

NRSP_TEMP10: National Database Resources for Crop Genomics, Genetics and Breeding Research

Status: Under Review

Duration 10/01/2019 to 09/30/2024

Admin Advisors: [\[George Smith\]](#) [\[Steven Lommel\]](#) [\[Scot H Hulbert\]](#)

NIFA Reps:

Statement of Issues and Justification

Prerequisite Criteria

How is the NRSP consistent with the mission?

Technological innovation is propelling crop science into an era of discovery driven by Big Data, where scientists routinely generate and analyze larger and ever more complex genomic, genetic and breeding (GGB) datasets for both model and non-model crop species. The value of Big Data increases significantly when it is curated and made available in community databases created by crop domain experts. When organized, annotated, integrated with other data, and made available to browse, query and analyze, value is added and data becomes a re-usable, enriched **critical resource that will enable new and accelerated research discovery and application**. Dealing with the complexity of new data types and sheer volume of data generated has challenged established crop databases (1). Excellent resources exist for these databases, but they were developed in isolation, without access to a common database platform. Most have a crop or clade-specific focus and customized database schemas, making it hard to accommodate new data types, and requiring major resources from within this crop community to manage or interface with other crops or organisms.

Faced with these issues in 2008 when creating a Cacao database and in 2011 with the established Rosaceae and cotton databases, the bioinformatics team led by Drs. Dorrie Main and Sook Jung at Washington State University chose to use the Tripal database platform, an open-source web front-end for the generic Chado database schema (2, 3, 4). During this process, the team developed core and extension Tripal modules, in collaboration with other researchers in the Tripal community, transformed Tripal from a platform to (mostly) analyze transcriptomes to a more comprehensive and extendable open-source genomics and genetics database platform. In addition, the team used Tripal to efficiently build community databases for other underserved crop groups (citrus, Vaccinium and cool season food legumes). In 2018, the five databases of NRSP10 – the Genome Database for Rosaceae (GDR, 5), CottonGen (4, 6), the Citrus Genome Database (CGD, 7), the Cool Season Food Legume Genome Database (CSFL, 8) and the Genome Database for Vaccinium (GDV, 9) – are internationally recognized as the community databases for their crops. Tripal databases currently serve data on over 4000 crop and wild relative species.

The mission of the first NRSP10 project was to: 1) expand database resources for underserved fruit, nut, legume and cotton crops and 2) further develop standardized database protocols for use by other communities. The specific objectives and brief summary of results (See Appendix 1 for more details) follow:

Objective 1 - Expand online community databases currently housing high quality genomics, genetics and breeding data for Rosaceae, citrus, cotton, cool season food legumes and Vaccinium crops. All five NRSP10 databases have been significantly expanded in terms of the amount of data, new data types, new interfaces and analysis tools, usage and citations in publications (Tables 1 and 2).

Objective 2 - Develop a tablet application to collect phenotypic data from field and laboratory studies. The NRSP10 fruit, nut, legume and cotton research communities have adopted the Field Book App to collect phenotypic data.

Objective 3 - Develop a Tripal Application Programming Interface for building breeding databases. The Tripal Breeding Information Management System (BIMS) is compatible with Field Book and provides breeding management, data import, search, download and statistical analysis. Several peach, apple, pea, lentil and cotton breeders are using BIMS for their breeding programs, and many more are testing it.

Objective 4 - Convert GenSAS, the community genome annotation tool, to Tripal. The web-enabled genome annotation platform GenSAS has been further developed to include new annotation tools and functionality, providing a comprehensive resource for genome annotation and community curation, with publication-ready outputs. It is in the final stages of being fully integrated as an annotation module for Tripal databases.

Objective 5 - Develop Web Services to promote database interoperability. Tripal sites can now use RESTful web services to support remote client queries and ElasticSearch for cross-site data querying. Data collections can now be exported to a Galaxy workflow for analysis.

NRSP10 Renewal: Input systematically obtained from academic and industry scientist stakeholders has provided guidance about future types and volume of data likely to be collected for our target crops through 2024. In addition to much more whole genome sequence and RNASeq expression data, we expect petabytes of resequencing, epigenomic, pan-genomic and phenomics data. Integrating new types of data into legacy databases often requires changing schema, costing significant time, money and expertise. However, Tripal databases are built on Chado, the generic, modular and ontology-based database schema, which accommodates new data types easily. In addition, if interfaces and tools for new data types are developed as Tripal extension modules, they can be shared with other databases. We are already using tools developed by other Tripal databases, such as the Synteny Viewer (Fei Lab, BTI), Expression module (10) and ElasticSearch (10). We have released four extension modules (11) and BIMS (12) will soon be released. In the next NRSP10, we propose to develop new or enhanced Tripal modules to handle new types of data such as pan-genomic, epigenomic and phenomic data. In addition, we will also incorporate various analysis tools in modules like BIMS. Individual research labs, as well as large collaborative research groups, regularly generate high-throughput data they need to analyze. These analyses often require access to high-performance computing and include large, publicly available datasets and/or the results need to be examined using a graphical interface along with existing data. In addition, output of the analysis may need to be stored and made available for reuse in further research. While some data need to be kept private, results can be used by other researchers if certain aspects of the data are kept private. For example, breeders can share the relationship between genotype, phenotype and environmental conditions without jeopardizing IP as long as the pedigree and accession information are kept private. Use of community database resources in analysis, and not just in the design and submission phases, can promote data standardization as well as re-use. The need for user-driven high-throughput analyses, data re-use, integration, standardization and submission can be facilitated when community databases provide appropriate analysis tools with computing resources that can access the data stored in the database as well as the user's data. We will invest resources in a professional assessment to identify sustainability models for core Tripal and the NRSP10 databases.

In 2019-2024 we propose to extend the research-enabling mission of NRSP10 to:

- Objective 1: Expand the online community databases for Rosaceae, citrus, cotton, cool season food legumes and Vaccinium crops
- Objective 2: Develop a Tripal module for visualization of epigenomics data
- Objective 3: Enhance TripalMap to integrate genomic and genetic data
- Objective 4: Enhance TripalBIMS to (a) support phenomics data, (b) add GWAS analysis, and (c) global performance prediction capability
- Objective 5: Identify sustainability models and provide additional tools and resources as required by the community

Providing tools for standardized database construction and continuing development of efficient and widely used target crop databases for GGB research are goals fully consistent with the stated NRSP missions:

- *Development of enabling technologies and/or support activities (such as to collect, assemble, store and distribute materials, resources and information)*
- *Sharing of facilities to accomplish priority research*

How does the NRSP pertain as a national issue?

Rationale

Priority Established by ESCOP/ESS

This NRSP proposal targets five of the seven grand challenge priorities:

Grand Challenge 1: "Enhance the sustainability, competitiveness, and profitability of U.S. food and agricultural systems."

Grand Challenge 2: “Adapt to and mitigate the impacts of climate change on food, feed, fiber, and fuel systems in the United States.”

The negative impact of climate change on crop production is well-documented (19, 20, 21, 22). The specific impact on a crop or region can be devastating. For example, Georgia’s disruptively warm winter in 2017 led to the loss of an estimated 85% of the peach crop (23). The warm and dry California winter in 2014 caused severe damage to the cherry crop, and atypical low temperatures during bloom in 2012 destroyed up to 90% of the Michigan apple and tart cherry crops. Recent climate trends, such as changes in temperature and precipitation, decreased cotton yield (24). Providing access to germplasm evaluation and other breeding data by location and weather/climate is crucial in developing new cultivars less sensitive to unusual field conditions. Integrating performance and environmental data with genomic and genetic information, such as expression and trait loci data will also help identify genes that are responsible for traits sensitive to environmental change. Adding new types of pan-genome data and epigenome data will help in discovering new genes and/or new gene regulation patterns that affect traits that can reduce the impacts of weather/climate. In addition, high-throughput phenotyping technology that uses reliable screening tools and platforms to measure expression of physiological traits in realistic field environments will aid in QTL/gene discovery for both drought and heat resistance. This renewal will allow us to build infrastructure to accommodate high throughput phenotyping data, which will facilitate the development of strategies to mitigate the impact of climate change in crop production. In addition, the performance prediction tool will enable participating breeders around the world to anonymously share their data from vastly different environments so that they can predict the performance of their material in a wide range of environments. If this works, it will accelerate crop breeding, especially for perennial crops.

Grand Challenge 3: “Support energy security and the development of the bioeconomy from renewable natural resources in the United States.”

The provision of an easy-to-use, manageable, standardized platform for GGB database development will facilitate database development for organisms relevant to energy security based on renewable natural resources. Funded by NSF DIBBS and PGRP, we are collaborating with TreeGenes and Hardwood Genomics Project to build further Tripal database construction tools. This includes development of web services and Elastic searches enabling users to have seamless access from their home Tripal database to data in all the Tripal databases. We are also planning to perform synteny analysis among tree genomes and display the results using the Tripal Synteny Module in each database with hyperlinks to gene pages in specific databases. This will facilitate sharing data among tree genomes and open new opportunities to advance our understanding. Many scientists working on potential biofuel crops lack any access to well-curated genome databases. New Tripal databases will provide access to breeding-decision tools, enabling comparison of parental and selection evaluation data that cannot be readily done in any other free breeding data management software. Outcomes of this proposal include the essential infrastructure and expertise to impact on breeding programs for high-yielding, low input bioenergy feedstocks.

Grand Challenge 4: “Play a global leadership role to ensure a safe, secure, and abundant food supply for the United States and the world.”

The fruit, nut, legume and cotton crops of NRSP10 are important for food, feed and fiber supply for the U.S. and the world. These openly-accessible databases enable domestic and international research programs to readily exploit GGB resources. While the databases themselves are mostly crop-specific, the software infrastructure and expertise to be further developed in this proposal are explicitly intended to be exportable to other crops or organisms. We estimate that more than 100 Tripal databases are under development, mostly for crop plants but including databases for microbes, insects and other animals. Supporting core development and help desk support, NRSP10 is a critical infrastructure resource for these.

Grand Challenge 5: “Improve human health, nutrition, and wellness of the U.S. population.”

All the crops housed in the Rosaceae, Citrus and Vaccinium and cool season food legume databases contribute significantly to a health-giving and nutritious human diet. Breeding programs in the U.S. will gain considerably in efficiency and effectiveness through ready access to a Breeding Information Management System integrated with the high quality publicly available GGB data in their community databases. In addition, the availability of BIMS as a Tripal extension module will encourage adoption of the platform by other crop communities, benefiting a much wider audience.

Relevance to Stakeholders

The stakeholders of the databases described in this NRSP include biologists, breeders, bioinformaticists, educators, consumers, funding agencies and the industries based on the participating crops. The databases will continue to store and integrate data from various research projects, funded by government and industry, accelerating knowledge discovery from the integrated information and maximizing the return on the GGB investment. The outreach activity of this NRSP project benefits

researchers working on other agricultural databases as well. As detailed in the management plan, primary stakeholders from research institutions and industry will continue to participate in project development and assessment as members of the steering committee for the databases. Their participation will help ensure the development effort is prioritized to meet the needs of stakeholders. The ultimate stakeholders of this NRSP are consumers and U.S. taxpayers. Below is a detailed description of the benefits for each type of stakeholder.

Biologists

The curated genomic and genetic data and analysis tools available in the databases developed in this NRSP will help basic biologists who are interested in the structure and evolution of genomes, in gene expression, gene function and genetic variability, and in the mechanisms underlying various traits. Integrated genomic, genetic and phenomic data will help translational scientists who are interested in further QTL and marker discovery and genetic mapping studies. The integrated genomic, genotypic and phenotypic databases will also help applied scientists who are interested in developing methods for marker-assisted breeding. The molecular diversity data and germplasm data will help scientists go beyond well-known gene pools to explore other ways to achieve their goals. All databases supported by this NRSP contain data for multiple species, which will enable transfer of knowledge among related species as well as studies on genome evolution. In addition to integrating data from multiple species within each database, one goal is to develop tools to integrate data from diverse major tree databases by providing access to conserved syntenic regions, web services and cross-search searching. Consistent interfaces will also promote cross-utilization between communities by decreasing the learning time required to navigate each new database.

Breeders

The extensive breeding database in GDR enables breeders to search integrated GGB datasets for apple, peach, cherry and strawberry in a fully targeted manner and to retrieve and compare performance data from multiple varieties and seedlings, years and sites. This facilitates the streamlining of selection decisions and output data needed for variety release, publications and patent applications. In the current NRSP, we developed BIMS in Tripal so that breeders can load their data themselves, allowing them to truly 'own' and use their databases. BIMS also is compatible with Field Book App, an Android application for data collection that significantly helps breeders and scientists reduce the time and cost of phenotypic evaluation. While using BIMS in their community database, breeders have the option to keep their data in a private database but also link to all relevant public data. We will continue to add more functionality, including further analysis tools and breeding decision tools, to help breeders manage and utilize their data more efficiently and effectively. The new performance prediction tool in BIMS will aid breeders to predict how their material will perform in various environmental conditions. In addition, using the tablet application to collect data and upload to BIMS is expected to facilitate standardization of phenotypic evaluation methods and development of collaborative research projects on cultivar evaluation and breeding.

Bioinformaticists and Agricultural Database Developers/Managers

The renewal of this NRSP project will further develop Tripal as a freely available GGB database construction platform. This NRSP will contribute to adding new components to Tripal to handle high-throughput phenomics data, develop new extension modules for Tripal, and further develop BIMS functionality. These improvements will greatly benefit developers as they build databases housing large-scale GGB data for their respective research communities. The web services will help bioinformaticists who require programmatic access to large-scale data to integrate with other databases or perform further analyses. Access to updates and upgrades of the core Tripal package is essential for all Tripal databases. This includes ensuring core Tripal is updated to be compatible with new versions of Drupal and providing guidance and help in ensuring extension modules are also updated to remain compliant with the Tripal code. The NRSP project team is part of AgBioData, a consortium of more than 35 agricultural biological databases and allied resources. Our combined effort to ensure standards and best practices for acquisition, display and retrieval of genomic, genetic and breeding data will be highly beneficial to all agricultural database developers. One benefit to the IT workforce will be an increase in their transferable skills, since skill in Tripal will be useful to many databases.

