

Statement of Research

Nicholas Santantonio

The chile pepper is perhaps New Mexico's most important cultural icon, prompting the foundation of the official state question, "Red or Green?". Important in this identity is the cultivar type, where the peppers are grown, how hot they are, and what kind of heat they have. Increasingly, chile cultivation and consumption has expanded, with the aromas of fresh roasted green chile reaching grocery store parking lots at least as far away as Ithaca, NY.

I intend to develop the foundational capacity to implement 21st century breeding technology in the Chile Pepper Breeding Program at NMSU to help support an industry so steeped in New Mexican culture. This will include the collection, storage and accessibility of genome-wide information, pedigrees, high throughput phenotypes collected through proximal sensing, and traditional phenotypes from field trials. Ground-breaking breeding technologies will be implemented to demonstrate the potential of these new resources and pique public interest in crop improvement at NMSU.

Short-term gains: Trait stability

Variability in heat, within a single cultivar, year, field, or even a single plant is a well known, but puzzling phenomenon. Given that the genetics do not vary within an inbred variety, only the environment can change this trait. However, this does not hold true for all peppers. When it comes to heat, jalapeños and green chiles can excite, or disappoint, while other peppers, such as the habañero, seem more consistently spicy. This apparent $G \times E$ suggests that there may be genetic factors controlling the variability in heat. We may be able to predict and select for environmental stability in these dispersion traits. The production of varieties with stable characteristics is of high priority for farmers and consumers, who desire predictability in flavor, flesh and especially heat.

Hierarchical generalized linear models can leverage genome-wide information to allow for modeling genetic effects in the variability, or dispersion, of a trait. Using a diverse panel of jalepeño and New Mexico chile cultivars, with habañero varieties as a control, genetic heterogeneity for dispersion parameters of capsaicin production will be assessed in irrigated and managed drought stressed field trials. Genomic predictability of trait stability will be determined using genome-wide prediction models, and a scan for important loci effecting variability (i.e. vQTL) will be performed. If found to be heritable, genomic selection can be deployed to improve and release a jalepeño and a New Mexico chile with improved heat stability. Cayenne types may also be used to investigate and select for stability of carotenoid production in a similar fashion.

Mid-term improvement: Proximal sensing and machine learning

Due to changes in labor availability, new varieties with adapted plant architectures that enable mechanical harvest are needed to maintain a viable industry. Proximal sensing has the potential to help solve this problem by monitoring growth and fruit development. Selections would then be made on branching and fruit set patterning that facilitate mechanical harvesting. Modification of canopy architecture can also change quantitative resistance to pathogens such as *Phytophthora capsici* through avoidance, narrow flowering time for uniform maturity, and increase resilience to heat and drought stress.

In order to understand the impact of plant growth and development on quantitative traits such as canopy architecture, we must be able to monitor plant development to fit genotype specific growth curves through time. Proximal sensing with relatively inexpensive unmanned aerial vehicles and ground robots would regularly collect high-throughput phenotypes related to growth, and would be used to construct three dimensional models of canopy architecture.

Machine learning algorithms, such as convolutional neural networks, will be trained to extract useful predictors of phenotypic traits from proximal sensing data, which will in turn be used in quantitative genetics models to predict canopy development. Proximal sensing can augment other phenotypic and genotypic information to predict genetic values when measurement is too costly or otherwise infeasible. Collection and public availability of dense phenotypic and genotypic information, when combined with weather data, will allow experimental trials in multiple breeding programs to be linked. Researchers can then exploit shared information to build predictive genetic models for complex, multidimensional traits such as canopy and root architecture, flavor profiles, and biotic and abiotic stress tolerance.

Long-term performance: Implementing a 21st century breeding program

To accelerate improvement of complex traits, the breeding program must take advantage of the latest breeding technologies. While there has been much discussion on genomics-assisted breeding efforts, very little literature focuses on the actual implementation. Because pepper is an annual self-pollinated diploid, it makes a relatively simple system for evaluation of new breeding technologies. As a proof of concept, I intend to use pepper as a model organism to ask important questions about how a traditional breeding program can be adapted to become a 21st century breeding program.

The infrastructure to support regular aerial and ground proximal sensing must be developed before the program is overwhelmed by incoming data. A public *Capsicum* database will need to be hosted to store phenotypic trial data, pedigrees, genotypes, and proximal sensing images that can be easily accessed and queried. Quality control of genetic markers, pedigrees and phenotypes must be implemented and standard operating procedures must be developed.

A low cost, moderate density genotyping platform will need to be developed to enable genotyping of all materials represented in recent field trials, as well as all new material that enters into the phenotypic pipeline. Genotyping will allow for interconnecting trials to build large training sets for genomic prediction. Field experimental design will then be optimized to leverage all phenotypic information available, using genotypic information to link otherwise unrelated trials and traits. Eventually, rapid-cycle genomic selection using mathematical optimization will be implemented for expedited improvement of quantitative traits.

Research Philosophy

In the era of big data, the sheer amount of testable hypotheses is seemingly limitless. A shift away from small designed experiments to large observational studies at the breeding program or whole organism scale is inevitable. A traditional breeding program generates a plethora of phenotypic data that is used to make yearly breeding decisions, and subsequently discarded. If genotyped, these materials become treasure troves of data for asking questions, as well as making breeding decisions. This does not mean that we should cease the design and execution of experiments to address specific hypotheses, but we cannot ignore the valuable resource of observational data being collected, typically at great expense. Genotyping at this scale is feasible given the drastic reduction in costs and availability of third party services, and can be offset by clever experimental design that trades replication at an individual level for replication at a genetic level.

I believe in the collaborative model, where breeding programs do not operate in isolation. They share germplasm, resources, expertise, and most importantly ideas. Unlike germplasm, ideas also have the merit of being species flexible. I intend to build a collaborative effort at NMSU to aid all breeding programs to build foundational capabilities to increase efficiency of varietal development. New Mexico provides an important environment for plant breeding that is likely to be experienced by more farms throughout the globe as climate change progresses. Heat, drought, intense storms and hard frosts will be the new norm, and we must work together to do our part in defending our food security through accelerated genetic improvement.