

AG2PI SEED GRANT PROPOSAL Title of Proposal:

Delivering Resource Allocation Guidelines for Optimizing High-Throughput Phenotyping and Genotyping in Modern Breeding Programs.

Lead PI (Name, Title, Affiliation(s), email)

Mitchell Feldmann, Postdoctoral Researcher, Department of Plant Sciences, University of California, Davis, mjfeldmann@ucdavis.edu

Co-PI (Name, Title, Affiliation(s), email)

Daniel Runcie, Associate Professor, Department of Plant Sciences, University of California, Davis, deruncie@ucdavis.edu

Hao Cheng, Assistant Professor, Department of Animal Sciences, University of California, Davis, qtlcheng@ucdavis.edu

Collaborator(s) (Name, Title, Affiliation(s), email):

Steven Knapp, Professor and Director of the UC Davis Strawberry Breeding Program, Department of Plant Sciences, University of California, Davis, sjknapp@ucdavis.edu

Michael Gore, Chair and Professor, School of Integrative Plant Science Plant Breeding and Genetics Section, Cornell University, mag87@cornell.edu

Vance Whitaker, Associate Professor and Director of the University of Florida Strawberry Breeding Program, Institute of Food and Agricultural Sciences, University of Florida, vwhitaker@ufl.edu

Luis Diaz-Garcia, Assistant Professor and Director of the UC Davis Grape Rootstock Breeding Program, Department of Viticulture and Enology, University of California, Davis, diazgarcia@ucdavis.edu

Patricio Muñoz, Associate Professor and Director of the University of Florida Blueberry Breeding Program, Institute of Food and Agricultural Sciences, University of Florida, p.munoz@ufl.edu

Kelly Robbins, Assistant Professor, School of Integrative Plant Science Plant Breeding and Genetics Section, Cornell University, krr73@cornell.edu

Grant Administrator:

Sponsored Programs, UC Davis Office of Research, 1850 Research Park Drive, Davis, CA 95618-6153 Phone: 530-754-7700 Email: awards@ucdavis.edu

Keywords:

Genomic prediction, phenomic prediction, resource allocation, high-throughput phenotyping, indirect phenotyping

Project Description

1. Objectives/Aims

This proposal aims to provide guidelines and computational tools to help plant and animal breeding programs decide when and how to invest in new phenomic and genomic technologies. To accomplish our goal, we will organize a webinar series to discuss and catalog costs and constraint parameters of modern breeding programs (**Aim #1**), optimize the UC Davis Strawberry Breeding Program's allocation of effort between direct phenotyping, High Throughput Phenotyping (HTP), and genomic evaluations (**Aim #2**), and build a general optimization strategy to help deploy new technologies (**Aim #3**).

Aim #1 Document the variation of resource constraints and designs in breeding programs.

Today's crop and livestock breeding programs have access to advanced genotyping and high-throughput phenotyping (HTP) technologies to characterize their genetic resources with unprecedented depth and precision and enable data-driven schemes to develop varieties faster and with greater reliability [1]. However, deploying HTP in a breeding program requires considerable resource investments, drawing resources from other breeding activities such as genotyping or variety evaluations. Given limited resources, researchers may over-(or under)-commit to HTP and/or genomic selection, potentially leading to reduced gains due to smaller ground-truth datasets to use these technologies effectively. We hypothesize that (a) resource constraints will be the primary factor limiting the adoption of genomic and HTP evaluations and (b) optimizing resource allocations can help resource-limited programs achieve greater genetic gains more reliably. To learn and disseminate how breeding programs are deploying genomic and HTP evaluation technologies, we propose to organize a webinar series to give researchers and breeders a venue to discuss and share experiences on deploying genomic and HTP evaluations and how they make decisions about resource allocation in plant and animal breeding programs.

The webinar series will be organized around identifying breeding program costs, constraints, and the accuracy of various selection strategies, including mating designs and relatedness estimates, trial design at different stages, how many entries can be directly phenotyped or evaluated by genomics or HTP technologies per year/stage, the accuracy of genomic and HTP evaluations, and trial timelines (Figure 1). This will inform simulation studies to reflect real-world breeding programs (**Aim #3**). Before each webinar, we will give each speaker a questionnaire to ensure all necessary program design details are documented for future use in our research and by others. We will curate and provide diagrams and parameters of these breeding programs so that simulations (**Aim #3**) may be better connected to authentic program designs and constraints. Our collaborators have agreed to speak in this series and share details from their breeding programs.

