Genomic analysis of natural variation for seed and plant size in maize

Dr. Shawn Kaepler
University of Wisconsin

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)

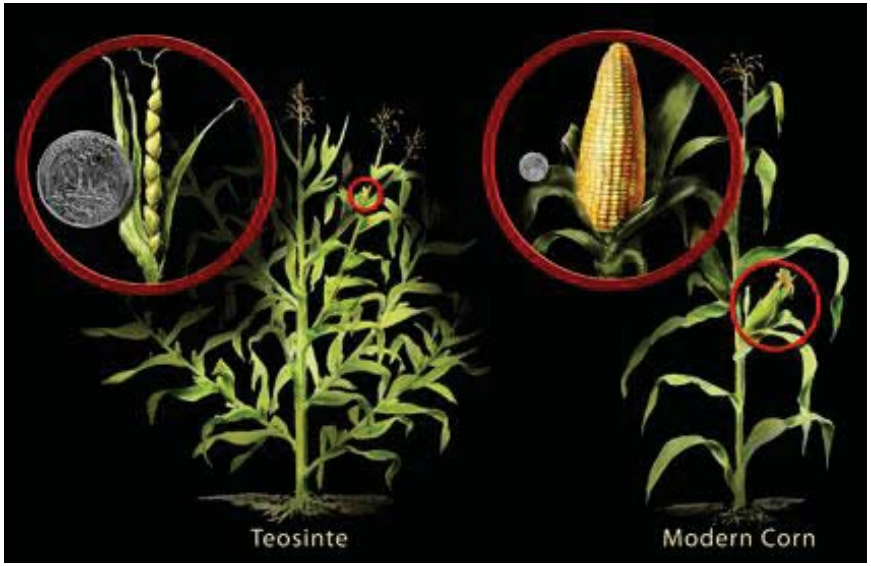
Overview

- 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

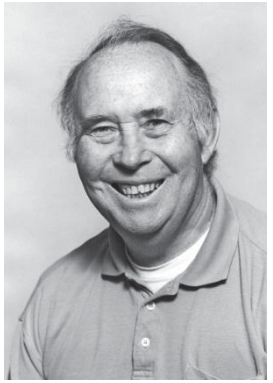
- Yield = Product of:
  - Plants / unit area
  - Ears / plant
  - Seeds / ear
  - Seed weight
- Target of domestication

Affected by or determined by planting density



National Geographic image

Krug Seed Size Selection Program


William Compton

- 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

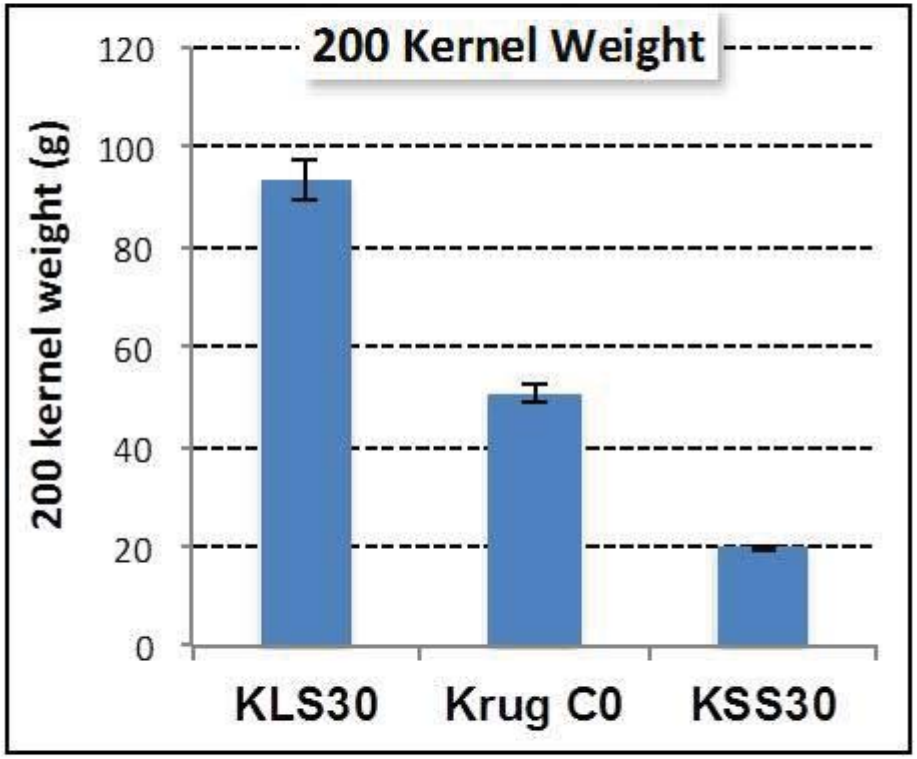
Krug Seed Size Selection Program

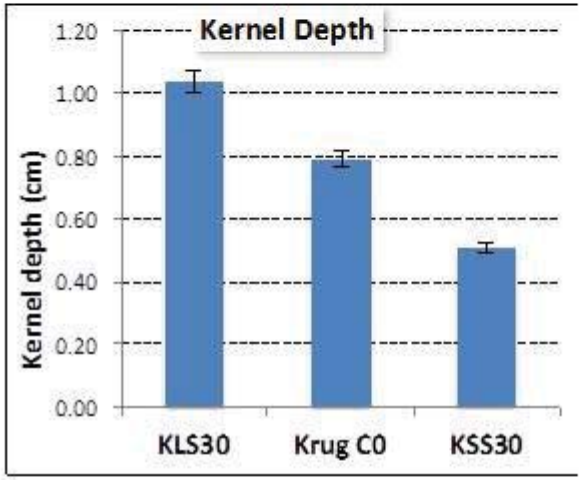
- Selection method: Phenotypic Mass Selection
- Year 1: 3500 plants at 40,000 plants/ha
    - Divergent pools selected based on visual observation of ears with uniformly large and small seeds
- Subsequent generations: 100+ ears selected from 1200 to 1500 plants
    - Selection based on size / weight 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

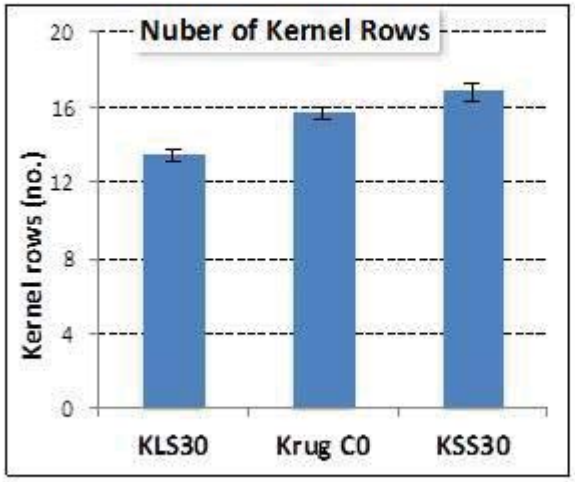
Genome-wide analysis of the Krug Seed Size divergent selection program

Krug C_{30} Large Seed (KLS) ← Krug C_0 (KC0) → Krug C_{30} Small Seed (KSS)

