Genome-Wide Selection

Brian S. Yandell, UW-Madison

January 2017

evolution of QTL models

original ideas focused on rare & costly markers
models & methods refined as technology advanced

- single marker regression
- QTL (quantitative trait loci)
  - single locus models: interval mapping for QTL
  - QTL model search: QTLs & epistasis
- polygenes (association mapping)
  - adjust for population structure
  - capture "missing heritability"
- **genome-wide selection**

what is genomic selection?

use statistical modeling to predict how a plant will perform before it is field-tested

- genomic selection (GS)
  - marker assisted selection (MAS)
  - genome-wide selection (GS)
- other uses of word (relevent to systems genetics)
  - natural selection: survival of the fittest
  - model selection: search for QTLs
  - selection bias: overestimate of QTL effects

why use genomic selection?

- trait is highly polygenic (genetically variable)
  - influenced by a few key genomic regions
  - high heritability (low environmental variation)
- measuring trait is costly
  - difficult or expensive process (technology)
  - measuring tool may be highly variable
  - time-consuming (plant has to grow first)
  - desire to streamline multi-year selection

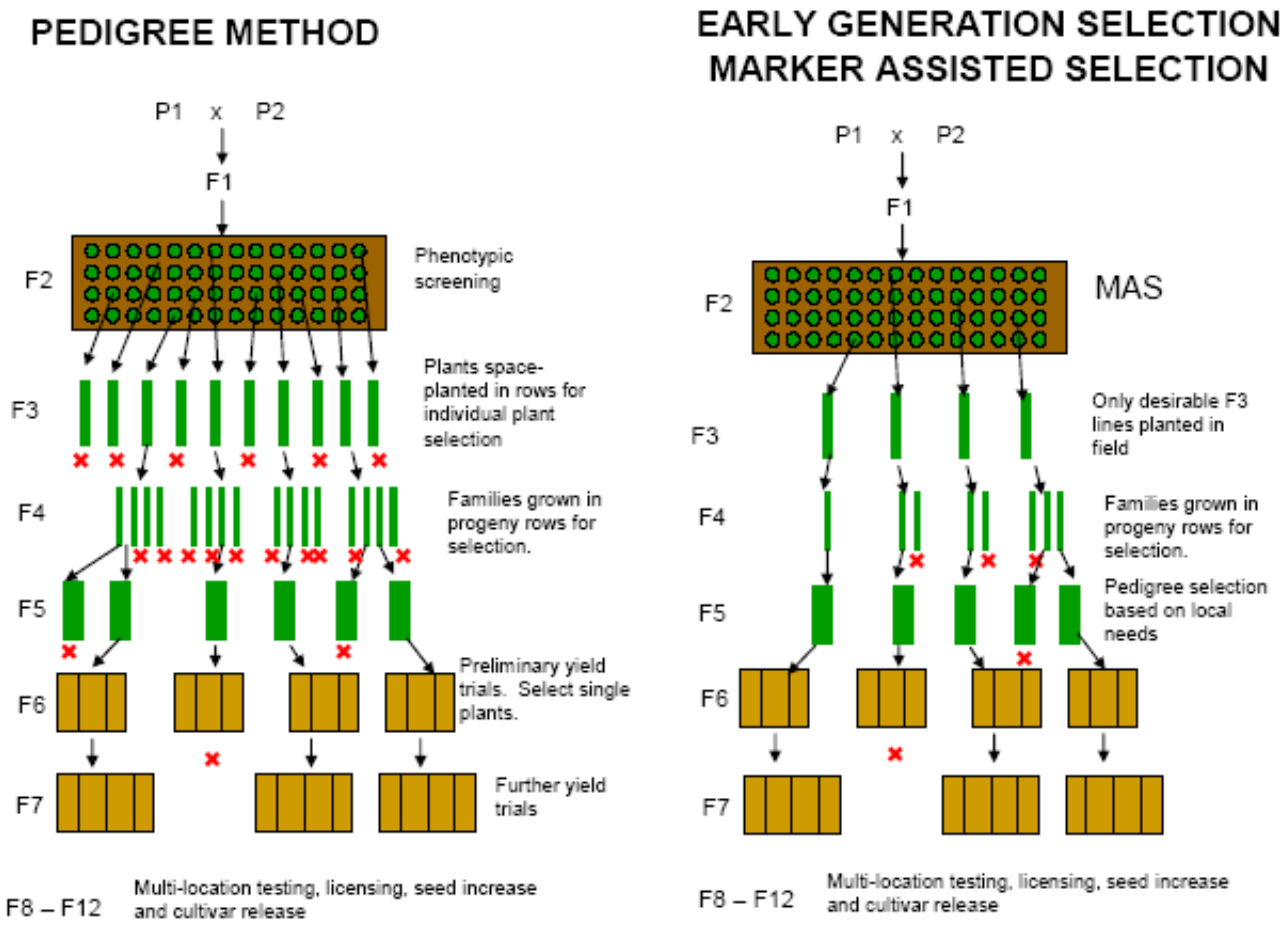
what is genomic selection?