Aim #2 Estimate statistical parameters for the UC Davis Strawberry Breeding Program.

Choosing which genomic and HTP technology to deploy in which selection stages, on which target traits, and on which entries is not straightforward. By imagining breeding programs as resource allocation problems using quantitative genetics principles, we can increase the response to selection by optimizing the trade-off between selection intensity and accuracy. We hypothesize that (a) HTP technologies will provide accurate predictions of disease resistance, plant productivity, and fruit quality traits in strawberries, (b) genomic and HTP evaluation are most advantageous in early selection stages, with direct phenotyping increasing in value in later selection stages, and (c) increasing evaluation capacity by HTP or genotyping will lead to an increased rate of genetic gain and improved

UC Davis Strawberry Breeding Program

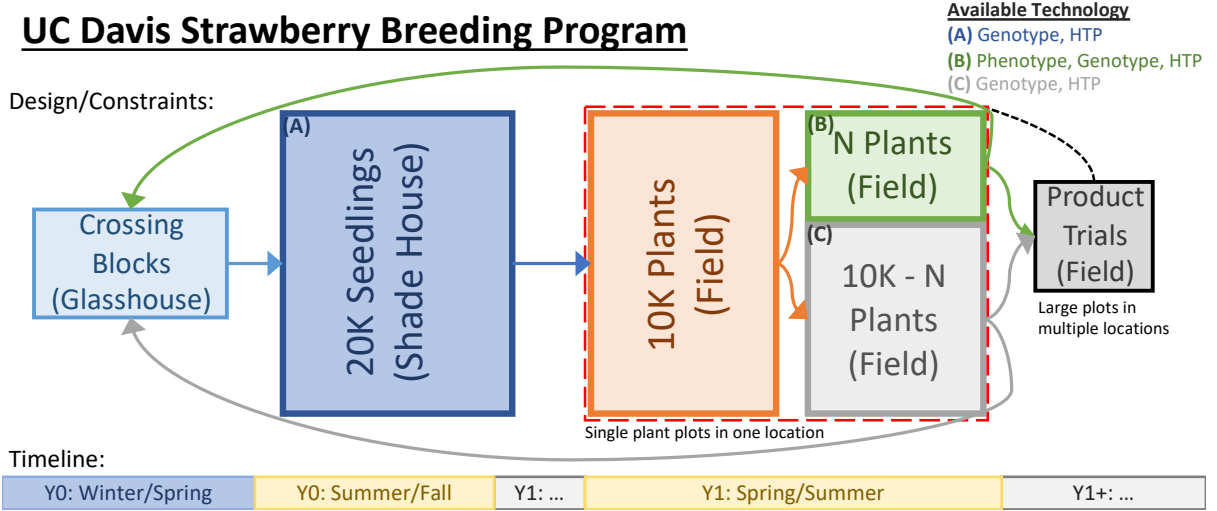


Figure 1: A basic schematic of one breeding cycle for the UC Davis Strawberry Breeding Program indicates scale capacity, points where decisions can be enhanced or optimized by adopting genomic and HTP strategies. (A) In the early stages, it may be possible to use genotyping and HTP on pre-transplant entries to improve the field-transplanted candidates. (B) For some traits, aroma by GC/MS, only $N \ll 10,000$ plants may be evaluated. (C) For the remaining $10,000 - N$ entries, genomic or HTP evaluations may be used to select the best candidates for product trials and crossing.

product advancement. To estimate the necessary statistical parameters, we propose to conduct extensive field data collection, including direct measurements of target traits, HTP measurements such as UAV-mount multi-spectral cameras and NIR spectroscopy, and high-throughput genotyping, and use these parameter estimates in simulations studies to inform future breeding decisions and designs [2].

