Factors Influencing the Composition/Metabolic Profile of Food from Plants

Sherry Flint-Garcia, PhD
USDA Agricultural Research Service
Sherry.Flint-Garcia@ars.usda.gov
Conflict of Interest Statement

As a federal employee, I have no conflicts of interest to declare.
Outline

- Domestication
- Plant Breeding and Biotechnology
- Maize Case Study
  - Domestication and Diversity
  - Genetics of Kernel Composition
  - Composition and Metabolites
- Appendix Slides
  - Overview of Genetics, Terms, and Methods
1. Domestication

Domestication: “Adaptation of plants and animals to intimate association with human beings”
Domestication Syndrome

- More robust plants
- Synchronous flowering
- Increased apical dominance
- Non-shattering
- Loss of seed dormancy
- Larger fruits/grains
Wheat

Complex evolutionary history due to polyploidization and distinct market classes

- **Ae. speltoides (BB)**
- **T. urartu (AA)**
- **T. tauschii (DD)**
- **T. spelta (AABBDD)**
- **T. turgidum (AABB)**
- **T. urartu (AA)**
- **T. turgidum (AABB)**
- **T. aestivum (AABBDD)**

- **Einkorn wheat**
- **Emmer wheat**
- **Durum (pasta) wheat**
- **Bread wheat**

USDA-ARS
Tomato

- Details of domestication are unclear: likely Ecuador/Peru and Mexico.
- “Pre-domesticate” was a cherry tomato.
- Brought to Europe by Columbus or conquistadors <1544.
- Model system for genetic studies.

Wild Ancestor: *Solanum pimpinellifolium*

Pre-domesticate: *S. lycopersicum* ssp cerasiforme

Domesticate: *S. lycopersicum* ssp lycopersicum

PI 634844  
PI 647555  
PI 347239
2. Plant Breeding and Biotechnology

Plant Breeding: “The art and science of plant improvement”
Biotechnology: “The use of living systems and organisms to develop or make useful products”
First Plant Breeders

- Thousands of years ago
  - Likely women who selected plants that they liked.
- Hundreds of years ago
  - Farmers had their favorite “family” variety = “heirloom.”
- 1926 Pioneer Hi-Bred was founded
  - First commercial hybrid corn seed company.
Plant Breeding

- Selection (artificial) of plants with desirable traits.
  - Often done without knowledge of the genes controlling the trait(s).

- Decreasing Unfavorable Alleles.
  - Susceptibility to diseases and insects, susceptibility to drought, unadaptedness.

- Increasing Favorable Alleles.
  - Increased yield and better quality, resistance to disease, insects and environmental stresses.

- This process has typically produced safe food for over 100 years.
Plant Breeding

• Changing the phenotypic mean of a population; two methods

Select the “best” individuals to form the next generation

Original population

Select the “best” individuals to form the next generation

Progeny population

Parent 1
Low yield
Insect resistant

Parent 2
High yield
Susceptible

F1

F2

New Inbred Line
High yield
Insect resistant
Plant Biotechnology

- Requires knowledge of the gene underlying the trait.
- Gene(s) excised from donor species, combined with promoter of choice to drive gene expression.
- The construct is transformed into recipient plant.
- Backcrossed six times to recover 99.2% recurrent parent.

**Traditional Plant Breeding**
DNA is a strand of genes, much like a strand of pearls. Traditional plant breeding transfers many genes at once.

**Plant Biotechnology**
Using plant biotechnology, you can add a single gene to the strand without the extra baggage (linkage drag).
3. Maize Case Study

What Can We Learn From Maize Diversity Using Breeding, Genetics, and Genomics?

- Maize is an important crop!
- Domestication and Diversity
- Inbred Lines
- Teosinte, the Wild Ancestor
Maize diversity is greater than the difference between human and chimps.

Silent Diversity; Tenallion, et al. (2001) *PNAS*
Maize Domestication

Domesticated from *Zea mays* ssp. *parviglumis* ~9,000 years ago in SW Mexico. Intermediate form of landraces, i.e., populations adapted to specific microclimates and/or human uses.
The Maize Genome

- Ten chromosomes.
- Ancient tetraploid – whole genome duplication ~ 5-12 mya.
- B73 is the reference inbred for the genome.
- 32,500 predicted genes in B73 – more likely ~40,000.
- Nearly 85% of the B73 genome sequence is annotated as transposable elements.
- B73 may capture only ~70% of the low-copy gene fraction of all maize inbred lines.

