

Integration of Genomic Selection into Hybrid Maize Breeding Schemes

Jose Osorio
AgReliant Genetics

Genomic Selection (GS) is a quite recent approach for MAS (Marker-Assisted Selection) that allows selecting favorable individuals based on GEBVs (Genomic Estimated Breeding Values): a set of reference individuals both phenotyped and genotyped is used to estimate the effect of markers on phenotype; the marker effects are then added along the genome to compute GEBVs of genotyped-only candidate individuals. Previous studies have suggested that GS is particularly useful for capturing QTL (Quantitative Trait Loci) having a small-effect on the trait of interest. It is becoming a popular methodology in the field of plant breeding as information for its practical use becomes available. Basic concepts used in GS methodology are addressed to illustrate its integration into maize breeding schemes. Furthermore, several points are discussed including and GS efficiency and its components: selection intensity, prediction accuracy and cycle length. GS is an approach complimentary to phenotypic selection that depends on phenotypic evaluation to build the prediction model. This method will be integrated into many practical breeding programs in the near future with further advances and the maturing of its theory. It is anticipated that GS will be successful and that the corn breeding companies will commercialize more and more GS-derived hybrids.

Integration of Genomic Selection (GS) into Hybrid Maize Breeding Schemes

Jose Osorio



Outline

- **GS Methodology**
- **GS Breeding Schemes**
- **GS Efficiency**
- **Practical Considerations**
- **Conclusions**

GS Methodology

GS is a 2 step procedure:

→ 1. **Training Step**



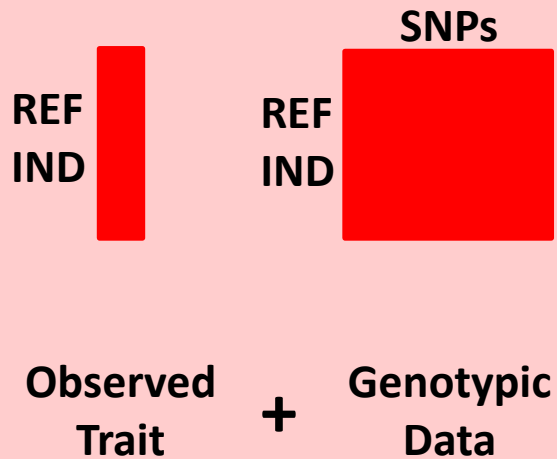
REF
IND

Observed
Trait

GS Methodology

GS is a 2 step procedure:

→ **1. Training Step**

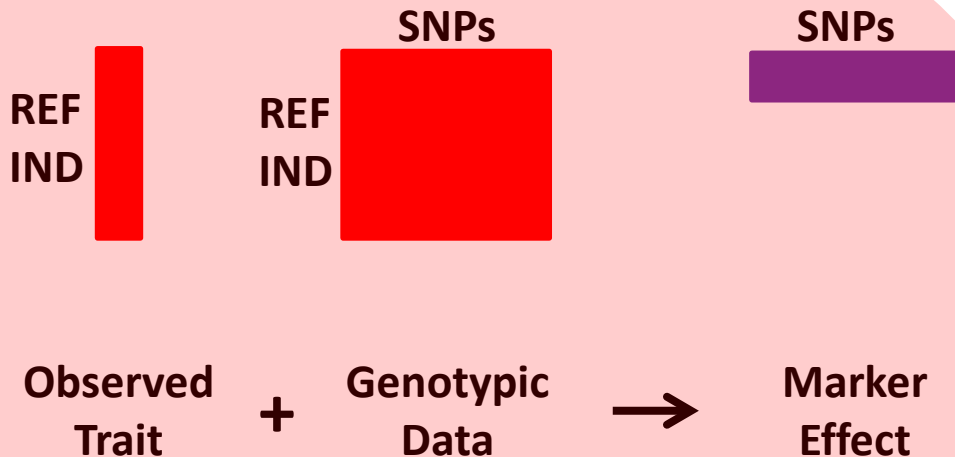


GS Methodology

GS is a 2 step procedure:

→ **1. Training Step**

→ **Marker Effects**



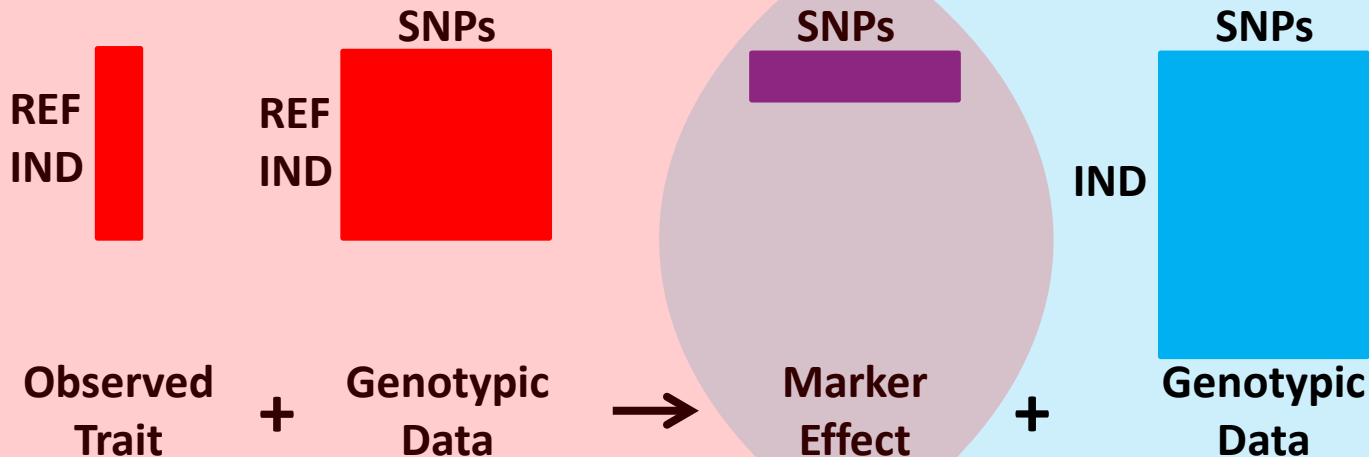
GS Methodology

GS is a 2 step procedure:

