Awards issued prior to January 20, 2025, were funded under previous Administrations and may not reflect the priorities and policies of the current Administration.

United States Department of Agriculture
Project Initiation

Project Data for Accession Number 1031452

Project Title

NIFA AG2PI Collaborative: Improving Causal Gene Detection across Crop and Livestock Species

Project Details

Sponsoring Institution National Institute of Food and Agriculture	**Project Status** ACTIVE
Funding Source NIFA Non Formula	**Grant Year** 2023
Grant No. 2023-70412-41087	**Cumulative Award Amt.** $1,132,877.00
Proposal No. 2023-06073	**Multistate No.** (N/A)
Project Start Date Sep 15, 2023	**Project End Date** Sep 14, 2027
Program Code [AG2PI] Agricultural Genome to Phenome Initiative	

Project Director

Schnable, P.

Recipient Organization

IOWA STATE UNIVERSITY OF SCIENCE AND TECHNOLOGY

1350 BEARDSHEAR HALL

AMES, IOWA 500112103

Performing Department

AGRONOMY - CALS/AES

Non Technical Summary

In the face of 21st-century agricultural challenges, our mission is clear: we must produce more food, feed, and fiber for a growing population with evolving dietary preferences, while dealing with limited rural labor and agricultural land, and the need for bio-energy sources. Moreover, climate change introduces more frequent biotic and abiotic stresses. While global crop productivity has matched these challenges, we must intensify our efforts to sustain this progress. This is especially vital as we navigate the rest of the century.To address these pressing needs, our team of experts, spanning crop and livestock breeding, genetics, biochemistry, and data science, is forging ahead. We're developing innovative tools to decode the genetic basis of traits in crops like maize, soybean, sorghum, and in pigs. Our advanced statistical models, enhancing methods like GWAS, TWAS, and eQTL mapping, empower biologists to explore data in groundbreaking ways, uncovering new insights.We are bridging the gap between genetics and traits, from crop yields to Vitamin B levels in maize. Our research probes the interplay of genetics, weather, and environment using diverse

data. This newfound knowledge will steer enhancements in crucial U.S. crops and livestock.Our ambitious endeavor extends beyond discovery. It entails crafting novel statistical tools to comprehend essential genes in both livestock and crops, applicable across species. Aligned with the USDA's strategic goals, our work contributes to an equitable, resilient, and prosperous U.S. agricultural system, ensuring accessible, wholesome food for all. Through education and outreach, we'll empower crop and livestock breeders and cultivate the human capital needed to fulfill these aspirations.

Animal Health Component

Animal Health Component: 0%

Research Effort Categories

Basic	100%
Applied	0%
Developmental	0%

Classification

Knowledge Area (KA)	Subject of Investigation (SOI)	Field of Science (FOS)	Percent
201	2410	1060	25%
304	7310	2090	25%
304	3910	1060	25%
201	7310	2090	25%

Knowledge Area

[201] Plant Genome, Genetics, and Genetic Mechanisms

[304] Animal Genome

Subject Of Investigation

[2410] Cross-commodity research--multiple crops

[7310] Experimental design and statistical methods

[3910] Cross-commodity research--multiple animal species

Field Of Science

[1060] Biology (whole systems)

[2090] Statistics, econometrics, and biometrics

Keywords

eQTL, gene discovery, livestock breeding, team science, crop breeding, gene trait association

Goals / Objectives

United States Department of Agriculture
Project Initiation

The overarching goal of this project is to develop and support new statistical tools for the breeding of superior individuals or cultivars in genetic populations with the long-term goal of enhancing the production, sustainability, and climate and disease resilience of crop and livestock species. One way to enhance livestock and crop breeding strategies is by better understanding gene-trait associations and prioritizing causal genes of diverse agriculturally important phenotypic traits. Towards this end, the project will bring together researchers from variety of disciplines, including phenomics, genomics, genetic diversity, and data science. Biologists will bring their own biological questions and datasets from different crop and livestock species. Statisticians, with extensive experience in collaborations with biologists, will build statistical models and methodologies to analyze these datasets. The models and analyses will be updated iteratively following feedback from the biologists.Our objectives in this project are: (1) to build powerful multi-locus methods for combined GWAS, TWAS, and expression quantitative trait loci (eQTL) mapping, (2) to develop user friendly open-source R packages and Python libraries with detailed manuals, vignettes, and video tutorials, and (3) interweave research and education through the integration of training and cross-disciplinary research toward producing a skilled STEM agricultural workforce. We plan to achieve these research objectives by pursuing the following three specific objectives:Objective #1: Develop methods to combine GWAS, TWAS, and eQTL mapping of quantitative traits.Our working hypothesis is that a hierarchy of high-dimensional partial-linear and linear models, with appropriate shrinkage on SNP and gene expression effects, will be able to mitigate the confounding effects. In this objective, we will focus on traits for which the responses can be assumed to be univariate or multivariate Gaussian (normally distributed), possibly after a suitable transformation (e.g., log).Objective #2: Develop methods to combine GWAS, TWAS, and eQTL mapping of ordinal traits.The non-Gaussian traits we will focus on are ordinal scores (e.g., disease and root lodging scores). Our working hypothesis is that we will be able to improve the association results by properly accounting for the nature of the non-Gaussianity through an appropriate hierarchical generalized partial-linear multi-locus model. We will also retain the advantages of Objective #1.Objective #3: Develop methods to combine GWAS, TWAS, and eQTL mapping of functional data traits.Here, we will develop multi-locus methods for traits that are measured by a smooth curve (e.g., repeatedly measured phenotypes such as growth rates, time-series, light curves, A/Ci curves). Our working hypothesis is that we will be able to improve the understanding of the genetic basis of variations in the whole trait curves instead of being limited to univariate analyses of summary measurements or independent analysis of individual time points.Alongside the research outcomes, these initiatives will enhance expertise in agricultural genome-to-phenome research through education and outreach activities. Existing ISU courses will be improved by accommodating GWAS and TWAS methods in the syllabus. Outreach programs will provide education and support the research and training a broad range of crop and livestock scientists at multiple types of U. S. institutions. Hybrid workshops will be organized to facilitate training students and scientists.

Project Methods

The statistical methods include the development of hierarchical Bayesian models for combining GWAS, TWAS, and eQTL mapping. Latent indicator variables will be assumed, and model size will be penalized through Bernoulli priors on these latent indicator vectors. Theoretical results will be developed for choosing the right shrinkage to accurately detect associated genes. Fast scalable computational algorithms based on delayed Cholesky factorization, sparse-matrix algebras will be developed and implemented in C++ programming language. These models and methods will be extended to accommodate more general phenotypic responses through link functions.Three methods will be used to assess the validation of association studies: cross-validation with independent datasets from literature, biological pathway analysis, and network analysis with functional enrichment (GO or gene ontology terms) analysis. In addition, simulation experiments will also be conducted based on the literature.Two hybrid workshops will be organized yearly to disseminate research and the software and train the broader scientific community. Feedback will be sought from the workshop participants to assess the overall effectiveness of the workshops and to improve the accessibility of the software, manuals, and vignettes.