Changes in Seed Traits

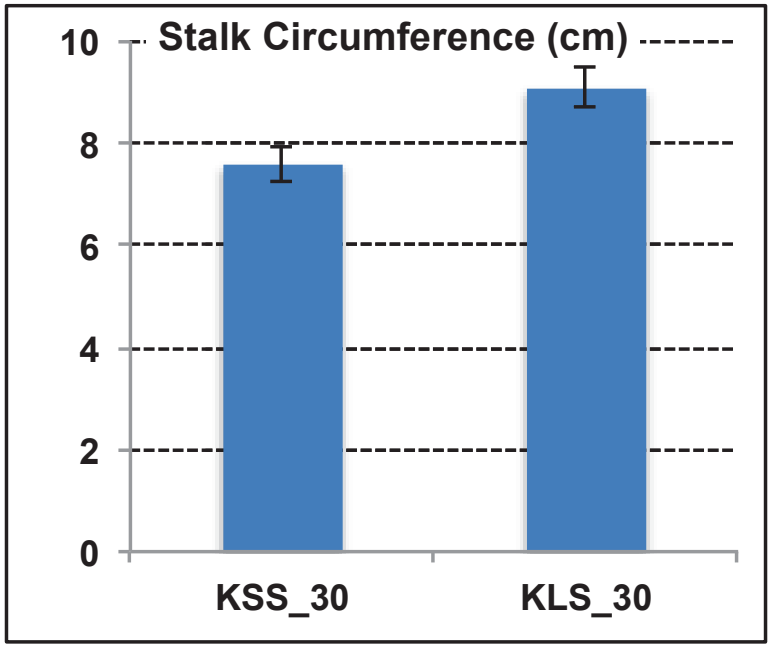






KLS plants are larger

KSS30 C0 KLS30

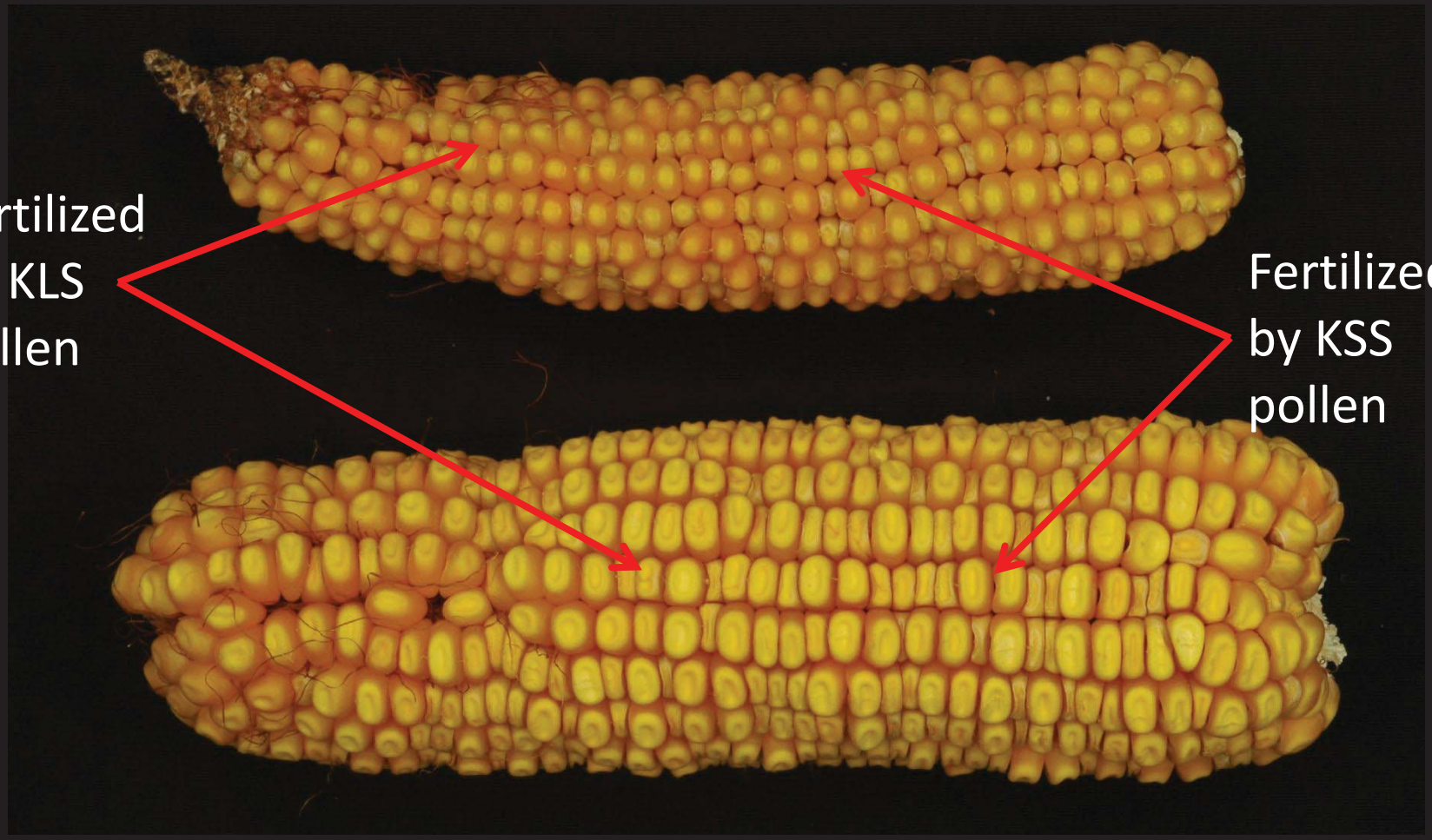


Stalk girth

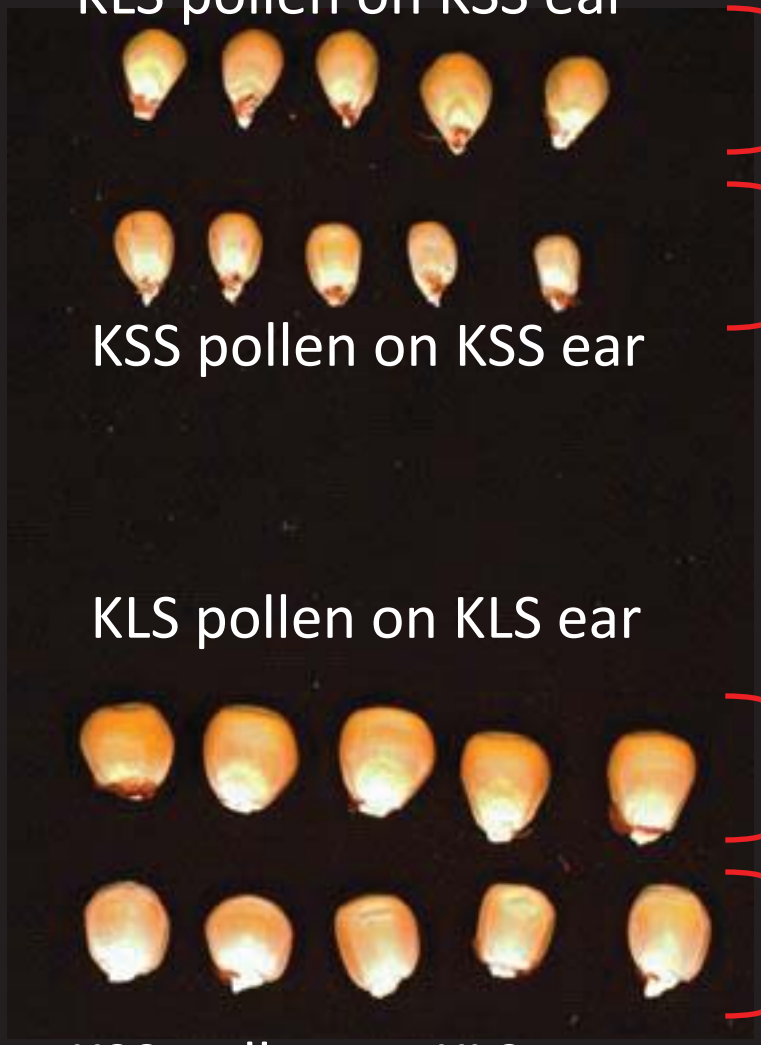
“My experiment is to prove that…”

