Updating Training Data for Long-Term Genomewide Selection

Jeff Neyhart
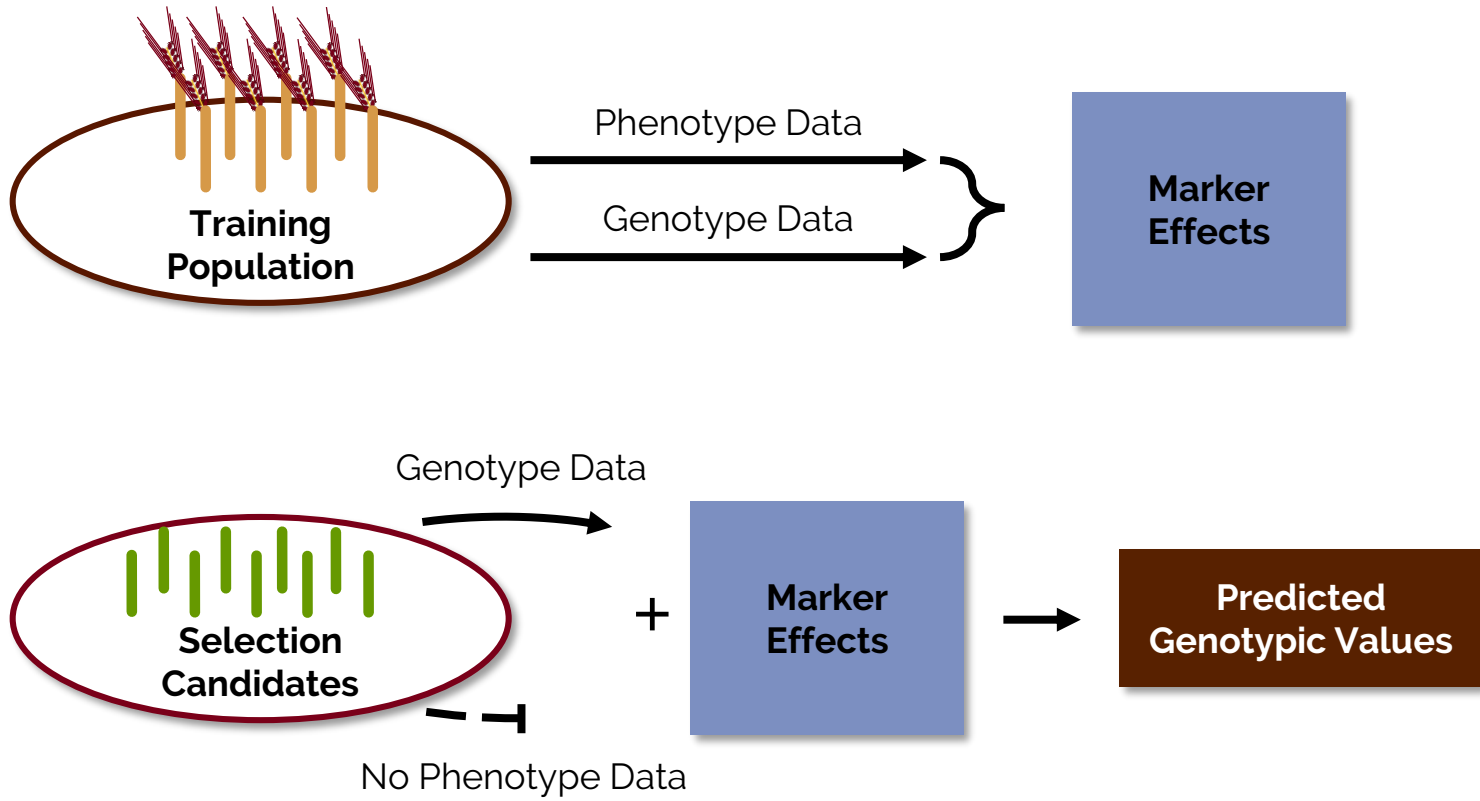
@neyhartje
slides: z.umn.edu/jnucdpss

