|  |  |
| --- | --- |
| * **P:\LOGOS\ISA\ISALogo.jpg**
 | * **Research Progress Report**
 |

|  |  |
| --- | --- |
| * **Project Title:**
 | * **Application of High throughput function discovery to soybean improvement**
 |
| * **Principle Investigator(s):**
 | * **Dr. Karen Hudson**
* **Dr. Brian Dilkes**
 |
| * **Report Type:**
 | * **Final**
 |
| * **Due Date:**
 |  |
| * **Current Project Period:**
 | * **June 1, 2015 – May 31, 2016**
 |
| * **Date Final Report Due:**
 |  |
| * **Purdue Project Number:**
 | * **208715**
 |
| * **ISA Project Number:**
 |  |
| * **1. Outputs - Explain what you did, what was discovered, and what was learned as a result of the research project.**
* Report outputs completed during the reporting period that contribute to the goals and objectives of the project (do not include publications here, they are to be reported separately in the block below).
* Do not include findings or conclusions that have been reached; these are to be reported separately as changes in knowledge in the outcomes section.
* Include a description of how the results have been disseminated to communities of interest or how the product is being shared. This report narrative is required of all projects.
* For a project just initiated, please note that status.
* Narrative is limited to 3,200 characters and spaces.
 |
| * This project was a continuation of a collaboration between the Dilkes and Hudson labs initiated in June 2014, and the project was awarded additional funding title “Application of High throughput function discovery to soybean improvement”. This is the final report of progress for that project with a focus on the products from 2016 and 2017. The project is complete and no additional work is expected and all unexpended funds should be returned to the funder.
*
* The long term goal of the project was to accelerate the discovery of genes in soybeans. We chose to identify the genes responsible for mutations in soybean lines with altered seed composition, particularly seed oil. In the service of this, we combined the genetic resources in the Hudson lab and the bioinformatics expertise in the Dilkes lab and developed bioinformatics approaches to efficiently filter out unimportant background mutations to improve the accuracy and reduce the cost of identifying causative mutations in soybean lines with altered seed composition. During the project period there were three objectives, The project had three objectives: 1. to sequence the genome of the crossing partner cultivar “Prize” to identify polymorphisms useful for mapping between Prize and Williams-82; 2. Sequence mutant and non-mutant siblings of Prize-mutant F2 populations to map mutant traits; and 3. Identify the location of causative mutations using bioinformatics. We were successful in building the bioinformatics approach. Sequencing of mutants and Prize and the calling of mutations was also successful. For the majority of mutants identified by HPLC, we identified candidate genes in the sequencing data. Remarkably, for all of the mutants that had strong mendelian (3:1) segregation ratios that were identified from the mutant population, we identified strong candidates for the causative mutations. All remaining mutants lacked a strong candidate gene, and when tested for segregation, they were inherited as weakly penetrant alterations of phenotype that gave a more continuous variation in oil content, rather than the mutant vs wildtype categorical changes in oil phenotype that we expected. Thus, we have now described all of the categorically heritable oil mutants identified by HPLC in the population to date.
* Novel alleles at previously identified loci were identified in this project, including multiple fatty acid desaturases. A manuscript describing novel alleles of the FAD3A gene has been submitted from this project and is currently under review.
* Briefly, by re-sequencing the genomes of thirteen mutants and combining these data with three other mutants sequenced by the Dilkes lab, we were able to identify and remove confounding sequence variation. Polymorphisms that are shared between different lines are a result of residual heterogeneity in the founding population or artifacts caused by the interaction between the complex polyploid genome of soybean and modern sequencing methods. Such positions, identified in error, confuse researchers and make determining the molecular cause of variations in phenotype costly and labor intensive. Using our approach, we were able to eliminate the vast majority of variants (up to 96%) discovered in each of these lines as not possibly causing the altered soybean seed composition by taking this approach. Thus we a were able to increase the precision of our approach by 25-fold when compared to sequencing of individual mutants. We then compared the possibly causative mutations with an extensive list of genes that are predicted to be involved in some aspect of oil biosynthesis or metabolism and were able to assign likely candidates to many of the lines directly, some of this was complete in the previous project period.
* We have completed the three objectives of this funding period and more as described below.
* 1. The cultivar Prize was resequenced. This is a valuable resource going forward for the development of molecular markers for genetic mapping of many traits under study in the Hudson lab, and was used for the analysis in Objective 2.
* 2. One aim of this project was to test the efficiency of whole-genome bulk segregrant resequencing, and compare it to similar methods such as genotyping by sequencing (GBS) as well as determining optimal sample sizes and preparations for soybean populations. Using F2 samples from 9 populations were sequenced. DNA from F2 individuals was prepared, purified, and pooled for sequencing. From 4 populations, we sequenced DNA from two contrasting phenotypic groups, and from 5 additional populations we sequenced DNA from only the mutant outliers. One of these populations corresponded to a line for which we had whole genome sequence from the first year of funding for the project. Through leveraged funding we obtained GBS data in parallel on these populations to aid in genetic mapping. In soybean, even with the use of winter nurseries and early phenotyping/genotyping, high resolution genetic mapping requires multiple crosses and growing seasons and larger population sizes than were anticipated for this project. We are still fine mapping these traits using both conventional and genomic methods. Figure 1 shows the most promising linkage data from one the lines. The data generally indicate that the remaining uncloned mutants will require larger populations and more complicated experimental designs to identify the molecular identities of any additional genes affecting altered seed composition.