→ 1. Training Step

→ Marker Effects

→ 2. Prediction Step



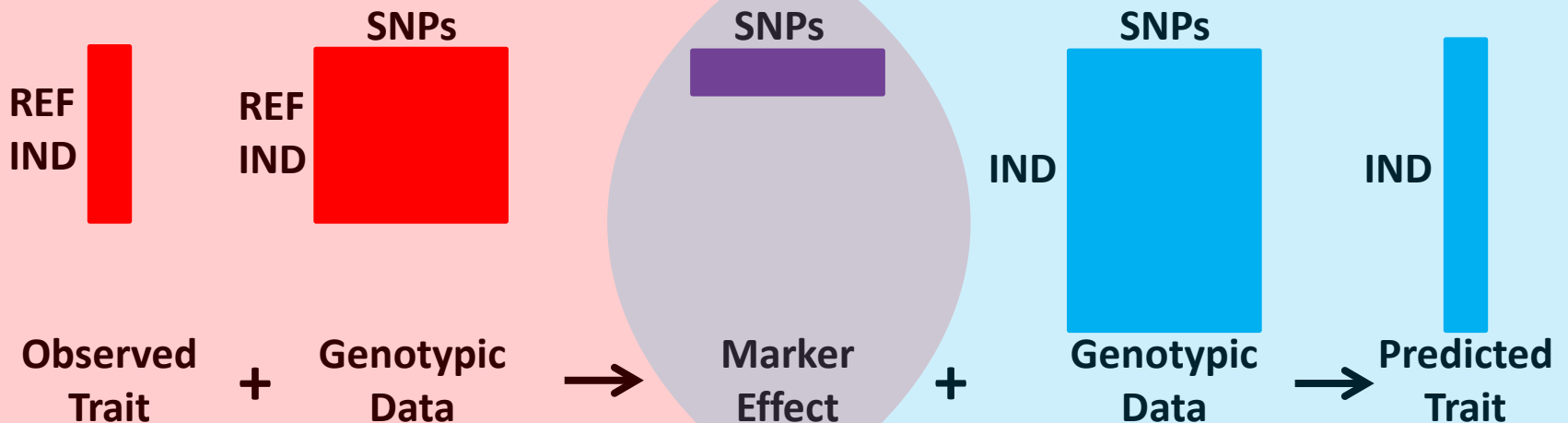
GS Methodology

GS is a 2 step procedure:

→ 1. Training Step

→ Marker Effects

→ 2. Prediction Step



GS Methodology

GS uses a training population of individuals both phenotyped and genotyped with many markers distributed genome wide

The effects that markers have on phenotype are estimated simultaneously using the RR-BLUP method, for instance

Marker effects are added across the genome to predict the value for the trait of interest for genotyped-only individuals

→ Selection of individuals **without phenotypic data**

GS Methodology

GS uses a training population of individuals both phenotyped and genotyped with many markers distributed genome wide

The effects that markers have on phenotype are estimated simultaneously using the RR-BLUP method, for instance

Marker effects are added across the genome to predict the value for the trait of interest for genotyped-only individuals

→ Selection of individuals **without** phenotypic data

→ Selection of individuals **with** phenotypic data

→ **2nd source of information**

GS Breeding Schemes: example

Year Season

Standard Scheme

1



A x B

1



F1 ← Inducer

2



Haploids

2



DH per se

3



DH x Tester

DH in Cold Room

3



DH per se TC1

GS model → GS

4

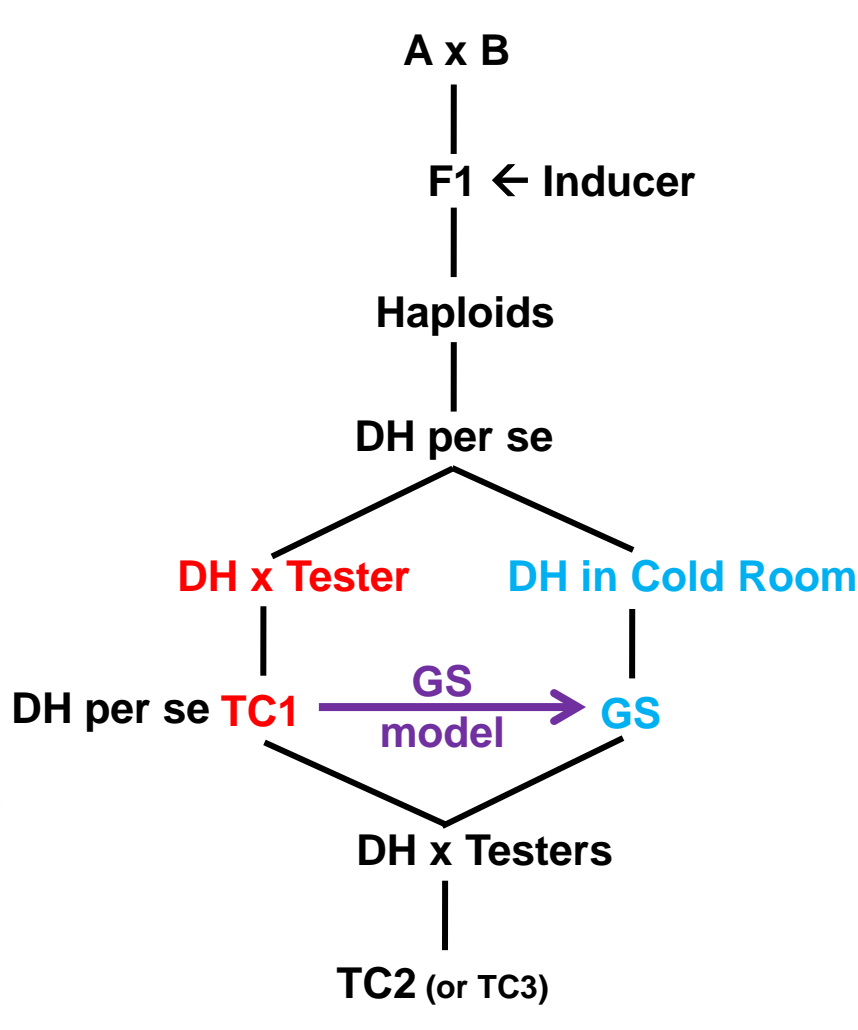


DH x Testers

4



TC2 (or TC3)



GS Breeding Schemes: example

Year Season

Standard Scheme

Accelerated Scheme

1



A x B

A x B

1



F1 ← Inducer

F1 ← Inducer

Haploids

2



Haploids

DH x Tester

DH in Cold Room

2



DH per se

DH per se

TC1

GS model → GS

3



DH x Tester

DH in Cold Room

3



DH per se

TC1

GS model

GS

DH x Testers

4



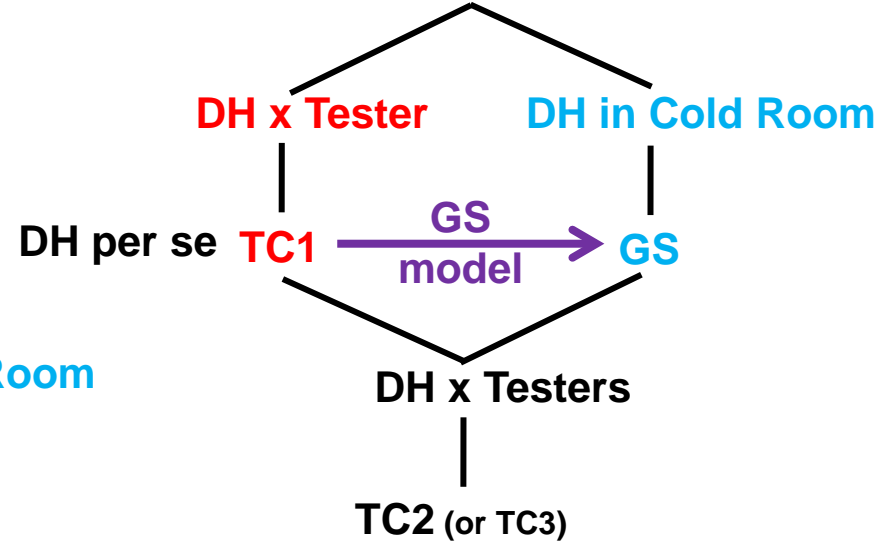
DH x Testers

TC2 (or TC3)