We will collect phenotypic data, including disease resistance, plant productivity, and fruit quality traits, on strawberry hybrids planted in unreplicated and partially replicated designs in Davis, CA, in collaboration with the UC Davis Strawberry Breeding Program (Figure 1) [3, 4, 5]. Direct phenotyping to train statistical models will be conducted on several hundred (aroma) or several thousand (fruit size) hybrids. We will genotype all hybrids with a medium-density platform ($m = 5,000$) to estimate the accuracy of genomic prediction for the target traits using multi-trait genomic BLUP. We will collect HTP data (UAV) on all hybrids to estimate the accuracy of HTP evaluations for the target traits by penalized regression. This data package will allow us to estimate statistical parameters for multiple target traits using MegaLMM [6] and to inform future decisions using simulations that incorporate relatedness estimates, population size constraints, and the costs associated with genotyping, HTP, and direct phenotyping considering concurrent and sequential design philosophy.

Aim #3 Guide the optimization of genomic and HTP evaluations in breeding programs.

The mathematical theory of quantitative genetics provides a clear framework for optimizing resource allocation in a breeding program when we assume a constant yearly budget that can be divided between creating new lines, running field trials, genotyping, and phenotyping entries using either direct or indirect measurements [7, 8, 9, 10]. For any selection strategy, the rate of genetic gain can be expressed as follows:

$$r = \frac{i\rho\sigma_a}{t} \quad (1)$$

where r is the rate of gain per time \times cost, i is the selection intensity, ρ is the accuracy of estimating the genetic value of an entry, σ_a is the additive genetic variance among entries, and t is the time per selection cycle. Because no breeding strategy can simultaneously optimize every parameter of the breeder’s equation (1), and allocating resources towards one component necessarily reduces the budget available to other breeding program components, breeders must decide which strategy to follow and where to allocate resources. We aim to provide resource allocation guidelines for breeders, and our collaborators, to direct informed decisions grounded in quantitative genetics theory. We hypothesize that (a) researchers will be able to increase the rate of gain r using strategic sampling and genomic and HTP evaluation and (b) not all programs will be well situated to deploy genomic, or HTP, evaluations because of severe resource constraints. We propose to simulate multiple real-world breeding programs with known resource constraints to determine which strategies, program, and population parameters and under what conditions HTP and genomic evaluations are deployed to improve gain rate.

Our simulations will take a comprehensive set of conditions to account for various program structures and maximize the gain rate for specific breeding objectives. We will use open-access statistical software packages [6, 11] to simulate breeding programs with population-specific parameters and program designs determined by conversations with our collaborators and webinar speakers (Aim #1). We believe that modeling our simulations on a broad set of modern breeding programs will make our results immediately helpful and allow us to learn the statistical laws governing the optimal deployment of genomic and HTP evaluations. The simulation frameworks will be available for our collaborators and the breeding programs they support and will be publicly maintained.

2. Furthering the aims of the AG2PI

This project follows the AG2PI topics supporting the cross-fertilization of ideas, education and training, and the development of tools and datasets that can be used across multiple crop species for integrated crop yield and livestock feed optimization. The proposed research and outreach strive to deliver a solution to a community-wide gap in capacity building for genomic and phenomic prediction strategies and the availability of easy-to-use solutions for resource allocation in crop breeding programs. In this proposal, we will collect, develop, and/or integrate phenotyping data, tools, and technologies to advance genome-to-phenome research, develop strategies for handling and integrating disparate data types (multi-scale, multimodal), and create new tools that are accessible to those with limited resources. The research will include three experts, Drs. Mitchell Feldmann, Daniel Runcie, and Hao Cheng will empower researchers to implement resource optimization models with guidelines learned from realistic simulations with concrete constraints.

This proposal cross-cuts organisms and geographies by bringing together investigators and collaborators in plant and animal breeding programs serving communities in North America, Africa, and Asia, primarily. The extensive network and range of influence that our and our collaborators’ programs represent substantiate the importance of our proposal and this line of research. We believe that our research has the potential to make significant impacts in small breeding programs operating under severe resource limitations but also in programs that operate with greater financial head space.

We believe rapidly disseminating these resources to the broader community is essential to this proposal. We will make all data and R scripts publicly available to benefit those in the community who seek to learn more about our approaches. We aim to create a network of researchers that can continue to share experiences with resource allocation

and program constraints, ask and answer questions, and form new research collaborations using the Slack communication platform advertised on social media (Twitter, Mastodon, LinkedIn) and disseminated to our professional networks and their home universities.