- (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



Maternal and Xenia Effects



0.21 g/seed

0.11 g/seed

Xenia/Dosage

0.39 g/seed

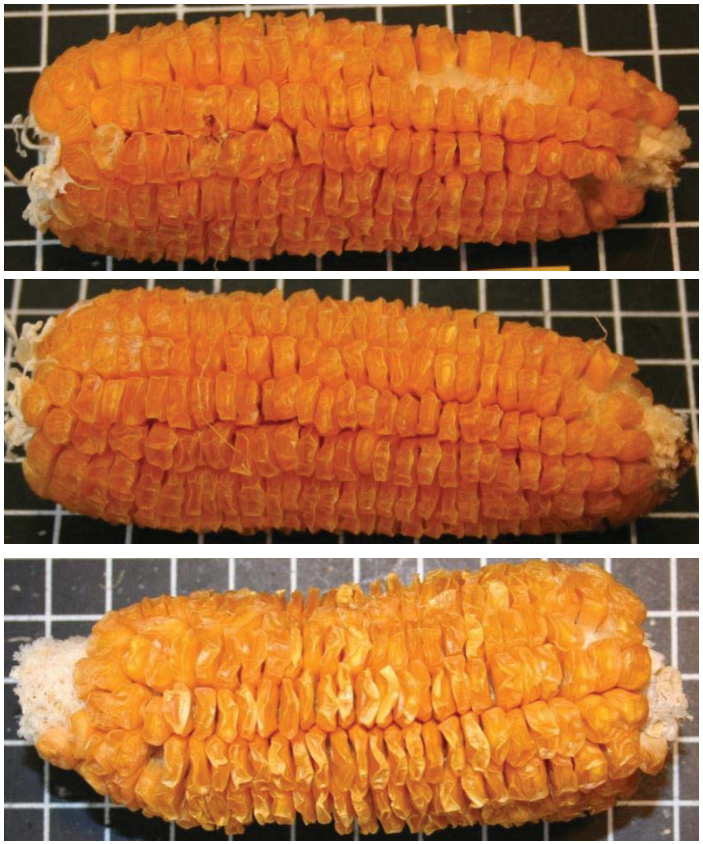
0.26 g/seed

Xenia/Dosage

Reciprocal/Maternal

Maternal influence on modifiers of su1 endosperm

Leah Viesselman



Sugary F_2 segregants

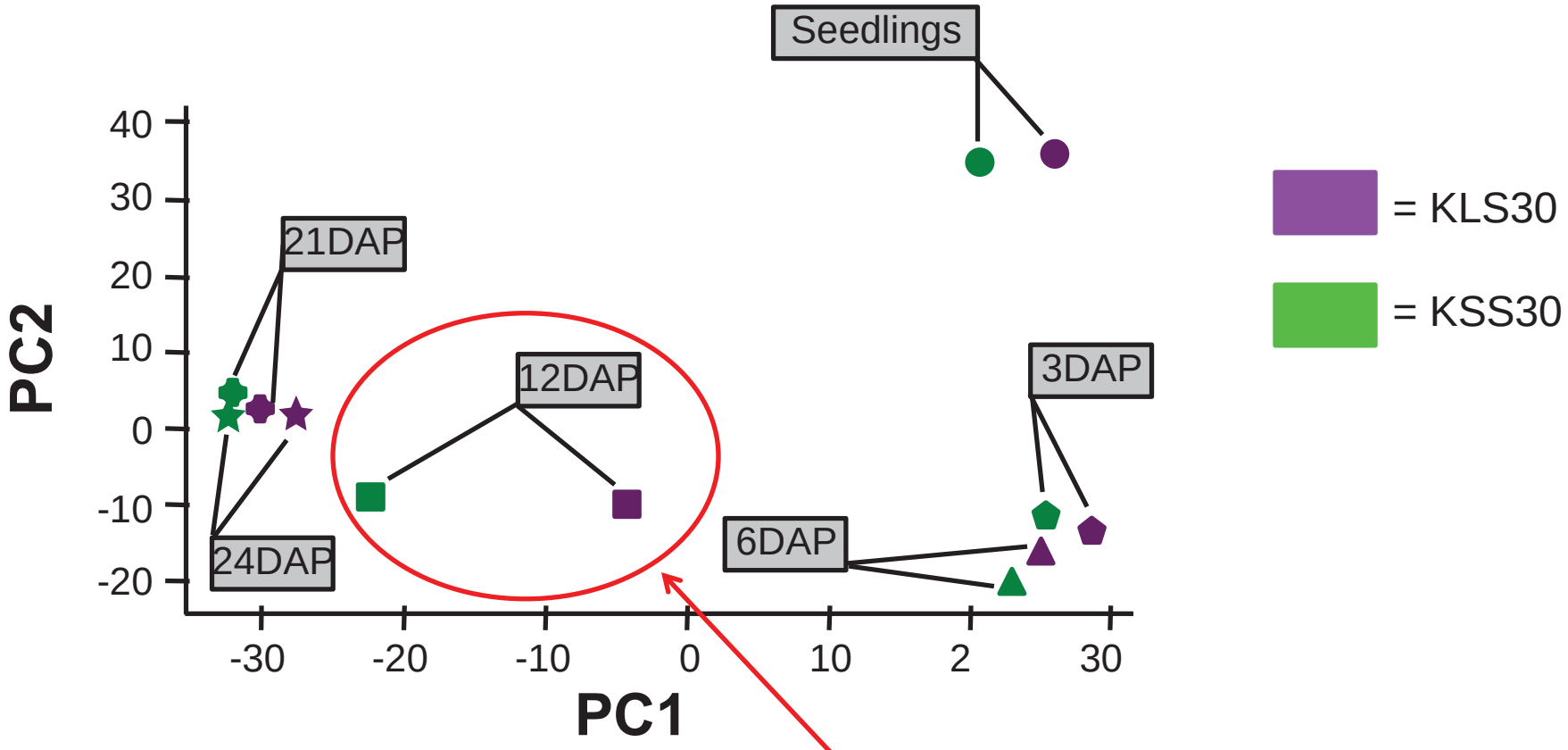
Pseudo starchy F_2 segregants

Expression Study

Rajan Sekhon

- 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

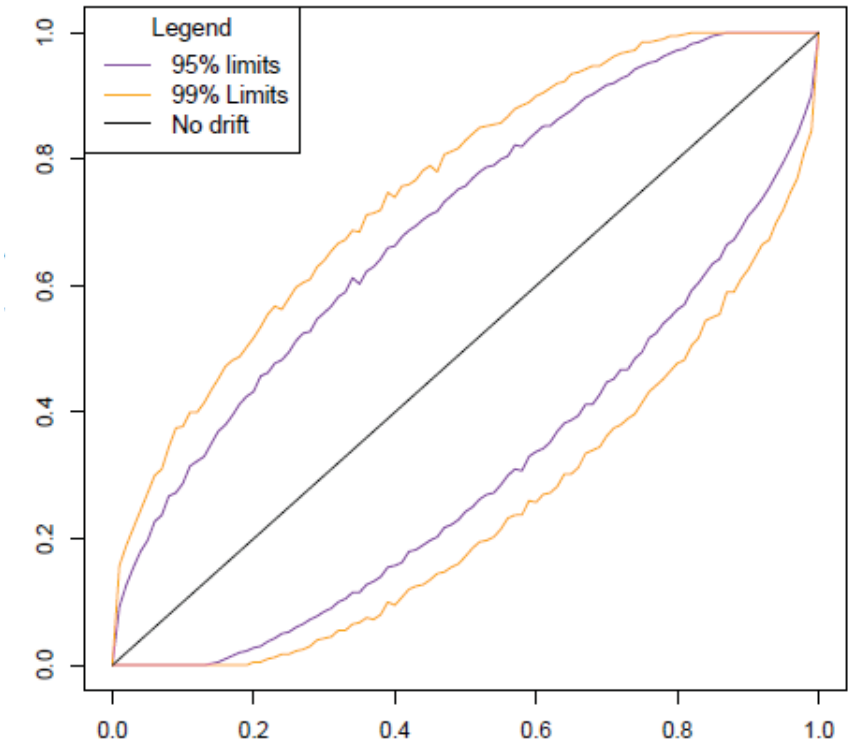


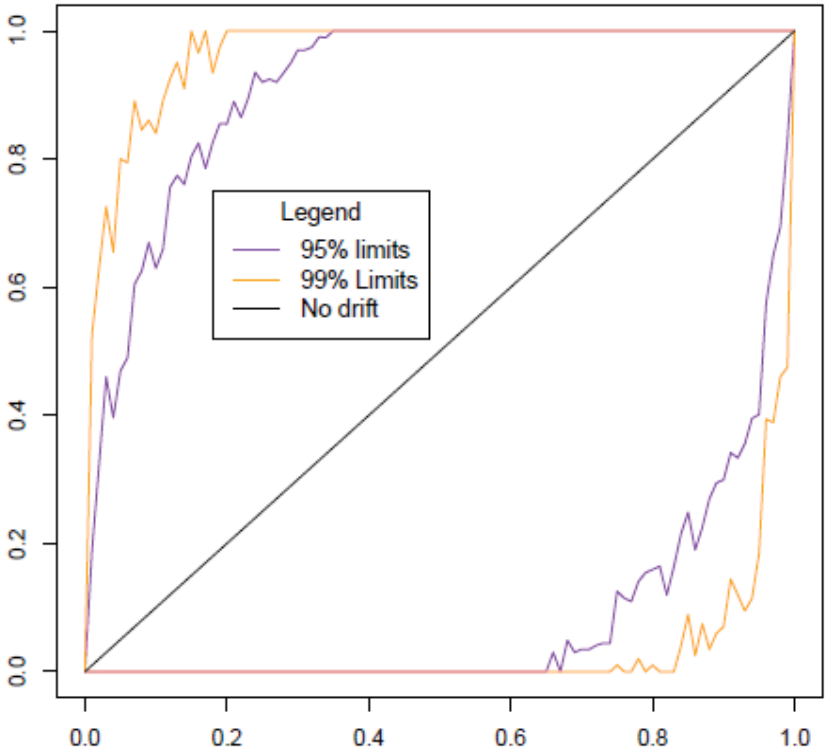
Consistent with early maturation of seeds in KSS

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

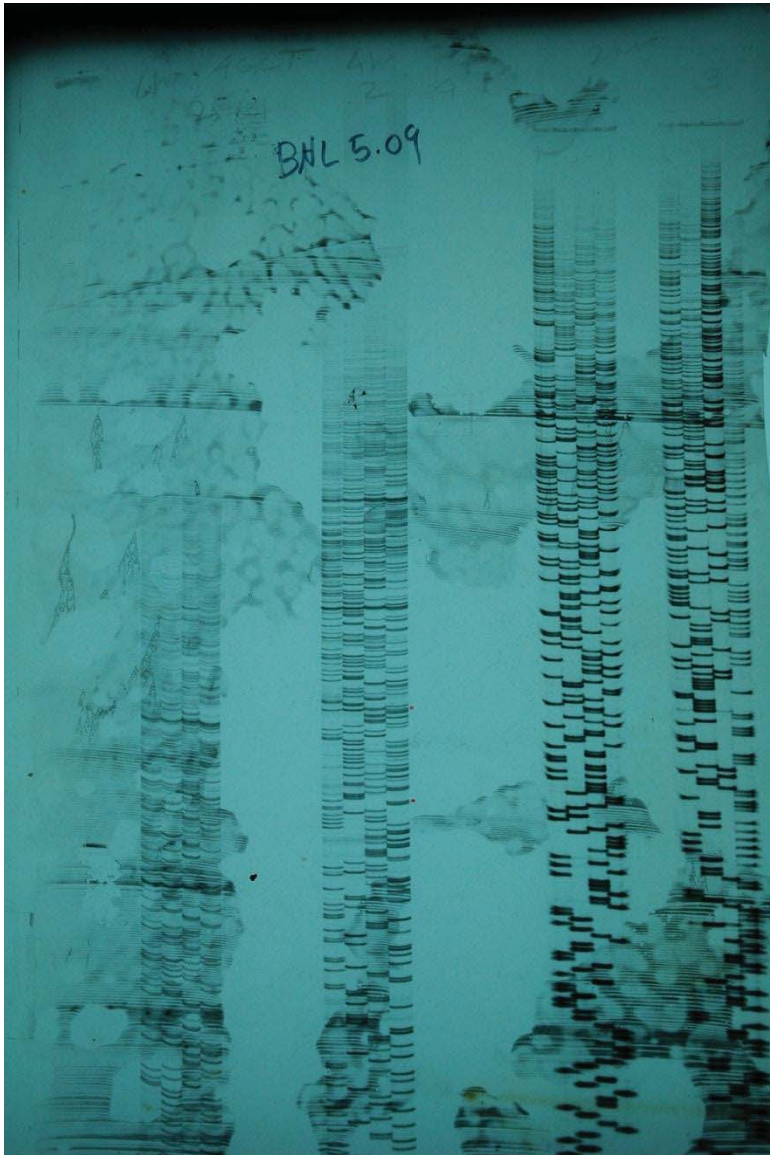


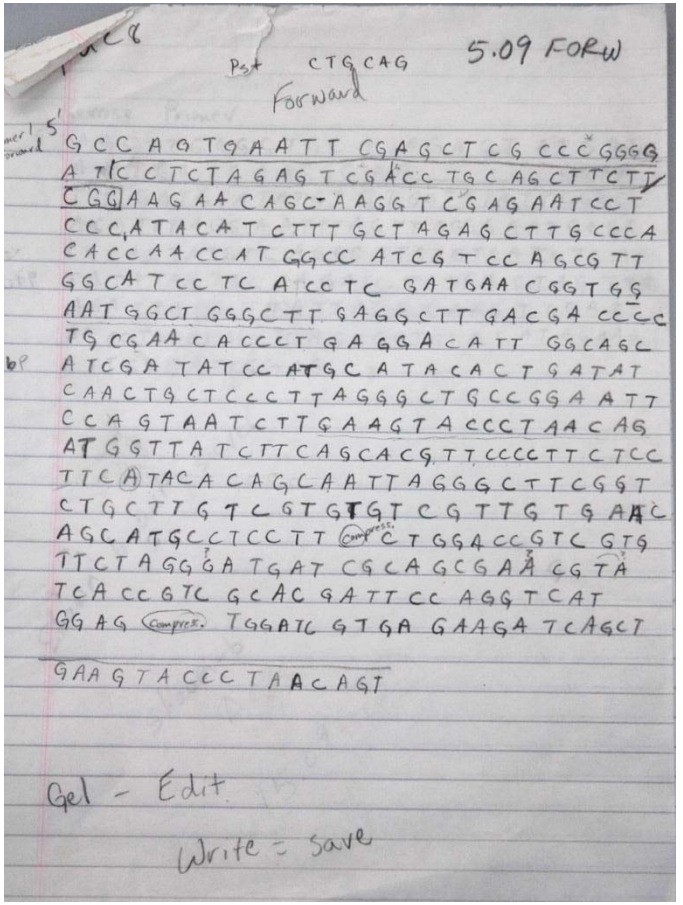


Krug / Golden Glow :
~ 200 females, 1000 + males
30 cycles

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


Candy Hansey

- 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

Steve Eichten

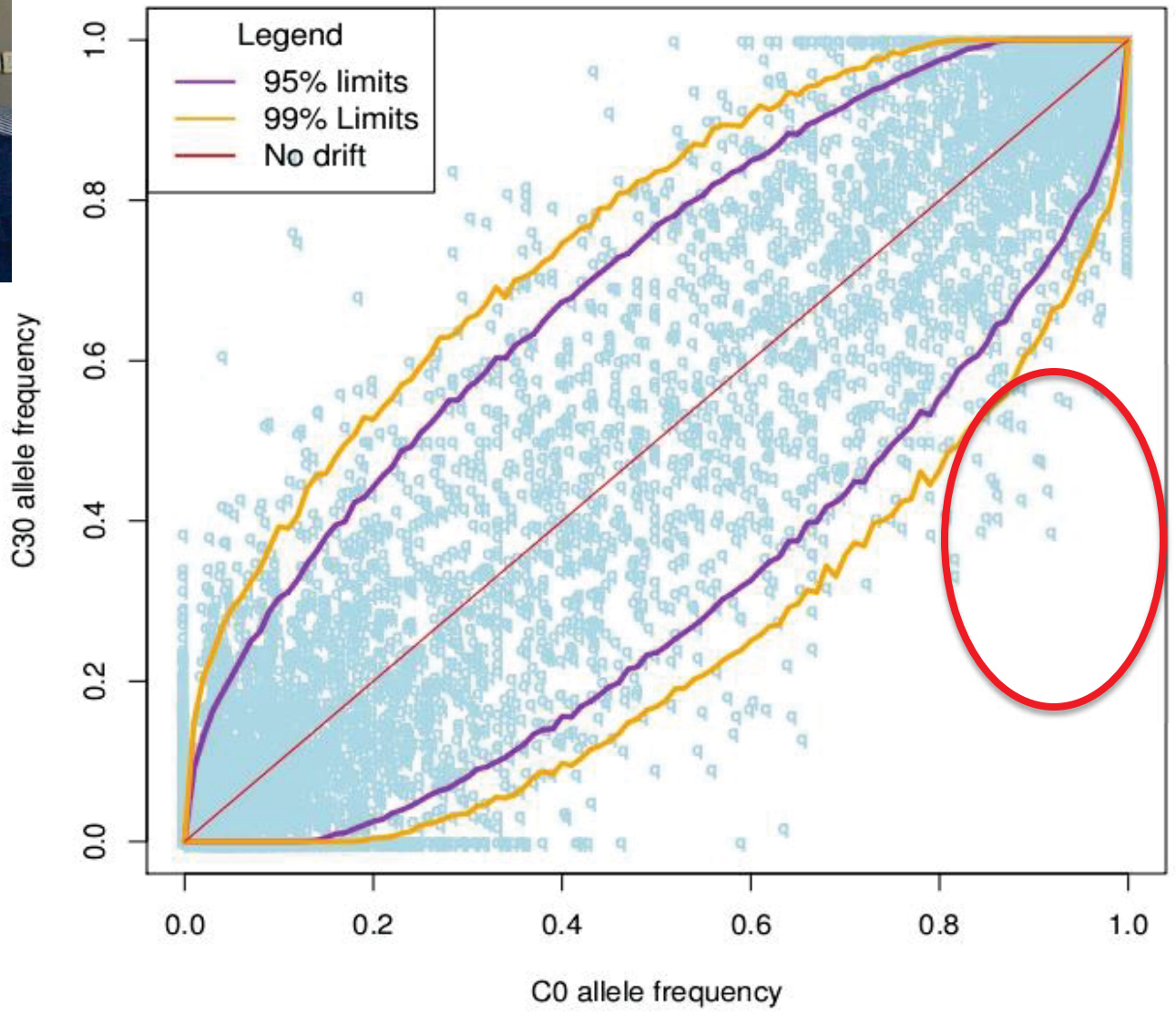
- Materials
  - 3 KLS inbreds, 3 KSS inbreds
    - Conceptually, 3 sampled gametes from each strain
- Comparative Genome Hybridization using a NimbleGen microarray
  - Details in Eichten et al. Plant Phys 156:1679
- Compared copy number of low copy regions across the genome to B73
- Computed difference in KLS average – KSS average