4



TC2 (or TC3)



GS Breeding Schemes: example

Year Season

Standard Scheme

Accelerated Scheme

1



A x B

A x B

1



F1 ← Inducer

F1 ← Inducer
Haploids

2



Haploids

DH x Tester

DH in Cold Room

2



DH per se

DH per se

TC1

GS model → GS

3



DH x Tester

DH in Cold Room

3



DH per se

TC1

GS model → GS

GS

DH x Testers

4



DH x Testers

4



TC2 (or TC3)

TC2 (or TC3)

Lower cost for hybrid seed production per testing unit

Higher cost for hybrid seed production per testing unit

GS Breeding Schemes: cycle length reduction

Phenotypic Selection

Year Season

1 

1 

2 

2 

3 

3 

4 

4 

5 

5 

A x B



F1 ← Inducer



Haploids



DH per se



DH x Tester



TC1



C x D



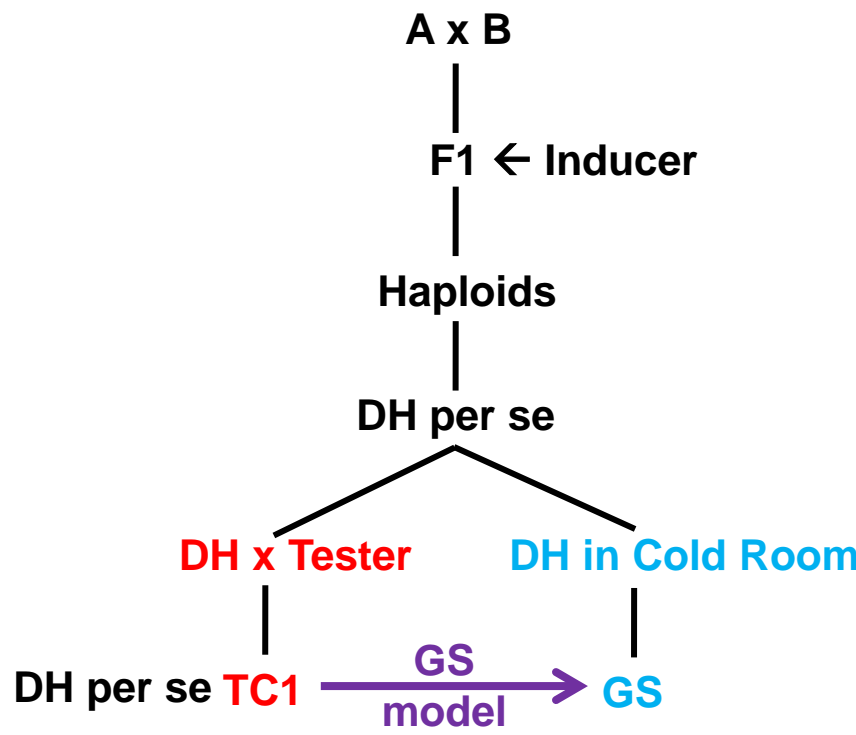
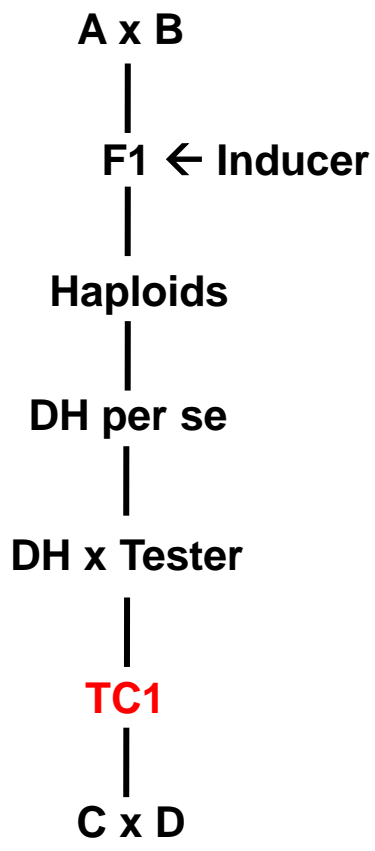
Average Cycle Length = 3 years

GS Breeding Schemes: cycle length reduction

Year Season

Phenotypic Selection

Phenotypic Selection & GS



Average Cycle Length = 3 years

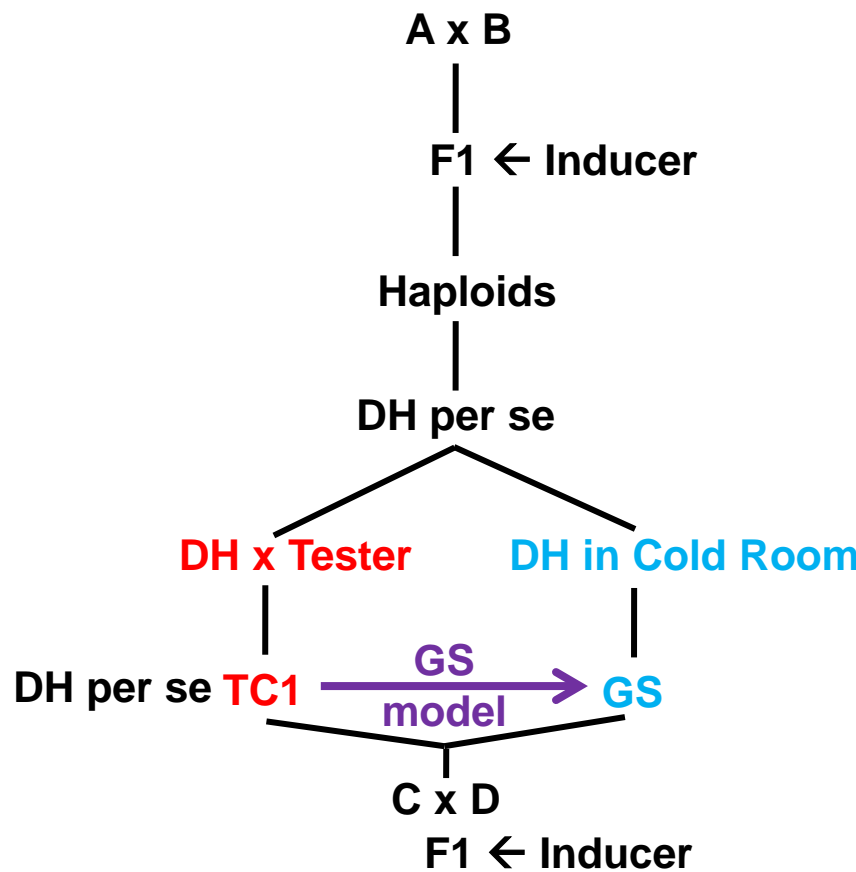
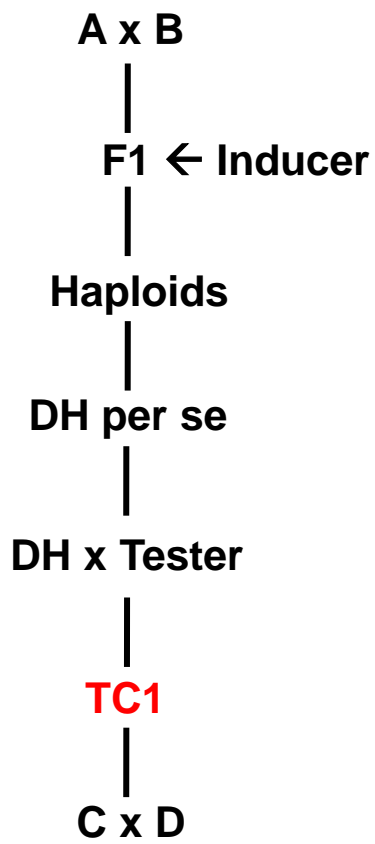
Lower cost for hybrid seed production per testing unit