* 3. We have been able to confirm that line 1877, with a low-linolenic acid phenotype, carries a mutation in the FAD3A and we have now submitted a publication describing this variant for publication in Crop Science (see below). For 5 of the lines, sequencing identified strong candidate genes with known functions in oil biosynthesis or lipid metabolism (Table 1). 8 of the lines did not, and these latter cases the inheritance of the trait was weaker, requiring larger populations sizes than we had expected to discover the causative variant. In one case we were able to demonstrate a non-recessive inheritance pattern (Table 2) that will need to be taken into consideration for future experimentation identifying the gene responsible for this mutant phenotype.
* Despite the loose ends, the work performed has comprehensively identified the variants present in all mutants discernably affected in seed composition in the population analyzed. The identification, and in some cases confirmation, of mutations in genes with known or expected roles in seed oil biosynthesis include novel variants of potential value. The loose ends, those mutants without mutations in genes with predicted roles in lipid metabolism, suggest that beyond the mutants we have cloned, there are genes with more modest effects on seed composition yet to be discovered. The weaker expressivity of the oil phenotype in some lines, apparent polygentic quantitative inheritance, the presence of mutants with dominant or additive inheritance and the lack of mutations in any of the known oil genes in some of the lines with weaker seed composition phenotypes all point to this. A greater focus on line-cross population generation and increased population sizes may yet uncover new loci that regulate seed oil in soybeans still lurking in this mutant population. The GBS approach used, bioinformatics tools, sequenced mutant data, and the sequence data from Prize will facilitate such future work.
 |
| * **2. Outcomes/Impacts - Explain the beneficial results (potential yield increase, financial benefits, new use, pollution or erosion reduction, change of behavior, etc.) of this project for farmers and other stakeholders.**
* Describe how findings, results, techniques, or other products that were developed from the project generated or contributed to an outcome/impact.
* Describe the results of the project evaluation. Indicate how resources and activities helped to produce project outputs and achieve project outcomes and impacts.
* This report narrative is required of all projects.
* For a project just initiated, please note that status.
* Narrative is limited to 3,200 characters and spaces.
 |
| * One outcome from this is the description of the molecular mechanism affecting low-linolenic line 18777. This resulted from a mutation in the FAD3A gene encodes one of the major enzymes that converts linoleic acid to linolenic acid, and when this enzyme is defective linolenic acid accumulates at a reduced level. This discovery has been submitted to Crop Science for publication and the draft is provided as an attachment with this document.
* This line could be made available to soybean breeders to incorporate the low linolenic acid trait into new germplasm. Reducing levels of linolenic acid reduce the need for hydrogenation and make a more stable, healthier soybean oil.
 |
| * **3. Publications/Extension/Outreach - Describe how findings and results were shared. Report number of website hits, number of meetings where results shared, number of people attending meetings, etc.**
* List publications, documents, meetings or events that are specific to the project during this reporting period.
* Include only those publications, documented meetings not previously reported.
* Include research and extension publications, handouts, electronic publications, websites, etc.
* If there are no publications, documents or meetings to report for the period, leave this field blank.
* Narrative is limited to 3,200 characters and spaces.
 |
| * Dr. Hudson presented data from this project at the following scientific meetings/stakeholder workshops:
* Oral presentation “Back and forth: NGS for forward and reverse genetics for soybean” at the Plant Genomics Congress, St. Louis, MO, 2015.
* Poster at Purdue Soybean Showcase 2015
* Oral presentation at Purdue Soybean Showcase 2017
* A manuscript was submitted to Crop Science in August 2017 describing the mutant allele of *FAD3A* that was identified in the first year of the project:
* **New alleles of *FAD3A* lower linolenic acid content of soybean seeds**
 |
| * **4. Project Modifications - Describe any significant changes to project content from original funded project proposal.**Select one of the following options:
* X Not applicable for this period, nothing significant to report.
* Report narrative entered in the box below.
* Describe major changes in approach, procedure, method, hypothesis, or timing and reason(s) for these major changes.
* Narrative is limited to 3,200 characters and spaces.
 |
|  |
| * **5. Completion Date - Describe any foreseen possibility of a no cost extension request. Be specific as possible as to why a no cost extension might be requested.**Select one of the following options:
* x Project expected to be completed on schedule.
* Project delay expected, report narrative entered in box below.
* Describe major changes in approach, procedure, method, hypothesis, or timing and reason(s) for these major changes.
* Changes in a project completion deadline require the submittal of a separate no-cost extension request via Purdue’s Agricultural Research Office.
* Note the expected target date for submittal of a request.
* Narrative is limited to 3,200 characters and spaces.
 |
|  |
| * **6. Attachments - Attach any copies of graphs, charts, publications, reports, field day flyers, etc. regarding project.**
 |

