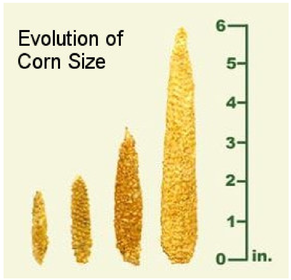
Software to Simulate Genomic Selection

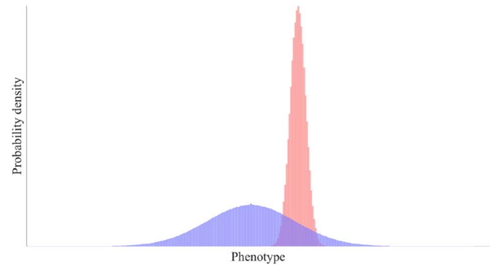
Takao Shibamoto, Dr. Lizhi Wang, Industrial and Manufacturing Systems Engineering, Iowa State University

1. Research Objective

- The objective of my research is to develop a tool to improve genomic selection methods.
- Plant breeding is the purposeful manipulation of plant species in order to create desired characteristics for specific purposes.

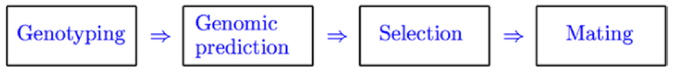


- Genomic selection (GS) is a new plant breeding approach that uses genome information.
- Phenotypic selection (PS) is a traditional approach which is slow and sometimes destructive. In the following figure, the red distribution is selected in PS, but we can efficiently select the rightmost of the blue distribution with GS.



Distributions of progeny phenotypes from two pairs of parents

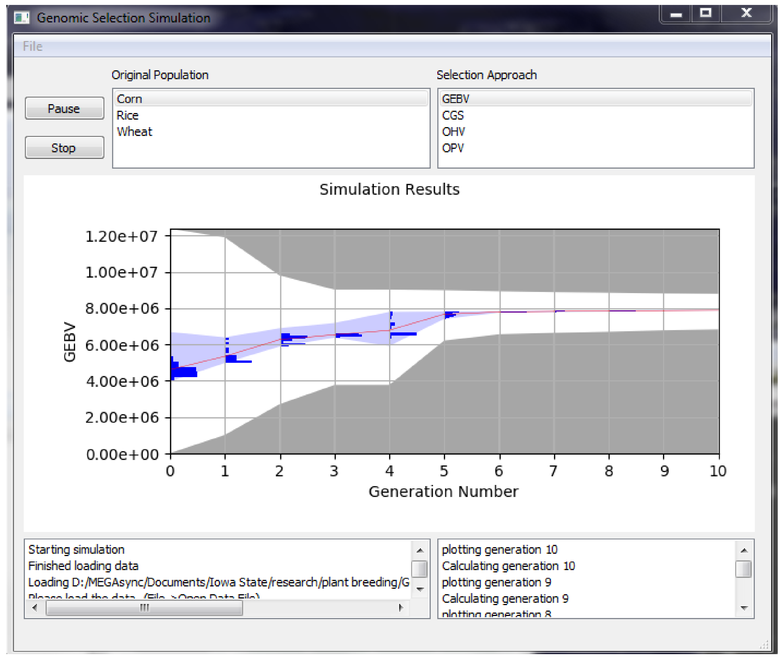
- Selection and mating steps are not well optimized yet.
- The key to GS is improving these steps.



Genomic selection process

2. GS Simulation Tool

- Informative
- Interactive
- Efficient
- Cross platform (written in Python)



- Dark blue sticks - histogram of breeding values
- Light blue filled part - maximum and minimum breeding values
- Grey part - the upper bound and lower bound
- Light blue filled part - maximum and minimum breeding values
- The breeding value will keep increasing and the change will become statistically insignificant after generation 6 in this case
- Mating function is well optimized (algorithm developed by Dr. Lizhi Wang)
- 3 sec. to mate one pair of genomes with 140 million loci on i5 6500.

3. Example

- GEBV is one of various breeding values which is used for selection.
- Suppose GEBV is determined by its genotype as follows

$$GEBV = \sum_{j=1}^{n} G_j e_j$$

- $GEBV$: Genomic Estimated Breeding Value
- $G_j \in \{0, 1\}$: Genetic variant of allele j. (0 means useless and 1 means useful)
- e_j: effect of $G_j = 1$ over $G_j = 0$.

Given

$$G = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}, e = \begin{bmatrix} 0.5 \\ 11.2 \\ 1.2 \\ 2.3 \\ 0.0 \end{bmatrix}$$

$$GEBV = 1 \cdot 0.5 + 1 \cdot 11.2 + 1 \cdot 1.2 + 2 \cdot 2.3 + 1 \cdot 0.0 = 17.5$$

4. Future research

- Some of the possible things to do in the future
  - Testing various selection methods at one time and automatically comparing how good each method is quantitatively (how fast and how high the breeding value reaches at the end.
  - Stopping the simulation automatically when the breeding value's change is not significant.
  - Automatically creating selection methods using deep learning.
- Breeders can develop a better selection method using this tool.
- Improvement of selection methods will lead to more efficient plant breeding, which is very important in solving global environmental problems and energy/food shortages.