Educators

Comprehensive tutorials, screencasts and videos developed for users of each database will be useful in formal or informal, classroom or distance learning contexts. Development of databases offers an ideal opportunity for graduate student education. Specific objectives include: 1) establishing cooperative Community-of-Practice-like interactions between the crop-specific curators and the core personnel that encourages the appropriate development and extension of resources; 2) small-group, face-to-face workshops in Pullman, WA to facilitate transfer of expertise between participants and common establishment of

priorities; 3) focused teaching modules for use in classroom or web-based delivery platforms that familiarize users and potential users with available resources and provide appropriate training.

Consumers

Consumers will be provided with higher quality, more nutritious and environmentally friendly fruits, vegetables and staple crops as a result of the use of the output of this NRSP.

Funding Agencies

Researchers increasingly rely on access to community databases to enable their research. Before NRSP10 was funded, its component databases largely relied on short-term funding from federal agencies, making them very vulnerable to loss of highly skilled personnel, valuable data and functionality, if/when there was a break, or potential break, in funding. The results from the proposed sustainability study will help identify the best economic and labor models to support core TriPal and the NRSP10 databases and reduce reliance on State Agricultural Experiment Stations and federal funding. The results of this analysis by Phoenix Bioinformatics, the group who successfully moved TAIR to a subscription model, will provide ideas and guidance for other GGB databases faced with similar challenges.

Crop Production Industry

Several industries, led by key commodity associations (rosaceous tree fruit, cotton, citrus, and dry pea and lentil), have provided funding and are participating in building and populating our target databases. This NRSP will ultimately benefit each of these crop industries since it will significantly enhance current, highly-utilized databases and expedite the development of cultivars that are competitive, profitable, sustainable and climate-adaptable. Our core strategy will allow additional crops to be added to the list on demand as support for further crop-specific efforts is generated. Continuation of this NRSP should substantially lower the expenses needed for a commodity to establish a database and the associated analytical tools.

Implementation

Objectives

1. Expand the online community databases for Rosaceae, citrus, cotton, cool season food legumes and Vaccinium crops
Comments: We will continue curation and integration of GGB data, adding the ability to handle new data types such as pan-genomic, epigenomic, and phenomic data, and providing tools for further analysis of new/updated whole genome sequencing data to connect to genes in other databases, find conserved syntenic regions and orthologous genes, and identify metabolic pathways. We will continue to develop utilities for QTL and marker identification, analysis of expression and methylation data as well as variation data from SNP array, resequencing and pan-genome analysis. In addition to publication/submitted data, we will add trait evaluation data from GRIN integrated with breeding data.
2. Develop a TriPal module for visualization of epigenomics data
Comments: Studies to understand epigenetic mechanisms underlying the basis of important traits are increasing for NRSP10 crops. To maximize utility of these data, we will develop a TriPal Epigenome module so users can search for genes/genomic fragments and view the level of epigenetic modification in various tissues and conditions. Integration with other genomic, transcriptomic, expression and genetic data will further increase the value and utility of this data.
3. Enhance TriPalMap to integrate genomic and genetic data
Comments: We will expand TriPalMap, our genetic map comparison and visualization viewer, to display genes and markers in chromosomes, as well as markers and QTL in linkage groups of genetic maps, similar to the NCBI map viewer. Chromosomes and linkage groups will be linked by shared markers, allowing users to explore the genomic features around QTL, even when only the genetic position is available. The genes in chromosome view in TriPalMap will hyperlink to JBrowse and other graphical viewers, allowing exploration of expression, methylation and sequence variation data.
4. Enhance TriPalBIMS to (a) support phenomics data, (b) add GWAS analysis, and (c) global performance prediction capability
Comments: Phenomics data is becoming increasingly available NRSP10 crops. BIMS will be enhanced to accommodate storage of high-throughput phenotyping, and integrate environmental and genotypic data. BIMS will be further enhanced by the addition of GAPIT for GWAS and genome prediction to allow users to identify genetic variations for important traits. A module will be developed that can combine national and international data for global performance predictions. Germplasm can perform differently in different environments and the module will enable breeders to predict the performance of their material under various conditions. In this tool, the phenotypic and genotypic data from various environmental conditions can be compiled in an anonymous database, allowing the effect of the environment on replicated genomic segments to be tracked. Users will be able to input the genotypic and phenotypic data of their materials then view the predicted performance under various conditions. Without revealing any proprietary information, breeders can contribute their data to the anonymous database to increase the accuracy of the prediction tool. We will also implement BrAPI web services (<https://brapi.docs.apiary.io/#>) so BIMS users can exchange their data with breeding tools in other systems using BrAPI.
5. Identify sustainability options and provide additional tools and resources as required by the community
Comments: Using data they collect in the last 6 months of the first NRSP10, Phoenix Bioinformatics will analyze and report on the best models for core TriPal and NRSP10 databases sustainability in years 1 and 2, with pilot implementation implemented and assessed in years 3 to 5. As current practice we will continue to add new tools or resources as requested by the community.

Projected Outcomes

- Outcome/Impact 1: Database resources that facilitate utilization and exchange of data among researchers across disciplines for Rosaceae, citrus, cotton, cool season food legumes and Vaccinium. Comments: ● GDR and citrus, cotton, cool season food legume and Vaccinium databases containing integrated up-to-date curated genomic, genetic and breeding data and new data visualization, mining and analysis capabilities ● Pan-genome database, epigenome viewer and advanced TriPalMap integrated with other genomic, genetic and breeding data. ● Enhanced BIMS integrated with databases where breeders can use tools for

performance prediction, GWAS analysis and other tools in remote systems through BrAPI.

- Outcome/Impact 2: An integrated genomics, genetics and breeding open-source database construction platform for building other biological databases. Comments: ● Up-to-date open-source platform for database construction, Tripal, with enhanced functionality ● Tripal epigenome viewer and enhanced TripalMap available for any Tripal database ● BIMS with enhanced functionality including phenomics data integration and GWAS and performance analysis tools, fully compatible with Tripal databases ● Increased provision of online resources, training and support to meet growing needs of the escalating adoption of Tripal implementation and module development
- Outcome/Impact 3: Resources that facilitate user-driven building of community databases through tools for direct data submission by users Comments: ● BIMS that allows direct data import for private use with options for releasing a portion of data to an anonymous database for analysis purposes
- Outcome/Impact 4: Community databases promote community building by acting as communication hubs Comments: ● Enhanced collaboration and coordination of specific research and extensions programs facilitated by access to data and communication tools in the target crop databases ● More exchange of ideas, data and tools within the individual communities and among communities through the standardization of databases. ● Steering committee meetings for each database facilitating communication and collaboration among researchers
- Outcome/Impact 5: Sustainability options identified for Core Tripal and NRSP10 databases Comments: ● Sustainability analysis enables progress toward stable self-sufficient platform and target databases enables retention of highly skilled staff ● Sustainable database promotes Stakeholder confidence in long term availability of research enabling resources, promoting use and provision of their data.

Management, Budget and Business Plan

Management

Thanks to the stable funding provided by the first NRSP10, and related federal and industry support, our *core team* remains mostly the same as the first project: the project director (Dr. Dorrie Main); lead curator/software designer (Dr. Sook Jung); three Tripal software developers (Taein Lee - *BIMS and GenSAS*; Chun-Huai Cheng - *Epigenome viewer, other software, Tripal core support, Tripal help desk support*; Katheryn Buble – *MapViewer and other software*); a database/system administrator (Heidi Hough); a data analyst (Dr. Ping Zheng); and two or more curators (Dr. Jodi Humann, Dr. Jing Yu, student/staff curators). In addition, we are formally adding Dr. Stephen Ficklin (co-initiator of Tripal with Dr. Meg Staton, UTenn), hired as Assistant Professor at WSU from postdoctoral position in the MainLab as lead core Tripal developer and Shawna Spoor - Tripal core development and Tripal Help Desk Support) to help meet the need of growing need for Tripal support and training. This is a direct result of increasing adoption of Tripal by the GGB database community.

The structure of our effort is rather complex, with NRSP10 (and components of other USDA and NSF grants) supporting infrastructure development, but working closely with crop-specific applications in order to develop valued infrastructure that can be implemented in other databases. As appropriate, effort will be shared with other projects, depending on how much of the task is infrastructure development and how much is specific to the associated database and supported by industry or federal funding directed to the crops involved. Each curator is assigned to a specific crop database(s). The curators will meet regularly with any newly hired or collaborating curators to train them on scientific curation procedures and check progress. Tripal data curation and analysis procedures are relatively independent of the crop/database so this management system will ensure that data curation expertise continues to be transferred efficiently as new crop databases arise in the community such as the Carrot Genome Database.

The project director (Main) will oversee the project in general and extension module development for the project. The lead curator (Jung) will design extension module functionality in consultation with other curators in the lab and in the wider Tripal community, and the lead core Tripal developer (Ficklin) will oversee core Tripal development and help desk support. The GAPIT package (25) and the performance prediction tool (26) will be implemented in BIMS in close collaboration with the authors of the tools, Dr. Zhiwu Zhang (WSU) and Dr. Craig Hardner (University of Queensland), respectively. Further development of BIMS functionality will continue to be in close collaboration with Dr. Ksenija Gasic (peach breeder, Clemson University), Dr. Todd Campbell (cotton breeder, USDA-ARS Florence) and Dr. Rebecca McGee (legume breeder, USDA-ARS Pullman). In addition, we will continue to seek input from participating breeders through in-person and teleconference visits/meetings with individual programs, monthly BIMS seminars, and NRSP10 workshops and training sessions. Each programmer will be responsible for using Drupal coding styles and the existing Tripal API. The lead curator and programmers will continue to have bi-weekly developers' meetings to discuss the progress of tool development and to share any practical information in developing tools. When a beta version of software is finished, the appropriate curators and other programmers in the lab will test the software. The entire project team will meet weekly using online teleconferencing to discuss tasks, timelines and progress.

The steering committees for GDR and CottonGEN will continue to meet regularly to review progress and provide feedback on tasks for the next quarter. Specific benchmarks will be set in consultation with the steering committees for each database at the start of the project and reviewed quarterly. The steering committees will also facilitate the standardization of various data to make the data findable, accessible, interoperable, and reusable (FAIR, 27). RosEXEC, who serve as the steering committee for GDR, created a subcommittee that considered and published principles for standardizing gene names (28) that were incorporated into GDR. We are in the process of establishing a steering committee for the Vaccinium database (chaired by Dr. Hamid Ashrafi, NC State University), meeting with the International Citrus community in January at PAG 2019 regarding a steering committee, and establishing one for the cool season food legume database (chaired by Dr. Rebecca McGee).

The adoption and growth of Tripal as a common platform for GGB database construction over the last 4 years has been very encouraging and suggests the model we proposed is working well. With over 150 downloads of the platform, Tripal databases have been constructed for a range of plant, animal and microbial organisms. The adoption trend continues, with more engagement as more functionality is added by the developer community. Tripal (<https://tripal.info>) and extension modules are being developed by 12 institutions, including Washington State University (Stephen Ficklin and Dorrie Main programs), University of Tennessee (Meg Staton program), University of Connecticut (Jill Wegrzyn program), Iowa State University (Ethalinda Cannon), National Center for Genome Research (Andrew Farmer), USDA National Agricultural Library (Monica Poelchau/Chris Childers), USDA-ARS (Steven Cannon), University of Saskatchewan (Lacey Sanderson), Boyce Thompson Institute (Zhangjun Fei program), Stowers Institute (Sanchez program), Bioversity International (Max Ruaz program), and Clemson University (Alex Feltus program). As a key part of the NRSP10 renewal we will continue to develop core Tripal, provide support for module development and implementation, and assess sustainability options for core Tripal and assist groups seeking to develop new databases. Tripal has a Project Management Committee (PMC) for code approval and feature requests and plans are underway to add a steering committee of representative stakeholders to help guide the continued growth and sustainability of Tripal.

We will continue to hold monthly Tripal community and developer meetings and yearly 2 day hackathons. The community meetings involve discussion between users and developers of Tripal, providing invaluable feedback on usability and functionality of Tripal, reducing duplication of effort, and promoting collaborations. The developer meetings alternate between a Help Desk open call and a training topic.

As core members of the AgBioData consortium, we will continue to ensure our work aligns with its goal to identify common issues in database development, curation, and management, with the goal of creating database products that are more FAIR compliant (Findable, Accessible, Interoperable and Reusable). To this end we have submitted a \$1M USDA NIFA FACT Coordinated Innovation Network proposal for the AgBioData Consortium (PI: Main) to advance the efforts outlined in the White Paper (1).

Timeline

In Year One: 1) Collect and curate genomic, genetic and breeding data for all databases and incorporate FAIR practices developed in AgBioData; 2) Collect gene expression data and implement Tripal expression module in GDR; 3) Collect pan-genome data for GDR; 4) TripalMap: Design graphical viewer for genome data in Chado; 5) BIMS: Obtain phenomics data from cotton breeders and design storage methods; 6) Global Prediction Tool in BIMS: Develop training dataset and implement prediction algorithm; 7) Begin analyzing a range of potential funding sources including voluntary membership models, data deposit fees, subscriptions, freemium models, crowdfunding, corporate support, and philanthropy; 8) Update all database webinars and tutorials; 9) Hold at least twice yearly individual database steering committee meetings and one annual project wide meeting; 10) Present database and tools at conferences and meetings.

In Year Two: 1) Collect and curate genomic, genetic and breeding data for all databases and incorporate FAIR practices developed in AgBioData; 2) Collect expression data and implement Tripal expression module in CottonGEN; 3) Develop and implement tools to display pan-genome data in GDR; 4) Design data storage method and search/display tool for epigenome data using any available data from Rosaceae; 5) TripalMap: Add functionality to compare features in chromosomes and genetic linkage groups; 6) BIMS: Develop data loaders for phenomics data; 7) Global Prediction Tool in BIMS: Design interface for BIMS users; 8) Report results of sustainability analysis study with recommendations; 9) Update all database tutorials; 10) Hold at least twice yearly individual database steering committee meetings and one annual project wide meeting; 11) Present database and tools at conferences and meetings.

In Year Three: 1) Collect and curate genomic, genetic and breeding data for all databases and incorporate FAIR practice developed in AgBioData; 2) Collect any available expression data and implement Tripal expression module in CSFL and CGD; 3) Develop and implement tools to display pan-genome data in CottonGEN; 4) Develop data loader and search/display tool for epigenome data in GDR; 5) TripalMap: Continue to develop TripalMap and add connectivity to other graphic viewers such as JBrowse; 6) BIMS: Develop search/download page for phenomics data; 7) Global Prediction Tool in BIMS: Design interface for non-BIMS users; 8) GWAS tools in BIMS: design interface to GWAS tool (GAPIT); 9) Begin pilot implementation of sustainability recommendations; 10) Hold at least twice yearly individual database steering committee meetings and one annual project wide meeting; 11) Present database and tools at conferences and meetings.