3. Expected outcomes & deliverables

Our webinar series will be developed by inviting keynote speakers who lead premier plant and animal breeding programs to speak about their breeding programs and how they have changed their standard operating procedures to accommodate, integrate, and pay for the routine deployment of modern technologies (Aim #1). As part of our webinar series, a survey will be disseminated through social media and other platforms to assess how researchers and breeders think about adopting new technologies and how those adoptions require changing strategies for resource allocation.

We will collect and analyze genotypic data, indirect measurements, and direct observations on three soil-borne disease resistance, plant productivity, and fruit quality on hundreds or thousands, depending on the target trait, of strawberry hybrids in Davis, CA, in collaboration with the UC Davis Strawberry Breeding Program to estimate the statistical and population parameters affecting genomic and HTP deployment (Aim #2). This data package will contain genetic data, direct measurements, and indirect HTP measurements and will be made publicly available upon manuscript submission.

We will conduct rigorous simulations to test HTP and genotyping implementations in breeding scenarios built to reflect real breeding programs with parameters supplied by our collaborators and additional webinar speakers (Aim #3). This data package will contain program design parameters (population size, experiment designs, relatedness), statistical parameters for HTP and genomic evaluations (prediction accuracy), and other resource limitations necessary to represent specific programs accurately. This second data package will be publicly available as designs and parameters are collected and curated.

Our award will fund two international graduate students working with Dr. Feldmann and Dr. Runcie for the award period. One student, who is expected to graduate soon after the award period to enter the plant breeding workforce, will be leading the HTP and genomic evaluation from the UC Davis Strawberry Breeding Program (Aim #2) is planned as part of this student's Ph.D. Dissertation. The other student, who has not yet advanced to candidacy, is expected to play a significant role in designing and analyzing the simulations using AlphaSim-R (Aim #3) as part of their Ph.D. dissertation. This award is expected to play a significant role in the framing and success of two graduate degrees from the University of California, Davis, and will *seed* further research collaborations between the PIs and collaborators.

4. Qualifications of the project team

Dr. Feldmann (PI) received his Ph.D. from the University of California, Davis, in 2020 in Horticulture and Agronomy. His work is focused on how genomics and phenomics can be implemented in breeding programs, developing cost-effective genotyping solutions, and quantitative evaluation of complex traits in strawberries. In March of 2023, Dr. Feldmann will start at UC Davis as an assistant professor and director-elect of the Strawberry Breeding Program at UC Davis. Dr. Feldmann will be responsible for designing and analyzing genomic and phenomic studies in strawberries and performing statistical analyses.

Dr. Runcie (co-PI) received his Ph.D. in genetics at Duke University alongside an MS in Statistics in 2013. His lab at UC Davis focuses on genetic variation and plasticity to

identify molecular pathways underlying phenotypic differences among genotypes, learn about forces that shape the evolutionary histories of natural populations, and identify critical systems that limit plant responses to climate change. Dr. Runcie will be responsible for designing and implementing resource allocation studies through realistic simulation studies.

Dr. Hao Cheng (co-PI) received his Ph.D. in Animal Science and Statistics from the University of Iowa in 2017. His lab at UC Davis develops statistical and computational methods for population (genetic) improvement through more accurate, efficient, and biologically meaningful analysis. Dr. Cheng will be responsible for providing expertise around design elements that separate animal and plant breeding programs and parameter optimization expertise. All PIs are genetic researchers who utilize genomics and phenomics and share an enthusiasm for developing analytical tools for the crop and livestock research communities.