Meuwissen *et al.* 2001 *Genetics*
Heffner *et al.* 2009 *Crop Science*
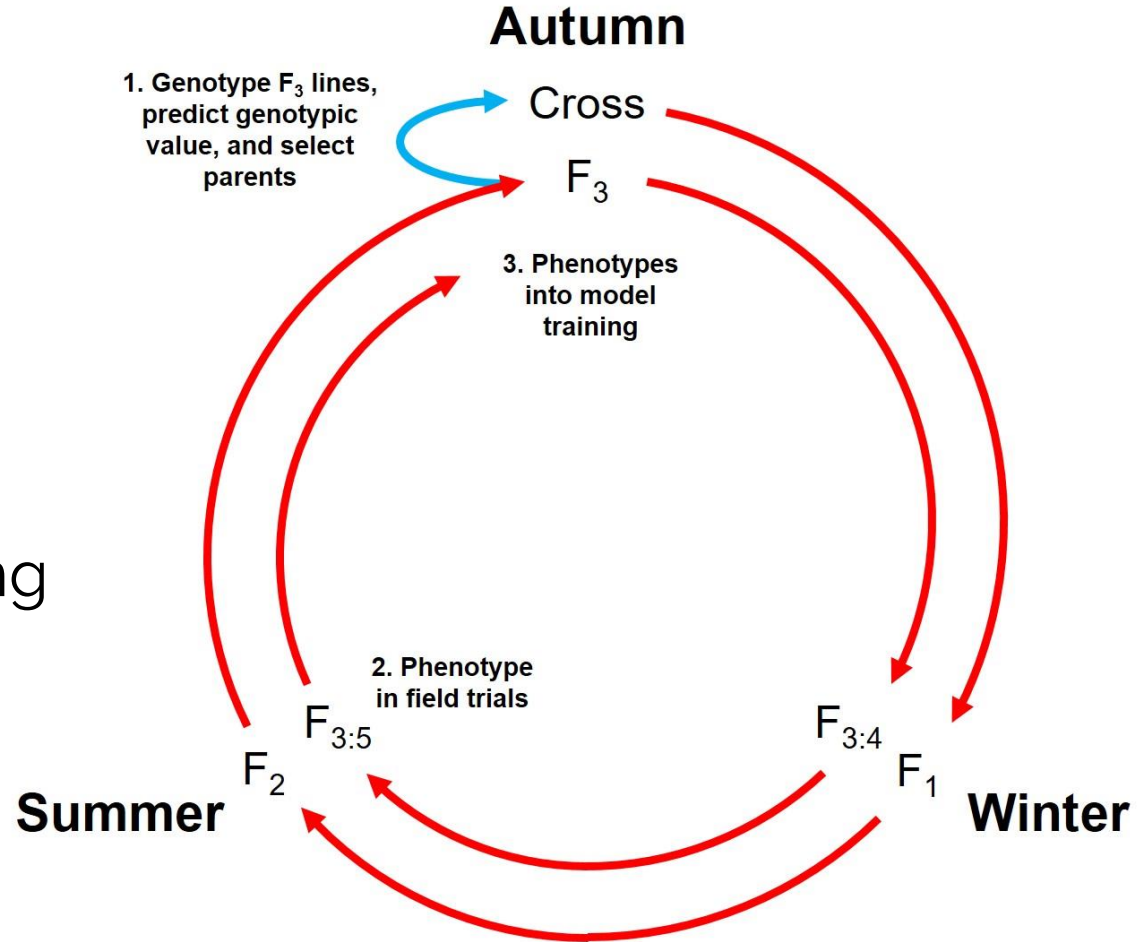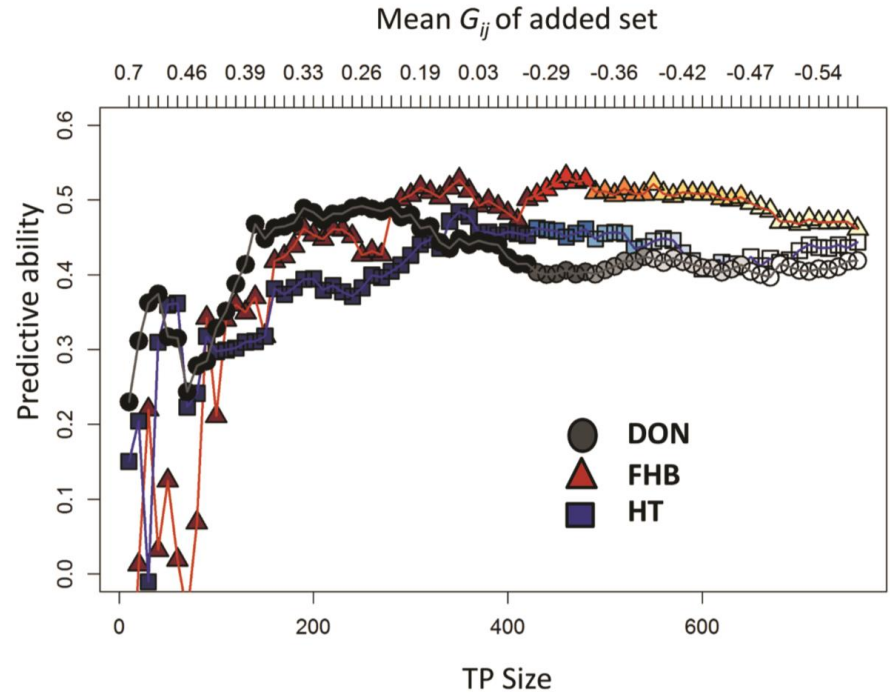
Genomewide selection to rapidly improve Minnesota spring two-row barley