In Year Four: 1) Collect and curate genomic, genetic and breeding data for all databases and incorporate FAIR practices developed in AgBioData; 2) Develop interface to connect expression/pan-genome/epigenome viewers; 3) Develop and implement tools to display pan-genome data in CSFL; 4) Implement data loader and search/display tool for epigenome data in

CottonGEN; 5) TripalMap: Further develop interface for administrator to reflect new functionality; 6) Implement BrAPI in BIMS; 7) Implement Global Prediction Tool in other databases when the training datasets are available; 8) GWAS tools in BIMS: Implement interface to GWAS tool (GAPIT) in GDR; 10) Assess success of sustainability pilot implementation; 11) Hold at least twice yearly individual database steering committee meetings and one annual project wide meeting; 12) Present database and tools at conferences and meetings.

In Year Five: 1) Collect and curate genomic, genetic and breeding data for all databases and incorporate FAIR practice developed in AgBioData; 2) Continue to refine epigenome viewer and the connectivity with other graphic viewers; 3) Implement pan-genome data viewer and epigenome viewer in other databases when data become available; 4) TripalMap: Further refine the functionality and tutorial; 5) Update all webinars and tutorials; 6) BIMS: Develop tools to produce output for other breeding related tools; 7) BIMS: Further develop any necessary functionality for underlying Global Prediction Tool and GWAS tool; 8) Continue to assess success of sustainability implementation; 9) Hold at least twice yearly individual database steering committee meetings and one annual project wide meeting; 10) present database and tools at conferences and meetings.

Budget

A total of \$2,449,789 (Table 3) is requested over five years for this NRSP10 (2019-2024) to support the development activities described in this proposal. Additional funding from aligned objective support (Table 3) is projected over the course of this project from WSU (\$3,035,527) and Industry (\$741,889). Our two major SCRI (\$2.74 M) and NSF (\$2.99 M) grants will be completed by the start of this renewal project, but we will be submitting a renewal for the GDR SCRI project and are requesting a Centre of Excellence designation, as well as submitting to NSF PGRP. We will also seek further support from the Washington Tree Fruit Research Commission for tree fruit data analysis and curation; the USA Dry Pea and Lentil Council and Northern Pulse Growers for pea and lentil data analysis and curation and the Citrus Research Commission for data curation support. Support of ~\$180,000 per year from Cotton Incorporated, the cotton industry and USDA-ARS is available for CottonGen through 2021. Known support for the period of renewal (\$2,069,983) combined with funds requested from NRSP (\$2,449,789), totals \$4,519,772 for the complete project.

Table 3: NRSP10 Budget Requested Summary

Cost	Year 1	Year 2	Year 3	Year 4	Year 5	Total
Salaries	394,140	409,906	426,302	443,353	461,088	2,134,789
Publications	4,000	4,000	4,000	4,000	4,000	20,000
Goods & Services	15,000	15,000	15,000	15,000	15,000	75,000
Equipment	40,000		40,000			80,000
Consultancy	15,000	15,000	5,000	5,000		40,000
Travel	20,000	20,000	20,000	20,000	20,000	100,000
Total	488,140	463,906	510,302	487,353	500,088	2,449,789

Salaries: 4 Tripal developers, 50% Database Administrator/System Admin, 15% PI Main, 7.5% Ficklin

Publications: 2 per year

Goods and Services: server room fees, service contracts, data storage and onsite and offsite backup

Equipment: for new database/web servers in years 1 and 3

Consultancy: Phoenix Bioinformatics to analyze and help implement sustainability options

Travel: to present activities at conferences/meetings for all 5 crop database communities, travel to meet with stakeholders for individual training on using BIMS, attend yearly Tripal hackathons

Table 4: NRSP10 Aligned Support Cost

Cost	Year 1	Year 2	Year 3	Year 4	Year 5	Total
Salaries	241,142	245,888	240,401	245,019	124,872	1,097,322
Benefits	177,049	182,414	188,149	193,948	156,232	897,792

Good and Services	8,548	11,148	11,214	13,959		44,869
Travel	6,000	7,000	8,000	9,000		30,000
Total	432,739	446,450	447,764	461,926	281,104	2,069,983

Salaries: 50% Sys Admin, 30% Dorrie Main, 120% Data Curator, 30% programmer, 25% Data Analyst

Benefits: paid by WSU for all positions

Funding from other sources: CottonInc, USADPLC, NPG, WTFRC

Business Plan

In the first NRSP10, we sought successfully (see Appendix 1 for more details) to replace prior *ad hoc* funding of critical crop databases and model database platform construction with a more sustainable two-level model. This two-level model obtains support for core database activities from NRSP10 but engages funding from industry stakeholders and regional/federal grant competitive sources for data curation and analysis activities, as well as other infrastructure development. Totalling over \$10M (Appendix 1) these aligned funds represented a **direct return on NRSP10 investment of more than 5 to 1**. This exceeded the projected aligned funding from the first proposal by more than \$6M. In this renewal we propose continuing and extending our two-level approach by assessing and implementing models toward creating a sustainable ecosystem for core Tripal and NRSP10 databases.

In this renewal proposal development activities will be supported by funds from the NRSP and include costs associated with database administration, development, data storage, IT server room space rental, server service contracts, data backup, and support desk help for other Tripal adopters. It will also fund updates to computational database and web servers to meet escalating demand for fast, efficient database access. The second major area of activity -- data curation and analysis support will be sought through industry and researcher stakeholder support. Extending this model we will add (a) assessment of potential models for sustainability in years 1 and 2, and (b) implementation of the most promising ones as pilot projects in years 3 and 4, with the goal of eliminating or reducing the need for a third NRSP10. To this end, we have solicited the services of Phoenix Bioinformatics (<http://www.phoenixbioinformatics.org/>), a company founded in 2013 by the staff of the Arabidopsis Information Resource (TAIR, 29), who successfully pioneered a subscription based sustainable funding model to support TAIR. Their nonprofit mission is to help other projects achieve sustainable support using the tools and expertise developed for TAIR's transition. As we have some funds available in this last year of the current NRSP10 project, we will redirect them to Phoenix to begin an assessment of core Tripal and NRSP10 crop databases. This study will capture cost of operations, staff level, sources of funding, usage level, data types, species and strains, and what types of researchers and others (educators, students) are being served. In the renewal project Phoenix will analyze a range of potential funding sources including voluntary membership models, data deposit fees, subscriptions, freemium models, crowdfunding, corporate support, and philanthropy. The results of this study will be shared with AgBioData member databases. If our USDA NIFA FACT Coordinated Innovation Network grant (PI Main) is successful, all 30 plus member databases will also be assessed by the Phoenix team, with more in depth analysis for representative databases.

Recognizing that the sustainability solution for our Tripal databases may involve starting a non-profit or for-profit company, Drs. Dorrie Main, Stephen Ficklin and Mike Kahn have audited the WSU NSF iCorps course (<https://research.wsu.edu/icorps/>) and will enroll in the class formally in Fall 2019. The I-Corps 8-week program engages faculty, student and staff entrepreneurs to transform their ideas into successful business products. This program advises on appropriate grants to apply for such as the Small Business Innovation Research (SBIR) program. We see potential for offering Tripal services to businesses as a possible mechanism to help fund our public research database efforts.

Integration

NRSP10 is highly integrated with academic and government research programs and is stakeholder-driven (as evidenced by the project participant list and supporting letters). Providing continued access to collated, curated, and integrated public genomics, genetics and breeding data will enable unanticipated scientific advances well beyond the research for which the

data were originally collected. The enhanced and connectable databases provide a catalytic environment where genomicists, geneticists, bioinformaticists, breeders and growers can share data and ideas to elevate trans-disciplinary understanding of their crops, to suggest compelling directions for new methods and research, and to produce efficient and focused practical steps toward the common goal of crop improvement. Specifically, integration of the genetic, genomic and breeding data will enable members of the crop improvement community to develop new scientific hypotheses and theoretical models and test these using appropriate databases and tools. The results will improve understanding of the fundamental biology underlying the crops and their valuable traits. By integrating genetic data (such as quantitative trait loci, genetic markers, and pedigrees) used to make and populate genetic maps with genomic data (genome sequences, chromosomal physical arrangements, sequence variants from large populations, and gene expression measurements), phenomics and environmental data, genomics-genetics-breeding translational biologists will be able to develop better tools and knowledge for developing improved cultivars.

In addition to providing integrated databases for individual research programs worldwide, the database team has also been involved with several extension and academic programs in more direct ways. GDR has been supported from 2014-2019 through an USDA SCRI-funded \$2.74 million award. The GDR team was also integral to the 2009-14 and 2015-19 USDA NIFA SCRI-funded \$10 million community-wide RosBREED projects providing that project with genomic and genetic data analysis including synteny analysis for data transfer among crops and development of Breeding Information Management System to enable marker-assisted breeding in rosaceous crops. CottonGEN is supported from 2011-2021 by a combined grant of \$1.2 M from Cotton Incorporated, Bayer, Corteva, Southern Association of Agricultural Experiment Station Directors and USDA-ARS. Our team has participated in other major research initiatives, providing both sequence analysis and database support to projects such as the International Pea and Lentil Genome Sequence Consortia and analysis of unculturable *Ca. Liberibacter crescens*, causal agent of citrus greening, with culturable *Liberibacter crescens* to identify candidate genes that may be missing in the unculturable strain. We are also embedded in the Vaccinium CAP, Peach and PolyPloids projects with planned submission to the 2019 SCRI program.

Through this NRSP, we hope to systematize the effort and use the experience acquired by these diverse crop research communities to maintain and improve existing databases and lower barriers to entry for the construction of new ones. Industry stakeholders have been directly involved with database development as funding agencies and as users and members of advisory groups. The Washington Tree Fruit Research Commission (WTFRC) funded development of an apple and cherry cultivar performance database and toolbox built upon GDR resources originally funded by the NSF. Stakeholders from university, government and industry sectors from all the production and research regions of the U.S. are well represented in RosEXEC, the steering committee for GDR. The CottonGen steering committee is also composed of representatives from universities, government and industry. It meets quarterly to communicate the current and emerging database needs of the cotton research community and other stakeholders for development, implementation and dissemination of CottonGen. CottonGen is home to registration and submission of abstracts for the biennial International Cotton Genome Initiative conferences.

NRSP10 is also developing sharable tools for database construction so that the research on crop database development can be integrated and shared. All the tools we developed and will develop are and will be available as extension modules, so that other TriPal databases for other crops and organisms can readily incorporate them. Some examples include the Legume Information System (30) and TreeGenes, who converted to TriPal, and new databases include PeanutBase, the Carrot Genome Database and i5K NAL (31).

In addition, we have been core members of the AgBioData Consortium, a consortium of agricultural biological databases and associated resources working together to ensure standards and best practices for acquisition, display and retrieval of genomic, genetic and breeding data.

Outreach, Communications and Assessment

Outreach

Our outreach effort will focus on (1) crop communities that the databases serve; (2) communities who may be in need of TriPal databases; (3) the TriPal community; and (4) participation in the AgBioData consortium to increase: (a) standardization of protocols and practices across agricultural GGB databases to enhance GGB research outcomes, (b) researcher compliance in FAIR principles, (c) journal compliance with availability of peer-reviewed publication data, and (d) collaboration with funding agencies to help facilitate data deposit of meta, raw and analyzed data by funded researchers.

NRSP10 workshops will continue to be held yearly at the International Plant and Animal Genome (PAG), American Society of Horticultural Science, and Cotton Beltwide Conference annual meetings, as well as presentations in appropriate workshops at PAG, at the Crop, Soil and Agronomy annual meeting, and more crop specific meetings such as the International Rubus and Ribes, International Rosaceae Genomics, International Cotton Genome Initiative, and Citrus HLB Conferences. The NRSP10

workshops provide training that primarily involves stakeholder researchers demonstrating how they use component NRSP10 resources. We have found this to be a very effective method of training, with time for a discussion session to solicit feedback with core NRSP participation to keep the development aligned with stakeholder interests. Participants at the workshops and computer demonstration sessions are composed of researchers from many other crop communities, so presentations at PAG, in particular, provide good opportunity to present the utility of the Tripal platform and the support we provide in implementing or converting other databases.

In addition to presentations at conferences, we will also continue to update tutorials on how to use databases, host webinars and make them available online, maintain mailing lists and continue to publish when significant development has been made.

To reach database developers, Tripal will continue to be presented and training provided at yearly Tripal workshops at PAG and other meetings such as the GMOD, Galaxy and Bioinformatics conferences. We will continue to provide tutorials, mailing list responses, and Help Desk support in collaboration with other Tripal developers and update modules as new Drupal versions are released. We will continue to provide tutorials and publish when we produce new modules and applications.

Additionally, it is very important that breeders, stakeholders and the general public are trained about the usefulness of their community databases and functionalities that might be incorporated from other databases. We will add a public component to our website highlighting the utility of the databases in successful research stories that impact consumers, explaining the terminology at an appropriate level and including database training from the workshops. These extension and outreach activities will greatly enhance the research and extension programs of cotton, legumes and horticultural specialty crops. Using the database in the process of breeding superior plant selections will distinguish the United States amongst others in the world market for those crops. In addition, in association with the continued growth of cotton, legumes and horticultural specialty crop production, it will lead to more jobs with higher incomes, which in turn will create economic development and prosperity, enhancing the quality of life in rural areas.

We will continue to be good citizens of the AgBioData Consortium of Agricultural-related databases. This will include continuing to serve on the Steering Committee, helping to organize yearly workshops, engaging in sustainability studies, hosting the AgBioData website (<https://www.agbiodata.org>), participating in monthly meetings and specific workgroups, as well as adopting the recommendations of the WhitePaper (1), and applying for funding to further the research enabling mission of this group (currently a USDA NIFA FACT Coordinated Innovative Network proposal for AgBioData is pending – PI Main)

Communication

In addition to the workshops mentioned above, other conduits to facilitate communication between the database developers and the users are needed. While we envision that implementation of each database will have significant individuality, each database that associates with NRSP10 will have a steering committee, composed of representatives from universities, government and industry for each crop, which will meet quarterly or biannually by online teleconferences to communicate the current and emerging database needs of their research community with other stakeholders and the NRSP staff to guide the development, implementation and dissemination of resources for the database. Any new major development will be extensively discussed in these committee meetings. The meeting minutes will be posted on the NRSP and crop-specific websites as well as all reports and work plans. We will also have regular newsletters, twitter and LinkedIn accounts to notify the users of any new developments in the database and the crop research community.

