Abstract: Crop productivity is a function of basic component traits. Grain yield in maize is determined by the product of the number of ears per hectare, the number of seeds per ear, and seed weight. Stover yield is a function of components including node number, internode length, stalk diameter, and leaf shape and number. We are using sequence-based expression and genotyping in structured populations, collections of diverse lines, and long-term selection populations to characterize genes and alleles underlying natural variation for productivity traits in maize used for food, feed, fiber, and raw materials such as for biofuel. As an example of the approaches that we are using, and as a basis to discusses synergies and challenges of various technologies, I will describe interpretations based on phenotypic and genetic analysis of seed size. Our analyses to date are consistent with 1) a significant pollinator effect on seed size, 2) an important role for the maternal plant in determining seed weight and synchronizing components of development, and 3) pleiotropic effects of some genes on overall plant and seed size.
Kaepler, Shawn
Illinois Corn Breeding School
March 5, 6
35 minutes
Genomic analysis of natural variation for seed and plant size in maize

Shawn Kaeppler, Department of Agronomy, University of Wisconsin-Madison
Key Collaborators: Natalia de Leon (UW), Robin Buell (MSU), Nathan Springer (U MN), Dan Rokhsar and Uffe Hellsten (JGI)
Overview

• QTL characterization using structured populations and analysis of long-term selection

• Example Trait: Seed size/weight
  – Experimental Platforms and Initial Results
    • Analysis of Krug Large and Krug Small populations
      – Expression
      – Allelic variation
    • Genetic mapping

• Discussion of Pros and Cons of Experimental Approaches
Seed Weight: A Core Yield Component

• Yield = Product of:
  – Plants / unit area
  – Ears / plant
  – Seeds / ear
  – Seed weight

• Target of domestication

Affected by or determined by planting density

National Geographic image
Krug Seed Size Selection Program

• Initial population: Krug
  – Combination of Iowa and Illinois Strains of Reid Yellow Dent and Goldmine

• 30 cycles of divergent selection for seed size/weight
  – Selection criterion in early cycles was visual assessment of seed size, transitioned to seed weight in later cycles
Krug Seed Size Selection Program

• Selection method: Phenotypic Mass Selection

• Year 1: 3500 plants at 40,000 plants/ha
  • Divergent pools selected based on visual observation of ears with uniformly large and small seeds

• Subsequent generations: 100+ ears selected from 1200 to 1500 plants
  • Selection based on size / weight of seeds in center of fully-pollinated ear
    – Note: Selection was based on uniform seeds on an ear (not segregants) and program was conducted without pollen control (in isolation)
  • ~7% selection intensity
Genome-wide analysis of the Krug Seed Size divergent selection program
Changes in Seed Traits

200 Kernel Weight

- **KLS30**
- **Krug C0**
- **KSS30**

**Kernel Depth**

- **KLS30**
- **Krug C0**
- **KSS30**

**Number of Kernel Rows**

- **KLS30**
- **Krug C0**
- **KSS30**
KLS plants are larger

Stalk girth
“My experiment is to prove that...”

• (My) Initial hypotheses
  – Low LD – high resolution
  – Sink strength driven primarily by starch accumulation will be important
  – Divergent allele change at important loci
  – Genomic analysis will have resolution to identify causal SNPs
    • Reinforced by structured populations and diverse inbred studies
Maternal and Xenia Effects

Fertilized by KLS pollen

Fertilized by KSS pollen
Maternal and Xenia Effects

KLS pollen on KSS ear
- 0.21 g/seed

KSS pollen on KSS ear
- 0.11 g/seed

KLS pollen on KLS ear
- 0.39 g/seed

KSS pollen on KLS ear
- 0.26 g/seed

Reciprocal/Maternal

Xenia/Dosage

Xenia/Dosage
Maternal influence on modifiers of su1 endosperm
Expression Study

• Samples
  – 8 day after emergence seedlings
  – 3, 6, 12, 21, 24 day after pollination developing seeds
  – 3 bioreps each composed of samples from 10 plants from the populations

• NimbleGen Array – all time points
  – Described in Sekhon et al. Plant J

• RNAseq – 12 dap developing seeds only
Microarray study: Greatest difference between KSS and KLS at 12 DAP

Consistent with early maturation of seeds in KSS
Genome-wide diversity analysis of selected populations

• Whole genome sequencing of pools of 46 individuals from each cycle (Illumina Hi-Seq)
  – Goal 50X depth