A barley breeding program using genomewide selection can complete a breeding cycle in 1 year

Autumn

1. Genotype F_3 lines, predict genotypic value, and select parents

Cross

F_3

3. Phenotypes into model training

2. Phenotype in field trials

$F_{3:5}$

F_2

$F_{3:4}$

F_1

Summer

Winter

Prediction accuracy depends on several factors, including *training population design*

Modified from Lorenz and Smith 2015 *Crop Science*

Lorenz and Smith 2015 *Crop Science*
Isidro *et al.* 2015 *Theor Appl Genet*

Properties of recurrent selection will alter prediction accuracy

Recombination breaks marker-QTL linkage disequilibrium

Training population becomes more distant to selection candidates

What are efficient methods to update the training population?

Modified from Habier *et al.* 2007 *Genetics*

Hill and Robertson 1968 *Theor Appl Genet*
Habier *et al.* 2007 *Genetics*
Jannink 2010 *Genetics Selection Evolution*
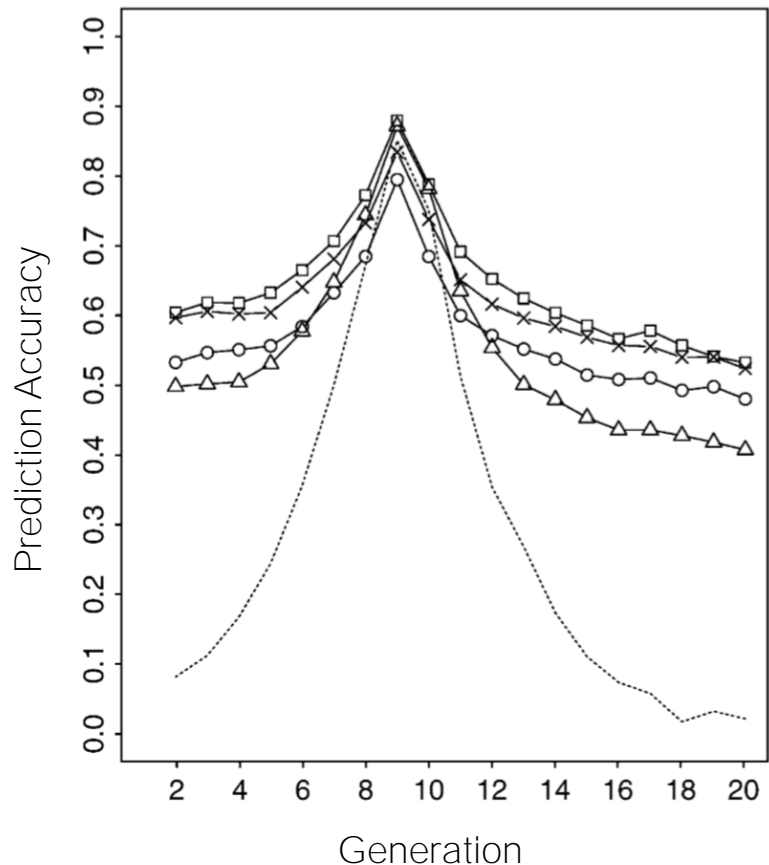
Objectives