GS Breeding Schemes: cycle length reduction

Year Season

Phenotypic Selection

Phenotypic Selection & GS



Average Cycle Length = 3 years

Average Cycle Length = 2 years

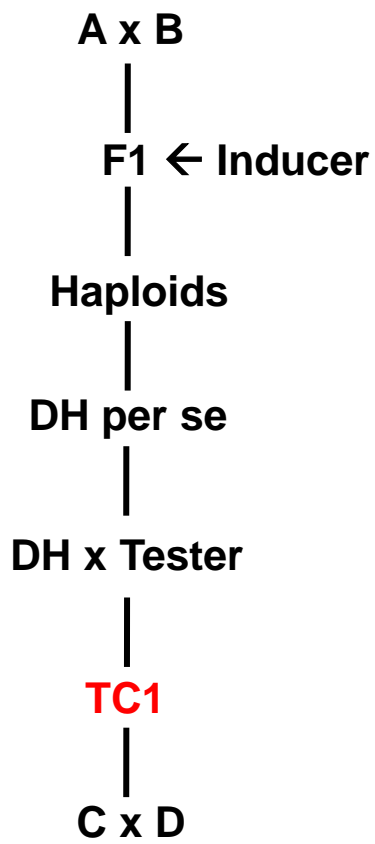
Much lower cost for hybrid seed production per testing unit