768 Illumina and whole-genome SNP results

- I will show you but, arghh…
  - 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

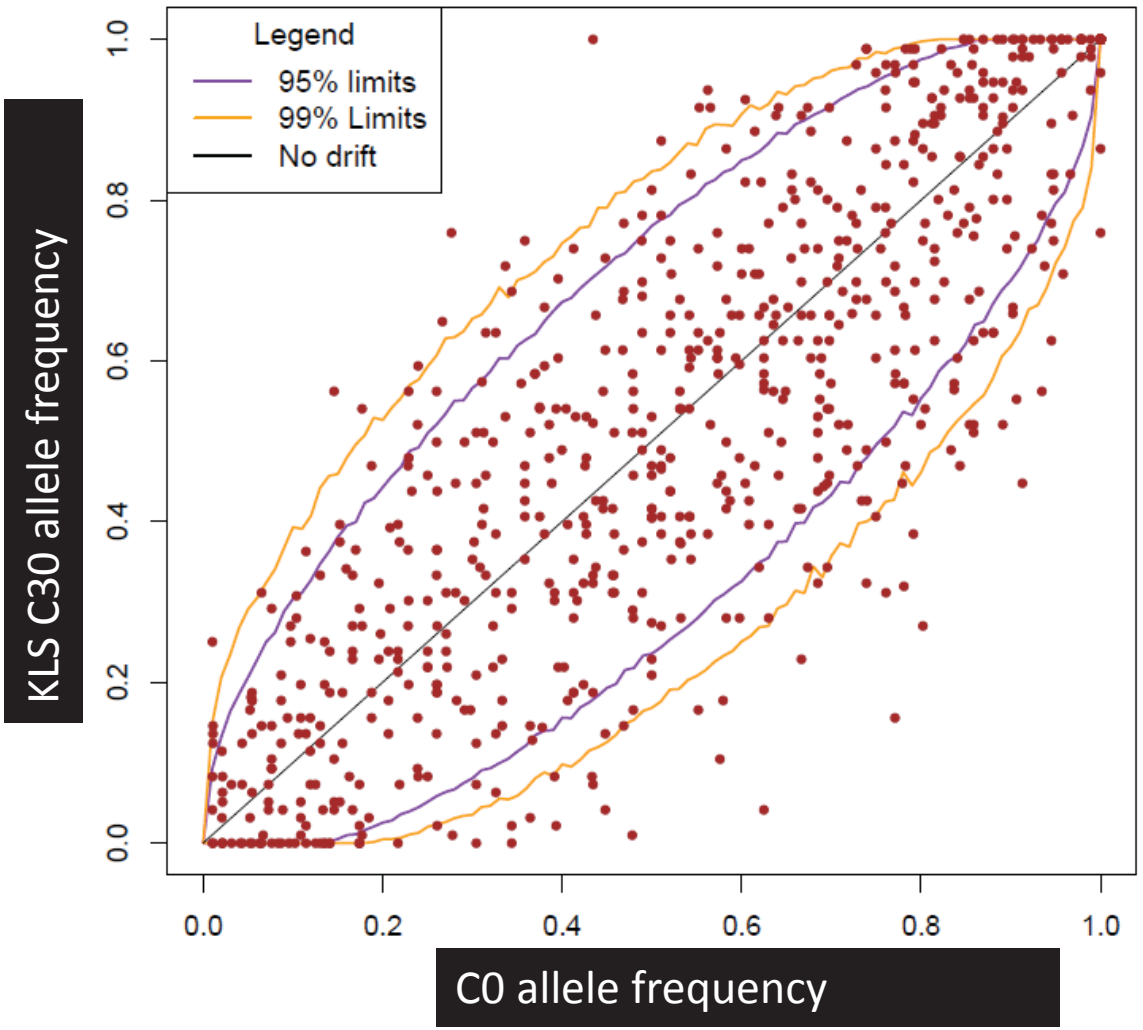
Allele Frequency change in KLS via Illumina Golden Gate:Chrom 1, 10,000 loci

Tim Beissinger



Expect loci with moderate initial frequency to respond the most to selection

Allele Frequency change in KLS via Illumina Golden Gate

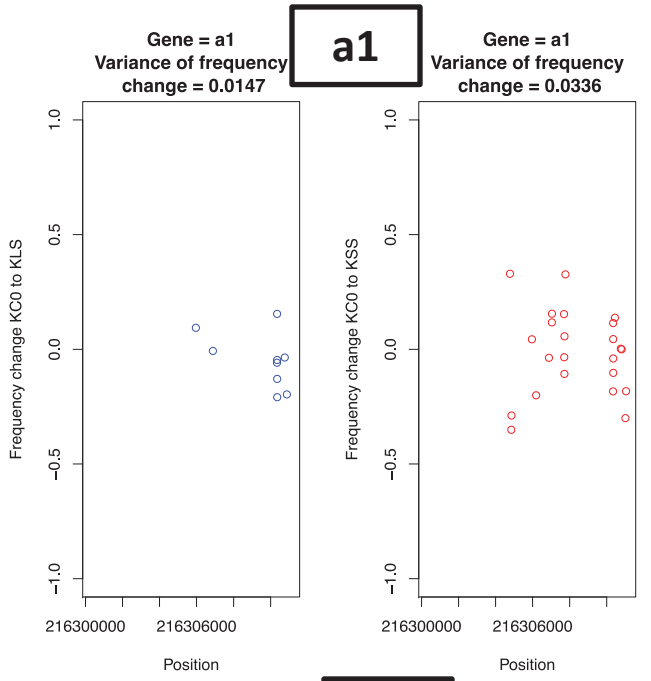


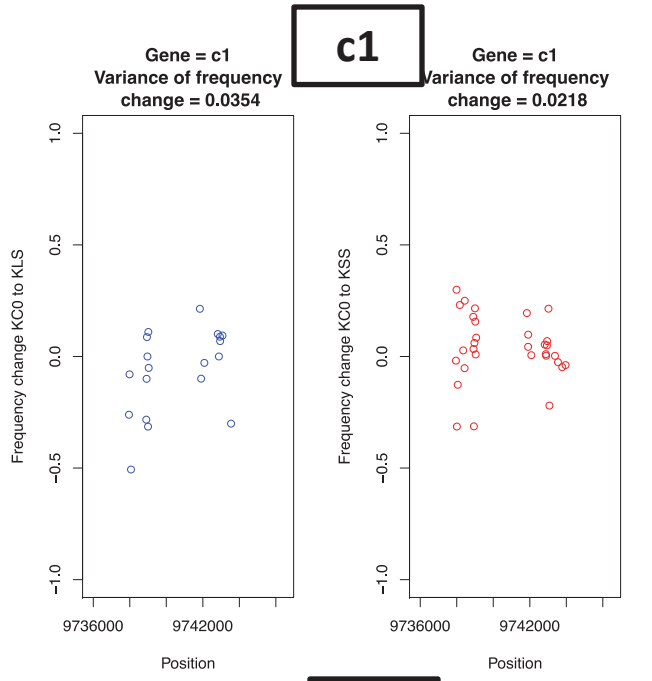
Some specific examples and observations

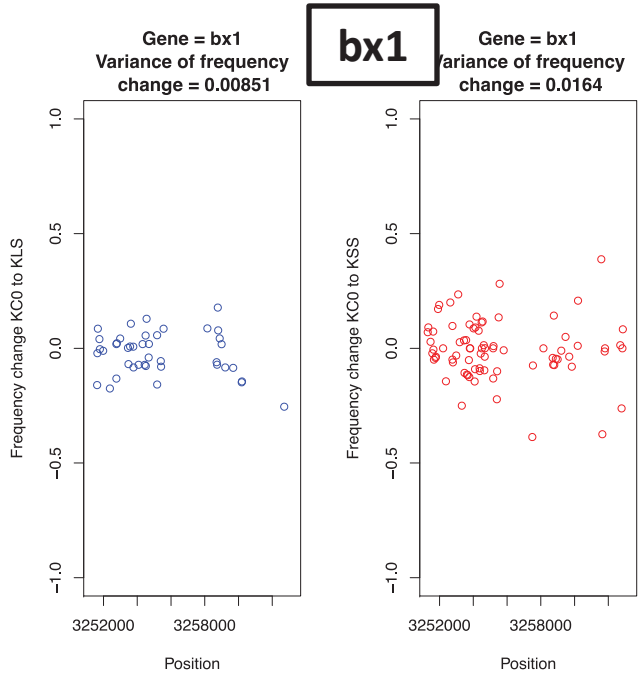
Candidate Genes – “Classics”

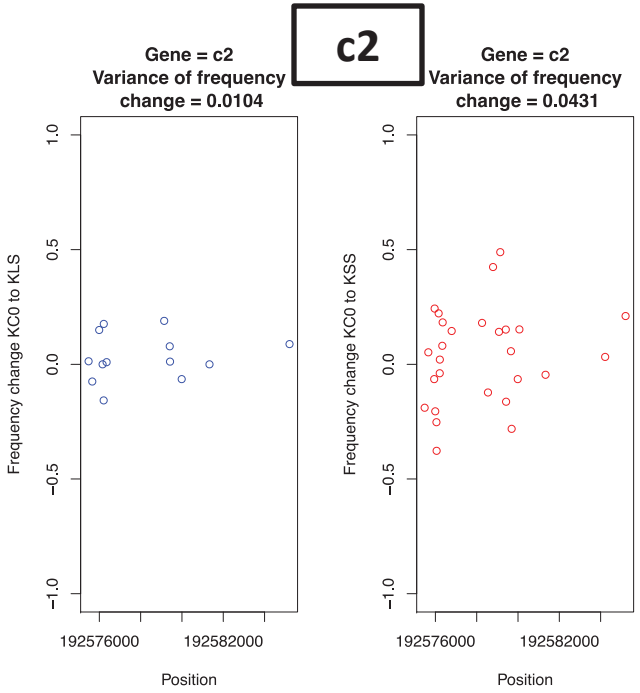
- Controls
  - a1, a2, bz1, c1, c2, p1
- Carbohydrate – starch
  - mn1, bt1, bt2, sh2, su, ae1, du1, …
- Imprinting / epigenetics
  - fie1, dmt101, rmr1