- Simulate a barley genomewide selection program

- Determine optimal methods to update training data

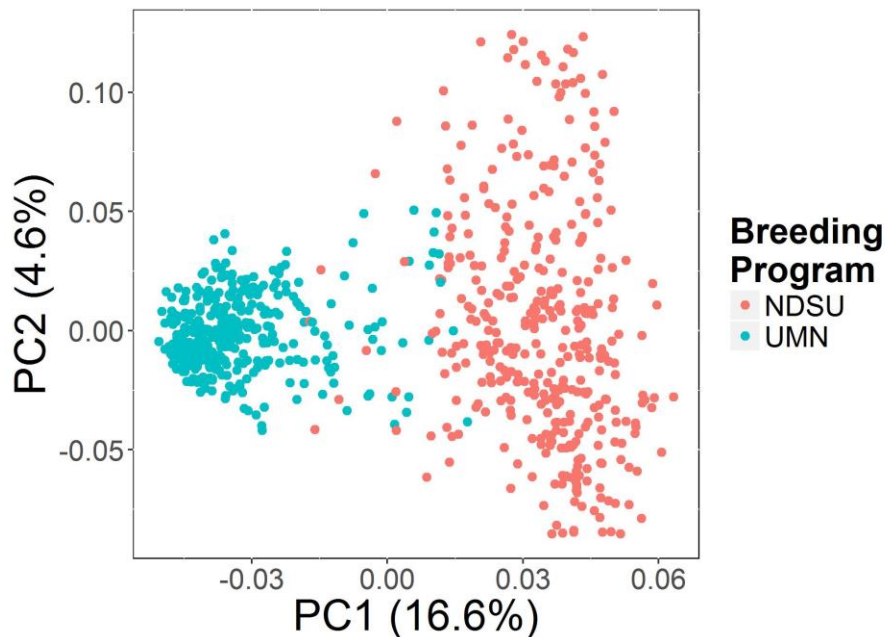- Suggest implications for a breeding program