5. Proposal timeline

Objective	Milestone	Q1	Q2	Q3	Q4
(1) Webinar and Parameter Collection	Finalize list of webinar speakers	■			
	Collect and curate program design parameters	■	■		
	Conduct ZOOM Webinar		■	■	■
(2) Strawberry Case Study	Collect Strawberry Genotypic and Phenotypic Data	■	■	■	
	Prepare and Analyze Genotypic and Phenotypic Data			■	■
(3) Simulation Experiments	Plan Simulation Experiment	■	■		
	Run and Analyze Simulations		■	■	■

6. Engaging AG2P scientific communities & underrepresented groups

One of our primary objectives is to create an open webinar series and user forum to discuss strategies and to help researchers plan the most useful experiments for genomic and HTP evaluations. To go along with the webinar series, we will initiate a user forum using the Slack communication tool, which will be advertised through our universities and those in our professional networks. This will enable users worldwide to interact with a commonly minded group of peers, where they can ask questions, troubleshoot problems, and form collaborations.

We will maintain open-source code bases for our simulations developed in the R statistical programming language using AlphaSim-R so interested parties worldwide can participate. The code will be made available and maintained through GitHub so that users can have a base to develop their case-specific modules. We will develop and release substantial data packages containing program parameters for simulations (Aim #1) and empirical data (Aim #2). These digital assets will be publicly available for the AG2PI community and others.

We will fund two international students attending UC Davis whose dissertation research will be significantly strengthened by this USDA AG2PI award. The students, PI, and co-PIs will attend the NAPPN conference in 2024 and present different aspects of this work.

References

- [1] Elliot L Heffner, Aaron J Lorenz, Jean-Luc Jannink, and Mark E Sorrells. Plant breeding with genomic selection: gain per unit time and cost. *Crop Sci.*, 50(5): 1681–1690, 2010.
- [2] Holly M Lane and Seth C Murray. High throughput can produce better decisions than high accuracy when phenotyping plant populations. *Crop Science*, 61(5):3301–3313, 2021.
- [3] Dominique DA Pincot, Michael A Hardigan, Glenn S Cole, Randi A Famula, Peter M Henry, Thomas R Gordon, and Steven J Knapp. Accuracy of genomic selection and long-term genetic gain for resistance to verticillium wilt in strawberry. *The Plant Genome*, 13(3):e20054, 2020.
- [4] Stefan Petrasch, Saskia D Mesquida-Pesci, Dominique DA Pincot, Mitchell J Feldmann, Cindy M López, Randi Famula, Michael A Hardigan, Glenn S Cole, Steven J Knapp, and Barbara Blanco-Ulate. Genomic prediction of strawberry resistance to postharvest fruit decay caused by the fungal pathogen botrytis cinerea. *G3*, 12(1): jkab378, 2022.
- [5] Nicolás P Jiménez, Mitchell J Feldmann, Randi A Famula, Dominique DA Pincot, Marta Bjornson, Glenn S Cole, and Steven J Knapp. Harnessing underutilized gene bank diversity and genomic prediction of cross usefulness to enhance resistance to phytophthora cactorum in strawberry. *The Plant Genome*, page e20275, 2022.
- [6] Daniel E Runcie, Jiayi Qu, Hao Cheng, and Lorin Crawford. Megalmm: Mega-scale linear mixed models for genomic predictions with thousands of traits. *BioRxiv*, 2020.
- [7] Gregor Gorjanc, R Chris Gaynor, and John M Hickey. Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *Theoretical and applied genetics*, 131(9):1953–1966, 2018.
- [8] Daniel Runcie and Hao Cheng. Pitfalls and remedies for cross validation with multi-trait genomic prediction methods. *G3: Genes, Genomes, Genetics*, 9(11):3727–3741, 2019.
- [9] Giovanni Covarrubias-Pazarán, Johannes WR Martini, Michael Quinn, and Gary Atlin. Strengthening public breeding pipelines by emphasizing quantitative genetics principles and open source data management. *Frontiers in Plant Science*, 12, 2021.
- [10] Giovanni Covarrubias-Pazarán, Zelalem Gebeyehu, Dorcus Gemenet, Christian Werner, Marlee Labroo, Solomon Sirak, Peter Coaldrake, Ismail Rabbi, Siraj Ismail Kayondo, Elizabeth Parkes, et al. Breeding schemes: what are they, how to formalize them, and how to improve them? *Frontiers in Plant Science*, 12:791859, 2022.
- [11] R Chris Gaynor, Gregor Gorjanc, and John M Hickey. Alphasimr: an r-package for breeding program simulations. *BioRxiv*, 2020.