Controls

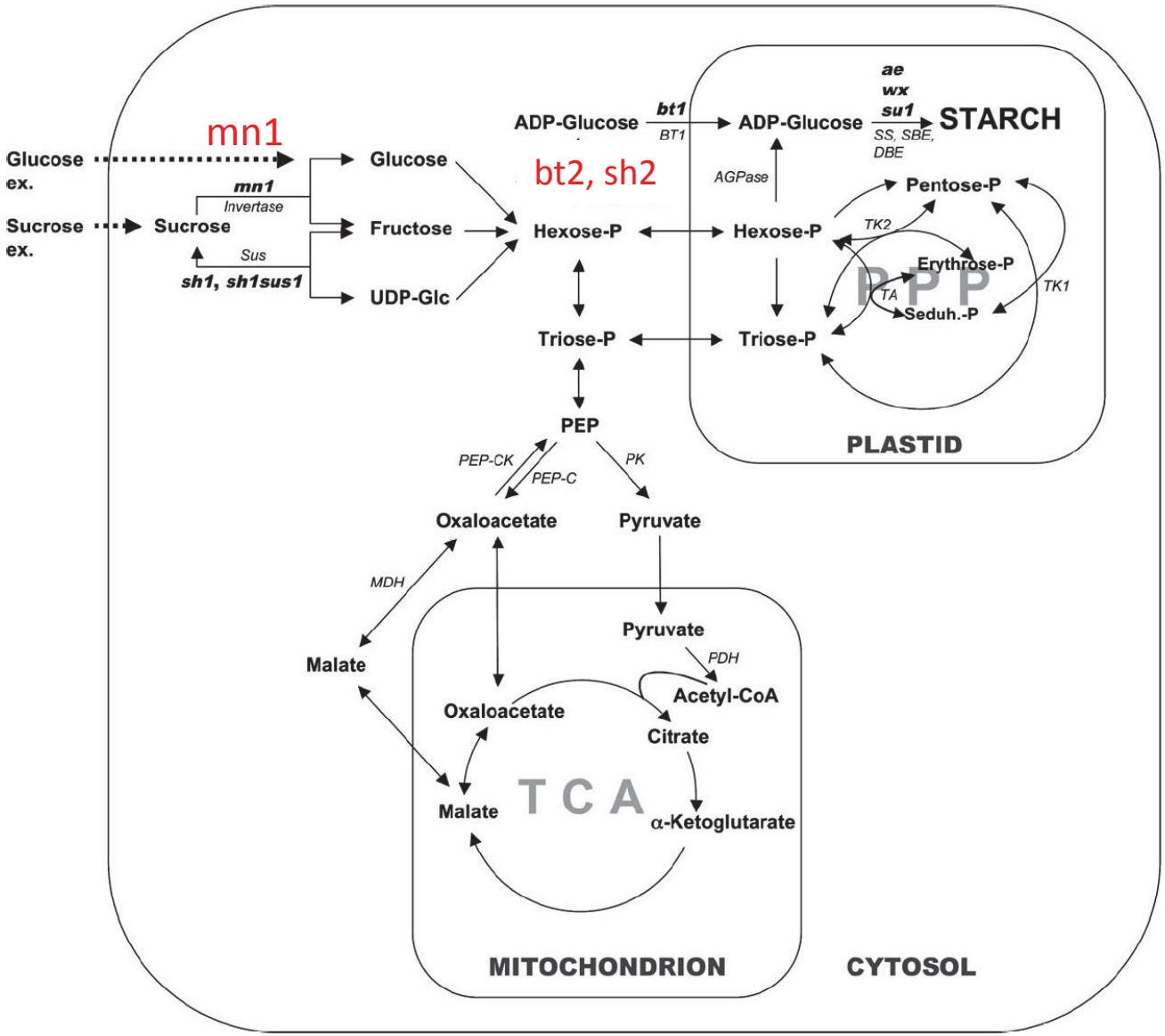

a1
Gene = a1
Variance of frequency
change = 0.0147
Gene = a1
Variance of frequency
change = 0.0336
Frequency change KC0 to KLS
Frequency change KC0 to KSS
1.0
0.5
0.0
−0.5
−1.0
216300000
216306000
Position


c1
Gene = c1
Variance of frequency
change = 0.0354
Gene = c1
Variance of frequency
change = 0.0218
Frequency change KC0 to KLS
Frequency change KC0 to KSS
1.0
0.5
0.0
−0.5
−1.0
9736000
9742000
Position


bx1
Gene = bx1
Variance of frequency
change = 0.00851
Gene = bx1
Variance of frequency
change = 0.0164
Frequency change KC0 to KLS
Frequency change KC0 to KSS
1.0
0.5
0.0
−0.5
−1.0
3252000
3258000
Position


c2
Gene = c2
Variance of frequency
change = 0.0104
Gene = c2
Variance of frequency
change = 0.0431
Frequency change KC0 to KLS
Frequency change KC0 to KSS
1.0
0.5
0.0
−0.5
−1.0
192576000
192582000
Position