GS Breeding Schemes: cycle length reduction

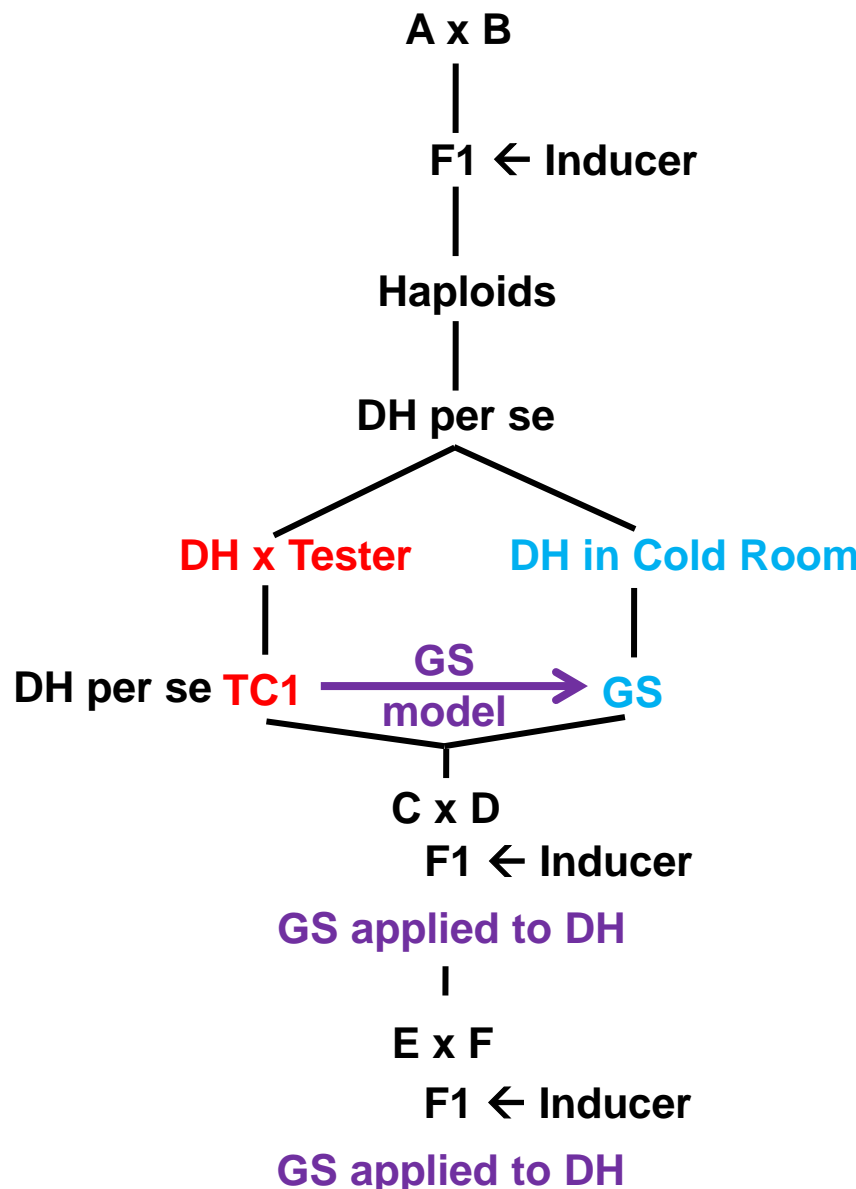
Year Season

Phenotypic Selection

Phenotypic Selection & GS



Average Cycle Length = 3 years



Average Cycle Length = 1.7 years

GS Breeding Schemes: classification

**Calibration and Prediction in the
Same Breeding Cycle**

**Calibration and Prediction in
Different Breeding Cycles**

recombination between cycles

GS Breeding Schemes: classification

**Calibration and Prediction in the
Same Breeding Cycle**

Longer average cycle length

**Calibration and Prediction in
Different Breeding Cycles**

recombination between cycles

Shorter average cycle length

GS Breeding Schemes: classification

Calibration and Prediction in the Same Breeding Cycle

Longer average cycle length

Higher cost for TC1 evaluations

Calibration and Prediction in Different Breeding Cycles

recombination between cycles

Shorter average cycle length

Reduced cost for TC1 evaluations

GS Breeding Schemes: classification

Calibration and Prediction in the Same Breeding Cycle

Longer average cycle length

Higher cost for TC1 evaluations

Larger contribution of epistasis

Less QTL x background interaction

Additive and Non-Add. gene effects

Calibration and Prediction in Different Breeding Cycles

recombination between cycles

Shorter average cycle length

Reduced cost for TC1 evaluations

Smaller contribution of epistasis

More QTL x background interaction

Additive gene effects only

GS Breeding Schemes: classification

Calibration and Prediction in the Same Breeding Cycle

Longer average cycle length

Higher cost for TC1 evaluations

Larger contribution of epistasis

Less QTL x background interaction

Additive and Non-Add. gene effects

Gain criterion is Genetic Value

Calibration and Prediction in Different Breeding Cycles

recombination between cycles

Shorter average cycle length

Reduced cost for TC1 evaluations

Smaller contribution of epistasis

More QTL x background interaction

Additive gene effects only

Gain criterion is Breeding Value

GS Breeding Schemes: classification

Calibration and Prediction in the Same Breeding Cycle

Longer average cycle length

Higher cost for TC1 evaluations

Larger contribution of epistasis

Less QTL x background interaction

Additive and Non-Add. gene effects

Gain criterion is Genetic Value

Relatedness and LD → GS accuracy

Calibration and Prediction in Different Breeding Cycles

recombination between cycles

Shorter average cycle length

Reduced cost for TC1 evaluations

Smaller contribution of epistasis

More QTL x background interaction

Additive gene effects only

Gain criterion is Breeding Value

LD and relatedness → GS accuracy

GS Breeding Schemes: type of crosses

**Biparental
cross**

**Selection
Within
Family**

**Multiple
crosses**

**Selection
Within & Between
Families**

**Half Sib
Families**

**Selection
Within & Between
Families**

**Synthetic
Population**

**Selection
Between
Individuals**

GS Breeding Schemes: type of crosses

**Biparental
cross**

**Selection
Within
Family**

**Multiple
crosses**

**Selection
Within & Between
Families**

**Half Sib
Families**

**Selection
Within & Between
Families**

**Synthetic
Population**

**Selection
Between
Individuals**



number of lines in the training set

GS Breeding Schemes: type of crosses

**Biparental
cross**

**Selection
Within
Family**

**Multiple
crosses**

**Selection
Within & Between
Families**

**Half Sib
Families**

**Selection
Within & Between
Families**

**Synthetic
Population**

**Selection
Between
Individuals**

number of lines in the training set

effect of epistasis

GS Breeding Schemes: type of crosses

**Biparental
cross**

**Selection
Within
Family**

**Multiple
crosses**

**Selection
Within & Between
Families**

**Half Sib
Families**

**Selection
Within & Between
Families**

**Synthetic
Population**

**Selection
Between
Individuals**

number of lines in the training set

effect of epistasis

LD block size

GS Breeding Schemes: type of crosses

**Biparental
cross**

**Selection
Within
Family**

**Multiple
crosses**

**Selection
Within & Between
Families**

**Half Sib
Families**

**Selection
Within & Between
Families**

**Synthetic
Population**

**Selection
Between
Individuals**

number of lines in the training set

effect of epistasis

LD block size

relatedness contribution to GS accuracy

LD contribution to GS accuracy

GS Breeding Schemes: type of crosses

**Biparental
cross**

**Selection
Within
Family**

**Multiple
crosses**

**Selection
Within & Between
Families**

**Half Sib
Families**

**Selection
Within & Between
Families**

**Synthetic
Population**

**Selection
Between
Individuals**

number of lines in the training set

effect of epistasis

LD block size

relatedness contribution to GS accuracy

LD contribution to GS accuracy

required marker density

GS Breeding Schemes: type of crosses

**Biparental
cross**

**Selection
Within
Family**

**Multiple
crosses**

**Selection
Within & Between
Families**

**Half Sib
Families**

**Selection
Within & Between
Families**

**Synthetic
Population**

**Selection
Between
Individuals**

number of lines in the training set

effect of epistasis

LD block size

relatedness contribution to GS accuracy

LD contribution to GS accuracy

required marker density

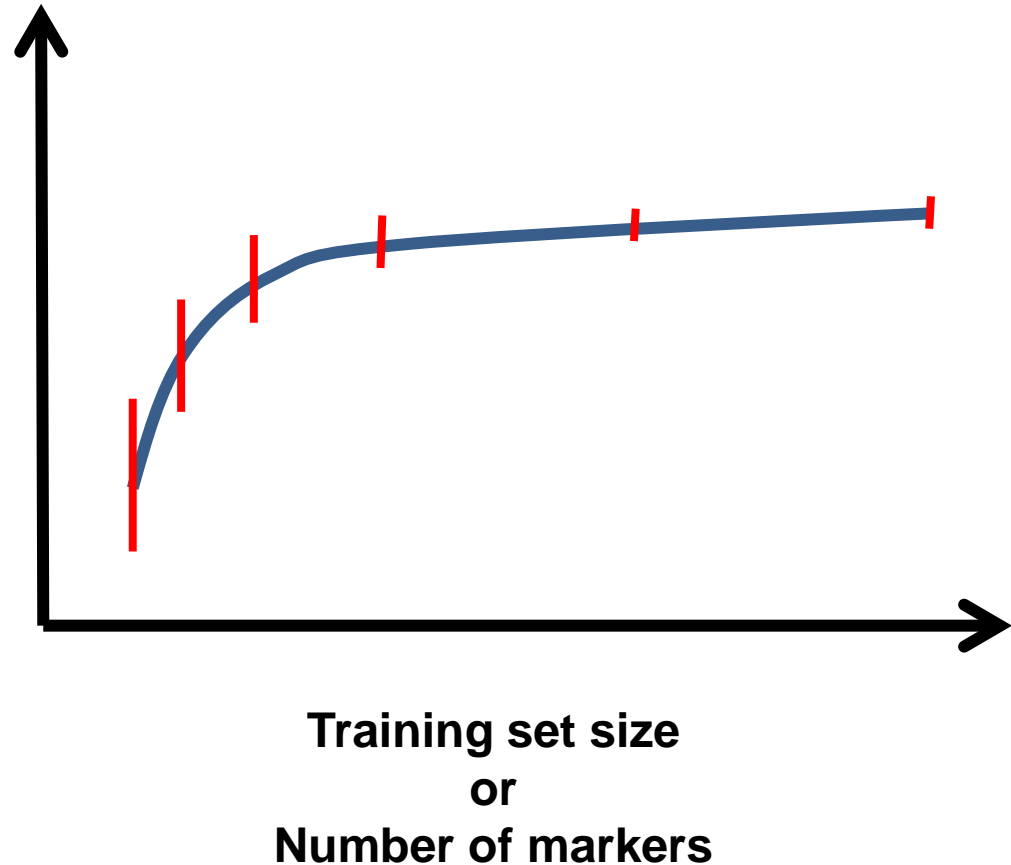
→ Perform pilot study with target genetic material and maturity zones