Schnable, et al. (2009) Science
Genome-Wide Diversity

- Maize HapMap.v1, HapMap.v2, HapMap.v3
  - 83 million SNPs
- Two historically important inbred lines in breeding (B73 and Mo17) differ by more than 20 million SNPs!
- Related species exhibit synteny (same gene order) along chromosomes.
- However, we see lots of non-collinearity in maize!

Domestication Genes in Maize

• *teosinte branched 1* – plant branching
• *teosinte glume architecture 1* - hard fruitcase

• Whole genome analysis of HapMap.v2
  – Stronger selection during domestication than improvement
  – > 1000 genes experienced selection

Impact of Artificial Selection on Diversity

- **Teosintes**
  - Unselected (Neutral) Gene
  - Domestication Gene
  - Improvement Gene

- **Domestication**

- **Artificial Selection**
  - Maize Landraces
  - Plant Breeding

- **Maize Inbred Lines**
  - 98% (~49,000) maize genes
  - 2% (~1,000) maize genes

- **Commercial Corn Breeding**
Diverse Maize Inbred Lines Nest ed Association Mapping (NAM)
Linking Phenotype to Genotype

- Can we identify the genes controlling phenotypic variation?

- QTL (Quantitative Trait Locus) Mapping
  - Identify genomic regions that contribute to variation
  - Very loose definition of a large region

- Association Mapping
  - Test specific genes for their involvement in the trait
  - Much more precise way to test (if done correctly!)
Association Panel

- 302 Inbred Lines from around the world

Flint-Garcia, et al. (2005) *Plant J.*
NAM Founders

- 25 Diverse Lines chosen to maximize diversity

Nested Association Mapping

- NAM combines the statistical power and mapping resolution from two different QTL mapping methods
  - Explore diversity from around the world
- Used to study many different traits in maize

Yu, et al. (2008) *Genetics*
McMullen, et al. (2009) *Science*
NAM Kernel Composition

Jason Cook

Kernel Composition QTL Mapping in Maize

“QTL for kernel oil content somewhere in here.”

Genome Wide Association (GWAS)

1.6 Million HapMap.v1 SNPs projected onto NAM

Getting closer to saying “This gene controls oil content”

Teosinte - The Wild Ancestor
Development of Teo Introgression Libraries

Backcrossing of the F1 with B73 to create BC1, BC2, BC3, BC4; Each backcross removes 50% of the teosinte genome.

Liu, et al. (2016) *Plant Genome*
Kernel Composition QTL in Teosinte

Teosinte NILs

Allelic effects (% starch)

- Z029: 0.62
- Z030: -0.21
- Z031: -0.93
- Z032: 0.02
- Z033: -0.06
- Z034: 0.82
- Z035: 0.41
- Z036: -1.58
- Z037: -0.04
- Z038: -2.56

Karn et al. (2017) G3
**Application: Reintroduce Variation**

- **Biological hypothesis:** A loss of genetic variation results in a loss of phenotypic variation.
- **Breeding hypothesis:** We can improve modern corn by introducing variation.

### Selected Genes in Maize Starch Pathway

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene Name</th>
<th>Possible Target Trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRMZM2G348551</td>
<td>Su2; Sugary 2</td>
<td>Starch</td>
</tr>
<tr>
<td>GRMZM2G394450</td>
<td>Ivr1; Invertase 1</td>
<td>Starch</td>
</tr>
<tr>
<td>GRMZM2G089836</td>
<td>Ivr2; Invertase 2</td>
<td>Starch</td>
</tr>
<tr>
<td>GRMZM2G110175</td>
<td>Bm1; Brown midrib 1</td>
<td>Starch</td>
</tr>
<tr>
<td>AC196475.3_FG004</td>
<td>Bm3; Brown midrib 3</td>
<td>Starch</td>
</tr>
<tr>
<td>GRMZM2G098298</td>
<td>Ccp1; Cysteine protease 1</td>
<td>Protein</td>
</tr>
<tr>
<td>GRMZM2G138727</td>
<td>Zp27; 27-kDa zein protein</td>
<td>Protein &amp; Amino Acids</td>
</tr>
<tr>
<td>GRMZM2G087612</td>
<td>SDP1; Sugar dependent1</td>
<td>Oil</td>
</tr>
</tbody>
</table>

