AG2PI SEED GRANT - PROJECT FINAL REPORT

PROJECT NAME	Harnessing Ag Genomics Data to link genotype to phenotype

PROJECT PRINCIPAL INVESTIGATOR	TODAY'S DATE	PROJECT START DATE	DATE OF COMPLETION
James Koltes	05/28/23	12/01/21	02/30/23
TEAM MEMBERS (co-PI, co-I, personnel)		COLLABORATORS	
Chris Tuggle, Peter Harrison, Alenka Hafner, Kyu-Sang Lim, and Bruna Petry		Zhiping Weng, Leonore Reiser, Damarius Fleming	

ACCOMPLISHMENTS

Please provide a short summary of the conclusions (both successes and failures) made from your project. Include a description of how this project will provide benefits to the agricultural genome to phenome community and, possibly, to a broader audience. You should include both qualitative and quantitative details, as necessary, to support your conclusions. Include a short accomplishment statement in non-technical language and do not include names.

This is not a technical report. Please keep to no more than 6-8 sentences (e.g., 1-2 sentences per point, above).

Successes: A virtual workshop entitled, "Harnessing the Ag Genomics Data torrent: A community-driven discussion on best practices for using and re-using genomics data" was held in February 2022 to identify opportunities and challenges in data reuse. A working group of researchers was also developed in collaboration with the AgBioData consortium in late 2022/ early 2023. The working group has drafted a data reuse white paper tentatively entitled, "Present and future challenges to data reuse in agricultural research" that we plan to submit by the end of June or early July 2023. We also developed a survey to share with the community to add additional plant and animal genomic data feedback to this white paper. This project identified a number of common challenges and needs across the plant and animal genomics community, including the need to improve metadata quality included with sequence data and the need to lower the barrier in submitting data to be publicly available, among other important issues. A grant proposal was also developed to promote genomics data resue that was submitted to USDA AFRI. The outcomes of this data reuse seed grant project will benefit the animal and plant genomics communities by identifying areas for improvement in data reuse and has resulted in the planning of a future software to assist in linking genotype to phenotype by reusing public data to annotate genome function.

Products

Please list any products from this project. This may include (but not limited to) publication, concept/white paper, workshop, conference presentation, website, publicly available data or pipelines, etc. Reminder: you are required to make your prodcuts available to the broader stakeholder community using standard USDA practices, open source, FAIR, or other models. Metrics may include number of participants or times accessed, etc. Include links to recordings, DOI, etc. when possible. For presentations and posters, provide authors, date, location and presentation title.

ACTIVITY / PRODUCT	**DESCRIPTION (include URL, if applicable)**	**OUTCOME / METRICS**
Virtual Workshop	Five presentations on data reuse and a roundtable discussion were held. More details are provided at this link, including speaker names, seminar titles and recordings of the session. https://www.ag2pi.org/workshops-and-activities/community-workshop-2022-02-09/	A total of 113 participants attended the virtual meeting t and more than 60 people attended the discussion to provide feedback on bottlenecks and opportunities in data reuse.
Poster Presentation	Citation: Chris Tuggle, Peter Harrison, Alenka Hafner, Kyu-Sang Lim, Zhiping Weng, Leonore Reiser, Chris Elsik, Damarius Fleming, and James Koltes. Harnessing Ag Genomics Data to link genotype to phenotype. USDA-AG2PI principal investigators (PI) meeting. Sept. 9, 2022 Ames, IA, USA.	Poster was presented to attendees of the AG2PI data reuse PI meeting.
Data Reuse Working Group (AgBioData)	A working group was developed with the intent of discussing issues that prevent data reuse, ideas to improve data reuse and examples where data reuse has been successful and helpful to the research community. A description of the working group and it's members is provided here: https://www.agbiodata.org/working_groups/data_reuse	A total of 24 members were recruited in the data reuse working group as a part of the AgBioData Consoritum. The work group has held biweekly meetings to discuss data reuse topics and has drafted a data reuse survey and whitepaper.
Workshop Presentation (AgBioData)	A talk entitled, "Data reuse working group: First steps and goals" was presented at the AgbioData meeting in Chicago, IL, USA on May 1, 2023 by Alenka Hafner. Authors: Alenka Hafner, Chris Elsik, Boas Pucker, Cecilia Deng, Peter Harrison, Ted Kalbfleisch, Elsa Herminia Quezada Rodriguez, Vicotoria DeLeo, Bruna Petry, Anne Thessen, and James Koltes A link to the presentation (at ~47:00) is available here: https://www.youtube.com/watch?v=FzkoGIRX2Ng&list=PL7PHQZk3QU4s_B1yAfve6MiGiizsTfPe_&index=1	This talk was presented at the AgBioData Meeting to all attendees. The talk is also available online and has been viewed 18 times. The talk was followed by a work group discussion that asked participants if additional topics should be covered in the white paper or if additional survey questions should be asked regarding data reuse.
Seminars presented at PAG30	The following 3 in-person presentations on data reuse were presented at PAG30 (authors in parentheses) and a link in included below to each of the meeting sessions where talks were presented. 1) Encode registry of candidate cis-regulatory elements (cCREs) for annotating the human genome (Zhiping Weng): https://plan.core-apps.com/pag_2023/event/3cee484866e6d00c5e492240c4d56f00 An additional two talks were presented in the session described here: https://plan.core-apps.com/pag_2023/event/3cee484866e6d00c5e492240c4d5637f 2) FAANGMine: Tools for Exploring Functional Annotation of Animal Genomes (Chris Elsik) 3) Risks and Rewards of Data reuse: Methylome Re-Analysis Yields Novel information (Alenka Hafner)	A seminar was presented at the, "US National Animal Genome Research Program (NRSP8)" workshop (Weng) and 2 seminars were presented in the session, "Big Data: Manage your data before your data kills you," (Elsik, Hafner) workshop to promote data reuse at PAG30.

Audience

With whom has this work been targeted to and shared? Please describe how this project and its products have been disseminated to a community of interest. Include any outreach activity or information sharing as well as training or professional development opportunities provided in this project.

Work from this project has been presented to plant and animal genomics researchers and industry at the PAG meeting, a virtual workshop, AG2PI workshops, an AgBioData workshop and is available to the general public. Products have been made available online, at workshop seminars, or posters presented in person and online. Professional development opportunities were provided for a PhD student who is a CoPI on the project (Alenka Hafner) who has been invovled in the grant objectives, presented data at the AgBioData conferece and the PAG 30 conference. In addition, Drs. Kyu-Sang Lim and Bruna Petry were invovled in the Data Reuse workshops, working group meetings and/ or Data reuse workshop and have helped take notes for workshops and meetings, assisted with writing, and have provided feedback as part of the project.

CONTINUATION OF WORK

Next steps

How do you/your team plan to continue moving this project forward? Include how AG2PI can assist in your forward momentum.

We are currently working to edit and complete our data reuse white paper as part of the activities invovled in the data reuse project.
We plan to continue the AgBioData data reuse working group to discuss phenomics/ precision sension data reuse and standardization as this was selected as a next foucus area for our working group by it's members.
We also plan to submit a follow up grant to extend this work. The major need from AG2PI is new avenues for larger funding to support the next steps of our research.