GS Breeding Schemes: pilot study

→ Perform low scale pilot studies for trait of interest with appropriate genetic material in target maturity zones

→ **Several random samplings**

Mean correlation between observed and predicted values



→ Mean GS accuracy increases with # plants (markers) and reaches a plateau
The standard deviation is reduced with # plants (markers)

GS Efficiency : response to selection

GS is a MAS approach which goal is to increase response to selection:

Response to selection with **phenotypic selection** $\rightarrow R = i r_{P,G} \sigma_A$ with $r_{P,G} = h$

Response to selection with **genomic selection** $\rightarrow R = i r_{M,G} \sigma_A$ with $r_{M,G} = r_{M,P} / h$

GS Efficiency : response to selection

GS is a MAS approach which goal is to increase response to selection:

Response to selection with **phenotypic selection** $\rightarrow R = i r_{P,G} \sigma_A$ or $R = i h \sigma_A$

Response to selection with **genomic selection** $\rightarrow R = i r_{M,G} \sigma_A$ with $r_{M,G} = r_{M,P} / h$

$$GS \text{ Efficiency} = \frac{i_{GS} r_{M,G} \sigma_A}{i_{PS} r_{P,G} \sigma_A}$$

$$GS \text{ Efficiency} = \frac{i_{GS} r_{M,G}}{i_{PS} r_{P,G}}$$

GS Efficiency : response to selection

GS is a MAS approach which goal is to increase response to selection:

Response to selection with **phenotypic selection** $\rightarrow R = i r_{P,G} \sigma_A$ or $R = i h \sigma_A$

Response to selection with **genomic selection** $\rightarrow R = i r_{M,G} \sigma_A$ with $r_{M,G} = r_{M,P} / h$

$$GS \text{ Efficiency} = \frac{i_{GS} r_{M,G} \sigma_A}{i_{PS} r_{P,G} \sigma_A}$$

$$GS \text{ Efficiency} = \frac{i_{GS} r_{M,G}}{i_{PS} r_{P,G}}$$

$$GS \text{ Efficiency} = \frac{i_{GS} r_{M,G} / GS \text{ cycle length}}{i_{PS} r_{P,G} / PS \text{ cycle length}}$$



GS Efficiency : response to selection

GS is a MAS approach which goal is to increase response to selection:

Response to selection with **phenotypic selection** $\rightarrow R = i r_{P,G} \sigma_A$ or $R = i h \sigma_A$

Response to selection with **genomic selection** $\rightarrow R = i r_{M,G} \sigma_A$ with $r_{M,G} = r_{M,P} / h$

$$GS \text{ Efficiency} = \frac{i_{GS} r_{M,G} \sigma_A}{i_{PS} r_{P,G} \sigma_A}$$

$$GS \text{ Efficiency} = \frac{i_{GS} r_{M,G}}{i_{PS} r_{P,G}}$$

$$GS \text{ Efficiency} = \frac{i_{GS} r_{M,G} / \text{GS cycle length}}{i_{PS} r_{P,G} / \text{PS cycle length}}$$

Intensity
or
"Quantity"

Accuracy
or
"Quality"

Length
or
"Speed"



GS Efficiency : response to selection

GS is a MAS approach which goal is to increase response to selection:

Response to selection with **phenotypic selection** $\rightarrow R = i r_{P,G} \sigma_A$ or $R = i h \sigma_A$

Response to selection with **genomic selection** $\rightarrow R = i r_{M,G} \sigma_A$ with $r_{M,G} = r_{M,P} / h$

$$GS \text{ Efficiency} = \frac{i_{GS} r_{M,G} \sigma_A}{i_{PS} r_{P,G} \sigma_A}$$

$$GS \text{ Efficiency} = \frac{i_{GS} r_{M,G}}{i_{PS} r_{P,G}}$$

$$GS \text{ Efficiency} = \frac{i_{GS} r_{M,G} / \text{GS cycle length}}{i_{PS} r_{P,G} / \text{PS cycle length}}$$



Cost

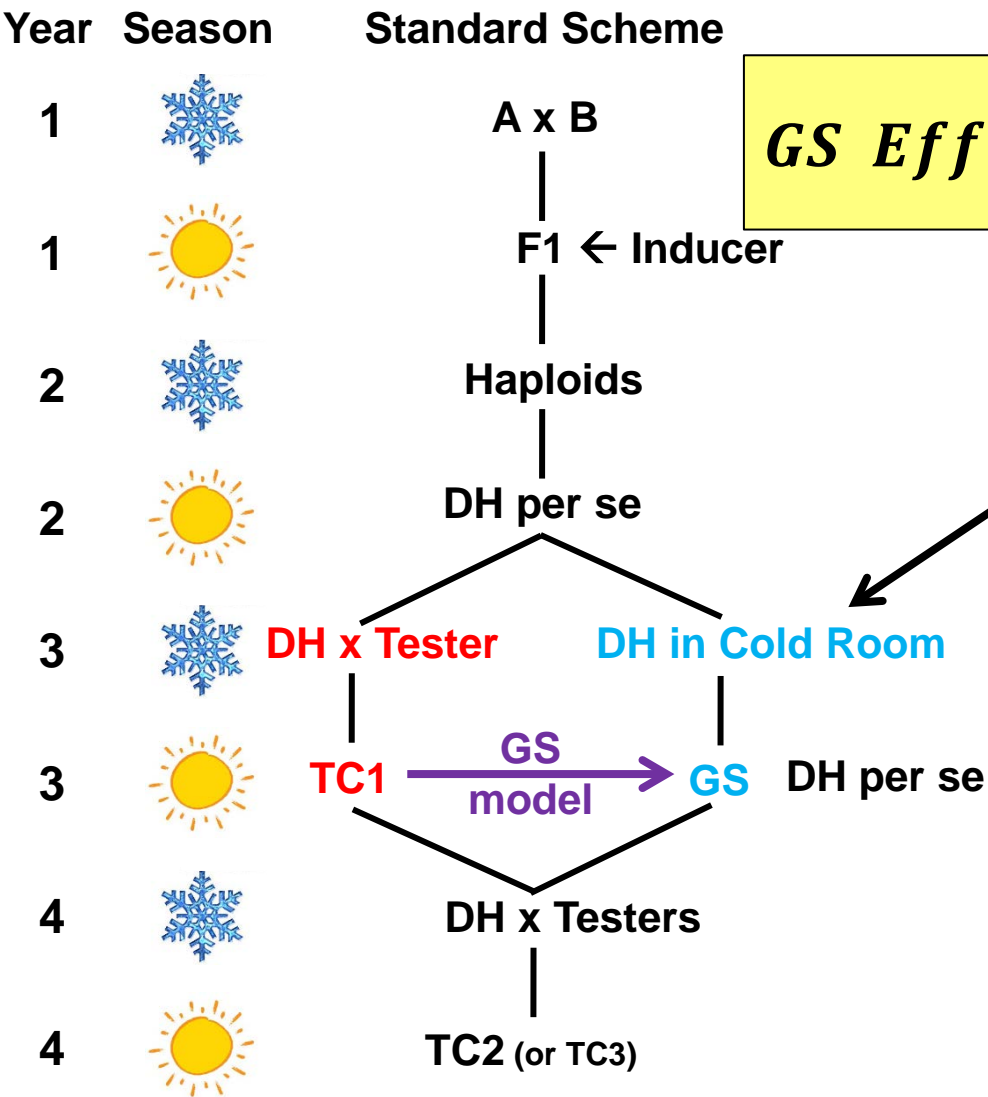
Intensity
or
"Quantity"

