Capturing Genetic Variation in Cottonseed Constituents to Enhance Value

Kent D. Chapman
University of North Texas
BioDiscovery Institute and Department of Biological Sciences
Denton, TX

**Seed yield is ~1.5X fiber but less than half of seed is processed - Place to add value - Oil >> protein

2019 Cotton Breeder’s Tour - Cotton Incorporated
College Station, TX -- July 22, 2019
Capturing Genetic Variation in Cottonseed Constituents to Enhance Value

Variation in Cottonseed Oil and Protein Content- Non-destructive measurements by Time-Domain $^1$H-NMR. Quantifying variation across Gossypium germplasm and genotypes

Compositional Differences-- Identification of a naturally-occurring mutant allele for increased oleic acid in cottonseeds.

A Potential Future with a Holistic Focus on Seed Products?
Targeted and untargeted strategies-- Ultra-low gossypol stacked with altered seed oil traits (high oleic, low saturate FA, cyclic FA, Vitamin E) or GWAS studies with NAM populations to identify important loci. Identity preservation considerations.
Oil and protein are synthesized in the embryo of the seed during development and maturation, primarily in the cotyledons of cotton seed (bulk of seed mass at maturity).

Fiber elongated from the seed coat surface during seed development and fills fruit (boll) pubs.caes.uga.edu/caespubs/pubcd/B1252.htm
Quantifying Seed Oil and Protein Content -
Non-destructive, time-domain $^1$H-NMR

In 2009-2011 worked with scientists at Buker Optics to develop NMR-based assay for quantification of protein and oil in ~3g cottonseed (single seed for oil). Horn et al., 2011, JAOCS (AOCS method for oil AOCS Cd 16b-93)

Since then, we have worked with the cotton breeding community to provide analyses where possible (T. Campbell, J. Snider, C. Main, and many others)

New service facility being set up by Cotton Inc to quantify seed oil (and protein) for public and private breeders, like is done already for fiber.

Also used this methodology to survey variation across U.S. Gossypium germplasm collection (with L. Hinze, R. Percy).
Seed Characteristics for 2250 accessions (US *Gossypium* Collection)

**Oil**
- Mean = 21.39%
- Std Dev = 2.46%
- Max = 27.2%
- Min = 8.2%
- Skewness = -0.86***
- Kurtosis = 1.10***

**Protein**
- Mean = 21.62%
- Std Dev = 2.36%
- Max = 35.9%
- Min = 9.9%
- Skewness = -0.25***
- Kurtosis = 2.49***

**Size**
- Mean = 9.13g
- Std Dev = 2.76g
- Max = 18.2g
- Min = 0.8g
- Skewness = -0.26***
- Kurtosis = 0.03

Distribution of oil, protein, and size for 2,256 cotton accessions in the *Gossypium* Diversity Reference Set.

**Protein** -- 10% to 36%

**Oil** -- 8% to 27%

**Size** -- 1g/100 to 18g/100

Hinze et al., Crop Science, vol. 55 2015
Seed Characteristics for 2250 accessions (US *Gossypium* Collection)

Distribution of protein and oil by genome in the *Gossypium* Diversity Reference Set

Wide range is captured in the tetraploid genomes-

Significant variation in seed oil and protein exists within most contemporary breeding programs.

Hinze et al., Crop Science, vol. 55 2015
Relationship of Seed Oil to Protein-
Gossypium hirsutum (AD-genome) “Texas” accessions

Strong negative correlation between protein and oil—often seen and reported by others. Outliers evident.

Variation suggests opportunities for selection.
Relationship of Seed Oil to Protein- 
Gossypium hirsutum (AD-genome) “Stoneville” accessions

Strong negative correlation between protein and oil—often seen and reported by others.
Relationship of Seed Oil to Protein-
*Gossypium barbadense* accessions

<table>
<thead>
<tr>
<th>% Oil</th>
<th>% Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>15.0</td>
<td>15.0</td>
</tr>
<tr>
<td>20.0</td>
<td>20.0</td>
</tr>
<tr>
<td>25.0</td>
<td>25.0</td>
</tr>
<tr>
<td>30.0</td>
<td>30.0</td>
</tr>
</tbody>
</table>

Less correlation between protein and oil—

But variation/outliers suggests opportunities for selection- e.g. GB0331 (high oil/ high protein)
**G.hirsutum** Elite Varieties in the Pee Dee Collection

- 82 varieties, 2-5 locations over three years.
- Strong **negative correlation** between protein and oil.
- Also **significant environmental component**, especially for protein.

**Outliers suggest relationship can be disrupted**

Hinze et al., Crop Science, vol. 55 2015
Examination of some outliers for oil/ protein content (US Collection)