Starting material

Historical genotype data on 764 barley breeding lines

Simulated a trait controlled by 100 QTL and h^2 = 0.5

15 cycles of genomewide recurrent selection
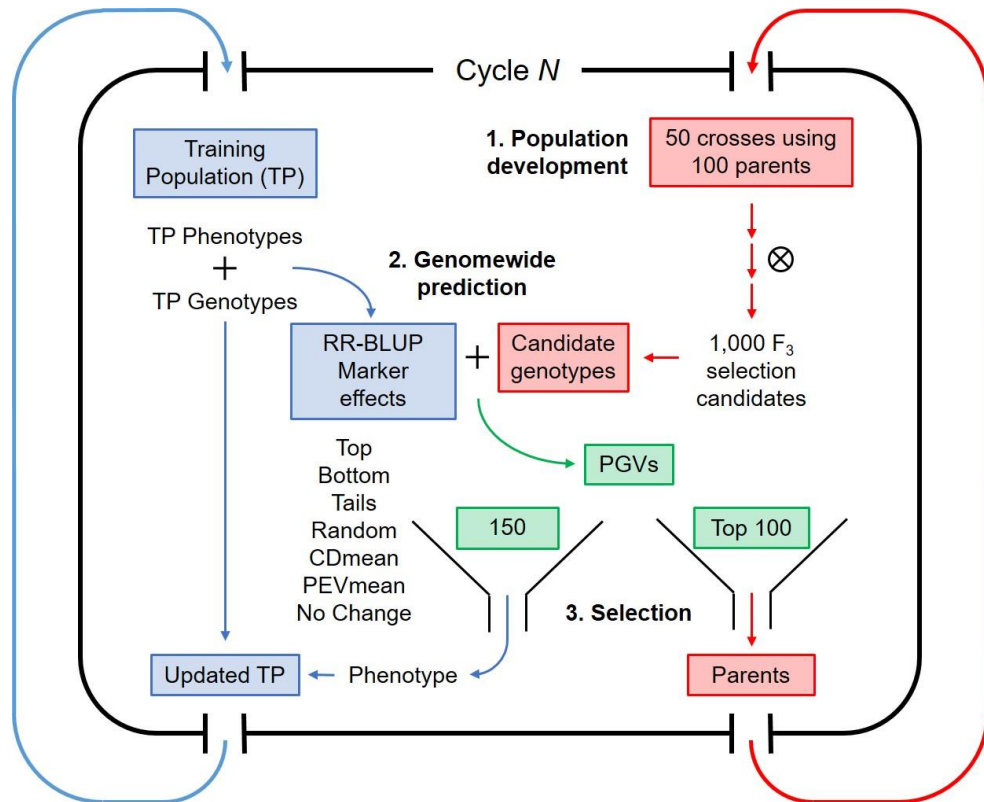
Three streams per cycle

Blue: steps to manage the training population
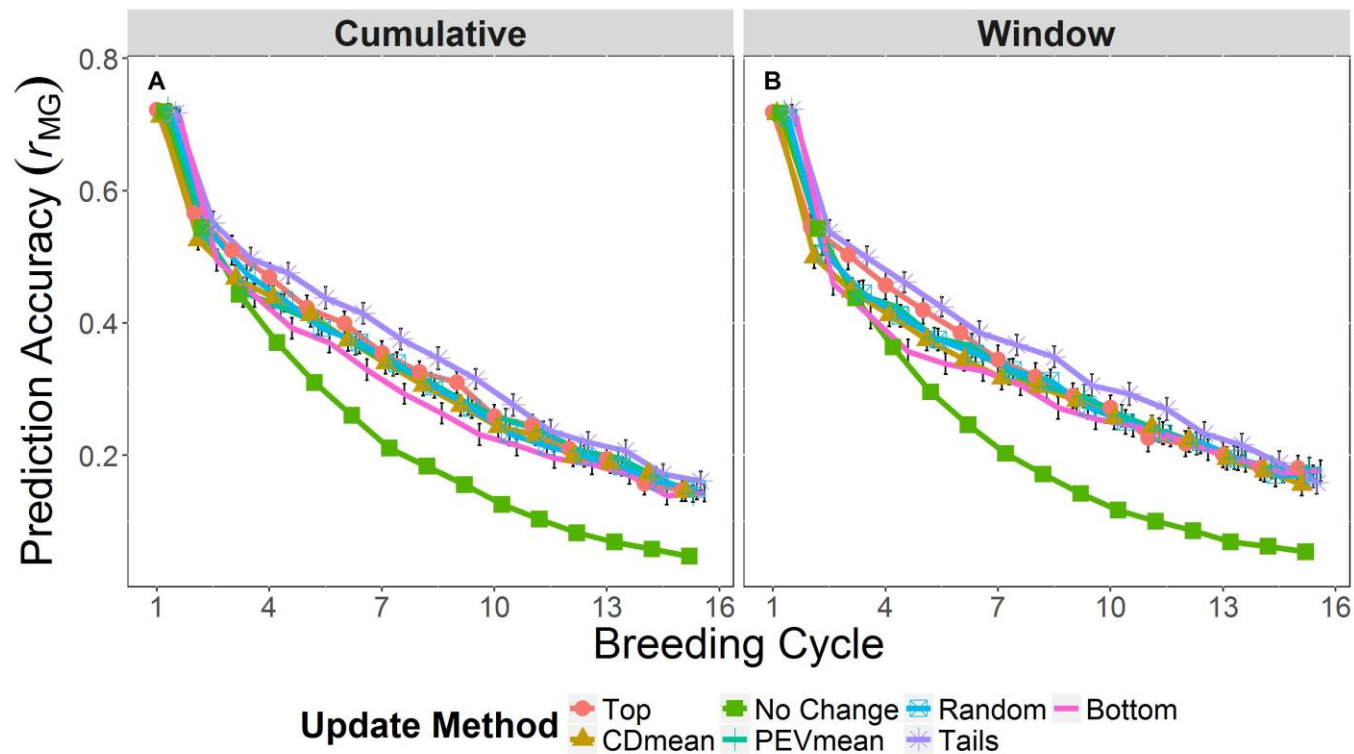
Red: steps to create populations