Assessment

Various metrics will be used to assess the impact of the proposed project. For each database this will include usage statistics as measured by google analytics, feedback from the steering committee, annual online surveys of each community, number of publications, number of publications citing the databases and feedback via the online forms in the databases. Creation of new Tripal databases, number of projects adopting Tripal, number of species being served by through Tripal databases, and number of active developers with this project will all be indicators of the success of the use of Tripal. Very relevant measures of participation will be the number of curators and crops associated with the NRSP and the level of investment in NRSP by the users.

Literature Cited

1. Harper, L., Campbell, J., Cannon, E.K.S., Jung, S., Poelchau, M., Walls, R., Andorf, C., Arnaud, E., Berardini, T.Z., Birkett, C. *et al.* (2018). AgBioData consortium recommendations for sustainable genomics and genetics databases for agriculture. *Database (Oxford)*, bay088
2. Sanderson, L.-A., Ficklin, S.P., Cheng, C.-H., Jung, S., Feltus, F.A., Bett, K.E., and Main, D. (2013). Tripal v1.1: a standards-based toolkit for construction of online genetic and genomic databases. *Database (Oxford)*, bat075.
3. Ficklin, S.P., Sanderson, L.-A., Cheng, C.-H., Staton, M.E., Lee, T., Cho, I.-H., Jung, S., Bett, K.E. and Main, D. (2011). Tripal: a construction toolkit for online

genome databases. *Database (Oxford)*, bar044.

4. Jung, S., Lee, T., Yu, J., Ficklin, S.P., and Main, D. (2016). Chado use case: storing genomic, genetic and breeding data of Rosaceae and Gossypium crops in Chado. *Database (Oxford)*, baw010.
5. Jung,S., Lee,T., Cheng,C.H., Buble, K., Zheng, P., Yu,J., Humann,J., Ficklin,S., Gasic, K., Scott, K, Frank, M., Ru, S., Hough, H., Evans, K., Peace, C., Olmstead, M., DeVetter, L.W., McFerson, J., Coe, M., Wegrzyn, J.L., Staton, M.E., Abbott, A.G. and Main, D. (2018) 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. *Nucleic Acids Research*
6. Yu,J., Jung,S., Cheng,C.-H., Ficklin,S.P., Lee,T., Zheng,P., Jones,D., Percy,R.G. and Main, D. (2014) CottonGen: a genomics, genetics and breeding database for cotton research. *Nucleic Acids Research*, **42**, D1229–D1236.
7. Humann, J.L., Piaskowski, J., Jung, S., Cheng, C.H., Lee, T., Frank, M., Scott, K., Zheng, P., Flores-Gonzales, M., Saha, S., et al. (2017). Resources in the Citrus Genome Database that enable basic, translational, and applied research. 5th International Research Conference on Huanglongbing: March 14-17, 2017, Orlando, FL, USA.
8. Jung,S., Humann,J., Cheng,C.H., Lee,T., Zheng, P., Frank, M., McGaughey, D., Scott, K., Buble, K., Yu, J., Hough, H., Coyne, C., McGee, R., Main, D. (2017b) Updates to the Cool Season Food Legume Genome Database: Resources for pea, lentil, faba bean and chickpea genetics, genomics and breeding. Proceedings of the North American Pulse Improvement Association Biannual Meeting: November 2017, East Lansing, MI, USA
9. Bassil, N., Jung, S., Cheng, C-H., Lee, T., Zheng, P., and Main, D. (2017). NRSP10 Resources for Small Fruit Research. Proceedings of the ASHS Annual Conference; September 19-22, Waikoloa, HI
10. Chen,M., Henry,N., Almsaeed,A., Zhou,X., Wegrzyn,J., Ficklin,S. and Staton,M. (2017). New extension software modules to enhance searching and display of transcriptome data in Tripal databases. *Database (Oxford)*, bax052
11. Jung,S., Lee,T., Cheng,C.-H., Ficklin,S., Yu,J., Humann,J. and Main,D. (2017) Extension modules for storage, visualization and querying of genomic, genetic and breeding data in Tripal databases. *Database c*
12. Jung, S., Lee, T., Cheng, C-H., Gasic, K., Campbell B.T., Main, D. (2018). Using the Tripal Breeding Information Management System (BIMS) to Enable Efficient Management of Phenotypic and Genotypic Data. Abstracts of the International Plant & Animal Genome Conference XXVI, January 13-17, 2018, San Diego, CA.
13. National Agricultural Statistics Service (2018). Crop Values Annual Summary, 02.23.2018
14. Humann, J.L., Lee, T., Ficklin, S.P., Cheng, C-H., Hough, H., Jung, S., Wegrzyn, J.L., and D.B. Neale. 2017. Using GenSAS for Specialty Crop Community Genome Annotation. Proceedings of the ASHS Annual Conference; September 19-22, Waikoloa, HI.
15. Rife, T., Poland, J.A. (2018). Integrating Free Mobile Apps into Specialty Crop Breeding and Horticultural Programs. Proceedings of the ASHS Annual Conference; July 31-Aug 4, 2018, Washington, D.C
16. Buble, K., Yu, J., Jung, S., Humann, J., Cheng, C.H., Lee, T., Hough, H., McGaughey, D., Frank, M., Main, D. (2018). Using TripalMap for Genetics Research. Proceedings of the International Cotton Genome Initiative (ICGI) Research Conference, May 31 - June 4, 2018, Edinburgh, Scotland, United Kingdom.
17. Groß,A., Pruski,C. and Rahm,E. (2016) Evolution of biomedical ontologies and mappings: Overview of recent approaches.*Computational and Structural Biotechnology Journal*, **14**, 333–340.
18. Falk, T., Herndon, N., Grau, E., Buehler, S., Richter, P., Zaman, S., Baker, E. M., Ramnath, R., Ficklin, S., Staton, M., Feltus, F. A., Jung, S., Main, D., and Wegrzyn, J.L. (2018) Growing and cultivating the forest genomics database, TreeGenes. *Database (Oxford)*, bay084.
19. FAO et al. The State of Food Security and Nutrition in the World (FAO, Rome, 2017).
20. Deutsch, C.A., Tewksbury, J.J., Tigchelaar, M., Battisti, D.S., Merrill, S.C., Huey, R.B., and Naylor, R.L. (2018) Increase in crop losses to insect pests in a warming climate. *Science*, **261**(6405), 916-919.
21. Riegler, M. (2018). Insect threats to food security. *Science*, **361**(6405), 846.
22. Walthall CL, Hatfield J, Backlund P, Lengnick L, Marshall E, Walsh M, Adkins S, Aillery M, Ainsworth EA, et al. (2012) Climate Change and Agriculture in the United States: Effects and Adaptation. USDA Technical Bulletin 1935. Washington, DC. 186 pages.
23. Koeze, E (2017) 'How A Warm Winter Destroyed 85 Percent Of Georgia's Peaches', FiveThirtyEight, Filed under Local Climates, Published Sep. 14,2017 (<https://fivethirtyeight.com/features/how-a-warm-winter-destroyed-85-percent-of-georgias-peaches/>)
24. Chen,C., Pang,Y., Pan,X. and Zhang,L. (2015). Impacts of climate change on cotton yield in China from 1961 to 2010 based on provincial data *Journal of Meteorological Research*, **29**, 515–524.
25. Lipka,A.E., Tian,F., Wang,Q., Peiffer,J., Li,M., Bradbury,P.J., Gore,M.A., Buckler,E.S. and Zhang,Z. (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics*, **28** (18): 2397-9.
26. Hardner, C., Satish, K., Main, D., Hayes, B., Peace, C. (2018). Global Genomic Prediction of Performance. Proceedings of the 9th International Genomics Conference, June 26-30, 2018. Nanjing, China.
27. Wilkinson,M.D., Dumontier,M., Aalbersberg,I.J., Appleton,G., Axton,M., Baak,A., Blomberg,N., Boiten,J.-W., da Silva Santos,L.B., Bourne,P.E., et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, **3**, 160018.
28. Jung,S., Bassett,C., Bielenberg,D.G., Cheng,C.-H., Dardick,C., Main,D., Meisel,L., Slovin,J., Troglio,M. and Schaffer,R.J. (2015). A standard nomenclature for gene designation in the Rosaceae. *Tree Genetics & Genomes*, **11**:108.
29. Reiser, L., Subramaniam, S., Li, D., Huala, E. (2017) Using the Arabidopsis Information Resource (TAIR) to Find Information About Arabidopsis Genes *Current Protocols in Bioinformatics*, 8:60.
30. Dash, S., Campbell, J. D., Cannon, E. K. S., Cleary, A. M., Huang, W., Kalberer, S. R., Karingula, V., Rice, A. G., Singh, J., Umale, P. E., Weeks, N. T., Wilkey, A. P., Farmer, A. D., Cannon, S. B. (2016). Legume information system (LegumelInfo.org): a key component of a set of federated data resources for the legume

Outreach Plan

Outreach

Our outreach effort will focus on (1) crop communities that the databases serve; (2) communities who may be in need of Tripal databases; (3) the Tripal community; and (4) participation in the AgBioData consortium to increase: (a) standardization of protocols and practices across agricultural GGB databases to enhance GGB research outcomes, (b) researcher compliance in FAIR principles, (c) journal compliance with availability of peer-reviewed publication data, and (d) collaboration with funding agencies to help facilitate data deposit of meta, raw and analyzed data by funded researchers.

NRSP10 workshops will continue to be held yearly at the International Plant and Animal Genome (PAG), American Society of Horticultural Science, and Cotton Beltwide Conference annual meetings, as well as presentations in appropriate workshops at PAG, at the Crop, Soil and Agronomy annual meeting, and more crop specific meetings such as the International Rubes and Ribes, International Rosaceae Genomics, International Cotton Genome Initiative, and Citrus HLB Conferences. The NRSP10 workshops provide training that primarily involves stakeholder researchers demonstrating how they use component NRSP10 resources. We have found this to be a very effective method of training, with time for a discussion session to solicit feedback with core NRSP participation to keep the development aligned with stakeholder interests. Participants at the workshops and computer demonstration sessions are composed of researchers from many other crop communities, so presentations at PAG, in particular, provide good opportunity to present the utility of the Tripal platform and the support we provide in implementing or converting other databases.

In addition to presentations at conferences, we will also continue to update tutorials on how to use databases, host webinars and make them available online, maintain mailing lists and continue to publish when significant development has been made.

To reach database developers, Tripal will continue to be presented and training provided at yearly Tripal workshops at PAG and other meetings such as the GMOD, Galaxy and Bioinformatics conferences. We will continue to provide tutorials, mailing list responses, and Help Desk support in collaboration with other Tripal developers and update modules as new Drupal versions are released. We will continue to provide tutorials and publish when we produce new modules and applications.

Additionally, it is very important that breeders, stakeholders and the general public are trained about the usefulness of their community databases and functionalities that might be incorporated from other databases. We will add a public component to our website highlighting the utility of the databases in successful research stories that impact consumers, explaining the terminology at an appropriate level and including database training from the workshops. These extension and outreach activities will greatly enhance the research and extension programs of cotton, legumes and horticultural specialty crops. Using the database in the process of breeding superior plant selections will distinguish the United States amongst others in the world market for those crops. In addition, in association with the continued growth of cotton, legumes and horticultural specialty crop production, it will lead to more jobs with higher incomes, which in turn will create economic development and prosperity, enhancing the quality of life in rural areas.

We will continue to be good citizens of the AgBioData Consortium of Agricultural-related databases. This will include continuing to serve on the Steering Committee, helping to organize yearly workshops, engaging in sustainability studies, hosting the AgBioData website (<https://www.agbiodata.org>), participating in monthly meetings and specific workgroups, as well as adopting the recommendations of the WhitePaper (1), and applying for funding to further the research enabling mission of this group (currently a USDA NIFA FACT Coordinated Innovative Network proposal for AgBioData is pending – PI Main)

Communication

In addition to the workshops mentioned above, other conduits to facilitate communication between the database developers and the users are needed. While we envision that implementation of each database will have significant individuality, each database that associates with NRSP10 will have a steering committee, composed of representatives from universities, government and industry for each crop, which will meet quarterly or biannually by online teleconferences to communicate the current and emerging database needs of their research community with other stakeholders and the NRSP staff to guide the

development, implementation and dissemination of resources for the database. Any new major development will be extensively discussed in these committee meetings. The meeting minutes will be posted on the NRSP and crop-specific websites as well as all reports and work plans. We will also have regular newsletters, twitter and LinkedIn accounts to notify the users of any new developments in the database and the crop research community.

Assessment

Various metrics will be used to assess the impact of the proposed project. For each database this will include usage statistics as measured by google analytics, feedback from the steering committee, annual online surveys of each community, number of publications, number of publications citing the databases and feedback via the online forms in the databases. Creation of new Tripal databases, number of projects adopting Tripal, number of species being served by through Tripal databases, and number of active developers with this project will all be indicators of the success of the use of Tripal. Very relevant measures of participation will be the number of curators and crops associated with the NRSP and the level of investment in NRSP by the users.