*Schnable and Freeling (2011) Plos One

*Schnable and Freeling (2011) Plos One

*Whitt et al. (2002)*

---

**Sucrose**

- **Sh1**: sucrose synthase

**UDP Glucose**

- **Sh2, Bt2**: ADP-glucose pyrophosphorylase

**ADP Glucose**

**Amylose**

- **Wx1**: GBSS

**Amylopectin**

- **Su1**: DBE
- **Ae1**: SBE-IIB

*Schnable and Freeling (2011) Plos One*
Do Teosinte Alleles Have Value for Improving Today’s Corn?

Traits that were relevant 9000, 1000, or 100 years ago may not be as relevant today.

Therefore the alleles selected 9000, 1000, or 100 years ago that persist in modern germplasm may not be optimal today.
Conclusions: Impact of Selection in Maize

- Selection on the genome
  - Thousands of years during domestication
  - Over one hundred years during plant breeding
  - 25+ years during plant biotechnology

- Modification of seed composition
  - Traditional breeding
  - Biotechnology

- Never identified a deadly/toxic variety of corn
  - Despite tremendous genetic and phenotypic variation
  - Long history of safe use
Surveys of Metabolites in Diverse Maize Germplasm

27 Nested Association Mapping (NAM) founder inbreds
27 Landrace (LR) inbreds created by selfing landraces
Both sets crossed with B73 to make hybrid seed
Genetic Relationships Among the Maize Lines

NAM Hybrids

Landrace Hybrids

Compositional Variation in Diverse Maize

Compositional Variation in Diverse Inbred Lines

- Wide range in proximates among inbred lines

Compositional Variation in Diverse Inbred Lines

- Two fold difference in some fatty acids among inbred lines

Inbred lines versus landraces: Gamma-Tocopherol

- Nearly six-fold difference in gamma-tocopherol across inbred lines and landraces

Inbred lines versus landraces: Beta-Carotene

- Huge range of variation for beta-carotene

Metabolite Profiles by GC-MS and NMR

- Same samples as previous study
- GC-MS and NMR were used to examine metabolite profiles
- 675 total metabolites collected by GC-MS
  - 504 met QC checks; 174 were annotated
- 33 total metabolites collected by NMR
  - 17 amino acids, 7 organic acids, 5 sugars, 4 others
Partial least-squares discriminant analysis (PLS-DA) of the GC-MS data

Another source of variation: Xenia Effect (Pollen)

- Corn is a cross-pollinated species (pollen travels in the air)

---

R-locus  
Y-locus  
Sh2-locus (sweet corn)  
Ornamental Corn
Xenia Effect (Pollen Donor) on Amino Acids

Zeins are the storage proteins in the endosperm.

Zein Profiles

β  γ  α family  δ

Flint-Garcia et al. (2009) *Theor App Genet*

Xenia Experiment

Unpublished data, Flint-Garcia (USDA) and Angelovici (University of Missouri) labs
To Summarize

- Maize is extremely genetically diverse.
- Maize is extremely phenotypically diverse.
  - Plant-level $\rightarrow$ $\rightarrow$ metabolite level
- Genetics, Environment, GxE contribute to phenotypes.
- Many genes contribute to grain composition.
- Even pollen source affects grain composition.
To Summarize

- For geneticists and breeders, the more variation the better!

- When a farmer plants their corn, every hybrid seed is genetically the same. When the farmer harvests the seed, every seed is different.
References

- Chia, et al. (2011) Nature Genetics 40:803-807
- Fu and Dooner (2002) PNAS 99:9573–9578
- Liu, et al. (2016) Plant Genome 9
Appendix Slides: Overview of Genetics & Breeding

- Genetics: “The study of heredity and the variation of inherited characteristics”
- Breeding: “The art and science of the genetic improvement of plants”
Genetics Terms

- **Genotype** - The genetic constitution of an individual organism
- **Phenotype** - The set of observable characteristics of an individual resulting from the interaction of its genotype with the environment
- **Trait** - A genetically determined characteristic
- **Allele** - An alternative form of a gene
- **Genetic Variation** - Naturally occurring genetic differences among individuals of the same species
Genetics 101

(a) primary structure - amino acids in linear order
(b) secondary structure – helices or sheets of amino acids
(c) tertiary structure – folding of the protein into a shape
(d) quaternary structure – multiple proteins form a complex

Nucleotides
Base Pairs
Sugar-phosphate backbone

DNA Transcription RNA Translation Protein

Glutamine
Glycine
Leucine
Threonine
Histidine
Arginine
Serine

CAGGGA
CACGUA
CAGGUA
CGCAUA
CGCAUA
CGCAUA
CGCAUA
Where Does Genotypic Variation Come From?