• Pioneer Hi-Bred public 768 Illumina GoldenGate assay
Drift versus selection in Krug and Golden Glow

Krug / Golden Glow:
~ 200 females, 1000 + males
30 cycles

What if ...:
~ 20 females, 50 males
30 cycles
Shawn’s first sequencing experiment
ca. 1990

- 3 day experiment
- Lots of trauma
- ~300 bps
Whole Genome Sequencing

- Illumina Hi-Seq paired end 100 bp reads
  - KLS: 119 billion base pairs = 51.8X
  - KSS: 187 billion base pairs = 81.6X
  - KC0: 163 billion base pairs = 71.1X
- Single-end (results shown today) and paired-end read mapping
  - SNPs
  - Indels with paired-end
- ~40 million SNPs identified
  - Focused on 15,380,471
    - Were filtered to require at least two reads of each SNP base to reduce those due to sequencing error
Comparative Genome Hybridization

• Materials
  – 3 KLS inbreds, 3 KSS inbreds
    • Conceptually, 3 sampled gametes from each strain

• Comparative Genome Hybridization using a NimbleGen microarray
  – Details in Eichten et al. Plant Phys 156:1679

• Compared copy number of low copy regions across the genome to B73

• Computed difference in KLS average – KSS average
768 Illumina and whole-genome SNP results

• I will show you but, arghhh...
  – Too many points to visualize easily
  – Still interpreting
  – Need to understand some complexities before interpreting results

• Brief summary
  – Few (relatively) positions significant
    • Read-depth not uniform – affects significance threshold!
  – Few alleles fixed after 30 cycles
Allele Frequency change in KLS via Illumina Golden Gate: Chrom 1, 10,000 loci

Expect loci with moderate initial frequency to respond the most to selection.
Allele Frequency change in KLS via Illumina Golden Gate
Some specific examples and observations
Candidate Genes – “Classics”

• Controls
  – a1, a2, bz1, c1, c2, p1

• Carbohydrate – starch
  – mn1, bt1, bt2, sh2, su, ae1, du1, ...

• Imprinting / epigenetics
  – fie1, dmt101, rmr1
Controls

Gene = a1
Variance of frequency change = 0.0147

Gene = a1
Variance of frequency change = 0.0336

Gene = bx1
Variance of frequency change = 0.00851

Gene = bx1
Variance of frequency change = 0.0164

Gene = c1
Variance of frequency change = 0.0354

Gene = c1
Variance of frequency change = 0.0218

Gene = c2
Variance of frequency change = 0.0104

Gene = c2
Variance of frequency change = 0.0431
Carbohydrate, Starch

Gene = mn1
Variance of frequency change = 0.0243

Gene = mn1
Variance of frequency change = 0.0151

Gene = bt2
Variance of frequency change = 0.0998

Gene = bt2
Variance of frequency change = 0.188

Gene = sh2
Variance of frequency change = 0.0167

Gene = sh2
Variance of frequency change = 0.0224

AGPase subunits
Carbohydrate, Starch

Gene = bt1
Variance of frequency change = 0.018

Gene = bt1
Variance of frequency change = 0.0165

Gene = ae1
Variance of frequency change = 0.0176

Gene = ae1
Variance of frequency change = 0.0174

Gene = du1
Variance of frequency change = 0.01

Gene = du1
Variance of frequency change = 0.00978
Imprinting / Epigenetics

- **fie1**
  - Gene = fie1
  - Variance of frequency change = 0.0253

- **mez**
  - Gene = mez1
  - Variance of frequency change = 0.0266

- **rmr1**
  - Gene = rmr1
  - Variance of frequency change = 0.0421
  - Variance of frequency change = 0.187
Integrated Genome Viewer Explanation

4 Megabase pairs (Mb) – Scale in base pairs

- Frequency of minor allele in Cycle 0
- KLS – KC0 variance
- KSS – KC0 variance
- KLS – KC0 frequency
- KSS – KC0 frequency
- KLS – KSS relative copy number