Organization/Governance

Literature Cited

1. Harper, L., Campbell, J., Cannon, E.K.S., Jung, S., Poelchau, M., Walls, R., Andorf, C., Arnaud, E., Berardini, T.Z., Birkett, C. *et al.* (2018). AgBioData consortium recommendations for sustainable genomics and genetics databases for agriculture. *Database (Oxford)*, bay088
2. Sanderson, L.-A., Ficklin, S.P., Cheng, C.-H., Jung, S., Feltus, F.A., Bett, K.E., and Main, D. (2013). Tripal v1.1: a standards-based toolkit for construction of online genetic and genomic databases. *Database (Oxford)*, bat075.
3. Ficklin, S.P., Sanderson, L.-A., Cheng, C.-H., Staton, M.E., Lee, T., Cho, I.-H., Jung, S., Bett, K.E. and Main, D. (2011). Tripal: a construction toolkit for online genome databases. *Database (Oxford)*, bar044.
4. Jung, S., Lee, T., Yu, J., Ficklin, S.P., and Main, D. (2016). Chado use case: storing genomic, genetic and breeding data of Rosaceae and Gossypium crops in Chado. *Database (Oxford)*, baw010.
5. Jung,S., Lee,T., Cheng,C.H., Buble, K., Zheng, P., Yu,J., Humann,J., Ficklin,S., Gasic, K., Scott, K, Frank, M., Ru, S., Hough, H., Evans, K., Peace, C., Olmstead, M., DeVetter, L.W., McFerson, J., Coe, M., Wegrzyn, J.L., Staton, M.E., Abbott, A.G. and Main, D. (2018) 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. *Nucleic Acids Research*
6. Yu,J., Jung,S., Cheng,C.-H., Ficklin,S.P., Lee,T., Zheng,P., Jones,D., Percy,R.G. and Main, D. (2014) CottonGen: a genomics, genetics and breeding database for cotton research. *Nucleic Acids Research* **42**, D1229–D1236.
7. Humann, J.L., Piaskowski, J., Jung, S., Cheng, C.H., Lee, T., Frank, M., Scott, K., Zheng, P., Flores-Gonzales, M., Saha, S., et al. (2017). Resources in the Citrus Genome Database that enable basic, translational, and applied research. 5th International Research Conference on Huanglongbing: March 14-17, 2017, Orlando, FL, USA.
8. Jung,S., Humann,J., Cheng,C.H., Lee,T., Zheng, P., Frank, M., McGaughey, D., Scott, K., Buble, K., Yu, J., Hough, H., Coyne, C., McGee, R., Main, D. (2017b) Updates to the Cool Season Food Legume Genome Database: Resources for pea, lentil, faba bean and chickpea genetics, genomics and breeding. Proceedings of the North American Pulse Improvement Association Biannual Meeting: November 2017, East Lansing, MI, USA
9. Bassil, N., Jung, S., Cheng, C-H., Lee, T., Zheng, P., and Main, D. (2017). NRSP10 Resources for Small Fruit Research. Proceedings of the ASHS Annual Conference; September 19-22, Waikoloa, HI
10. Chen,M., Henry,N., Almsaeed,A., Zhou,X., Wegrzyn,J., Ficklin,S. and Staton,M. (2017). New extension software modules to enhance searching and display of transcriptome data in Tripal databases. *Database (Oxford)*, bax052
11. Jung,S., Lee,T., Cheng,C.-H., Ficklin,S., Yu,J., Humann,J. and Main,D. (2017) Extension modules for storage, visualization and querying of genomic, genetic and breeding data in Tripal databases. *Database c*
12. Jung, S., Lee, T., Cheng, C-H., Gasic, K., Campbell B.T., Main, D. (2018). Using the Tripal Breeding Information Management System (BIMS) to Enable Efficient Management of Phenotypic and Genotypic Data. Abstracts of the International Plant & Animal Genome Conference XXVI, January 13-17, 2018, San Diego, CA.
13. National Agricultural Statistics Service (2018). Crop Values Annual Summary, 02.23.2018
14. Humann, J.L., Lee, T., Ficklin, S.P., Cheng, C-H., Hough, H., Jung, S., Wegrzyn, J.L., and D.B. Neale. 2017. Using GenSAS for Specialty Crop Community Genome Annotation. Proceedings of the ASHS Annual Conference; September 19-22, Waikoloa, HI.
15. Rife, T., Poland, J.A. (2018). Integrating Free Mobile Apps into Specialty Crop Breeding and Horticultural Programs. Proceedings of the ASHS Annual Conference; July 31-Aug 4, 2018, Washington, D.C
16. Buble, K., Yu, J., Jung, S., Humann, J., Cheng, C.H., Lee, T., Hough, H., McGaughey, D., Frank, M., Main, D. (2018). Using TripalMap for Genetics Research. Proceedings of the International Cotton Genome Initiative (ICGI) Research Conference, May 31 - June 4, 2018, Edinburgh, Scotland, United Kingdom.
17. Groß,A., Pruski,C. and Rahm,E. (2016) Evolution of biomedical ontologies and mappings: Overview of recent approaches.*Computational and Structural Biotechnology Journal*, **14**, 333–340.
18. Falk, T., Herndon, N., Grau, E., Buehler, S., Richter, P., Zaman, S., Baker, E. M., Ramnath, R., Ficklin, S., Staton, M., Feltus, F. A., Jung, S., Main, D., and Wegrzyn, J.L. (2018) Growing and cultivating the forest genomics database, TreeGenes. *Database (Oxford)*, bay084.

19. FAO et al. The State of Food Security and Nutrition in the World (FAO, Rome, 2017).
20. Deutsch, C.A., Tewksbury, J.J., Tigchelaar, M., Battisti, D.S., Merrill, S.C., Huey, R.B., and Naylor, R.L. (2018) Increase in crop losses to insect pests in a warming climate. *Science*, **261**(6405), 916-919.
21. Riegler, M. (2018). Insect threats to food security. *Science*, **361**(6405), 846.
22. Walthall CL, Hatfield J, Backlund P, Lengnick L, Marshall E, Walsh M, Adkins S, Aillery M, Ainsworth EA, et al. (2012) Climate Change and Agriculture in the United States: Effects and Adaptation. USDA Technical Bulletin 1935. Washington, DC. 186 pages.
23. Koeze, E (2017) 'How A Warm Winter Destroyed 85 Percent Of Georgia's Peaches', FiveThirtyEight, Filed under Local Climates, Published Sep. 14, 2017 (<https://fivethirtyeight.com/features/how-a-warm-winter-destroyed-85-percent-of-georgias-peaches/>)
24. Chen,C., Pang,Y., Pan,X. and Zhang,L. (2015). Impacts of climate change on cotton yield in China from 1961 to 2010 based on provincial data *Journal of Meteorological Research*, **29**, 515–524.
25. Lipka,A.E., Tian,F., Wang,Q., Peiffer,J., Li,M., Bradbury,P.J., Gore,M.A., Buckler,E.S. and Zhang,Z. (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics*, **28** (18): 2397-9.
26. Hardner, C., Satish, K., Main, D., Hayes, B., Peace, C. (2018). Global Genomic Prediction of Performance. Proceedings of the 9th International Genomics Conference, June 26-30, 2018. Nanjing, China.
27. Wilkinson,M.D., Dumontier,M., Aalbersberg,I.J., Appleton,G., Axton,M., Baak,A., Blomberg,N., Boiten,J.-W., da Silva Santos,L.B., Bourne,P.E., et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, **3**, 160018.
28. Jung,S., Bassett,C., Bielenberg,D.G., Cheng,C.-H., Dardick,C., Main,D., Meisel,L., Slovin,J., Troglio,M. and Schaffer,R.J. (2015). A standard nomenclature for gene designation in the Rosaceae. *Tree Genetics & Genomes*, **11**:108.
29. Reiser, L., Subramaniam, S., Li, D., Huala, E. (2017) Using the Arabidopsis Information Resource (TAIR) to Find Information About Arabidopsis Genes. *Current Protocols in Bioinformatics*, 8:60.
30. Dash, S., Campbell, J. D., Cannon, E. K. S., Cleary, A. M., Huang, W., Kalberer, S. R., Karingula, V., Rice, A. G., Singh, J., Umale, P. E., Weeks, N. T., Wilkey, A. P., Farmer, A. D., Cannon, S. B. (2016). Legume information system (LegumelInfo.org): a key component of a set of federated data resources for the legume family. *Nucleic Acids Research*, 44(D1), D1181–D1188.
31. Poel, M., Childers, C., Moore, G., Tsavatapalli, V., Evans, J., Lee, C.-Y., Lin, H., Lin, J.W. Hacket, K. (2015). The i5k Workspace@NAL--enabling genomic data access, visualization and curation of arthropod genomes. *Nucleic Acids Research*, 43(Database issue), D714–D719. <https://doi.org/10.1093/nar/gku983>

Land Grant Participating States/Institutions

SC,GA,MI,TN,WA,NC

Non Land Grant Participating States/Institutions

Pacific West Area

Participation

Participant	Is Head	Station	Objective	Research						Extension	
				KA	SOI	FOS	SY	PY	TY	FTE	KA
Bassil, Nahla V		Pacific West Area	1	136	1115	1080	1.00	0.00	0.00	0	0
				201	1120	1081					
				202	1121	0					
				0	1122	0					
				0	1123	0					
				0	1129	0					
				0	1199	0					
				0	1210	0					
Evans, Kate M		Washington - Washington State University	1,2	201	1110	1081	0.10	0.00	0.00	0	0
				201	1115	1081					
Gasic, Ksenija		South Carolina - Clemson University	2,3	201	1114	1081	0.30	0.00	0.30	0.3	201
Iezzoni, Amy	Yes	Michigan - Michigan State University	1,2	201	1110	1080	0.10	0.00	0.00	0	0
Iorizzo, Massimo		North Carolina - North Carolina State University	1,2,3,4,5	201	1452	1081	0.00	0.00	0.00	0.03	201
				201	1010	1080					
				201	1020	1080					
				201	1120	1081					

Participant	Is Head	Station	Objective	Research						Extension	
				KA	SOI	FOS	SY	PY	TY	FTE	KA
Jackson, Scott	Yes	Georgia - University of Georgia	1,2,3,4	201 201 201	1419 1830 1820	1080 1080 1080	0.10	0.00	0.00	0	0
Kahn, Michael L		Washington - Washington State University	5	202	1640	1080	0.10	0.00	0.00	0	0
Milla-Lewis, Susana R		North Carolina - North Carolina State University		201 202 0	1620 1621 2130	1081 0 0	1.00	1.00	2.00	0	0
Missaoui, Ali M		Georgia - University of Georgia	1	201	910	1040	0.10	0.00	0.00	0	0
Schwartz, Brian M		Georgia - University of Georgia	1	201	910	1040	0.10	0.00	0.00	0	0
Staton, Margaret E		Tennessee - University of Tennessee	2,5	201	699	1040	0.05	0.05	0.00	0	0
van Nocker, Steven	Yes	Michigan - Michigan State University	1	205	1199	1050	0.10	0.00	0.00	0	0

Combined Participation

Combination of KA, SOI and FOS	Total SY	Total PY	Total TY
201-1114-1081	0.3	0	0.3
201-1419-1080	0.03	0	0
201-1820-1080	0.03	0	0
201-1830-1080	0.03	0	0
201-910-1040	0.1	0	0
201-910-1040	0.1	0	0
201-1110-1080	0.1	0	0
205-1199-1050	0.1	0	0
201-699-1040	0.05	0.05	0
202-1640-1080	0.1	0	0
201-1110-1081	0.05	0	0
201-1115-1081	0.05	0	0
0-1122-0	0.13	0	0
0-1123-0	0.13	0	0
0-1129-0	0.13	0	0
0-1199-0	0.13	0	0
0-1210-0	0.13	0	0
136-1115-1080	0.13	0	0
201-1120-1081	0.13	0	0
202-1121-0	0.13	0	0
201-1010-1080	0	0	0
Grand Total:	3.05	1.05	2.30

Combination of KA, SOI and FOS	Total SY	Total PY	Total TY
201-1020-1080	0	0	0
201-1120-1081	0	0	0
201-1452-1081	0	0	0
0-2130-0	0.33	1	2
201-1620-1081	0.33	1	2
202-1621-0	0.33	1	2
Grand Total:	3.05	1.05	2.30

Program/KA	Total FTE
201	0.1
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0
201	0.01
0	0
Grand FTE Total:	0.33

Budgets

MRF Funding 2019

Description	Dollars	FTE
Salaries	394140.00	4.65
Fringe Benefits	0.00	0.00
Wages	0.00	0.00
Travel	20000.00	0.00
Supplies	19000.00	0.00
Maintenance	15000.00	0.00
Equipment / Capital Improvement	40000.00	0.00
Totals	488140	4.65

Comments

Total = 488,140.00 / FTE = 4.65 Salaries: In year 1 salary funds of \$394,140 are requested for 4 full-time developers, 50% for a Database Administrator/System Administrator (Heidi Hough), 15% PI Main salary for each year of the project. Three of the developers (Taein Lee, Chun-Huai Cheng, and Katheryn Buble will work on Tripal module development and contribute to help desk support, while the 4th developer, Shawna Spoor will continue Tripal core development and help desk support. Travel: In year 1 travel funds of \$20,000 are requested present activities at conferences/meetings for all 5 crop database communities and Tripal related conferences, travel to meet with stakeholders for individual training on using BIMS, and attend yearly Tripal hackathons Supplies: In year 1 supply funds of \$19,000 are requested. \$15,000 is a consultancy payment for Phoenix Bioinformatics sustainability modeling and \$4000 for 2 open access publications. Maintenance: In year 1 maintenance funds of \$15,000 are requested to pay for server room fees, service contracts, data storage and onsite and offsite backup. Equipment/Capital Improvement: In year 1 equipment funds of \$35,000 are requested to replace database and web servers and \$2,000 for 2 new PCs/Laptops for 2 developers.

MRF Funding 2020

Description	Dollars	FTE
Salaries	409906.00	4.65
Fringe Benefits	0.00	0.00
Wages	0.00	0.00
Travel	20000.00	0.00
Supplies	19000.00	0.00
Maintenance	15000.00	0.00
Equipment / Capital Improvement	2000.00	0.00
Totals	465906	4.65

Comments

Total = 465,906.00 / FTE = 4.65 Salaries: In year 2 salary of \$ 409,906 is requested for 4 full-time developers, 50% for a Database Administrator/System Administrator (Heidi Hough), 15% PI Main salary for each year of the project. This includes a 4% yearly increase as mandated by WSU. Three of the developers (Taein Lee, Chun-Huai Cheng, and Katheryn Buble will work on Tripal module development and contribute to help desk support, while the 4th developer, Shawna Spoor will continue Tripal core development and help desk support. Travel: In year 2 travel funds of \$20,000 are requested present activities at conferences/meetings for all 5 crop database communities and Tripal related conferences, travel to meet with stakeholders for individual training on using BIMS, and attend yearly Tripal hackathons. Supplies: In year 2 supply funds of \$19,000 are requested. \$15,000 is for pay for continuing consultancy fees for Phoenix Bioinformatics Sustainability modeling and \$4000 for 2 open access publications. Maintenance: In year 2 maintenance funds of \$15,000 are requested to pay for server room fees, service contracts, data storage and onsite and offsite backup. Equipment/Capital Improvement: In year 2 equipment funds of \$2,000 are requested 2 PCs/Laptops for 2 developers.

