Selection for Yield and Seed Composition Traits with Lines from a Recombinant Inbred Line Population

Jay Gillenwater

9/7/2021

Core ideas here.

Author list + affiliations

Abbreviations (alphabetical order: Abbreviation, definition; next abbreviation, next definition; …)

CAS, Caswell research farm; CLA, central crops research station; PLY, Tidewater Research Station; Pop 201, oil mapping population 201; Pop 202, oil mapping population 202; RCBD, randomized complete block design; RIL, recombinant inbred line; SDWT, one hundred seed weight; SSD, single seed descent;

# Abstract

# Introduction

* Importance of yield.
* Importance of oil, protein coupled with oil. Stats on worldwide meal consumption, value contributed by high protein.
* Difficulty in simultaneously increasing yield, protein.
* Need for diverse germplasm with high protein, yield, good agronomic qualities.

Soybean is among the most valuable crops in the current national and international markets. Chief among it’s uses are for the production of vegetable oils, and protein meal for animal feed. In 2021, 70% of worldwide protein meal consumption and 59% of worldwide oilseed production came from soybean. While these seed composition traits contribute value to the crop, overall seed yield is of primary importance. As such, improving these seed composition traits without compromising seed yield is of great importance to breeders and farmers alike. Historically the task of improving these three traits simultaneously has been difficult due to inverse correlations which exist between seed protein and seed yield, and between seed oil and seed protein compositions(Burton, 1987). These inverse correlations are thought to arise from competing metabolic pathways which control the traits(Gupta et al., 2017).

While difficult, it is possible to improve all three of these traits simultaneously through conventional breeding techniques as evidenced by the recent release of several cultivars with improved seed composition traits, and with yield comparable to, or superior to that of existing check cultivars(Burton et al., 1999; Chen et al., 2011; Panthee & Pantalone, 2006). As such, it is important that breeders take every opportunity to both better understand the relationship between these traits, and identify new cultivars that have improved seed composition traits without sacrificing yield.

# Materials and Methods

## Phenotypic data collection

Genotypes were evaluated for both agronomic and seed composition traits. Agronomic data collected included seed yield, hundred seed weight (SDWT), lodging, height, maturity date, and a composite agronomic score. The agronomic score was a visual metric which sought to quantify traits not directly covered by the other phenotypes like pod load and plot uniformity to provide an estimate of a genotypes general desirability. Height was measured in centimeters from the soil to the top of the plant. Lodging was scored on a scale of 1 to 5 where a score of 1 indicates that all plants are erect and a score of 5 indicates that all plants in a plot are on the ground(I. Fehr, 1987). Agronomic score was also measured on a scale of 1 to 5 where a score of 1 indicates lines that are the most desirable in a population while a score of 5 indicates that a genotype is among the least desirable. Maturity was recorded at the R8 maturity date(W. R. Fehr & Caviness, 1977). Following harvest seed yield, SDWT, seed oil content, and seed protein content were measured after seed was allowed to dry to approximately 7% moisture content in a greenhouse. Seed yield and SDWT were measured in grams after seed had been sifted to remove debris and cracked seed. Seed protein and oil contents were measured with a Perten DA 7250 NIR instrument®.

## Pedigree

Lines used in this study were selected from two F5-derived recombinant inbred line populations that were segregating for seed protein and seed oil contents, denoted as populations 201 and 202 (Pop 201 and Pop 202). The first population, population 201 consisted of 273 recombinant inbred lines (RILs) derived from a cross between the high protein breeding line LMN09-119 and the high oil breeding line N09-09 in 2015. Population 202 consisted of 197 RILs derived from a cross between the high protein breeding line LMN09-19 and the high oil breeding line N13-47, also made in 2015. RILs were advanced to the F5 generation by single seed descent (SSD) at which point a single plant was harvested to represent each F5 derived RIL.

## Population development

In 2019, 80 RILs with the highest bulk weight, seed protein, and seed oil content with non-extreme maturity dates were selected from Pop 201 and Pop 202. The selections were used to make two yield trials per population, named Test 1 - Test 4 where Test 1 and Test 2 consisted of RILs from Pop 201 and tests 3 and 4 consisted of RILs from Pop 202. Yield check cultivars Osage(Chen et al., 2007), Dunphy**find citation**, and Roy(Burton et al., 2005) were used in tests 1 and 2 while Dunphy**find citation**, Dilday **find citation**, and Raleigh(Burton et al., 2006) were used for tests 3 and 4. Check cultivars were selected for each test to match the maturity date range of the RILs of each test. Following the collection of phenotypic data in the 2019 growing season, a subset of the lines was identified which exhibited yield comparable to the check cultivars, and also high seed protein, high seed oil, or a combination of high seed oil and seed protein content. In total, 30 genotypes were selected which met this criteria and were organized into two new yield trials, each consisting of 15 RILs with similar maturity dates. The Dunphy, Dilday, and Raleigh check cultivars were included in test 1 while the Osage, Ellis, and Dunphy check cultivars were used in test 2. These two trials were grown and evaluated in two locations in both the 2020 and 2021 growing seasons.

