

# Scripts and data files for simulation of UW potato breeding pipeline

Anil Adhikari

11 January 2021

The goal of this project is to model the potato breeding program of University of Wisconsin led by Dr. Jeff Endelman using AlphaSimR and see the effects of parents recycling at different stages of the pipeline. The potato breeding pipeline is in the file **UW\_potato\_breeding\_outline\_7.31.2020.xlsx**.

We are trying to answer these questions?

1. At what stage of the pipeline should we recycle parents (select parents for next cycle) for highest genetic gain?
2. Does genomic selection (GS) perform better than phenotypic selection?
3. At what stage should GS be implemented for maximum gain without depleting the genetic variance?

## Step 1: Filling the breeding pipeline and Burnin

In the first step, a breeding pipeline is created within script `UW_potato_burnin.R` and a burnin phase (20 years) is modeled with parents recycled at Stage 4 via phenotypic selection only (see pipeline for stage descriptions).

A training population for GS is created with inclusion of all individuals at Stage 3 in the last five years.

In addition, a “parent pool” that is a collection of parents from previous cycle and individuals at Stage4 of current cycle.

We also track these parameters for each year.

- **Mean:** Mean of population of Stage1 and Stage4.
- **Variance:** Genetic variance of population at Stage1.
- **Parents replaced:** Number of parents replaced in parent pool every cycle.

The burnin is iterated for 5 times and burnin datasets are saved in the “./Datasets” folder named as “BURNIN\_PP(1-5).RData”

## Step 2: Different parent recycling scenarios

After the burnin phase that runs for 20 years, parent recycling scenarios are implemented for additional 20 years and above mentioned parameters are tracked.

**Baseline model** This model keeps running the baseline model (parent selection via phenotypic selection at Stage4) for additional 20 years. This is run by script `UW_potato_baseline_model.R`. The script saves output as .RDS files for each iteration named as “Stage4\_sel\_(1-5).rds” in the “./Datasets” directory.

**Stage3 selection** In this model, parents are selected at Stage3 via phenotypic selection instead of Stage4. This model is run by script `UW_potato_stage3_selection.R`. The script saves output as .RDS files for each iteration named as “**Stage3\_sel\_(1-5).rds**” in the `**“./Datasets”*` directory.

**Stage2 selection** In this model, parents are selected at Stage2 via phenotypic selection instead of Stage4. This model is run by script `UW_potato_stage2_selection.R`. The script saves output as .RDS files for each iteration named as “**Stage2\_sel\_(1-5).rds**” in the `**“./Datasets”*` directory.

**Stage1 selection** In this model, parents are selected at Stage1 via phenotypic selection instead of Stage4. This model is run by script `UW_potato_stage1_selection.R`. The script saves output as .RDS files for each iteration named as “**Stage1\_sel\_(1-5).rds**” in the `**“./Datasets”*` directory.

**Stage3 genomic selection** In this model, parents are selected at Stage3 via genomic selection. The genomic selection model is trained on a training population that consists of individuals at stage3 from last 5 years which is saved at the burnin stage. This model is run by script `UW_potato_stage3_GS.R`. The script saves output as .RDS files for each iteration named as “**Stage3\_GS\_(1-5).rds**” in the `“./Datasets”` directory. In this model, GS accuracy is also tracked which is saved in the .RDS file.

**Stage2 genomic selection** In this model, parents are selected at Stage2 via genomic selection. The genomic selection model is trained on a training population that consists of individuals at stage3 from last 5 years which is saved at the burnin stage. This model is run by script `UW_potato_stage2_GS.R`. The script saves output as .RDS files for each iteration named as “**Stage2\_GS\_(1-5).rds**” in the `**“./Datasets”*` directory. In this model, GS accuracy is also tracked which is saved in the .RDS file.

### Step 3: Create plots

The script `UW_potato_plot_results.R` access the datasets saved in `“./Datasets”` folder to create the following plots.

1. **Genetic.gain.stage1.pdf** : This plot shows the genetic gain in different parent selection strategies measured at Stage1.
2. **Genetic.gain.stage4.pdf** : This plot shows the genetic gain in different parent selection strategies measured at Stage4.
3. **Genetic\_variance\_sel.stages.pdf** : This plot shows the change in relative genetic variance across 20 years of selection in different parent selection strategies.
4. **Parent\_replacement\_parentpool.pdf** : This plot shows the number of parents replaced in the parentpool. Parentpool for each cycle includes the selected parents from previous cycle and also the parents used in crossing in the cycle before that.
5. **GS\_accuracy\_plot.pdf** : This plot shows the GS accuracy tracked over the 20 years of breeding.