MRF Funding 2021

Description	Dollars	FTE
Salaries	426302.00	4.65
Fringe Benefits	0.00	0.00
Wages	0.00	0.00
Travel	20000.00	0.00
Supplies	9000.00	0.00
Maintenance	15000.00	0.00
Equipment / Capital Improvement	40000.00	0.00
Totals	510302	4.65

Comments

Total = 510,302.00 / FTE = 4.65 Salaries: In year 3 salary of \$ 426,302 is requested for 4 full-time developers, 50% for a Database Administrator/System Administrator (Heidi Hough), 15% PI Main salary for each year of the project. This includes a 4% yearly increase as mandated by WSU. Three of the developers (Taein Lee, Chun-Huai Cheng, and Katheryn Buble will work on Tripal module development and contribute to help desk support, while the 4th developer, Shawna Spoor will continue Tripal core development and help desk support. Travel: In year 3 travel funds of \$20,000 are requested present activities at conferences/meetings for all 5 crop database communities and Tripal related conferences, travel to meet with stakeholders for individual training on using BIMS, and attend yearly Tripal hackathons. Supplies: In year 3 supply funds of \$9,000 are requested. \$5,000 is to for pay for continuing consultancy fees for Phoenix Bioinformatics to help with pilot implementation of sustainability model/or costs associated with this implementation and \$4000 for 2 open access publications. Maintenance: In year 3 maintenance funds of \$15,000 are requested to pay for server room fees, service contracts, data storage and onsite and offsite backup. Equipment/Capital Improvement: In year 3 equipment funds of \$35,000 are requested to replace database and web servers.

MRF Funding 2022

Description	Dollars	FTE
Salaries	443353.00	4.65
Fringe Benefits	0.00	0.00
Wages	0.00	0.00
Travel	20000.00	0.00
Supplies	9000.00	0.00
Maintenance	15000.00	0.00
Equipment / Capital Improvement	2000.00	0.00
Totals	489353	4.65

Comments

Total = 489,353.00 / FTE = 4.65 Salaries: In year 4 salary of \$ 443353 is requested for 4 full-time developers, 50% for a Database Administrator/System Administrator (Heidi Hough), 15% PI Main salary for each year of the project. This includes a 4% yearly increase as mandated by WSU. Three of the developers (Taein Lee, Chun-Huai Cheng, and Katheryn Buble will work on Tripal module development and contribute to help desk support, while the 4th developer, Shawna Spoor will continue Tripal core development and help desk support. Travel: In year 4 travel funds of \$20,000 are requested present activities at conferences/meetings for all 5 crop database communities and Tripal related conferences, travel to meet with stakeholders for individual training on using BIMS, and attend yearly Tripal hackathons. Supplies: In year 4 supply funds of \$9,000 are requested. \$5,000 is to for pay for continuing consultancy fees for Phoenix Bioinformatics to help with pilot implementation of sustainability model/or costs associated with this implementation and \$4000 for 2 open access publications. Maintenance: In year 4 maintenance funds of \$15,000 are requested to pay for server room fees, service contracts, data storage and onsite and offsite backup. Equipment/Capital Improvement: In year 4 equipment funds of \$2,000 are requested 2 PCs/Laptops for 2 developers.

MRF Funding 2023

Description	Dollars	FTE
Salaries	461088.00	4.65
Fringe Benefits	0.00	0.00
Wages	0.00	0.00
Travel	20000.00	0.00
Supplies	4000.00	0.00
Maintenance	15000.00	0.00
Equipment / Capital Improvement	2000.00	0.00
Totals	502088	4.65

Comments

Total = 502,088.00 / FTE = 4.65 Salaries: In year 5 salary of \$ 461088 is requested for 4 full-time developers, 50% for a Database Administrator/System Administrator (Heidi Hough), 15% PI Main salary for each year of the project. This includes a 4% yearly increase as mandated by WSU. Three of the developers (Taein Lee, Chun-Huai Cheng, and Kathryn Buble will work on Tripal module development and contribute to help desk support, while the 4th developer, Shawna Spoor will continue Tripal core development and help desk support. Travel: In year 5 travel funds of \$20,000 are requested present activities at conferences/meetings for all 5 crop database communities and Tripal related conferences, travel to meet with stakeholders for individual training on using BIMS, and attend yearly Tripal hackathons. Supplies: In year 5 supply funds of \$4,000 are requested for 2 open access publications. Maintenance: In year 5 maintenance funds of \$15,000 are requested to pay for server room fees, service contracts, data storage and onsite and offsite backup. Equipment/Capital Improvement: In year 5 equipment funds of \$2,000 are requested 2 PCs/Laptops for 2 developers.

Table 1: NRSP10 Crop Database Summary

Database	Crops Serve	Production value in 2016* (\$billion)	2015-2018 Usage by community** <i>Number of citations***</i>
Genome Database for Rosaceae www.rosaceae.org	almond, apple, apricot, blackberry, cherry, nectarine, peach, pear, plum, raspberry, rose, strawberry	15.45	Users = 81,334 Countries = 185 Visits = 212,256 Pages = 1,390,627 <i>Citations = 707</i>
Citrus Genome Database www.citrusgenomedb.org	grapefruit, lemon, lime, orange, tangelo, tangerine/ mandarin	3.43	Users = 19,760 Countries = 160 Visits = 32,789 Pages = 179,987 <i>Citations = 160</i>
Cool Season Food Legume Genome Database www.coolseasonfoodlegume.org	chickpea, fava bean, lentil, pea	0.84	Users = 12,797 Countries = 151 Visits = 21,556 Pages = 119,389 <i>Citations = 151</i>
CottonGen www.cottongen.org	cotton	5.81	Users = 40,737 Countries = 160 Visits = 32,789 Pages = 704,944 <i>Citations = 222</i>
Genome Database for <i>Vaccinium</i> www.vaccinium.org	blueberry, cranberry	1.04	Users = 7,518 Countries = 111 Visits = 11,272 Pages = 61,494 <i>Citations = 111</i>
Total	25 crops	26.57	Users = 162,146 Countries = 185 Visits = 387,913 Pages = 2,456,411 <i>Citations = 1059</i>

* 2016 Crop Value of Production as reported by the National Agricultural Statistics Service 2017

** As reported by Google Analytics

*** As reported by Google Scholar from January 1, 2015 to October 30, 2018

Table 2: NRSP10 Crop Database Data Summary (Oct 15, 2018)

Data Type	NRSP10 Database				
	Rosaceae	Cotton	Citrus	Legumes	Vaccinium
Gene	597226	258462	243066	0	0
Genome	22	7	19	3	1
Genotype	10,806,569	64775	0	?	0
Germplasm	14,411	17640	1426	2429	355
Map	316	109	70	167	10
Marker	3,286,882	573381	48154	137268	4253
mRNA	659,987	503516	448144	85359	1157
Pathway	1797	750	901	673	0
Phenotype	389,191	383431	23070	0	429
Polypeptide	155,001	154,213	67598	90,216	0
Publication	7,515	16852	4045	4988	477
QTL/MTL	3,902	4923	579	3009	26
Reference Transcript	259,213	214,180	33,111	191,154	39,461
RNASeq read	9.31 billion	5.85 billion	1.2 billion	5.11 billion	757 million
Species	1967	73	67	33	15

Cost	Year 1	Year 2	Year 3	Year 4	Year 5
Salaries	241,142	245,888	240,401	245,019	124,872
Benefits	177,049	182,414	188,149	193,948	156,232
Travel	6,000	7000	8000	9000	0
Supplies	4548	7148	7214	9959	0
Maintenance	4000	4000	4000	4000	0
Total	432,739	446,450	447,764	461,926	281,104

Description

Salaries from WSU, Cotton Incorporated, SAEDS, Bayer Crop Science, Corteva, USA Dry pea and Lentil Council, Northern Pulse Growers, Washington Tree Fruit Research Commission, WA SCRI Block Grant (2.5 FTE). Benefits from WSU, Cotton Incorporated, SAEDS, Bayer Crop Science, Corteva, USA Dry Pea and Lentil Council, Northern Pulse Growers, Washington Tree Fruit Research Commission, WA SCRI Block Grant (6.7 FTE). Travel from Cotton Incorporated, SAEDS, Bayer Crop Science, Corteva.

Supplies from Cotton Incorporated, SAEDS, Bayer Crop Science, Corteva, USA Dry Pea and Lentil Council, Northern Pulse Growers, Washington Tree Fruit Research Commission, WA SCRI Block Grant.

Maintenance from Cotton Incorporated, SAEDS, Bayer Crop Science, Corteva, USA Dry Pea and Lentil Council, Northern Pulse Growers, Washington Tree Fruit Research Commission, WA SCRI Block Grant.

NRSP Review (Submitted)

NRSP Peer / Proposal Review Form

Project Number: National Database Resources for Crop Genomics, Genetics and Breeding Research

The following statement defines the mission of the National Research Support Projects (NRSP's):

"The activity of an NRSP focuses on the development of enabling technologies, support activities (such as to collect, assemble, store, and distribute materials, resources and information), or the sharing of facilities needed to accomplish high priority research, but which is not of itself primarily research."

Based on this mission statement, please rate the proposed NRSP using the following criteria:

Mission:

Consistency with the mission of an NRSP Excellent ▼

Relevance:

Addresses and supports a high priority national issue Excellent ▼

Demonstrates clear/tangible benefit to the scientific community as a whole Excellent ▼

Clearly identified sponsoring "stakeholders" /beneficiaries Excellent ▼

"Stakeholder" involvement in project development, project activities, review and/or management plans Excellent ▼

Technical Merit:

Overall technical merit (sound scientific approach, achievable objectives, appropriate scope of activity) Excellent ▼

Potential for significant outputs (products) and outcomes and/or impacts Excellent ▼

Implementation Plan:

Benchmarks for success clearly identified Excellent ▼

Management structure that adequately coordinates efforts of multiple participants Excellent ▼

Well developed business plan that links multiple sources of funding and leverages limited off-the-top research funds Excellent ▼

Funding plan that develops alternative funding sources to reduce off-the-top funding in future years Excellent ▼

Efforts integrated with extension and/or academic programs Excellent ▼

Outreach, communications and assessment plan that communicates the programs goals, accomplishments and outcomes/impacts Excellent ▼

Regional Committee or AA Recommendation:

Comments:

The goals of the project are very well aligned with the mission statement. The GDR framework, housed within Tripal, will leverage recent developments to continue to provide a robust platform for discovery. The project and associated team are experienced and the goals are aligned with researcher and breeder needs. Their goals include building, expanding, and implementing tools that will bring analytics to the data platform. They are leveraging multiple sources of funding and are moving towards a sustainable platform that is applicable across systems.

Recommendation:

Approve ▼

 Report a Bug

NRSP Review (Submitted)

NRSP Peer / Proposal Review Form

Project Number: National Database Resources for Crop Genomics, Genetics and Breeding Research

The following statement defines the mission of the National Research Support Projects (NRSP's):

"The activity of an NRSP focuses on the development of enabling technologies, support activities (such as to collect, assemble, store, and distribute materials, resources and information), or the sharing of facilities needed to accomplish high priority research, but which is not of itself primarily research."

Based on this mission statement, please rate the proposed NRSP using the following criteria:

Mission:

Consistency with the mission of an NRSP Excellent ▼

Relevance:

Addresses and supports a high priority national issue Excellent ▼

Demonstrates clear/tangible benefit to the scientific community as a whole Excellent ▼

Clearly identified sponsoring "stakeholders" /beneficiaries Excellent ▼

"Stakeholder" involvement in project development, project activities, review and/or management plans Excellent ▼

Technical Merit:

Overall technical merit (sound scientific approach, achievable objectives, appropriate scope of activity) Good ▼

Potential for significant outputs (products) and outcomes and/or impacts Good ▼

Implementation Plan:

Benchmarks for success clearly identified Excellent ▼

Management structure that adequately coordinates efforts of multiple participants Good ▼

Well developed business plan that links multiple sources of funding and leverages limited off-the-top research funds Good ▼

Funding plan that develops alternative funding sources to reduce off-the-top funding in future years Excellent ▼

Efforts integrated with extension and/or academic programs Excellent ▼

Outreach, communications and assessment plan that communicates the programs goals, accomplishments and outcomes/impacts Excellent ▼

Regional Committee or AA Recommendation:

Comments:

The project proposal was written in an excellent way and it describes the local and National needs on building a user-friendly informatics platform for the plant breeding and genomics community. The project plans to develop modules via the Tripal platform by adding, APIs, GenSAS, tablet version of UI, support major US crops and Breeding Information Management System. It is a big undertaking and the project team is well rounded to deliver on the proposed items. As of now the system works on storage, curation and analytics. However, it does not address on sharing the community's collected data for sharing with the National and International archives in standard formats, eg the GBS, genotyping and Phenotyping data to be submitted to EBI-EVA (mandated fro crops and plants instead of dbSNP), submissions to SRA, submission of genome assemblies and annotation to INSDC, Uniprot and Planteome. This aspect is a must to make sure the longevity of the data access is maintained and allows other international projects to intake data automatically from the major archives instead of coming to Tripal. Authors need to address how they plan on updating the legacy data to the new versions of genomes and

[Report a Bug](#)

annotations to help breeders always work with the most updated and recent annotations. Work with the GOBII and CGIAR Breeding platform portals already developed for integration/training into Tripal instead of developing a new one. Feasibility of the bringing a non-profit organization Phoenix Bioinformatics is a big distraction and suggest avoiding. It also sets a wrong precedence that in future all datasets and platforms have to go that route. The stake holders, public institutions and education courses for training and developing skills in future workforce needs free access and development. Overall project sounds great and shows use of community standards, but there is little investment for continued participation, contribution and improving these standards. Suggest the agencies to reach out to the partners for contributing and supporting other projects developing standards so that this project can also benefit.

Recommendation:

Approve ▼

[Back \(https://www.nimss.org/review/for_project/18619\)](https://www.nimss.org/review/for_project/18619)

 **Report a Bug**

NRSP Review (Submitted)

NRSP Peer / Proposal Review Form

Project Number: National Database Resources for Crop Genomics, Genetics and Breeding Research

The following statement defines the mission of the National Research Support Projects (NRSP's):

"The activity of an NRSP focuses on the development of enabling technologies, support activities (such as to collect, assemble, store, and distribute materials, resources and information), or the sharing of facilities needed to accomplish high priority research, but which is not of itself primarily research."