## Plot techniques

* Plot techniques: How many rows, row length, spacing, seeding rate (from lead sheets)
* How data was converted between three-row and four-row trials.

In 2019, each test was grown at the Tidewater Research Station in Plymouth, NC (PLY), and the Caswell Research farm in Kinston, NC (CAS). RILs were grown in a randomized complete block design (RCBD) with three replications at each location. The yield plots consisted of three rows on 96.5 cm row spacing with 6.7 m row length. The plots were end trimmed to 4.57 m prior to harvesting at or near maturity. Seeds were planted at a density of approximately 31 seeds per meter in these plots.

In 2020, each test was grow in CAS and the Central Crops Research Station in Clayton, NC (CLA). RILs were grown in a RCBD with four replications at each location. Plot techniques and dimensions were otherwise identical to those employed in 2019.

In 2021, each test was grown in the CAS and PLY locations. RILs were grown in a RCBD with four replications at each location. Yield plots consisted of four rows with a 96.5 cm row spacing with 5.5 m row length. The plots were end trimmed t0 3.05 m prior to harvesting at or near maturity. Best management practices were used throughout all growing seasons.

## Statistical analysis

* Statistical model used to analyze data, fixed vs random effects what software was used.
  + Model used.
  + ANOVA results for model term significance (mainly genotype) both within environment and in all environments.
  + Software used to fit the model
  + How were genotype means compared (LSD?)
* The following statistical model was used to analyze the data.
* Where is the overall mean, is the effect of location , is the effect of replication nested within location , is the effect of genotype , is the effect of the interaction between genotype and location . Genotype is assumed to be a fixed effect and all other effects are assumed as random.

# Results and Discussion

* Results of the ANOVA analysis
  + Were G and GE effects significant for traits in the individual and combined environments?
  + Present ANOVA tables for the combined analysis for each trait in main manuscript body (space allowing) and the individual environments in the supplementary section
* Pearson correlation coefficients between phenotype LSMeans for individual and combined analyses.
* Report agronomic performance of well performing lines.
  + Well performing lines are those with yield within a LSD of the check average.
  + Report yield, maturity, lodging, heinght, seed weight, seed protein, seed oil, and meal protein.
    - Meal protein maybe from the SPROC program(?) **look into how to calculate this**

# Acknowledgements

# References

Burton, J. W. (1987). Quantitative genetics: Results relevant to soybean breeding [Journal Article]. *Agronomy (USA)*.

Burton, J. W., Carter Jr, T. E., & Bowman, D. T. (2005). Registration of’NC-roy’soybean [Journal Article]. *Crop Science*, *45*(6), 2654–2655.

Burton, J. W., Carter Jr, T. E., Fountain, M. O., & Bowman, D. T. (2006). Registration of’NC-raleigh’soybean [Journal Article]. *Crop Science*, *46*(6), 2710.

Burton, J. W., Wilson, R. F., & others. (1999). Registration ofProlina’soybean [Journal Article]. *Crop Science*, *39*(1), 294–295.

Chen, P., Ishibashi, T., Dombek, D. G., & Rupe, J. C. (2011). Registration of R05-1415 and R05-1772 high-protein soybean germplasm lines [Journal Article]. *Journal of Plant Registrations*, *5*(3), 410–413. <https://doi.org/10.3198/jpr2010.05.0304crg>

Chen, P., Sneller, C. H., Mozzoni, L. A., & Rupe, J. C. (2007). Registration of ‘osage’soybean [Journal Article]. *Journal of Plant Registrations*, *1*(2), 89–92.

Fehr, I. (1987). *Soybeans: Improvement, production and uses,”* [Journal Article].

Fehr, W. R., & Caviness, C. E. (1977). *Stages of soybean development* [Journal Article].

Gupta, M., Bhaskar, P. B., Sriram, S., & Wang, P. H. (2017). Integration of omics approaches to understand oil/protein content during seed development in oilseed crops [Journal Article]. *Plant Cell Reports*, *36*(5), 637–652. <https://doi.org/10.1007/s00299-016-2064-1>

Panthee, D. R., & Pantalone, V. R. (2006). Registration of soybean germplasm lines TN03-350 and TN04-5321 with improved protein concentration and quality [Journal Article]. *Crop Science*, *46*(5), 2328.