularly assigned homework and hands-on labs are crucial for evaluating the pace and overall understanding of the material presented. Greenhouse and field facilities on campus provide opportunities to get students out of the classroom to see plant breeding in practice. Public databases provide resources where students can get experience working with real datasets. Computer simulations are useful tools to evaluate comprehension, where in order to simulate a system correctly, the student must understand that system well.

All courses I instruct would contain a term project of relevant complexity to augment exams, in-class labs and homework assignments. For quantitatively oriented courses, these projects would

be computational in nature, where students would use real or simulated data to explore the ideas covered in the course. They would then be asked to present their results in written and oral formats that mirror typical scientific communication. Projects may be team oriented to promote collaborative skills and project management.

Courses

I intend to develop a yearly, advanced undergraduate course in quantitative genetics to augment the undergraduate plant breeding course. This course would target 21st century plant breeding concepts with a focus on the use of genome-wide information to drive decision making. In addition, I intend to develop an advanced graduate course that would be held every other year, focusing on the methodology of complex quantitative ideas. Both courses would include a hands-on computational component to reflect the skills currently desired in the field.

The undergraduate course would start with basic probability theory and the single locus model, advancing through genome-wide association, genomic prediction and selection theory. Computational labs would be used to augment student understanding of course material. Students would use available computational tools to analyze small example datasets with a focus on interpretation of results. For the term-project, students will be split into groups, and given a breeding scenario and a dataset. They will determine the genetic architecture of their trait and propose a breeding strategy based on their scenario and what they can learn from the data. By the end of the semester, students will be able to demonstrate critical thinking of plant breeding methods and ideas, with the ability to synthesize when given new plant systems or breeding goals.

The graduate quantitative genetics course would shift focus onto the methodology of more complex ideas, including coalescent theory, hierarchical Bayesian models, spatial variation, G×E, multivariate selection indices, longitudinal models and optimal contribution. Students will be expected to write their own software to solve computational plant breeding problems presented in weekly computational labs. Assignments would be required to be submitted as typed documents in Markdown, LaTeX or similar format, to expose students to more effective modes of mathematical communication outside of Microsoft Office. To facilitate teamwork, the term project would be assigned to groups of two to four students, who would first be tasked with finding a genotype-phenotype dataset. They would then work to develop a genotype to phenotype map, assess genomic predictability and construct an optimized breeding scheme throughout the semester.

Curriculum

Currently, most life science students do not acquire in-depth statistics and programming skills until graduate school, impeding their progress while they learn to grapple with these new languages. Moving forward, I would like to work with faculty in SCS, statistics and computer science to build a quantitative/computational genetics undergraduate curriculum. Genetics is a vast field; it is imperative that students get exposure to and training in the rapidly changing environment in which they will soon be seeking jobs. Students in essentially all sub-fields of genetics will need to be able to deal with large datasets, using tailored algorithms to make inferences, predict the unobserved, and guide decision making. Big data management requires skills in programming, linear algebra, statistics and machine learning that must be incorporated into the curriculum at earlier stages.

Genomics and digital agriculture are only just starting to change the landscape of food production. Quantitative skills are one of the specializations imperative in plant breeding, and CSU must be at the forefront of best preparing the individuals who will usher in this new era.

Statement of Diversity, Equity and Inclusion

Nicholas Santantonio

> ***Note to CSU:*** *I have spent much time considering these issues. You can find a detailed version of my thoughts and the initiatives I intend to pursue here:* github.com/nsantantonio/ColoradoStateQG/blob/master/SantantonioDEIContributionLong.pdf

Everyone deserves an equal chance to show their potential. Often, it takes the right time, place, person or simply an opportunity to inspire someone. As a high school dropout, it took several tries for me. I struggled through the first two years of my undergrad at New Mexico State (NMSU), failing to meet a 2.0 GPA by the end of my fourth semester. With the help of several professors at NMSU and continued support from my family, I found a passion for genetics during my third year and everything changed.

How can I provide adequate support and opportunity to help create that spark in those far less privileged than I? As an educator, I can work to create opportunities for students from underrepresented groups to pursue further education and expose students to different cultural perspectives in and outside of the classroom. I can take active measures to include students as part of a healthy lab culture, where expectations are clear and everyone is given equal opportunity to succeed in their own pursuits. Importantly, I can continue my own education into diversity, equity and inclusion (DEI) so that I can adapt my own efforts to better serve the community.

I have worked toward learning about, and engaging in DEI concepts and initiatives to further my understanding and to help me become a better ally. To actively engage in increasing the diversity of graduate students in the plant sciences, I joined the Diversity Preview Weekend at Cornell as a co-leader, currently serving as fundraising chair. I have attended several workshops on the hidden curriculum to learn how I can identify and help shed light on the unspoken expectations in academia that may be foreign to those unfamiliar with the academic environment in the US.

It is important that we recognize the need for education in DEI concepts and ideas. To enact societal change, individuals need to be willing to be taught and engage in the conversation. During this learning process many will make mistakes, including me. Mistakes are okay! As long as we learn from them and strive to better ourselves and our understanding of and compassion for others. It is paramount that we cultivate an environment where people can learn, practice and ask questions about DEI concepts without fear of retribution if they misstep.

As faculty, we must educate ourselves on how to provide an inclusive environment and work to make resources equally available and expectations clear. We must reevaluate our own practices and implicit biases to ensure we have not unintentionally created an inequitable or exclusive situation. The curriculum needs to be built to reflect of the breadth of diversity, such that students recognize a bit of themselves, their culture or contributions that were made by people they can relate to. To address these needs, I want to work to develop a series of DEI related requirements for faculty, staff and graduate students, and engage with the community both on and off campus. Details of initiatives I aim to pursue can be found **here**.

The most effective teams with the best ideas come from diverse backgrounds and experiences. Women, minorities, LGBTQIA and members of other underrepresented groups continue to face less opportunity and numerous obstacles, especially in STEM. Unfortunately, this has lead to a lack of perceived academic "merit" by some, when these differences are purely environmental in nature and driven by legacy and current social biases. I am committed to helping reverse these trends by building a diverse team and an environment in which everyone is valued and given the resources and opportunities they need to succeed.