- Mutations
  - Point mutations – Change or delete base pairs
  - Chromosomal – Polyploidy and larger scale mutations

- Transposable Elements – “Jumping Genes”
  - Can completely disrupt gene function

- Hybridization
  - Two different genomes come together

- Recombination
  - Shuffling of genes during meiosis
DNA Sequence Variation

SNP: Single Nucleotide Polymorphism

InDel: Insertion/Deletion
Where Does Phenotypic Variation Come From?

- **Genotype**
  - Why fraternal twins look different
- **Environment**
  - Why identical twins look different
- **Genotype x Environment**
  - Complex interactions between genetics and the environment
How Does Variation Change?

- Usually variation (allele frequencies) is balanced, i.e., “in Hardy-Weinberg Equilibrium.”
- Genetic Drift
  - Sampling error – Rarer alleles can disappear completely and thereby reduce genetic variation.
- Selection – Natural or Artificial
  - Increased survival and reproduction of some individuals
- Gene Flow
  - Migration of individuals, bringing new alleles together
Types of Variation

Qualitative = Discrete

Quantitative = Continuous

Male sterile 8

Cob Color

Albino seedling

dwarf 8

Seed Composition

Native Rootworm Resistance

Carotenoid (provitamin A)

Yield

Fiber & Ash 13%
Oil 4%
Protein 8%
Water 14%
Starch 61%
Quantitative Trait Locus (QTL) Mapping

- “Genome Scan”
- Identify genomic regions that contribute to variation and estimate QTL effects

\[
\begin{align*}
\text{Parent 1} & \quad \text{F}_1 & \quad \text{Parent 2} \\
\text{F}_2 & &
\end{align*}
\]

\[
\begin{align*}
\text{Genotype} & \quad \text{Phenotype} \\
\text{Composite} & \quad \text{Interval Mapping}
\end{align*}
\]
Association Analysis

- “Genome Scan”
- Utilizes natural populations
  - Exploit extensive ancestral recombination

<table>
<thead>
<tr>
<th>Line</th>
<th>Gene ABC</th>
<th>1.3m</th>
<th>1.4m</th>
<th>1.5m</th>
<th>1.8m</th>
<th>2.0m</th>
<th>2.5m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line 1</td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Line 2</td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>A</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Line 3</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Line 4</td>
<td>C</td>
<td>T</td>
<td>A</td>
<td>G</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>A</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>2.0m</td>
<td>2.5m</td>
</tr>
<tr>
<td>Line 300</td>
<td>A</td>
<td>C</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Association Analysis**

- **“Genome Scan”**
- Utilizes natural populations
  - Exploit extensive ancestral recombination

<table>
<thead>
<tr>
<th>Line</th>
<th>Gene ABC</th>
<th>Sequences (Mbp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A C G A G</td>
<td>1.3m</td>
</tr>
<tr>
<td>2</td>
<td>A C G A T</td>
<td>1.4m</td>
</tr>
<tr>
<td>3</td>
<td>A T A A G</td>
<td>1.5m</td>
</tr>
<tr>
<td>4</td>
<td>C T A G T</td>
<td>1.8m</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>A T G G T</td>
<td>2.5m</td>
</tr>
</tbody>
</table>

Gene ABC: A C G A G
Nested Association Mapping (NAM)

**Linkage (QTL) Mapping**
- Genome scan
- Structured population
- High power
- Low resolution
- Analysis of 2 alleles

**Association Mapping**
- Candidate gene testing
- Unstructured population
  - Low power
  - High resolution
- Analysis of many alleles

Nested Association Mapping (NAM)
- Structured families nested within an unstructured population
- High power
- High resolution
- Analysis of many alleles