Accuracy
or
"Quality"

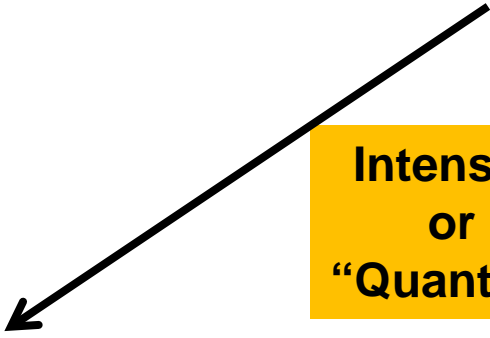
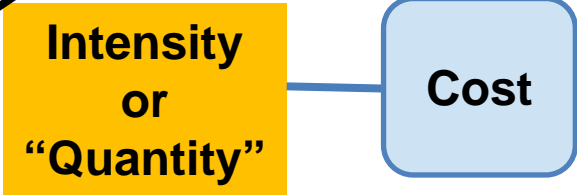
Length
or
"Speed"



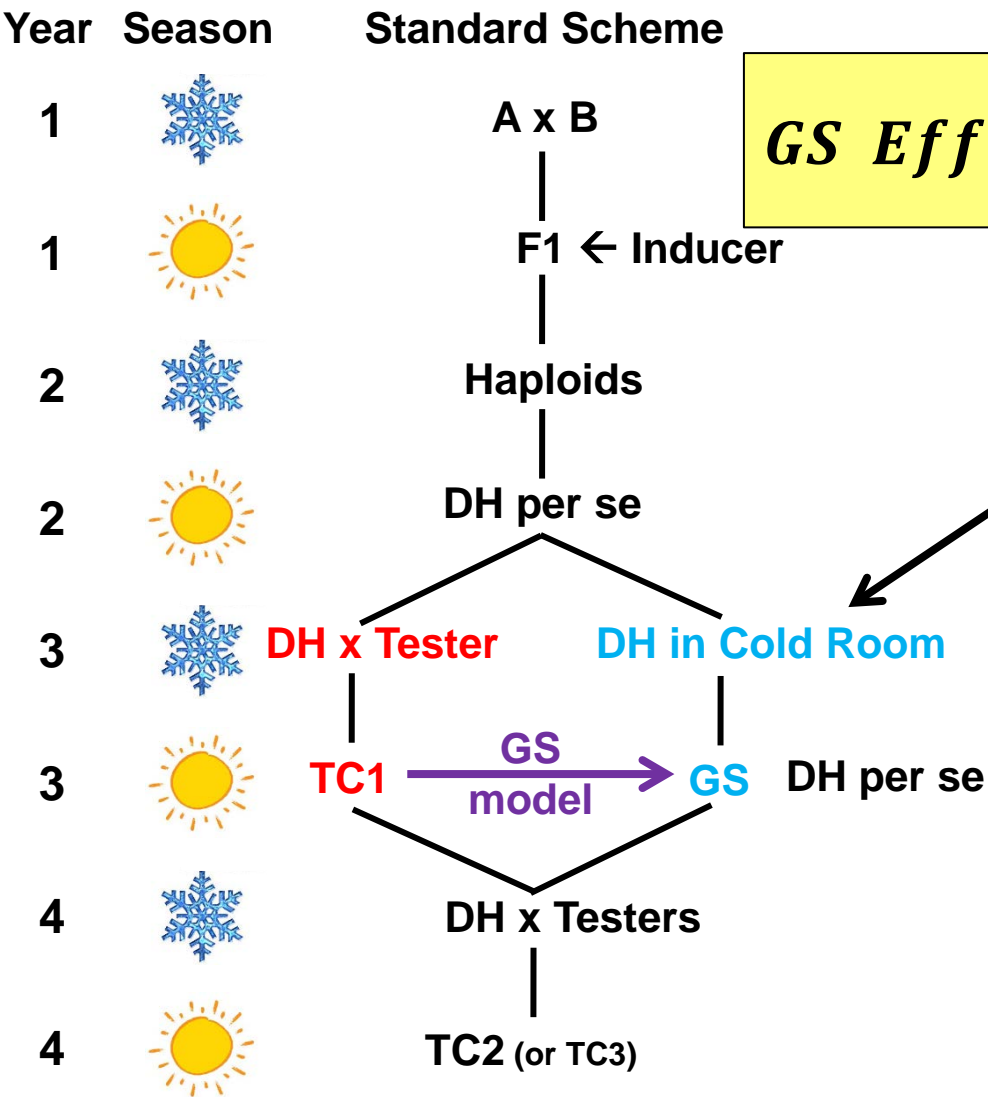
GS Efficiency: selection intensity



$$GS \text{ Efficiency} = \frac{i_{GS} r_{M, G} / GS \text{ cycle length}}{i_{PS} r_{P, G} / PS \text{ cycle length}}$$



GS Efficiency: selection intensity

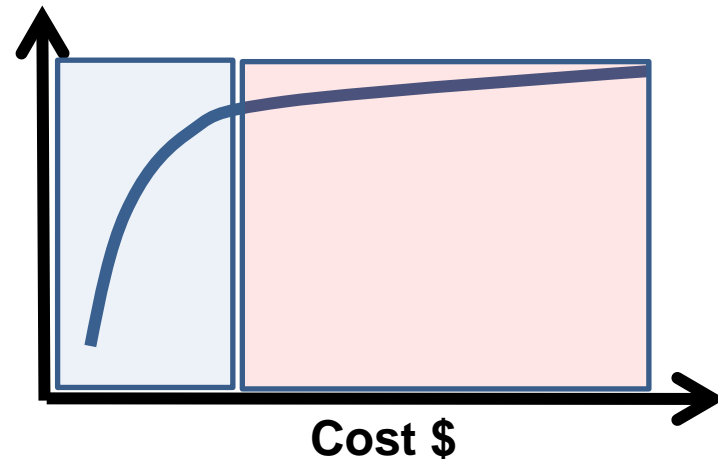


$$GS \text{ Efficiency} = \frac{i_{GS} r_{M, G} / GS \text{ cycle length}}{i_{PS} r_{P, G} / PS \text{ cycle length}}$$

Intensity or "Quantity"

