Final Report

MSMC Project #385

1. *Title:* **GWAS to Genes: A system to utilize association analyses to clone genes and develop markers to improve soybean breeding for variety development**
2. *Period Covered:* 6/1/2016-11/30/2018
3. *Project Leader:* Kristin Bilyeu, Co-Primary Investigators Trupti Joshi and Dong Xu
4. *Layman’s Summary:*

The goal of this project was to enhance existing bioinformatics tools and resources to enable broad and efficient identification of soybean genes that control phenotypes so that they can be directly and effectively utilized in soybean germplasm and variety development programs. The challenge we sought to address was the gap between the rapid generation of soybean genomic sequence information and the ability to utilize the “big data” in ways that could lead to improved soybean varieties. This project accomplished the goal, although it required more time than originally anticipated.

New analysis methods, collectively termed Genome Wide Association Studies, or GWAS, that take advantage of “big data” in both genotype and phenotype information hold the promise to ultimately identify hundreds of soybean genes that control phenotypes. However, these new methods of association analysis stop short of producing information that can be directly utilized in soybean breeding programs. The current methodologies available to generate GWAS results do not lead to the direct identification of genes that control the phenotypes. This project led to the development of a research pipeline and online tools to enable researchers and breeders to identify key alleles that control important soybean traits.

We designed and verified a strategy and developed bioinformatics tools to enable generalized “cloned” gene identification that can be immediately utilized in soybean germplasm and variety development programs. Identifying the genes that control key phenotypes leads to the development of molecular marker resources that allow perfect selection for the trait, a critical feature of the most effective soybean breeding programs.

The outcomes of this project are two manuscripts in preparation that fully describe the technical aspects of the work and present other researchers with the strategy (the GWAS to Genes pipeline) and online tools so that it can be widely adopted. The key tool we developed, SNPViz2, facilitates the exploration of allelic variation and diversity by its visualization in a web-based haplotype viewer. Major improvements were made to an existing tool that was a single nucleotide polymorphism (SNP) similarity-based clustering tool. The key features of the new SNPViz2.0 tool include increasing the available datasets to over 600 re-sequenced soybean genomes, incorporation of additional gene-based annotation and variant impact information, improved filtering of phenotype and sequence information, and the ability to link GWAS output information for use with the tool.

1. State your objectives in question form and discuss how your results answer these objectives:

*What strategy and tools are needed to utilize soybean genomic sequence information to develop allele-based markers that will enable improved soybean variety development?*

We tested different strategies while building and modifying bioinformatics tools to achieve the project goal. This work involved developing a working relationship between groups with differing expertise: molecular biology and data science/bioinformatics. Although the project took more time than we originally planned, the collaboration formed has expanded to address additional research challenges. The final strategy and tools of this project were validated by a case study on cloned genes where various types of causative mutations were analyzed. As a result, candidate genes for phenotypes of traits such as pubescence density and formation, seed coat luster and pod color were confirmed. Thus, the results provide a novel resource that facilitates cloning of prospective genes in soybean. We call the strategy GWAS to Genes, and the critical tool is SNPViz2.

1. Please answer the following:
2. *How do your results benefit Missouri Soybean Growers.*

Soybean growers in Missouri should see new varieties available in the future with traits that improve yield and crop value.

1. *Estimate financial return for the average Missouri soybean producer.*

Successful production of soybean varieties that have high yields and meet customer quality requirements will result in profitability for the average Missouri soybean producer.

1. *Do your results benefit the environment?*

There is no impact on the environment from this project and associated research.

1. *What products or processes can be commercialized from this research?*

The research results will be disseminated by peer-reviewed research publications and availability of an online tool.

1. *List disclosure(s) of inventions or plant varieties submitted to the MU Tech Transfer Office.*

None.

1. *Identify potential disclosure(s) of inventions or plant varieties. Please note that credit must be given to MSMC for any inventions or discoveries resulting from this research.*

None.

1. *How would you commercialize these products or processes?*

Ultimately, soybean breeding programs (public and private) would be responsible for commercializing new soybean varieties resulting from our research.

1. *If no specific products or processes were produced, how do you plan to make your results available to producers or industry?*

The research results will be disseminated by peer-reviewed research publications and availability of the online tool.

1. *Is additional time or research required before your results can be used by producers and industry?*

Two manuscripts are in preparation now. The experiments have been completed and the tool is receiving final adjustments and checks.

1. *Where does this research go from here? What are the next steps?*

The research is ready to be utilized by the soybean research and breeding community.

1. *List publications by type (popular press, thesis, journals, other) written or planned.*
2. **“GWAS to Genes” pipeline identifies causative genes underlying important soybean phenotypes**; M. Skrabisova, S. Zeng, C. Miranda, N. Dietz, C. Culp, J. Hyun, R. Combs, J. Kim, T. Joshi, and K.D. Bilyeu, to be submitted to BMC Genomics.
3. **SNPViz2, an online tool utilizing genome data to enhance gene cloning in soybean**; S. Zeng, M. Skrabisova, K.D. Bilyeu, and T. Joshi, to be submitted to Bioinformatics Application Notes.

1. *List cost of original project and actual expenditures.* (The U.S. Department of Agriculture requires that we ask for budget information, including the number of hours spent on the project, the number of dollars remaining on account, as well as a breakdown of expenses. You are required to provide this information in your report. Please also include names and titles/positions of those whose time has been charged to this project.)

The proposed budgets were:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | YR1 | YR2 |
| Salaries/Wages | | $39,000 | $29,000 |
| Fringe benefits | |  | $2,628 |
| Equipment | |  |  |
| Supplies |  | $5,000 | $2,000 |
| Travel |  | $3,500 | $3,000 |
|  |  |  |  |
| Total |  | $47,500 | $36,628 |

Actual expenditures were:

|  |  |  |
| --- | --- | --- |
| **Period: 6/1/2016-11/30/2018 (Year1: $47,500 Year2: $36,628) Total: $85,128** | | |
| **Salaries/Wages** | Graduate Students and Postdoc | $52,104.74 |
| Molecular Biologist | $22,540.75 |
| **Fringe benefits** |  | $2,560.96 |
| **Equipment** |  | $2,609 |
| **Supplies** |  | $1,783.60 |
| **Travel** |  | $3,528.95 |
| **Total** |  | $85,128 |

* Names and positions/titles of those whose time is being charged to this research project

Maria Skrabisova, Molecular Biologist contractor

Shuai Zeng, University of Missouri PhD student (0.5 FTE, 4 months)

Sadia Akter, University of Missouri PhD student (0.25 FTE, 7 months)

Zhen Lyu, University of Missouri PhD student (0.5 FTE, 7 months)

Siva Narisetti, University of Missouri PhD student (0.25 FTE, 5 months)

Ramya Payyavula, University of Missouri PhD student (10 Hours/week, 3 months)

Yang Liu, University of Missouri PhD student (0.25 FTE, 2 months)

Duolin Wang, University of Missouri Visiting Scholar (0.5 FTE, 6 months)

Juexin Wang, University of Missouri Postdoc (1 FTE, 2 months)

1. *List equipment purchased with MSMC funds, identifying inventory and serial number. (It is not considered equipment unless it costs $500 or more and has a life expectancy of at least 2 years.) Indicate current and future use of this equipment in support of soybean research.*

2 computer desktops and monitors were purchased.