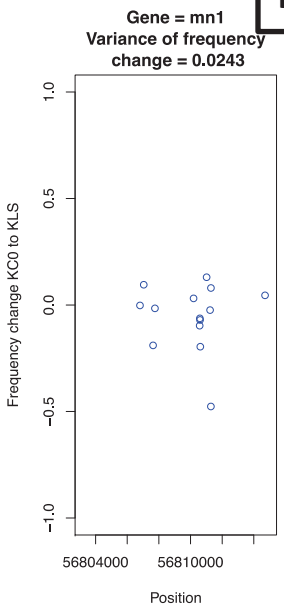
Carbohydrate, Starch

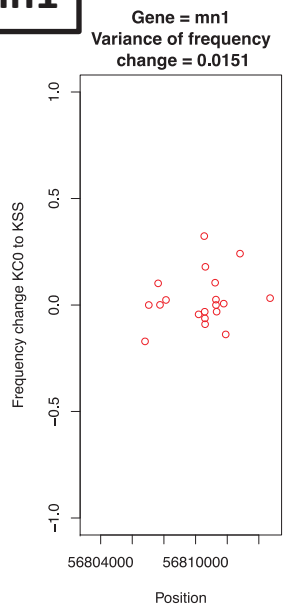


G. Spielbauer et al. / Phytochemistry 67 (2006) 1460–1475

Carbohydrate, Starch

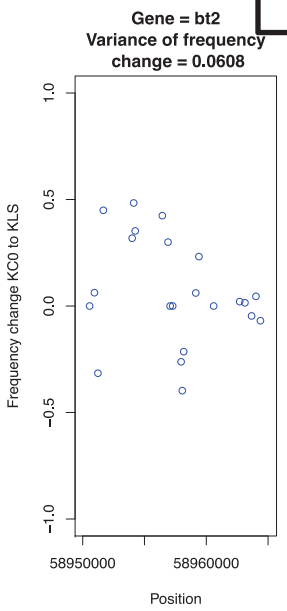
mn1

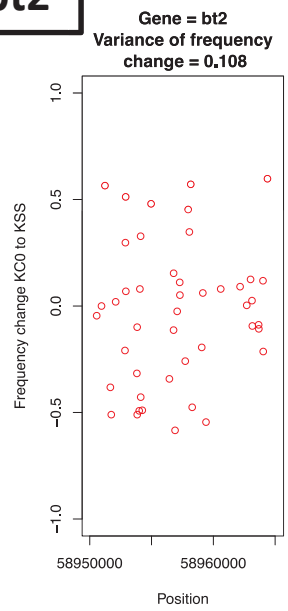




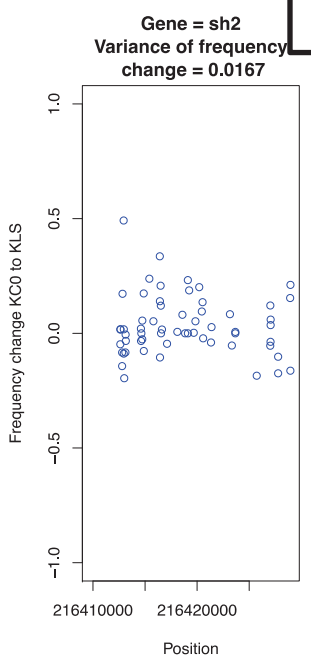
AGPase subunits

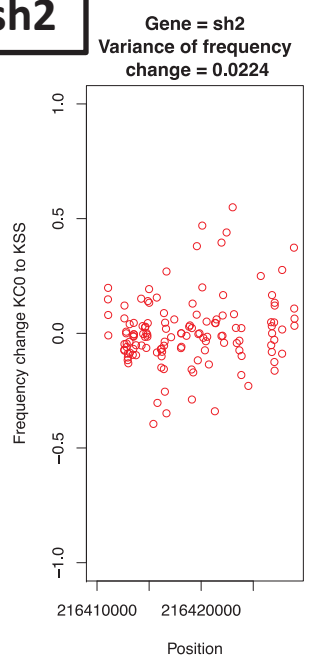
bt2





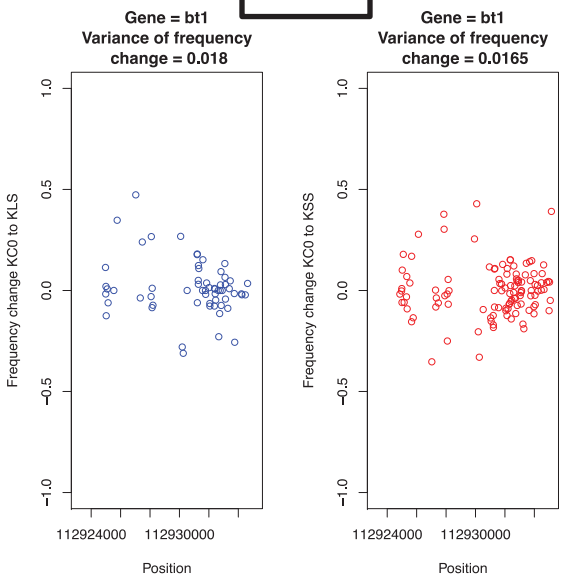
sh2



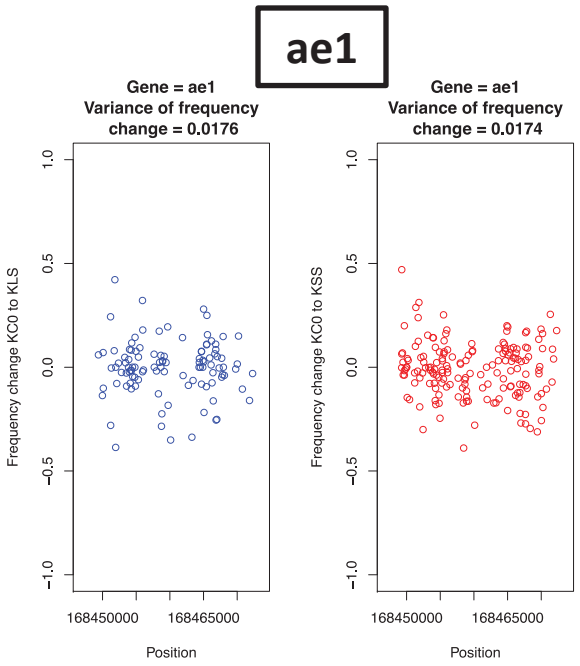


Carbohydrate, Starch