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Species</th>
<th>Genome/Race</th>
<th>Collected from</th>
<th>Seed Size (mg/seed)</th>
<th>Oil (%)</th>
<th>Protein (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B01-1</td>
<td>G. anomalum</td>
<td>B1</td>
<td>Africa</td>
<td>23.0 ± 1.5</td>
<td>12.8 ± 0.3</td>
<td>20.2 ± 0.3</td>
</tr>
<tr>
<td>D01-10</td>
<td>G. thurberi</td>
<td>D1</td>
<td>Mexico/Arizona</td>
<td>20.0 ± 3.0</td>
<td>16.6 ± 0.3</td>
<td>11.4 ± 0.3</td>
</tr>
<tr>
<td>E01-3</td>
<td>G. stocksii</td>
<td>E1</td>
<td>Arabia</td>
<td>23.8 ± 2.8</td>
<td>8.2 ± 0.3</td>
<td>23.8 ± 0.3</td>
</tr>
<tr>
<td>GB-0331</td>
<td>G. barbadense</td>
<td>(AD)2</td>
<td>South America</td>
<td>119.5 ± 10.1</td>
<td>26.2 ± 0.6</td>
<td>27.3 ± 0.8</td>
</tr>
<tr>
<td>PIMA-S6</td>
<td>G. barbadense</td>
<td>(AD)2</td>
<td>South America</td>
<td>129.9 ± 16.7</td>
<td>25.5 ± 0.3</td>
<td>16.3 ± 0.4</td>
</tr>
<tr>
<td>SA-1254</td>
<td>G. hirsutum</td>
<td>(AD)1</td>
<td>Central/North America</td>
<td>114.8 ± 12.9</td>
<td>20.0 ± 0.4</td>
<td>37.4 ± 2.1</td>
</tr>
<tr>
<td>TX-2236</td>
<td>G. hirsutum</td>
<td>(AD)1</td>
<td>Central/North America</td>
<td>68.1 ± 8.1</td>
<td>14.4 ± 0.1</td>
<td>31.0 ± 0.4</td>
</tr>
<tr>
<td>TX-2500</td>
<td>G. hirsutum</td>
<td>(AD)1</td>
<td>Central/North America</td>
<td>83.6 ± 7.1</td>
<td>24.8 ± 0.1</td>
<td>23.0 ± 0.5</td>
</tr>
<tr>
<td>Coker-312</td>
<td>G. hirsutum</td>
<td>(AD)1</td>
<td>Central/North America</td>
<td>110.0 ± 6.8</td>
<td>20.6 ± 0.3</td>
<td>24.0 ± 1.5</td>
</tr>
</tbody>
</table>

Genotypes with low oil/ high protein; high oil/ low protein; **high oil/high protein**.

** Outliers suggest relationship can be disrupted

Sturtevant et al., Planta vol. 245, 2017
** Genetic variation in Oleic Acid content (at expense of Linoleic) - suggested variation in Fatty Acid Desaturase 2 (FAD2)

18:1 ➔ 18:2

Sturtevant et al., Planta vol. 245, 2017
Molecular Identification of a Mutant Allele of FAD2-1 in GB-0331

The mutant *fad2-1D-1* sequence was only identified in the GB-0331 accession (not Pima S6 or GB0332).

Mutation is a 90 bp insertion (red font). Patent issued 2019. Also found by Michael Dowd and Jay Shockey– USDA-ARS (back to back publications; *Planta*, 2017).

Mutation results in a truncation at the C-terminus– gene is expressed but normal subcellular targeting is disrupted.

Lori Hinze– USDA-ARS, crossing into non-photoperiodic germplasm.
Examination of Some Outliers for Oil Distribution (US Collection)

*Variation Exists in Composition and Distribution within the Embryo* – Complexity of Seed Oil Metabolism

Sturtevant et al., Planta vol. 245, 2017
A Potential Future with a Holistic Focus on Seed Products?

**Targeted strategies**-- Ultra-low gossypol stacked with altered seed oil traits (high oleic oil, low saturate oil, cyclic FA, Vitamin E)-- conventional transgenics (many examples available)-- gene editing approaches.

**Untargeted Strategies**- Marker and Genome-wide association studies appearing in the literature-- also NAM (nested association mapping) populations offer opportunities to identify important loci.

**More research needed**- to understand 1) partitioning of resources between embryo (seed products) and seed coat (fiber) 2) complexity of synthesis and distribution pathways within the seed.

**Identity Preservation of Seed Traits**- To capture added value, needs to be a means to segregate traits during production.
Identity preservation will be required to capture enhanced cottonseed value

- In the U.S., gene editing requires minimal regulatory cost since USDA-APHIS process is established.
- Most of the world still considers gene editing to be GMO thus unlikely that a company will incur the cost for global GMO regulatory approval of a cottonseed trait.
- If growers, ginners and processors want gene-enhanced cottonseed value, the protein and whole seed must be kept out of specific export markets (EU) and risks need to be addressed.

Kater Hake & Tom Wedegaertner, Cotton Inc
Identity Preservation for Enhanced Seed Value - System already in place can be adapted to capture value and maintain stewardship
Concluding Remarks

> Cottonseed production on a mass basis exceeds fiber by about 50%, yet less than half of the seed is processed for added value.

> Considerable genetic variation exists in seed oil (and protein) content and composition. Likely other valuable components as well. Largely untapped opportunities.

> Seed trait development will require more focused research, more attention in selection (measurements), identity preservation for commercial success.

> Time is right- genomics resources increasingly available for cotton, promise of gene editing for targeted trait development, recent deregulation of ultra-low gossypol seed (new potential markets for meal), and some oil traits already available (GMO and non-GMO).
Acknowledgements

**University of North Texas**
Patrick Horn, Drew Sturtevant, Chris Kennedy, many undergrads

**USDA-ARS**
Todd Campbell
Michael Dowd, Jay Shockey
Lori Hinze, Richard Percy

**Cotton Incorporated**
Kater Hake
Tom Wedegaertner

**Funding**
Cotton Incorporated
Relationship of Seed Oil to Protein-
*Gossypium* D-genome accessions

Less of an obvious relationship in diploid genomes—**more variation**.
Relationship of Seed Oil to Protein-
*Gossypium* A-genome accessions

Less of an obvious relationship in diploid genomes—more variation.
Seed Oil Content of Selected Exotic Gossypium Germplasm/Genomes

Oil Content ranges from low of 8% in “E” to high of 24% in “K”

Additional opportunities?