

# 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. Interest in sustainable agriculture has brought renewed attention to forage and industrial crops, which can increase soil nitrogen, sequester carbon and reduce weed populations. These crops can also be adapted to marginal lands, leaving more flexibility for farmers and less competition with food crops. Implementation of 21<sup>st</sup> century breeding technologies will expedite improvement, but several challenges must be overcome due to the complexity of the genetics and the difficulty of phenotyping.

## Genomic selection for alfalfa

Alfalfa, known as the “Queen of forages” for its high protein content, was valued in 2018 at \$127 million in NY ([USDA, 2019](#)), and is key to the states \$2.5 billion dairy industry. Beneficial allele enrichment in alfalfa is arduous due to high inbreeding depression and the autotetraploid nature of the crop. As an obligate out-crosser, alfalfa must be bred on a population level, where varieties are released as synthetics to avoid inbreeding and exploit population-level heterosis. This has limited implementation of marker-based selection strategies because most methods are individual-based.

In pursuit of this research position, I obtained funding through a US Alfalfa Farmer Research Initiative (USAFRI) grant to develop a new population-level genomic selection framework (popGS). Tissue from many individuals within a variety or breeding population are bulked for DNA extraction and sequence-based genotyping. This allows the genomic prediction framework to operate on allele frequencies in the population as opposed to allele counts in an individual. PopGS will allow for prediction of additive effects for genetic gain, as well as dominance effects to exploit population level heterosis. This genotyping method will also be used to track allele frequency changes through time. Loci under natural selection for stand persistence will be identified and appropriately weighted in the popGS model. PopGS will require the development of affordable, medium density, sequence-based genotyping platform, presenting a potential collaboration with Breeding Insight.

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.

## Developing a forage phenome

Forages are typically harvested multiple times per year and are often perennial, requiring several years of evaluation. Forage quality must be measured to gauge animal nutrition and digestibility, but is expensive to phenotype. Leguminous cover crops are also highly effective for soil rejuvenation, but the genetic variability of these processes is largely unknown because looking underground is notoriously difficult. This makes for a heavy phenotypic burden on the breeding program that stands to benefit from high-throughput phenotyping platforms, with opportunities for collaboration with the Cornell Initiative for Digital Agriculture.

To mitigate the phenotypic burden of multiple years, cuts and the measuring of quality and root traits, multi-spectral imaging data will be collected regularly throughout each growth cycle to produce (HTP). Genetic correlations of low-throughput phenotypes (LTPs) and HTPs will be estimated and used to build genomic prediction models to predict unobserved LTPs. These HTPs will also be used to develop population specific growth curves to model development above and below ground under differing biotic and abiotic stressors. To pursue this goal, I initiated a pilot study as part of the USAFRI grant and in collaboration with the forage breeding program to evaluate the utility of multi-spectral imaging for predicting growth and biomass production.

Moving forward I intend to expand the size and number of field trials, using HTPs to limit the phenotypic burden. Instead of waiting for multi-year trial completion, rapid-cycle selection will be performed on a yearly basis using popGS and all available HTPs and LTPs from historical and in-field material. I also intend to help establish a public forage database using BreeDBase to facilitate collaboration and data sharing between forage programs.

## Synergistic systems breeding

Forage crops are only one part of an agronomic, biological system which includes soil microorganisms, the animals that feed on the forages 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*. Unfortunately, the signaling and infection process for nodulation is typically reduced or absent under moderate to high soil nitrogen levels. Genetic increases in 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, even when some nitrogen is available in the soil.

I also seek to establish a collaboration with the animal nutritionists, dairy farmers, and animal genetics companies to investigate the potential for synergistic forage and animal breeding. Instead of breeding animals independently of their feed, we can start to breed specialized animals to specialized feeds. Bringing animal genetics, nutrition, and farming communities together with forage breeders would set precedent for future long-term integrated breeding operations.

## 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, stored, 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. Genotyping at this scale is feasible given the drastic reduction in cost, and can be further offset by clever experimental design that trades 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 Cornell 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.