CGH (KLS-KSS): 3 to -3

CO frequency: 0 to 0.5
KLS variance: 0 to 0.3
KSS variance: 0 to 0.3
KLS frequency: 1 to -0.5
KSS frequency: 1 to -0.5

ae 1
Linkage blocks
Chromosome 5 – Example non-centromeric region

4 Megabase pairs (Mb)

- CO frequency: 0 to 0.5
- KLS variance: 0 to 0.3
- KSS variance: 0 to 0.3
- KLS frequency: 1 to -0.5
- KSS frequency: 1 to -0.5
- CGH (KLS-KSS): 3 to -3

ae 1
Chromosome 8 – 20 Mb centromere region

22 Megabase pairs (Mb)

CO frequency: 0 to 0.5

KLS variance: 0 to 0.3

KSS variance: 0 to 0.3

KLS frequency: 1 to -0.5

KSS frequency: 1 to -0.5

CGH (KLS-KSS): 3 to -3

Centromere
Chromosome 10 – 30 Mb centromere region

30 Megabase pairs (Mb)

- C0 frequency: 0 to 0.5
- KLS variance: 0 to 0.3
- KSS variance: 0 to 0.3
- KLS frequency: 1 to -0.5
- KSS frequency: 1 to -0.5
- CGH (KLS-KSS): 3 to -3

Centromere

Dull 1
Copy Number
Chromosome 5 – zoom in on ae1

0.3 Megabase pairs (Mb)

- C0 frequency: 0 to 0.5
- KLS variance: 0 to 0.3
- KSS variance: 0 to 0.3
- KLS frequency: 1 to -0.5
- KSS frequency: 1 to -0.5
- CGH (KLS-KSS): 3 to -3

ae 1
Chromosome 4 – High KLS, Low KSS

24 (Mb)

0.2 Megabase pairs (Mb)

CO frequency: 0 to 0.5

KLS variance: 0 to 0.3

KSS variance: 0 to 0.3

KLS frequency: 1 to -0.5

KSS frequency: 1 to -0.5

CGH (KLS-KSS): 3 to -3

Zoom In >>
Chromosome 8 – Strong selection response

C0 frequency: 0 to 0.5
KLS variance: 0 to 0.3
KSS variance: 0 to 0.3
KLS frequency: 1 to -0.5
KSS frequency: 1 to -0.5
CGH: 3 to -3

Zoom In >>
A few comments on technologies

• Illumina Golden Gate – by individual
  – Fixed SNPs / Limited number
  – Allows calculation of linkage
  – Allows estimation of heterozygosity

• Bulk sequencing – HiSeq – 50X depth
  – Unlimited SNPs
  – No linkage / zygosity calculation
  – Repeat analysis complex
  – 50X expected depth provides less than 20X median
    • Allele sampling a critical component of “drift”

• RNAseq – 30 plants bulked – 10 per rep
  – Combines expression and diversity
  – Confounds expression and diversity
Observed vs Expected Read Depth for Whole Genome Sequencing

Sequencing reads of the 452 SNPs that overlap with the filtered Pioneer 768 Public Plex

KC0 Reads

Expect 71.1 X

KLS Reads

Expect 51.8 X

KSS Reads

Expect 81.6 X
Initial trends / observations

• Carbohydrate “candidate” genes show little response to selection
  • Variation may have been reduced by domestication and previous selection

• Uncharacterized role of “maternal” genes in determining seed weight (and composition)
  • More difficult to screen for maternal effect mutants affecting whole ears than segregating seeds on an ear

• Timing of seed development appears to differ in small versus large-seeded types

• Plants larger in large-seed population
  • Drift, pleiotropy
Initial trends / observations

• Large effective size populations needed to separate drift from selection in long-term experiments
  • Not many existing selection populations with acceptable effective population size such as found in Krug and Golden Glow
  • Potential to genotype extreme individuals from very large populations
  • Sample depth of selected strains an important component of “drift”

• Technologies affect determination of parameters
  • Allele frequency
  • Linkage
  • Genotype Frequency / Zygosity
    – Inbreeding depression noted in KLS but not KSS
Collaborators and Sponsors

Robin Buell group – MSU:
Candy Hansey
Kevin Childs
Brieanne
Haining Lin

JGI:
Kerrie Barry
Dan Rokhsar
Uffe Hellsten

U of Minnesota:
Nathan Springer
Steve Eichten

DOE
USDA
American Seed Research Foundation
Pioneer Hi-Bred International, Inc.
Monsanto

UW Group – Seeds and Natural Variation
Natalia de Leon
Tim Beissinger
Jillian Foerster
Karl Haro von Mogel
James Johnson
German Muttoni
Rajan Sekhon
Bill Tracy, Leah Viesselman

United States Department of Agriculture

Great Lakes Bioenergy Research Center
National Institute of Food and Agriculture