- forms of genome-wide selection
  - marker-assisted: with phenotypes
  - marker-based: without phenotypes
- use markers to improve selection for complex traits
  - predict phenotype from marker genotype
  - select candidates based on best marker genotypes
  - use training set to predict test set of individuals

old paradigm: marker prediction

- 1990s & 2000s: markers were expensive
- economic strategy:
  - first identify significant markers (QTL analysis)
  - use best markers to genotype selection candidates
- estimate marker effects by multiple regression
  - treat genetic effects as fixed and few
  - $E(y) = \mu_q, q = (q_1, q_2, q_3)$

marker assisted selection (MAS)

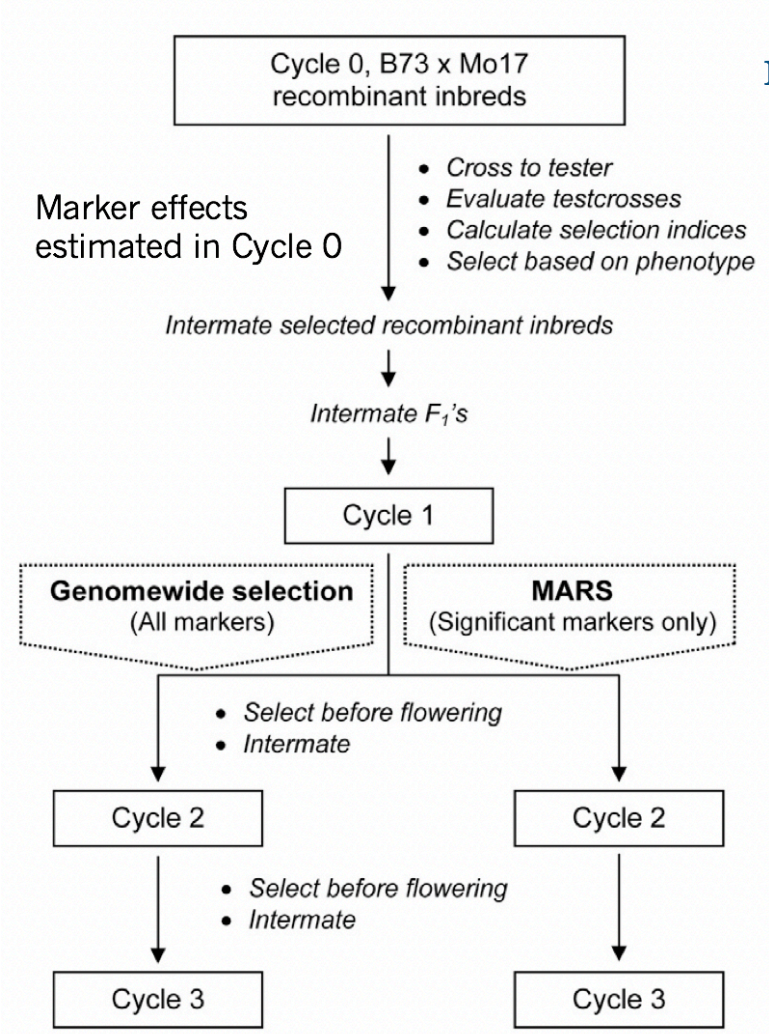


www.21stcentech.com/heard-marker-assisted-breeding/

new paradigm: use "all" markers

- new paradigm with technology advances
  - improved statistical methods and software
  - cheap markers
- using only significant markers to predict trait …
  - gives good estimates (maybe) of markers …
  - but does not maximize accuracy
- simple but effective approach
  - treat marker effects as random
  - use all markers (away from QTL if any)

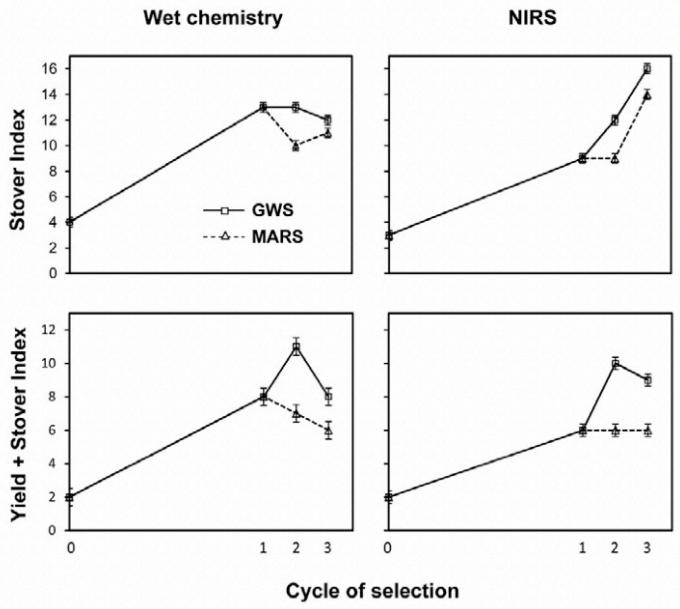
old vs new



Genomewide Selection versus Marker-assisted Recurrent Selection to Improve Grain Yield and Stover-quality Traits for Cellulosic Ethanol in Maize