Based on this mission statement, please rate the proposed NRSP using the following criteria:

Mission:

Consistency with the mission of an NRSP Excellent ▼

Relevance:

Addresses and supports a high priority national issue Excellent ▼

Demonstrates clear/tangible benefit to the scientific community as a whole Excellent ▼

Clearly identified sponsoring "stakeholders" /beneficiaries Excellent ▼

"Stakeholder" involvement in project development, project activities, review and/or management plans Excellent ▼

Technical Merit:

Overall technical merit (sound scientific approach, achievable objectives, appropriate scope of activity) Excellent ▼

Potential for significant outputs (products) and outcomes and/or impacts Excellent ▼

Implementation Plan:

Benchmarks for success clearly identified Excellent ▼

Management structure that adequately coordinates efforts of multiple participants Excellent ▼

Well developed business plan that links multiple sources of funding and leverages limited off-the-top research funds Excellent ▼

Funding plan that develops alternative funding sources to reduce off-the-top funding in future years Excellent ▼

Efforts integrated with extension and/or academic programs Excellent ▼

Outreach, communications and assessment plan that communicates the programs goals, accomplishments and outcomes/impacts Excellent ▼

Regional Committee or AA Recommendation:

Comments:

This is simply an outstanding project, and on all fronts. The mission and implementation of this project are vital to US global competitiveness in agriculture, central to the US research community, and important to community building and in fostering communication between different research groups and stakeholders.

Recommendation:

Approve ▼

 Report a Bug

[Back \(https://www.nimss.org/review/for_project/18619\)](https://www.nimss.org/review/for_project/18619)

 [Report a Bug](#)

NRSP Review (Submitted)

NRSP Peer / Proposal Review Form

Project Number: National Database Resources for Crop Genomics, Genetics and Breeding Research

The following statement defines the mission of the National Research Support Projects (NRSP's):

"The activity of an NRSP focuses on the development of enabling technologies, support activities (such as to collect, assemble, store, and distribute materials, resources and information), or the sharing of facilities needed to accomplish high priority research, but which is not of itself primarily research."

Based on this mission statement, please rate the proposed NRSP using the following criteria:

Mission:

Consistency with the mission of an NRSP Excellent ▼

Relevance:

Addresses and supports a high priority national issue Excellent ▼

Demonstrates clear/tangible benefit to the scientific community as a whole Excellent ▼

Clearly identified sponsoring "stakeholders" /beneficiaries Excellent ▼

"Stakeholder" involvement in project development, project activities, review and/or management plans Excellent ▼

Technical Merit:

Overall technical merit (sound scientific approach, achievable objectives, appropriate scope of activity) Excellent ▼

Potential for significant outputs (products) and outcomes and/or impacts Excellent ▼

Implementation Plan:

Benchmarks for success clearly identified Excellent ▼

Management structure that adequately coordinates efforts of multiple participants Excellent ▼

Well developed business plan that links multiple sources of funding and leverages limited off-the-top research funds Excellent ▼

Funding plan that develops alternative funding sources to reduce off-the-top funding in future years Excellent ▼

Efforts integrated with extension and/or academic programs Excellent ▼

Outreach, communications and assessment plan that communicates the programs goals, accomplishments and outcomes/impacts Excellent ▼

Regional Committee or AA Recommendation:

Comments:

This proposal is absolutely critical to the advancement of research in agriculture. The PI's do an excellent job of conveying and translating previous successes of the developed tools, platforms, databases to the present and future needs of our planet. The success of the communities of scientists, breeders, etc. is critically dependent on being able to quickly and efficiently interface with legacy data. GenBan and other non-crop specific databases differ in scale and ease of use and are simply not useful to query data. Having a key PI (Dr. Main) coordinate the development of key computational tools and databases establishes a common platform and data format which makes it so much easier to perform larger scale analyses among species. These formats and capabilities should eventually become a standard for genetic/genomic data. Furthermore, this proposal addresses the complexity of bioinformatic tools and relieves the standard research from the necessity of complex command line and increases capability through tool development. The objectives are clear and well linked

 Report a Bug

together. There must be a mechanism of persistence for these much needed resources.

Recommendation:

Approve ▼

[Back \(https://www.nimss.org/review/for_project/18619\)](https://www.nimss.org/review/for_project/18619)

 [Report a Bug](#)

NRSP Review (Submitted)

NRSP Peer / Proposal Review Form

Project Number: National Database Resources for Crop Genomics, Genetics and Breeding Research

The following statement defines the mission of the National Research Support Projects (NRSP's):

"The activity of an NRSP focuses on the development of enabling technologies, support activities (such as to collect, assemble, store, and distribute materials, resources and information), or the sharing of facilities needed to accomplish high priority research, but which is not of itself primarily research."

Based on this mission statement, please rate the proposed NRSP using the following criteria:

Mission:

Consistency with the mission of an NRSP Excellent ▼

Relevance:

Addresses and supports a high priority national issue Excellent ▼

Demonstrates clear/tangible benefit to the scientific community as a whole Excellent ▼

Clearly identified sponsoring "stakeholders" /beneficiaries Good ▼

"Stakeholder" involvement in project development, project activities, review and/or management plans Excellent ▼

Technical Merit:

Overall technical merit (sound scientific approach, achievable objectives, appropriate scope of activity) Excellent ▼

Potential for significant outputs (products) and outcomes and/or impacts Excellent ▼

Implementation Plan:

Benchmarks for success clearly identified Excellent ▼

Management structure that adequately coordinates efforts of multiple participants Excellent ▼

Well developed business plan that links multiple sources of funding and leverages limited off-the-top research funds Excellent ▼

Funding plan that develops alternative funding sources to reduce off-the-top funding in future years Good ▼

Efforts integrated with extension and/or academic programs Excellent ▼

Outreach, communications and assessment plan that communicates the programs goals, accomplishments and outcomes/impacts Excellent ▼

Regional Committee or AA Recommendation:

Comments:

This is a very important project that addresses a need for which public sector institutions have not otherwise met. The ongoing work is exceptional in providing stakeholders, well beyond those noted in the proposal, with a foundation for database development. The ongoing effort is of very high quality in both developing sound, very functional database resources well capable of handling the very large and rapidly growing datasets being generated, as well as being very approachable, easy-to-use resources for those users with all levels of expertise. Funding for this effort is strong now and will be for the near future. The longterm challenge will be to sustain funding into the distant future when database resources progress into the ranks of other infrastructure, but that is not a criticism for this project, but rather a larger concern for the future.

 Report a Bug

Recommendation:

Approve

[Back \(https://www.nimss.org/review/for_project/18619\)](https://www.nimss.org/review/for_project/18619)

[🐛 Report a Bug](#)

10 Dec 2018

Dr. Dorrie Main
Dept of Horticulture
Washington State University
Pullman, WA 99164-6414

Dear Dorrie:

I wish to express my organization's strongest possible support for renewal of your NRSP project: "National Database Resources for Crop Genomics, Genetics and Breeding Research". The U.S. rosaceous crop industry is quite diverse, including such crops as almond, apple, blackberry, peach, pear, plum, raspberry, rose, strawberry, sweet cherry, and tart cherry, with commercial production areas across the entire U.S. and local production of some importance in every single state. Despite such apparent agricultural and geographic diversity, each of those crops has a common strong interest in plant improvement via genomics, genetics, and breeding. Fortunately, each has benefitted immensely for the work you and your team have performed in the past the years to develop and build up the Genome Database for Rosaceae (GDR), as well as the work performed in the NRSP10 project. The same is true for the many other ag commodities that have become part of your work at the GDR and the NRSP10.

Now, we can see that stakeholders in these industries are benefitting from both these projects -- private and public investments have really begun to pay off. Our rosaceous crop research community now has access to high quality, well-curated genomic, genetic, and breeding data as well as the cutting-edge bioinformatics tools you and your team have developed. The GDR is now the acknowledged international resource for our crop groups. This is certainly an endorsement by your peers that the GDR is a solid scientific accomplishment. NRSP10 has significantly enhanced its growth and impact.

Looking back, it has been very encouraging to see the move from crop-specific information repositories to a vital and dynamic resource that helps translate massive DNA datasets into DNA information useful in our research communities. On this strong base, we foresee the more efficient and effective development and commercialization of superior new rootstock and scion cultivars. In addition, we strongly support the new resources GDR and NRSP10 offer to scientists working in genomics and genetics and many other biological disciplines.

The renewal proposal continues in its quest to procure the necessary funding to not just maintain, but to further enhance the GDR and its portfolio of tools. We are far beyond the “good old days” when simply sequencing part of a genome was foundational work and highly fundable by federal agencies. Further, federal funding for ag research is facing uncertain times. It seems obvious that we must vigorously pursue collaboration among stakeholder organizations and research communities of disparate crop groups to acquire necessary long-term support, as you have been doing in this project.

The separation of crop- specific and core research activities provides an opportunity for industry and research stakeholders to help set priorities and provide funding for bioinformatics work with outcomes most relevant to one or only a few crops. On the other hand, core research activities that build and enhance bioinformatics knowledge and applications of more general utility have worked well in your NRSP project and merit consideration for renewal by SAES Directors.

For example, your work has led to software platforms and protocols useful in all the crop groups involved in this proposal. The work you and your colleagues have led on Tripal software has paid off. Many existing databases, including GDR, have been converted to Tripal. Bioinformatics tools developed in one crop group have been applied to other crop groups. This is evidence your approach is both scientifically exciting and financially practical.

We cannot guess how funding for ag research will evolve, but we look forward to working with you to develop a funding model for the GDR that at least gives us a chance to maintain your momentum and create a hybrid model that offers our crop group and research communities the opportunity to work with a stable, robust, world-leading resource. Your business plan and two-level strategy seems logical and practical.

We continue to strongly support the work you propose. Should it be funded, we commit to working with you to advance both core and crop-specific activities. We believe it is reasonable and fair to expect industry and research stakeholders to participate in directly supporting the work you and your team propose. We also hope the SAES Directors give this renewal careful consideration and the opportunity to contribute to the mission of the NRSP.

Sincerely,



Ines Hanrahan, PhD
Executive Director
Washington Tree Fruit Research Commission

December 14, 2018

Dr. Dorrie Main
Professor of Bioinformatics
Washington State University
Pullman, WA 99164



Dear Dorrie,

The U.S. Rosaceae Executive Committee (RosEXEC) would like to offer our organizations' strongest support for your NRSP10 proposal renewal "National Database Resources for Crop Genomics, Genetics and Breeding Research".

The U.S. RosEXEC is an elected body of scientists, across disciplines, which represent the interests of U.S. Rosaceae academia, government and industry. It also includes three international liaisons. Our mission is to (1) serve as a communication and coordination focal point for the US Rosaceae genomics, genetics and breeding community (2) define research priorities based on input from the industry and research community (3) facilitate scientific interaction and foster dynamic research teams (4) promote research priorities and (5) coordinate educational efforts from the research community to the industry and the public. Rosaceae crops continue to grow in economic impact and consumer consumption. In 2016, crops from this important crop family contributed \$15.5 billion (value of production) to the US economy.

The oldest of the NRSP10 databases, the Genome Database for Rosaceae (GDR) has been an essential resource for facilitating Rosaceae research and helping to build our community into a well-organized, collegial body of scientists. Its continued development and support is critical to house the immense volume of genomic, genetic and breeding data the community continues to generate at an unprecedented scale through both large projects and individual laboratories. Support for the continuation of GDR is one of our highest priorities; it has served us very well for fifteen years both as a research resource and a communication portal. This includes hosting the annual RosEXEC elections through the website and the 2nd and 7th International Rosaceae Genomics Conference.

Since 2014, the *RosEXEC has served as the official advisory board for GDR*. In this role we advise on priorities, inform on new data and projects and assist in planning funding proposals for GDR. We receive quarterly reports and annual reports about work completed in that period and plans for the next quarter. These are provided in advance of our meetings so we have time to review them and can provide constructive feedback at each meeting including our annual in-person meeting at PAG. We are very satisfied with the progress made in GDR over the last 4 years of the current NRSP10 grant, with most of the deliverables exceeding stated objectives. The RosEXEC members as well as the GDR team also engage the larger community to gather

feedback and guidance in prioritizing activities. We have been integrally involved in developing the next set of objectives for this 5-year NRSP10 proposal. We recognize the need to develop a community-led and accepted sustainability plan for GDR as it is such a critical resource for our community. We will work diligently with the sustainability assessment team to identify what options have most potential for GDR and be active in helping implement them where possible.

Thank you for all your team has done for the community over the last fifteen years of GDR and the other underserved NRSP10 target crop databases. We wish you every success for this critically important proposal.

Sincerely,

A handwritten signature in black ink, reading "Mercy Olmstead". The signature is written in a cursive, flowing style with a large initial 'M' and a long, sweeping tail.

Mercy Olmstead, Ph.D.
Chair, U.S. RosEXEC
Watsonville, CA

December 17, 2018

Dorrie Main
Professor of Bioinformatics
Dept. of Horticulture
Washington State University
Pullman, WA 99163

Dear Dorrie;

I wish to express Cotton Incorporated's continued whole hearted endorsement for the renewal proposal for NRSP10 titled "National Database Resources for Crop Genomics, Genetics and Breeding Research". The effort your team has led in the past few years to develop genome databases for several crops, especially cotton, is of critical value to our research community. The impressive growth and use of CottonGen over the first four years of NRSP10 is testament to the quality of data and expanding tools available to the cotton research community. Breeders are now able to employ the benefits of the genomics revolution toward solving the increasingly complex production problems for growers. Housing the International Cotton Genome Initiative (ICGI) within CottonGen has enabled improved communication within the international cotton community. The database hosts biennial ICGI elections and helped organize and host submission of abstracts for the 2018 ICGI conference which was held in Edinburgh, Scotland.

The CottonGen Steering committee, representing industry, government and academic scientific stakeholders, is provided with quarterly progress reports and offers regular guidance on priorities for the database. Feedback from the Steering Committee and wider community has shaped the objectives for the next five years of NRSP10. We are supportive of all activities and will work with you toward establishing sustainability options for CottonGen.

As you know, in 2009 we decided that the two existing cotton community databases, CottonDB and Cotton Marker Database, did not meet the cotton research communities' needs largely because of the legacy systems both used. We investigated several crop database groups to decide where best to invest our grower provided funds. The community chose your team because of the efficiencies gained by using and further developing the standardized Tripal platform and your emphasis on developing tools for breeding. Your decision to hire the CottonDB curator and allow her to work remotely from Texas so she could have continued direct access to cotton researchers at USDA-ARS and Texas A&M has proven to be a wise decision. The CottonGen Steering Committee and the cotton community are very satisfied with the database, and it is our intention to continue to support it in the future. Between 2011 and 2018 we provided over \$800,000 dollars to support CottonGen with another \$363,000 committed for 2019-2020.