Green: steps to select parents and lines to update the training population
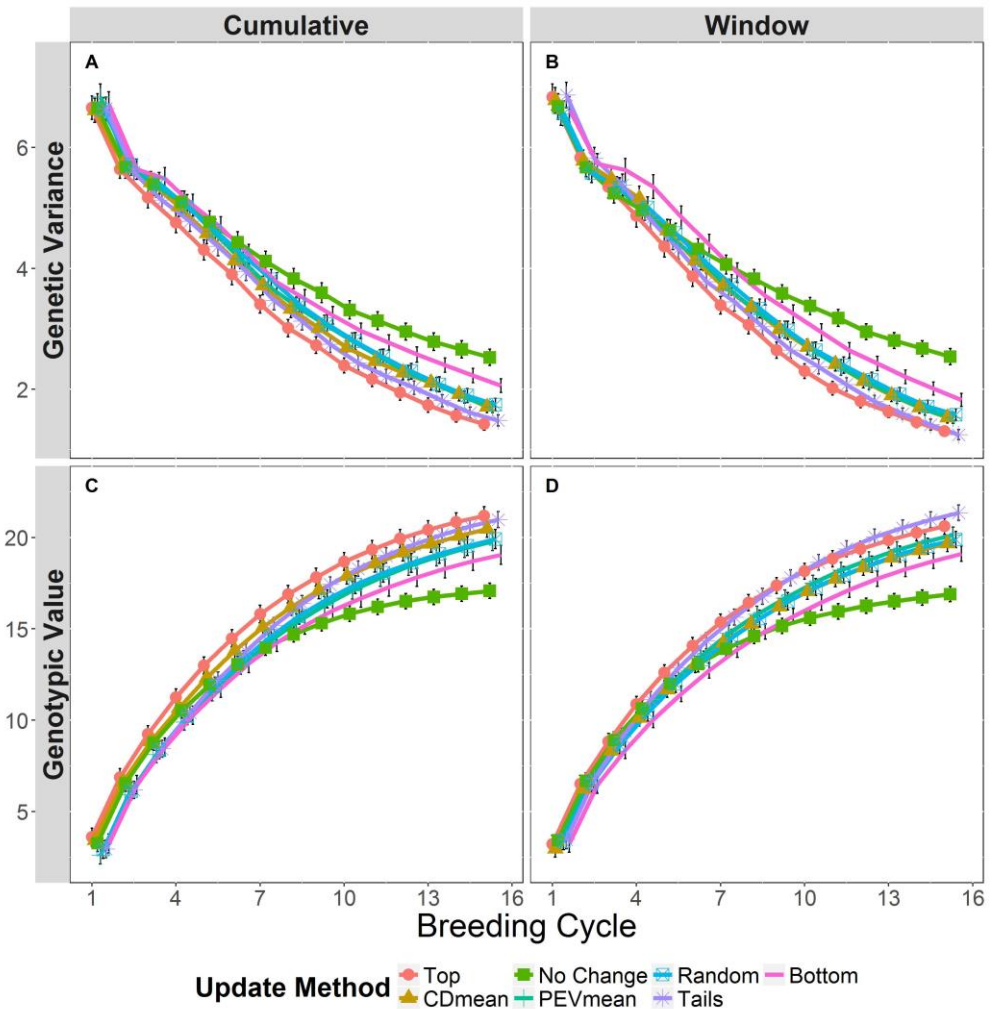
Two updating scenarios
- Cumulative
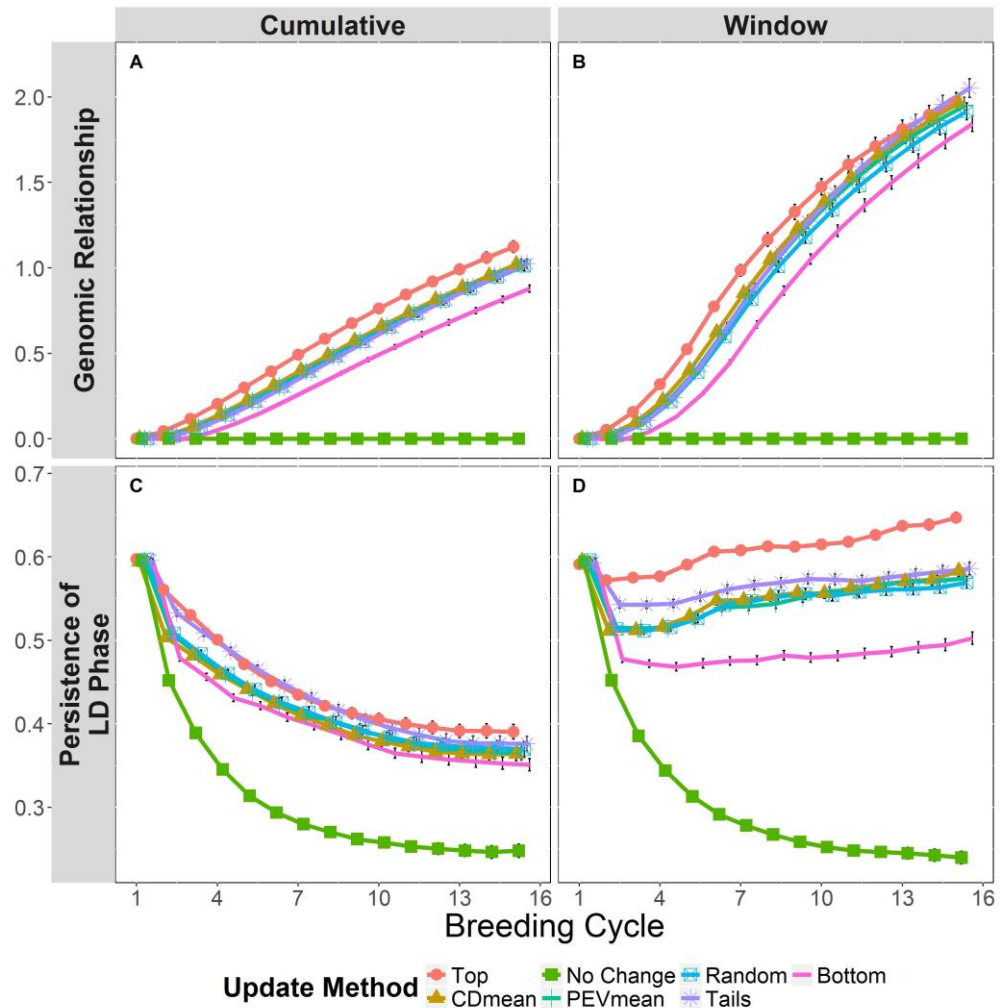- Window

Rincent *et al.* 2012 *Genetics*

Updating with something is better than nothing

Patterns of genetic variance and response to selection reflect prediction accuracy

Training populations updated via *Top* or *Tails* are more closely related to selection candidates and maintain LD phase

Implications

- Training data must be updated

- Breeders can update using data from typical selections

- More recent data is informative

Acknowledgements

Questions?

Early online paper in G3: Genes|Genomes|Genetics
doi: *10.1534/g3.117.040550*

Data access

github.com/UMN-BarleyOatSilphium/GSSimTPUpdate