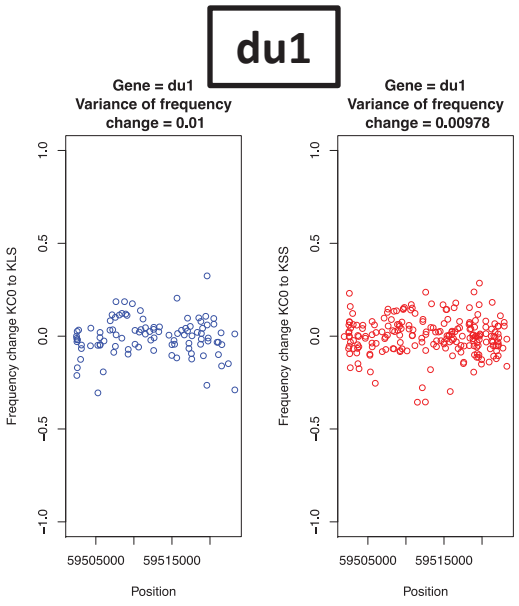
bt1



ae1

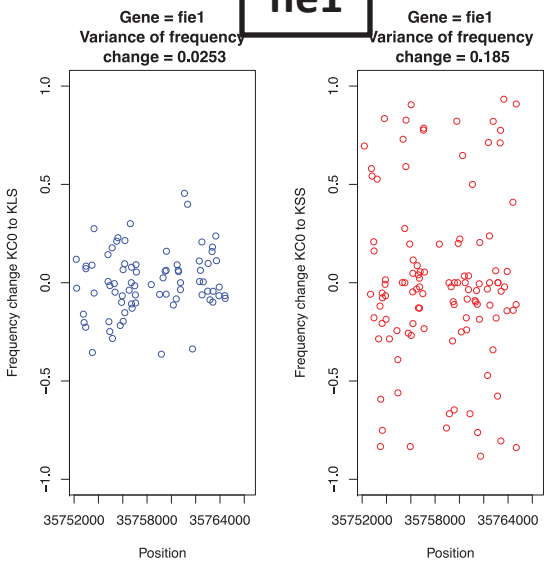


du1

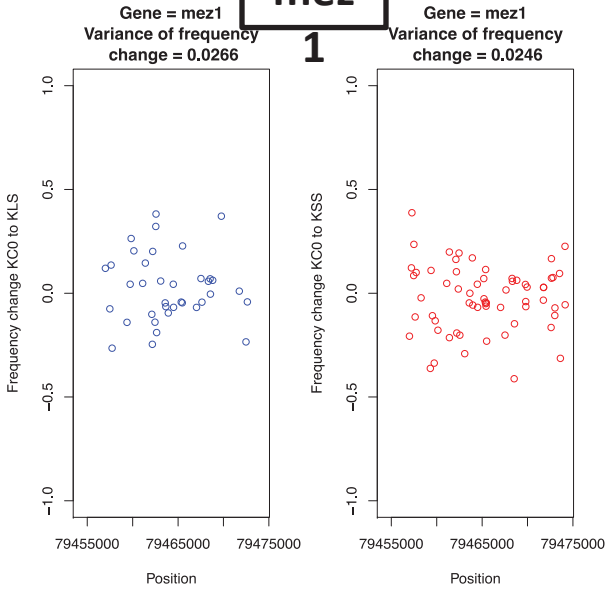


Imprinting / Epigenetics

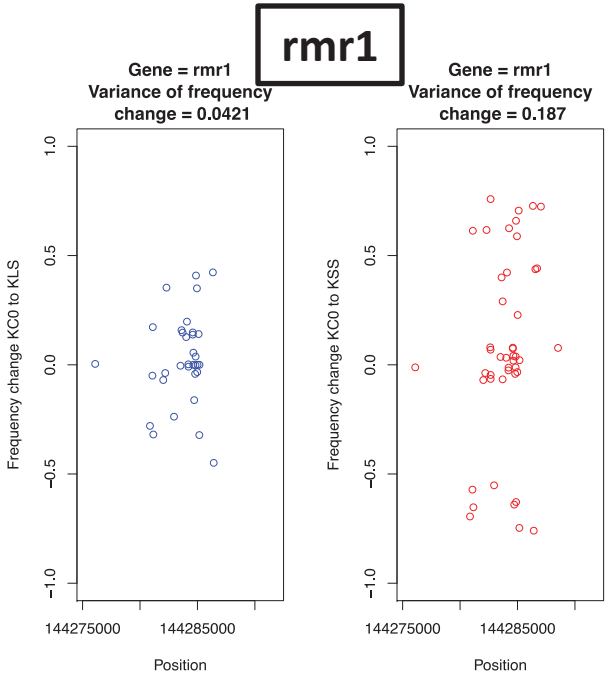
fie1



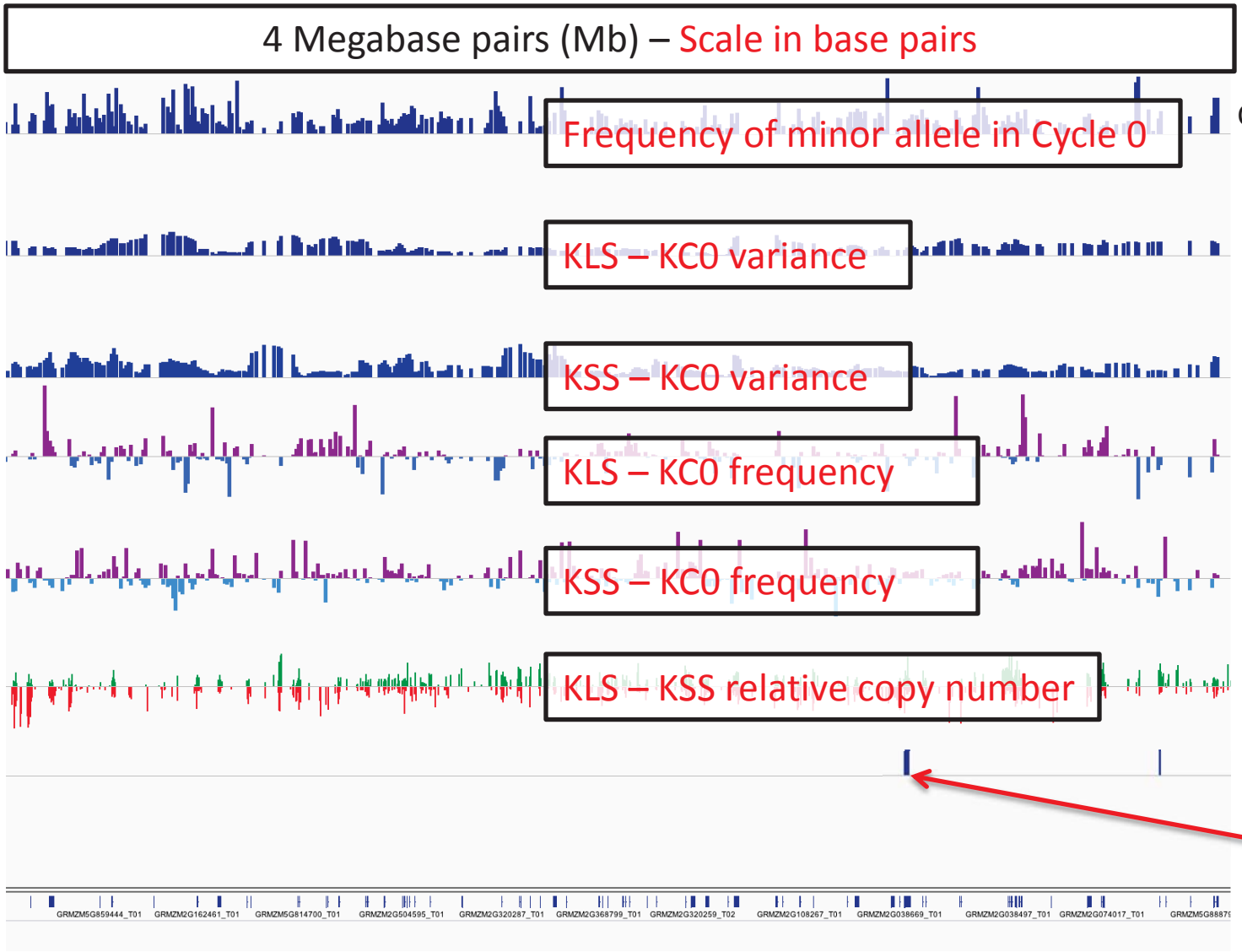
mez 1



rmr1

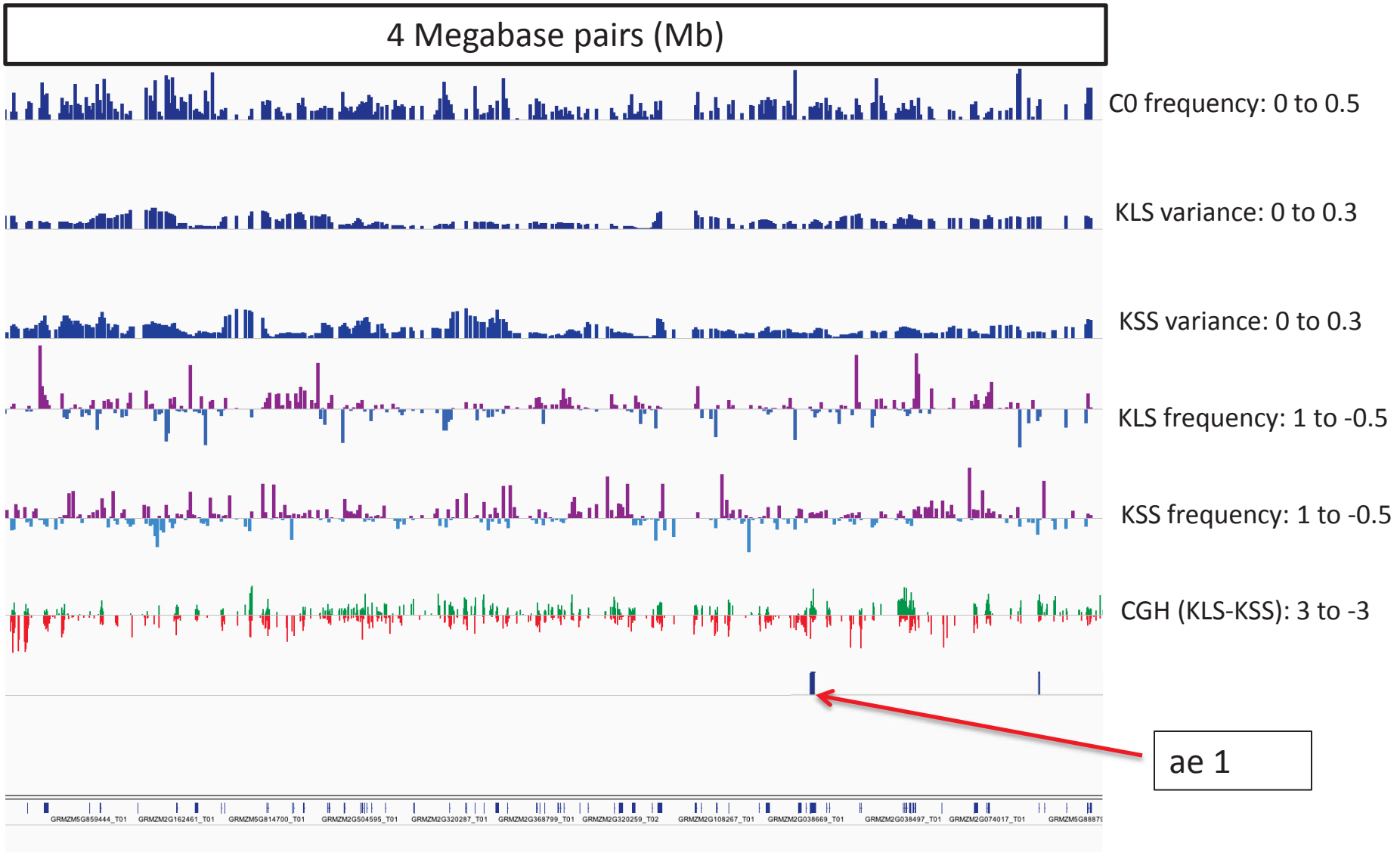


Integrated Genome Viewer Explanation

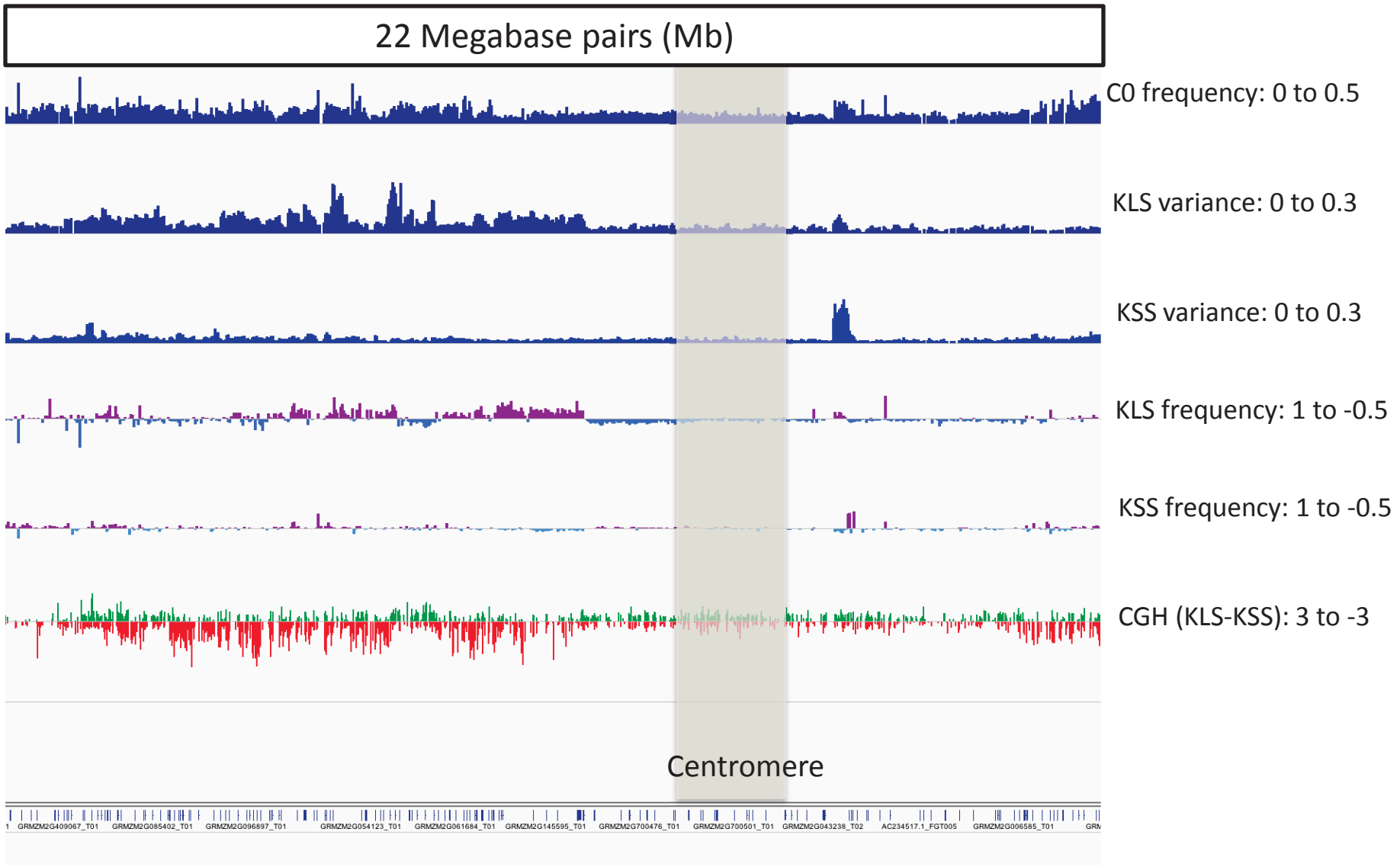


Linkage blocks

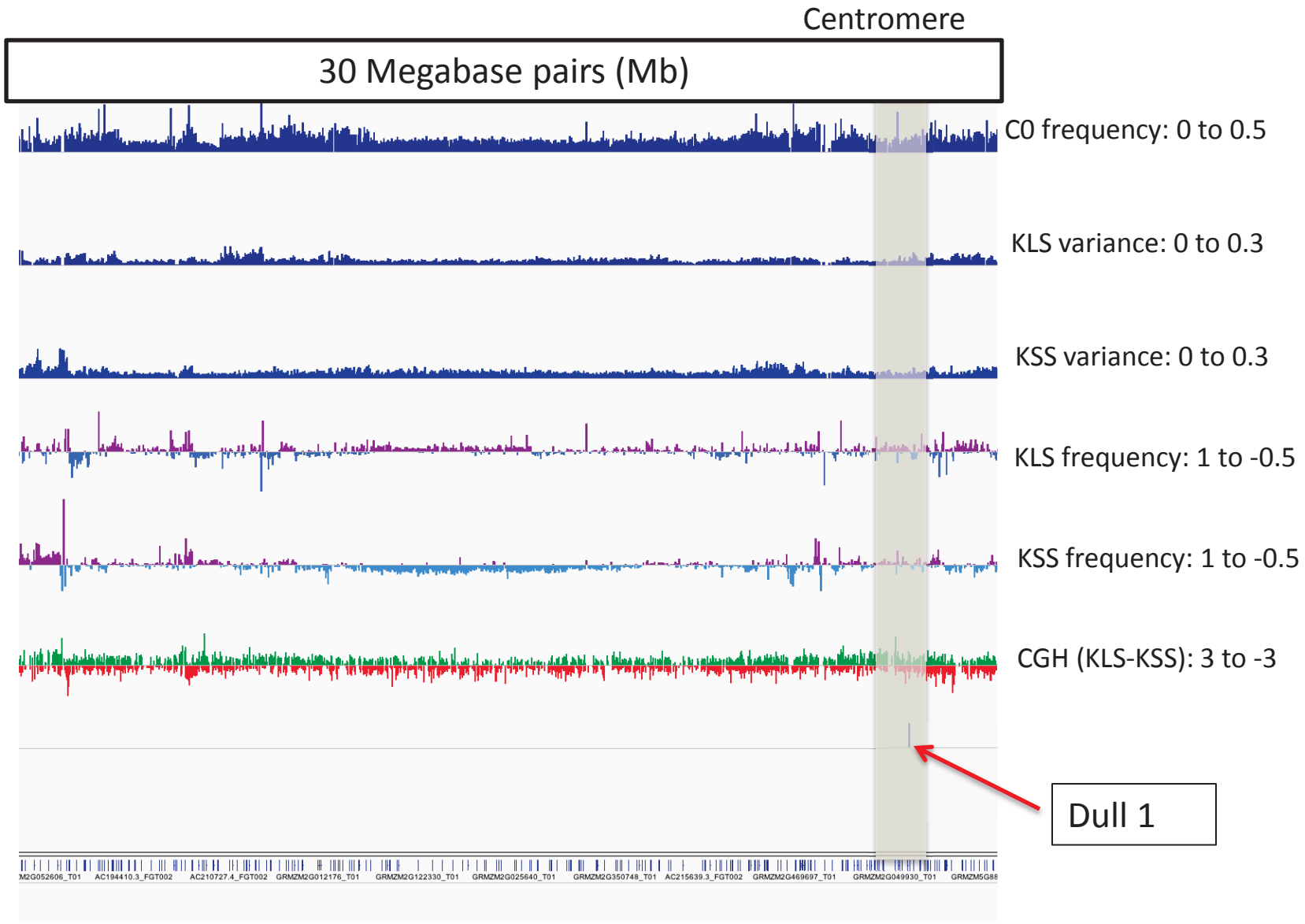
Chromosome 5 – Example non-centromeric region