Cotton
Incorporated

We continue to view your NRSP proposal as a natural evolution of the databases for the five crop groups targeted. While the crops are distinctly different, in aggregate they are vitally important across the entire country and feature in a range of research programs at land grant universities, USDA-ARS, and private industry partners. Despite the diversity, each of these crops has a need for a world-class, dynamic, accessible genome database. Your team has provided the cotton community this bioinformatics resource and in so doing, has created a unique software platform.

Please accept this letter as an indication of Cotton Incorporated's continued, firm commitment to support your NRSP renewal proposal.

Sincerely,



Don C. Jones, Ph.D.
Director of Agricultural Research



Cotton
Incorporated



United States Department of Agriculture

Research, Education, and Economics
Agricultural Research Service

10 December 2018

Dr. Dorrie Main
Department of Horticulture
Washington State University
Pullman, WA 99164-6414

Dear Dorrie

As the chair of the newly formed Cool Season Food Legume Genome Database Steering Committee, comprised of academic, government and industry stakeholders, I would like to express our strongest support for your NRSP10 renewal proposal. We have provided input on the objectives and activities in the proposal and support the final version included in the proposal.

Established in 2012, the Cool Season Food Legume Genome Database (CSFL, <https://www.coolseasonfoodlegume.org>) provides curated and integrated genomic, genetics and breeding data and analysis tools to enable basic, translational and applied research in pea, lentil, chickpea and fava bean crops. It's growth over the last 4 years has been impressive, it is current with all published genomic and genetic data (extracted from literature), provides easy to use data management and analysis tools for breeding programs at no cost, and delivers access to analyzed data distilled into useable information for genomicists, physiologists, molecular geneticists, and breeders.

We look forward to working with your team to ensure CSFL continues to enable pea, lentil, chickpea and fava bean research and in particular can help us utilize the volume of big data that is now being generated for our crops toward accelerated research discovery and crop improvement.

Sincerely

A handwritten signature in blue ink that reads "Rebecca McGee".

Rebecca McGee, Ph.D.
Chair of the CSFL Steering Committee



Pacific West Area – Grain Legume Genetics and Physiology Research

305 Johnson Hall Washington State University

Pullman, WA 99164

(509) 335-0300 • Fax (509) 335-7692

rebecca.mcgee@ars.usda.gov

College of Agriculture and Life Sciences
NC Research Campus
600 Laureate Way
Kannapolis, NC 28081
704.250.5400
704.250.5409 (fax)

December 18, 2018

To Dorrie Main, PhD
Department of Horticulture
Washington State University
45 Johnson Hall
Pullman, WA 99164-6414

Dear Dr. Main,

I am writing this letter to express my strong support for the NRSP Project entitled: “**National Database Resources for Crop Genomics, Genetics and Breeding Research**”.

At North Carolina State University my program focuses on developing genetic and genomic resources to study economically important traits with an emphasis on nutritional related traits (e.g. anthocyanin, carotenoid accumulation). I am coordinating a SCRI-CAP proposal focusing on Vaccinium crops, specifically blueberry and cranberry. The project will involve over 32 US Co-PI's and collaborators, including all blueberry and cranberry breeding programs, genomics scientists and is supported by 24 stakeholder organizations representing thousands of stakeholders from all major blueberry and cranberry production regions.

The Genome Database for Vaccinium that your team manage represents a strategic resource for any past, ongoing and future blueberry and cranberry genetic-genomic based projects. Considering the amount of data generated in genomic-breeding based projects and the financial investments involved (public and private), having a centralized computational infrastructure such as the GDV is a strategic need to avoid duplication of effort and efficiently translate such information to end users. As we plan new projects, member of our community can rely on this infrastructure and focus our efforts and resources to plan activities that can translate scientific knowledge that we will gain in the project to end users. Your involvement in the VacciniumCAP proposal represents a great example of such a partnership. As you know the primary objective of our VacciniumCAP proposal will be to expand DNA based resources to facilitate the development of blueberry and cranberry cultivars with improved fruit quality attributes. We have allocated funding to support data curation in the database to your team, while the GDV infrastructure is already available at no cost for our project. As discussed during our planning activities, such a partnership has already been established with many other blueberry and cranberry scientists, and provides an incredible added value to our projects and your proposal.

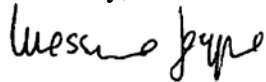
Looking at your objectives and planned activities, it is encouraging to see that in the NRSP10 renewal proposal you are planning to expand the Breeding Information Management System

(BIMS) to include more analysis capability to the already established BIMS implemented in GDV and the other NRSP10 databases Obj 4 of your proposal. These tools can be used to effectively translate massive DNA and phenotypic datasets into useful information for the research and breeding community. On this strong base, we foresee the more efficient development and release of new improved blueberry and cranberry cultivars. In addition, use of GenSAS for community curation of reference sequences, access to an epigenomics visualization tool and the ability to export data into other bioinformatics platforms (such as Galaxy) will facilitate data analysis for any scientists working in the Vaccinium crops.

Interaction and feedback that we will collect from our Vaccinium community will serve to advise priorities, inform on new data and projects and assist in planning funding proposal for GDV. In this regard, confirm that I commit to serve in the GDV Steering committee, to provide you such feedback and advise how the objectives of your proposal can effectively be used by our community.

I look forward to being involved in an advisory capacity through my attendance at both an annual meeting and conference calls. Good luck with your proposal.

Sincerely,

A handwritten signature in black ink, appearing to read "Massimo Iorizzo". The signature is fluid and cursive, with the first name "Massimo" written in a larger, more prominent script than the last name "Iorizzo".

Massimo Iorizzo, Ph.D.
Assistant Professor



United States Department of Agriculture

Research, Education, and Economics
Agricultural Research Service

Prof. Dorrie Main, PhD
Department of Horticulture
Washington State University
45 Johnson Hall
Pullman, WA 99164-6414

Dear Dr. Main,

I am writing this letter to express my strong support for the NRSP Project entitled: “**National Database Resources for Crop Genomics, Genetics and Breeding Research**”.

Located at the University of Wisconsin-Madison my USDA, ARS program focuses on carrot breeding and genetics and my program leads multiple efforts to develop DNA based tools that can facilitate carrot breeding programs in the U.S. and globally. I am coordinating a SCRI-SREP project focusing on develop genotyping and phenotyping tools to advance carrot breeding programs. The project involves 14 US Co-PI's and collaborators from public carrot breeding, outreach, and research programs, and 18 seed companies, processors, and growers that represent at least 80% of the commercial U.S. carrot crop.

As part of the project we are generating large amounts of data, that is critical to have at a centralized hub to store and make the data available to the carrot breeding, production, and research community. To store and organize our data we have adopted and implemented in our server the database platform that your team has developed as part of the NRSP10. We very much benefit from and appreciate your willingness to provide a robust and easy-to-use bioinformatics infrastructure, which have allowed us to allocate funding more towards the implementation of the tools rather than re-developing a tool that has already been developed for other crops.

Your input in our effort to develop a carrot database represents a clear example of how the tools developed as part of the NRSP10 represent a critical resource, even beyond the crops you have included in your proposal. This clearly show that the activities of your project are relevant at national level.

For these reasons I strongly support your project. Good luck with your proposal.

Sincerely,

A handwritten signature in blue ink, appearing to read "Philipp W. Simon".

Philipp W. Simon
Research Geneticist
and Professor of Horticulture

Vegetable Crops Research Unit, Midwest Area
Dept of Horticulture - University of Wisconsin
1575 Linden Drive - Madison, WI 53706
(phone) 608-262-7922 • (fax) 608-262-4743
USDA is an Equal Opportunity Provider and Employer

We very much appreciate the time taken by NRSP10 reviewers to review this proposal and their kind comments on its value. There were really only two areas that suggest a response is needed, which we try to address below.

Long term support for NRSP10 database and informatics resources is indeed a concern, just as it is for all such useful resources that do not have long term permanent support. In this renewal we have extended our successful past efforts to leverage other funding through industry and competitive programs to include the involvement of the experienced Phoenix Bioinformatics team. They will assess and recommend viable solutions for long term sustainability, which we will test and implement by 2024. Based on discussions with this team and the wider AgBioData community, sustainability will most likely involve a mixed model for support.

As a community database we make available the data we analyze from publicly available datasets in standard formats and work with contributors to ensure that their raw or analyzed data is submitted to the appropriate primary repository. We work with the communities to ensure we have the most up to date versions of genome sequences and annotations and plan to provide “mapping” of genes and gene name versions for all our databases (currently GDR is the only one that really does this). It is indeed important to have the data we make available in standard formats and we certainly endeavor to do so with the understanding that this data does not belong to us. Our good relationship with the crop communities served by NRSP10 helps us promote this effort.

NRSP Region Review - Blank Form

Status: Assigned

NOT FOR SUBMISSION. This form is for information purposes only. Your available reviews for submission can be found under Reviews > My Reviews.

NRSP Proposals Review Form

The following statement defines the mission of the NRSP program: "MISSION OF NATIONAL RESEARCH SUPPORT PROJECTS

The activity of an NRSP focuses on the development of enabling technologies, support activities (such as to collect, assemble, store, and distribute materials, resources and information), or the sharing of facilities needed to accomplish high priority research, but which is not of itself primarily research. Ideally, an NRSP would facilitate a broad array of research activities. The primary purpose of NRSPs shall not be solely to conduct research as there are other available mechanisms for creating these types of projects including the multistate research projects and the National Research Project (NRP) options. Examples of NRSP activities might include collection of data that are widely used by other research groups and efforts; development of databases; or development of critical technologies."

Based on the mission of NRSPs, all proposals will be evaluated using the following criteria:

Project:

A. Prerequisite criteria for NRSPs:

1. Mission: Is the NRSP consistent with the mission of an NRSP?

2. National Issue:

a. All NRSPs must involve a national issue, relevant to and of use by most, if not all regions. These projects draw on the best minds and resources within and outside the State Agricultural Experiment Station (SAES) system to address the issues. The proposal should discuss its support activities relative to other NRSPs.

b. For renewals, proposals must demonstrate direct relationship in support of continuing national priority need(s). The renewal application builds on the previous project and provides a logical progression.

Comments

B. These are the criteria addressing the rationale for the NRSP:

Total Points:

1. (20 points) Priority Established by ESCOP/ESS: Priority for funding will be given to NRSPs that address and support one or more of the national priority areas identified by ESCOP (see ESCOP Science and Technology Committee and Science Roadmap)

/20

2. (20 points) Relevance to Stakeholders:

/20

Comments

 Report a Bug

C. Criteria for implementing the NRSP proposal

Total Points:

1. (15 points) Management and Business Plan:

/15

a. Each NRSP should have a well-developed business plan that describes how the project will be managed and funded for a five-year period. This plan includes a management structure to adequately integrate the efforts of multiple participants. The plan should include provisions for linking multiple sources of funding and leveraging those sources with the limited off-the-top research funds. The plan should demonstrate that alternative funding sources have been explored. This plan should include efforts to bring in new agencies, organizations, industry, foundations, etc. to help address the issues and provide funding for the project. All project proposals must provide evidence of contributions from experiment stations across the nation beyond what is available through off-the-top funds.

b. The business plan for project renewals must include a funding plan including development of alternative funding for reducing off-the-top funding to a minimal level. Renewals will be judged as to the degree to which the project has been on task, had an impact, on time and within budget for the previous funding period. The renewal application should include a critical assessment of the original plan and address any shortcomings to ensure that the project will function more smoothly or effectively in the future. The proposal must indicate what additional resources have been generated or leveraged and indicate how those and any additional resources will be continued or sought.

Comments

2. (15 points) Objectives and Projected Outcomes:

/15

a. Objectives, milestones and deliverables should be described in sufficient detail such that progress can be measured. Indicate the prospects for meaningful impacts within the proposed duration of the project. The proposal must indicate what approaches will be used to assess outcomes including stakeholder use and how these assessments will be used in program planning.

b. For renewals, the proposal must address productivity, completion of original objectives and the relationship between projected goals and actual accomplishments. The proposal must include an assessment of the outcomes and/or impact of the previous project period. This assessment must include an evaluation of stakeholders' use of project outputs. The proposed objectives must reflect appropriate revision, e.g. evolution or building to greater depth, and/or capacity. All project revisions must incorporate stakeholder needs.

Comments

(15 points) Integration:

/15

a. Projects should indicate how efforts are integrated with extension or academic programs and how results might be of use by other potential stakeholders.

b. For renewals, the proposal should indicate any new partnerships built during the project period. The proposal should address the degree to which the full team is engaged in project planning and implementation. Discuss plans to correct any weaknesses that may have been identified.

Comments

4. (15 points) Outreach, Communications and Assessment:

/15

a. All projects must have a sound outreach, communications and assessment plan that seeks to communicate the programs goals, accomplishments and outcomes/impacts. The communication plan must detail how results will be transferred to researchers and other end users and contain the following elements:

i) Clear identification of the intended audience(s) of the NRSP. Since this is a Research Support Project, in most instances the primary beneficiary of the results will be other scientists. However, careful consideration should be given to other possible users of the information (such as consumers, producers, governmental agencies (local, state and federal), general public, etc.)

ii) Clear description of the engagement of stakeholders in the definition and/or conduct of the research support project.

iii) Thorough description of the methodology to measure the accomplishments and impacts of the National Research Support Project and effectiveness of the communication plan. Methods such as surveys, town meetings, conferences, analyses of reference data (e.g. citation index, etc.), and use of professional evaluators should be considered.

iv) Specific description for development of communication pieces describing the activities, accomplishments, and impacts of the NRSP. The communication pieces will be used with SAES/ARD directors, stakeholders and their organizations, funding sources and agencies, and congressional delegations.

v) Suggested mechanisms for distribution of the results of the research support project. Examples include sharing the results at annual meetings of stakeholders, providing material to the Budget and Advocacy Committee of the NASULGC Board on Agriculture Assembly and other appropriate committees within the SAES/ARD organization, and assisting CSREES in preparation of appropriate documents highlighting the impacts of the project.

Comments

[Back \(https://www.nimss.org/review/for_project/18619\)](https://www.nimss.org/review/for_project/18619)

Save

Submit

 Report a Bug