

# 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 21<sup>st</sup> 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 $G \times E$

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.  $G \times G$ ), despite an overall notion that they are important. These interactions can be thought of as a special case of the  $G \times E$  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  $G \times G$  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.