* Please find the draft of the manuscript submitted to Crop Science attached to the email containing this report.
* The following figures and tables were referred to in the text of this report

18291 x Prize

Blue = High oleic group

Red = Low oleic group

W82

Prize

Chromosome 19

Figure 1.

* Table 1. Resequenced Lines

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| * Line
 | * Phenotype
 | * Unique nsSNPs
 | * nsSNPS in oil genes
 | * Candidate
 | * Population status
 |
| * 18291
 | * High LLN
 | * 51
 | * 0
 |  | * F3 phenotyped,bulk seg. genotyped
 |
| * 17026
 | * High STE
 | * 69
 | * 0
 | * FAD8-like
 | * F3 from 2015 field
 |
| * 19828
 | * High STE/protein
 | * 37
 | * 1
 |  | * F3 from 2015 field
 |
| * 16007
 | * High STE
 | * 29
 | * 0
 |  | * F3 from 2015 field
 |
| * 22677
 | * High STE
 | * 195
 | * 13
 | * FAD2-like
 | * F1
 |
| * 14223
 | * High PAL
 | * 694
 | * 24
 | * FAT-A
 | * F1
 |
| * 19396
 | * High OLE
 | * 507
 | * 36
 |  | * F1
 |
| * 19618
 | * High OLE
 | * 138
 | * 4
 | * phospholipase
 | * F2 phenotyped, bulk seg
 |
| * 17171
 | * High OLE
 | * 237
 | * 4
 |  | * F2 phenotyped, bulk seg
 |
| * 15891
 | * Short internode
 | * 38
 | * NA
 |  | * F2 phenotyped, bulk seg
 |
| * 16851
 | * Short internode
 | * 305
 | * NA
 |  | * F1
 |
| * 15207
 | * Creeping morphology
 | * 122
 | * NA
 |  | * F1
 |
| * 18777
 | * Low LLN
 | * 352
 | * 6
 | * FAD3A
 | * F2 genotyped and phenotyped
 |

* Table 2. Pools

|  |  |  |
| --- | --- | --- |
| * Plant ID
 | * Phenotype
 | * notes
 |
| * 18291
 | * Low OLE
 | * 13% OLE, mutation is dominant
 |
| * 20324
 | * High OLE
 |  |
| * 19060
 | * High LLN
 |  |
| * 14726
 | * High LLN
 |  |
| * 19483
 | * High LLN
 |  |
| * 14552
 | * High OLE
 |  |
| * 19120
 | * High OLE
 |  |
| * 18190
 | * High STE
 | * 20+% STE, multiple genes
 |