Jon M. Massman, Hans-Joachim G. Jung, and Rex Bernardo*

Crop Science 2013



mixed model approach

MAS approach $y = \mu_q + e, V(e) = \sigma^2 I$

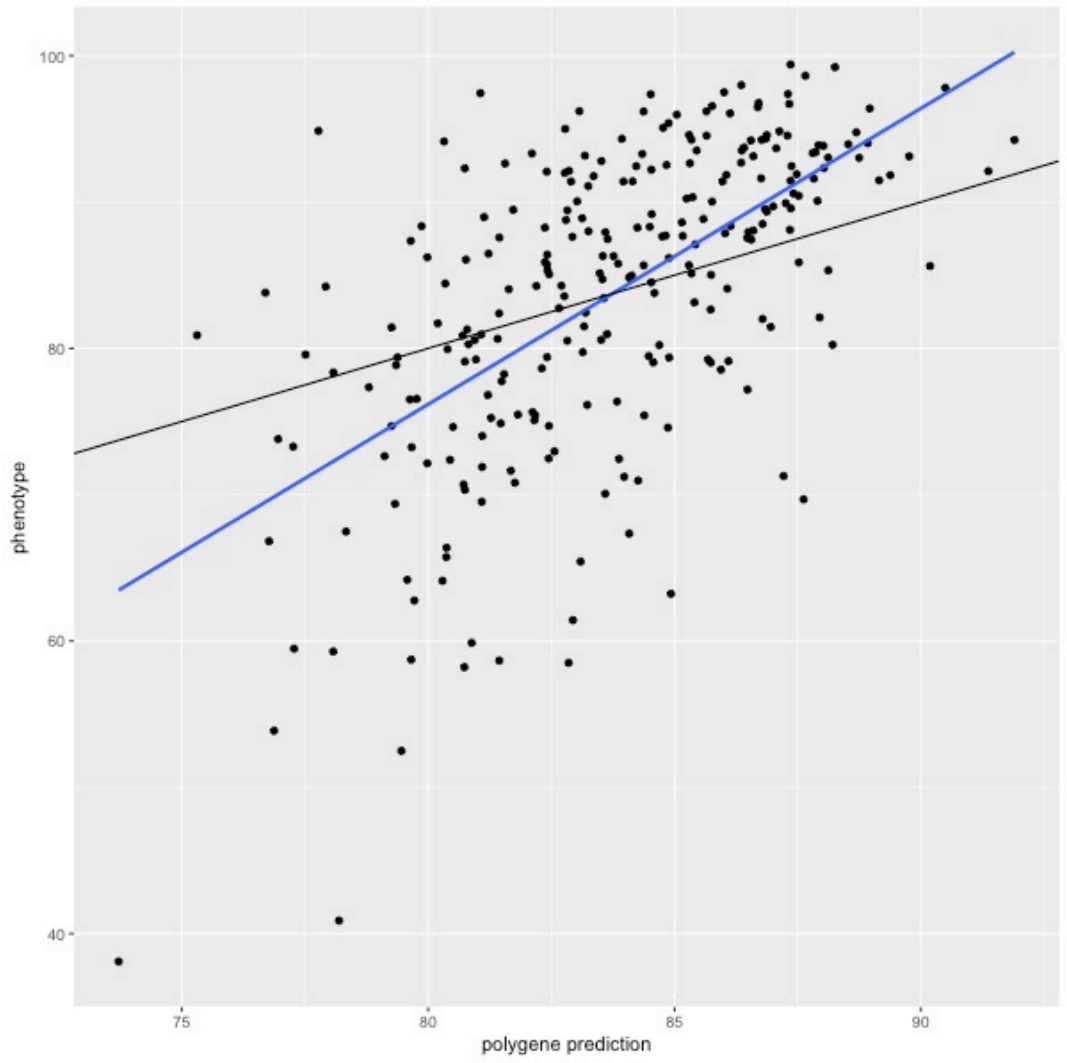
- estimate fixed QTL effects $\hat{\mu}_q$ (MLEs)
- predict phenotype using fixed effects $\hat{y} = \hat{\mu}_q$

GS approach $y = \mu + g + e, V(g) = \sigma_g^2 K$

- estimate kinship K from all markers M as for `poly`
- predict random effect $\hat{g}$ using BLUP
- predict phenotype $\hat{y} = \hat{g}$

genomic prediction

- DO example
- rrBLUP fit without QTL
- correlation 0.79



poly + QTL genomic prediction

- DO example
- rrBLUP fit with QTL
- correlation 0.74