Final Report on “Manipulating a major gene governing seed reserves as a means to maintain yield and oil while increasing protein”, October 2021.

Introduction:

Of the many QTL controlling soybean seed protein content, alleles of the cqSeed protein-003 QTL on chromosome 20 exert the greatest additive effect. The high protein allele exists in both cultivated and wild soybean (*Glycine soja* Siebold & Zucc.) germplasm. We fine mapped this and identified the underlying causative gene. An insertion / deletion variant detected in *Glyma.20G85100* was found to have near-perfect + / - concordance with a high / low protein allele genotypes inferred for this QTL in parents of published mapping populations. The indel structure was concordant with an evolutionarily recent insertion of a TIR transposon into the gene in the low protein lineage. For this project we developed and trialed transgenic plants designed to alter the expression level of *Glyma.20G85100.* Seed protein was significantly greater in greenhouse-grown soybean expressing an RNAi hairpin down-regulation element in two independent events relative to control null segregant lineages. We conclude that a transposon insertion within the CCT domain protein encoded by the *Glyma.20G85100* gene accounts for the high / low seed protein alleles of the cqSeed protein-003 QTL. Field trials for the transgenic plants have been completed for three growing seasons, and show pleiotropic effects of the transgene that include higher protein content along with other phenotypes in mineral content and maturity.

At the beginning of the project in October 2019 we stated the following milestones and KPIs. A brief progress report is given for each milestone and KPI. Our project to engineer protein levels using RNAi transgenics has proceeded according to plan and all milestones and KPIs were met.

Milestones:

1. Field trial completed for one growing season using RNAi transgenics

We have now completed field trials for three seasons at the transgenic field facility in Nebraska and we are still analyzing the data. Preliminary indications are that at least one transgenic event shows increased protein and free amino acid levels, and that oil content does not appear to be affected. However the results are preliminary and may not be statistically significant. The interpretation of yield data has been complicated by the fact that the RNAi transgenic constructs seem to affect maturity date, with the transgenic plants maturing later than the control, untransformed line. This is unexpected as the original pro/oil allele derived from PI468916 does not seem to affect maturity. The RNAi lines show very significant GxE and results vary between seasons and plots more than for PI468916-derived alleles.

1. The high protein allele for Glyma.20G085100 downregulated and the effect on protein and oil tested in both a greenhouse and field and for yield in the field.

We have shown downregulation of the Glyma.20G085100 gene and have good data from the greenhouse and extensive data from the field. The effect on protein appears to vary between different transgenic events, as expected. So far we are seeing a 1-2% increase in protein content in the greenhouse consistently in the best transgenic lines. We also see an increase under field conditions but it is more variable. Interestingly we are also seeing relatively large differences in leaf mineral content in some transgenic events, and also effects on maturity and flowering time that are not seen in germplasm with the naturally occurring pro/oil allele. Completion of the analysis for the third year of field data will hopefully solidify these results from the field trials.

1. CRISPR guide construct designed, created and tested in transgenic soybean roots

The CRISPR guide was designed, synthesized and placed in a plasmid vector. Restrictions on laboratory work prevented the testing of this vector using the soybean hairy root system through 2020. However, we have now developed a new, high throughput CRISPR guide testing system that allows us to test in leaf cells directly, as well as a software application that predicts functional guides very accurately, so we no longer need to test the construct in roots.

We aimed to complete the following Key Performance Indicators (KPIs) by the end of Year 1:

1. Data available for multiple RNAi transgenic events for yield trials in the field

By the end of Year 1 we had extensive field data for two RNAi events. We now have three field seasons of data for several events.

1. The impact of down regulating the high protein allele tested

We have demonstrated down-regulation of the mRNA for the allele in transgenic plants that are in the trials.

1. At least one plant transformation plasmid completed containing a guide RNA sequence targeted to the Glyma.20G085100 gene.

The plasmid was completed. Work is continuing to transfer it to soybean plants.

In addition, we have found several interesting research leads that were not anticipated when the initial proposal was written. Two key insights are

1. A finding that the transgenic RNAi construct is also affecting protein, but has pleiotropic effects on maturity and seed mineral content, where the original allele does not.
2. A discovery that the low protein gene found in high-yielding conventional soybean varieties may revert (at a very low frequency) back to the high protein gene. Thus, we may be able to identify revertant alleles within existing elite lines to facilitate the creation of high protein varieties.

In addition to these novel findings, which both require further research before they can be exploited in breeding and agronomy, we have developed straightforward markers that define the locus for the pro/oil trait that can be directly used by any breeders wishing to alter protein content in germplasm.