Chromosome 8 – 20 Mb centromere region

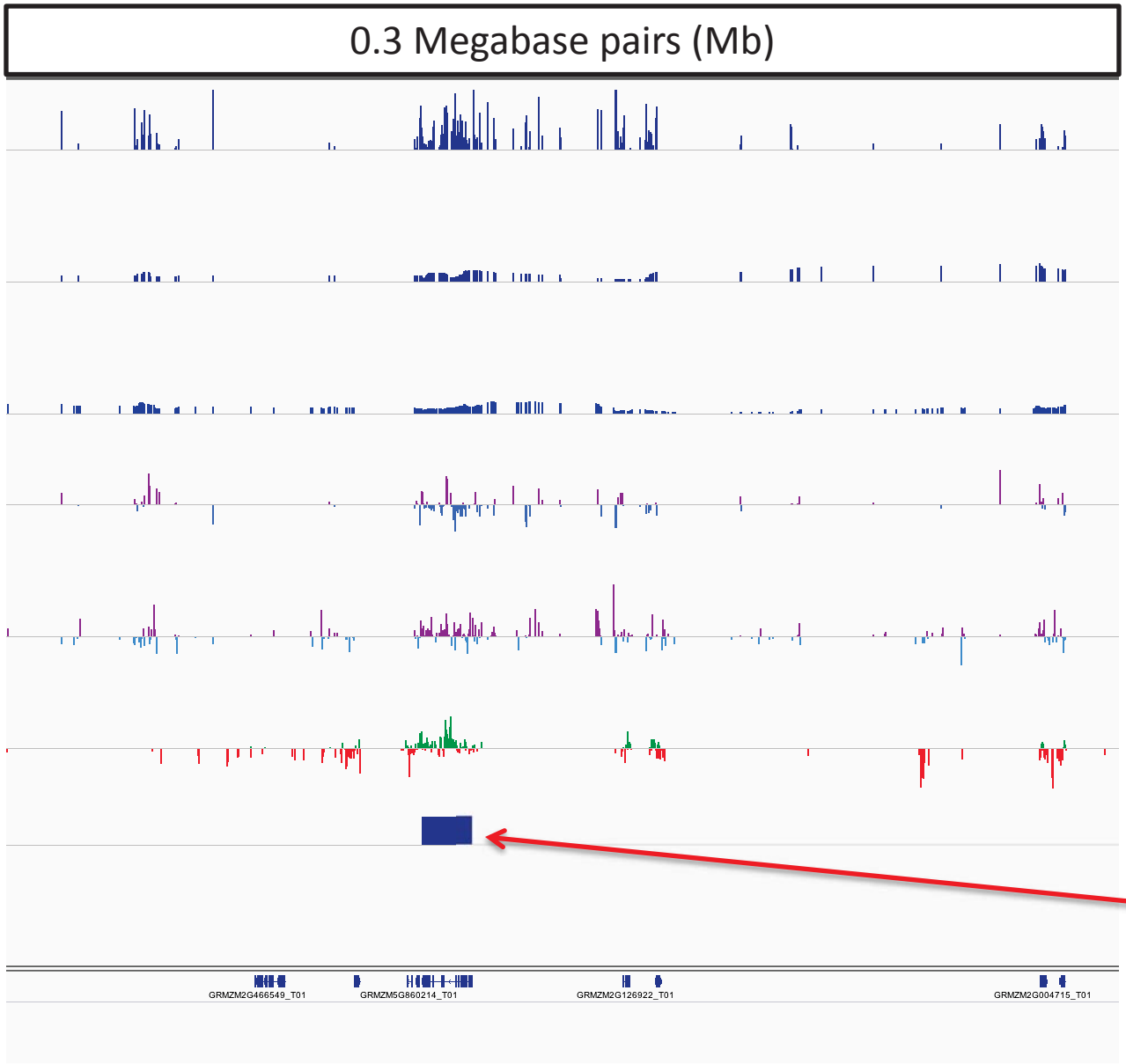


Chromosome 10 – 30 Mb centromere region

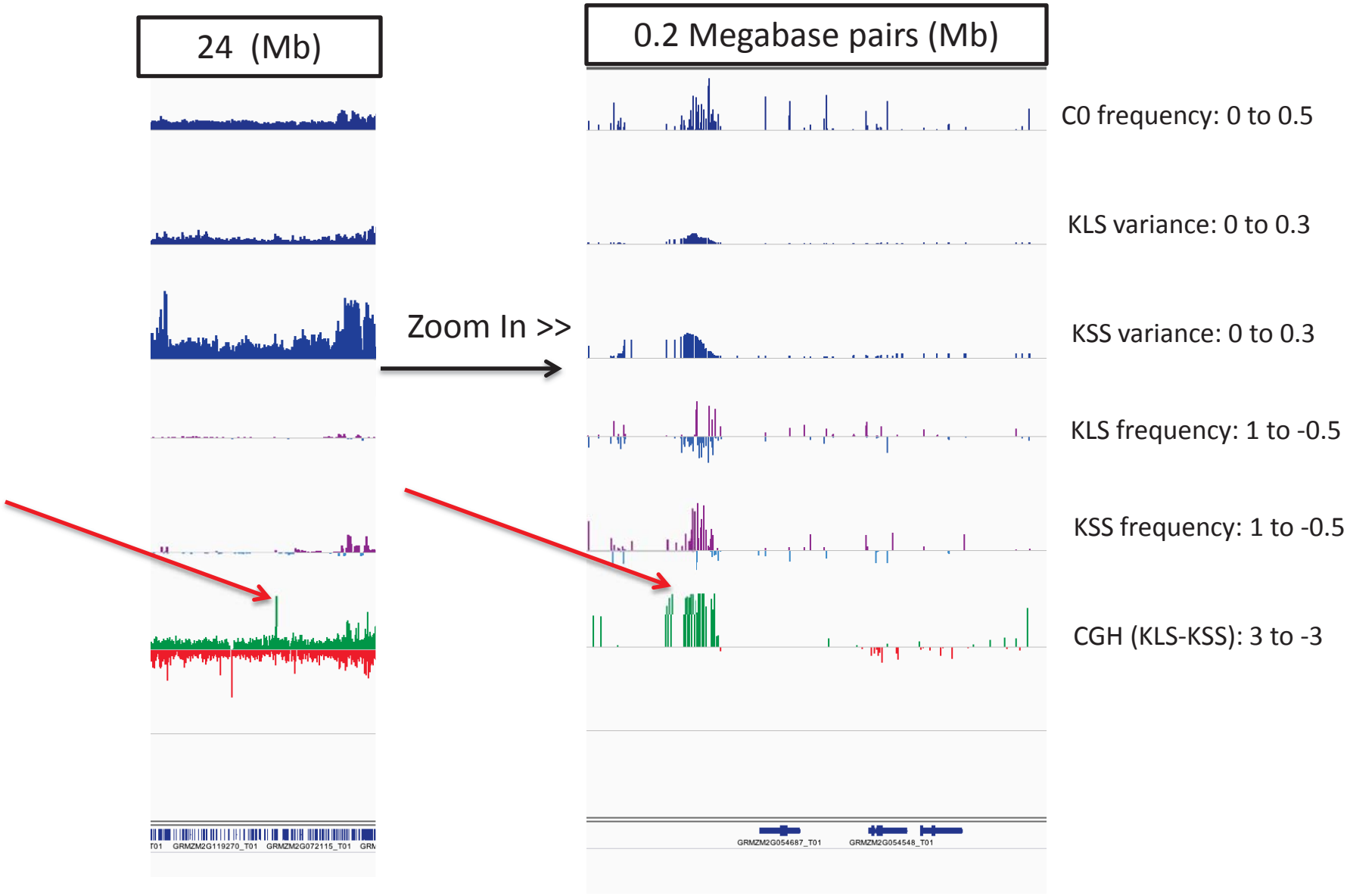


Copy Number

Chromosome 5 – zoom in on ae1

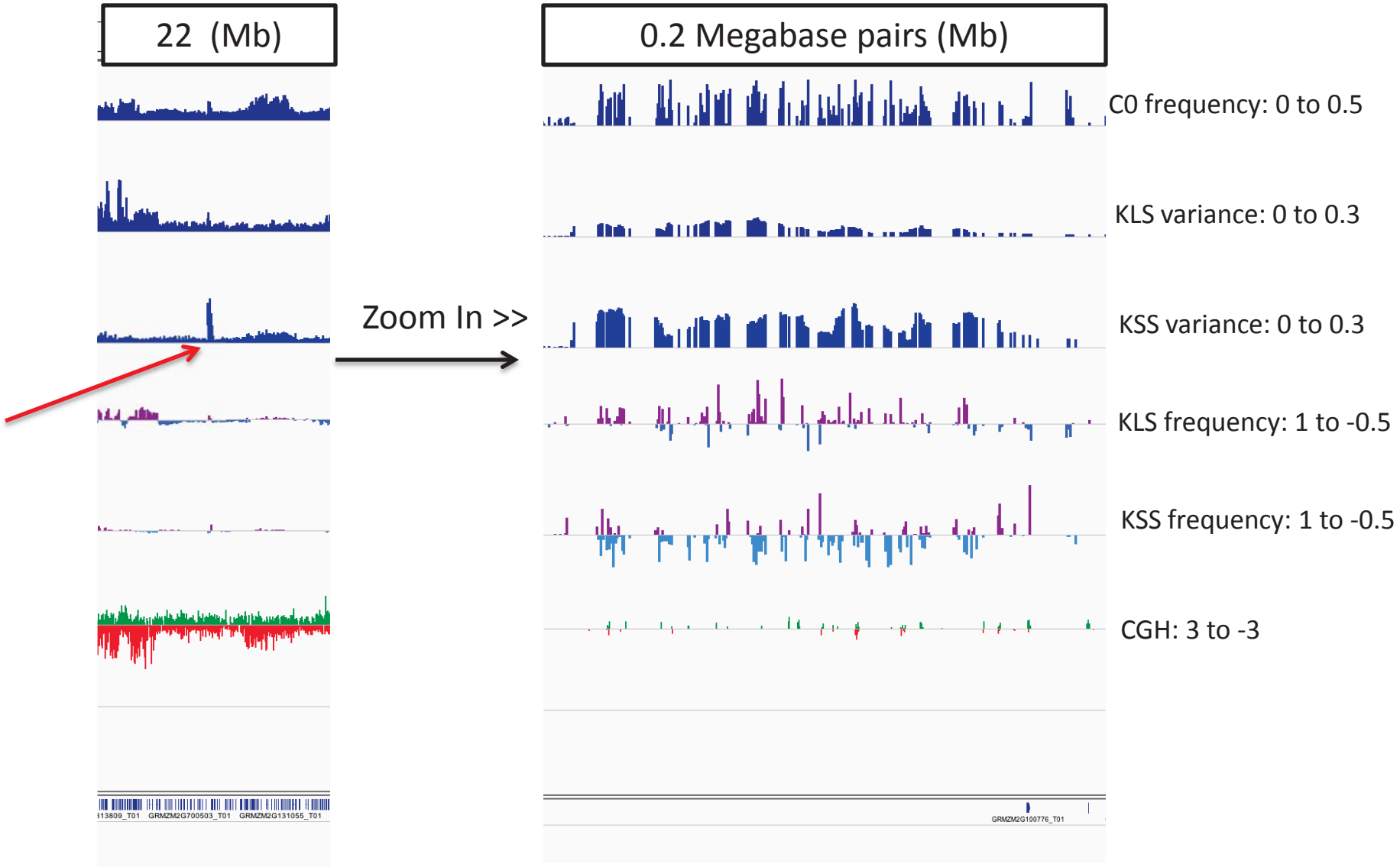


Chromosome 4 – High KLS, Low KSS



Novel “gene” discovery

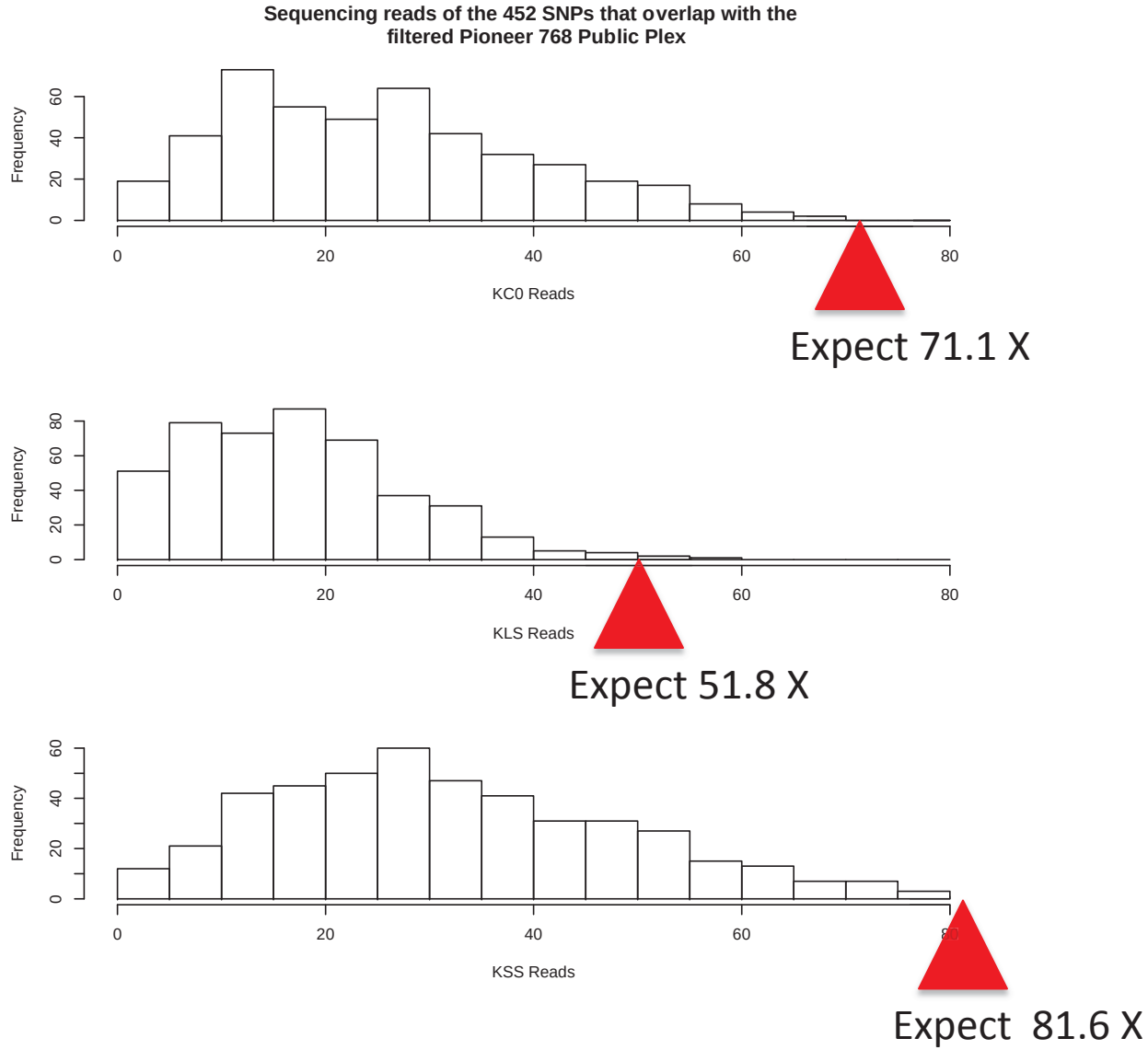
Chromosome 8 – Strong selection response



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



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

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

U of Minnesota:
Nathan Springer
Steve Eichten

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

U.S. DEPARTMENT OF ENERGY

GLBRC
Great Lakes Bioenergy Research Center

USDA
United States Department of Agriculture

National Institute of Food and Agriculture