Cost

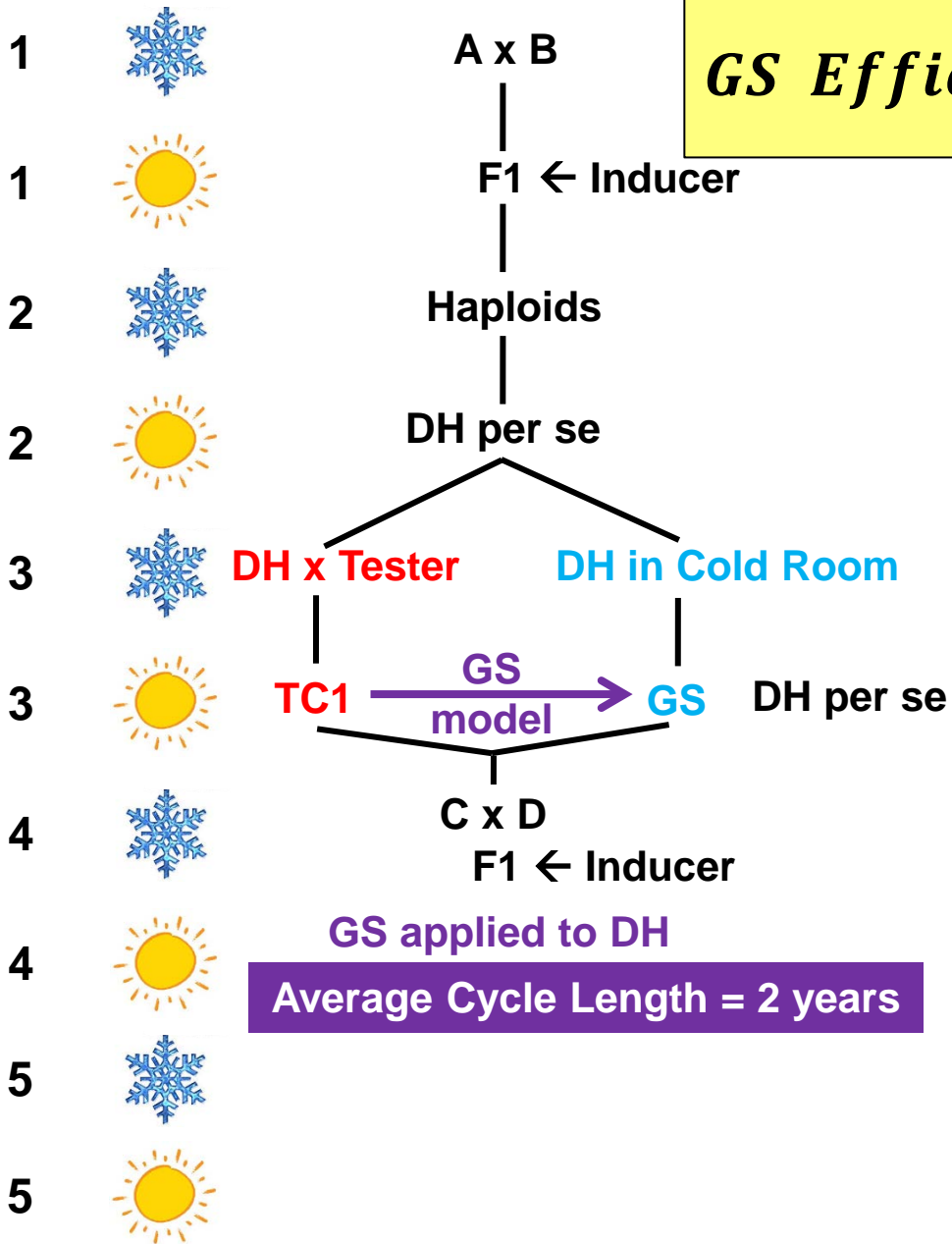
Selection Intensity



GS Efficiency: cycle length

Year Season

Phenotypic Selection & GS



$$GS \text{ Efficiency} = \frac{i_{GS} r_{M, G} / GS \text{ cycle length}}{i_{PS} r_{P, G} / PS \text{ cycle length}}$$

Length or "Speed"

GS applied to DH
Average Cycle Length = 2 years

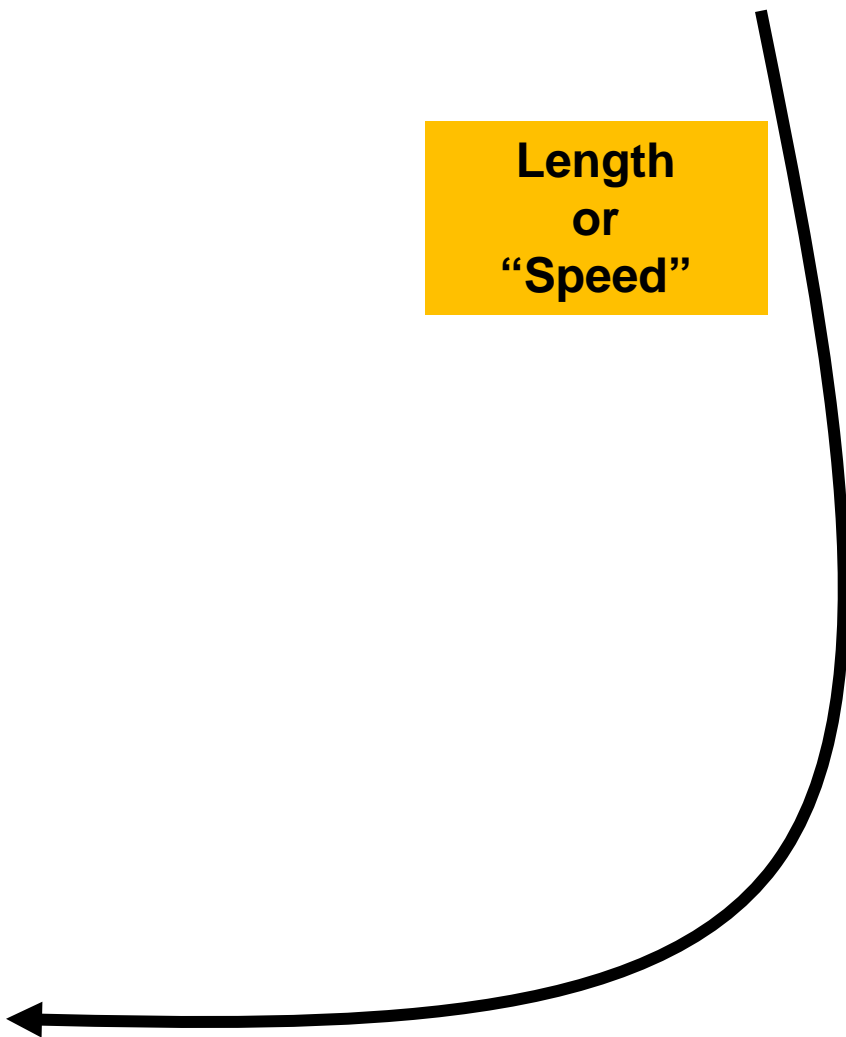
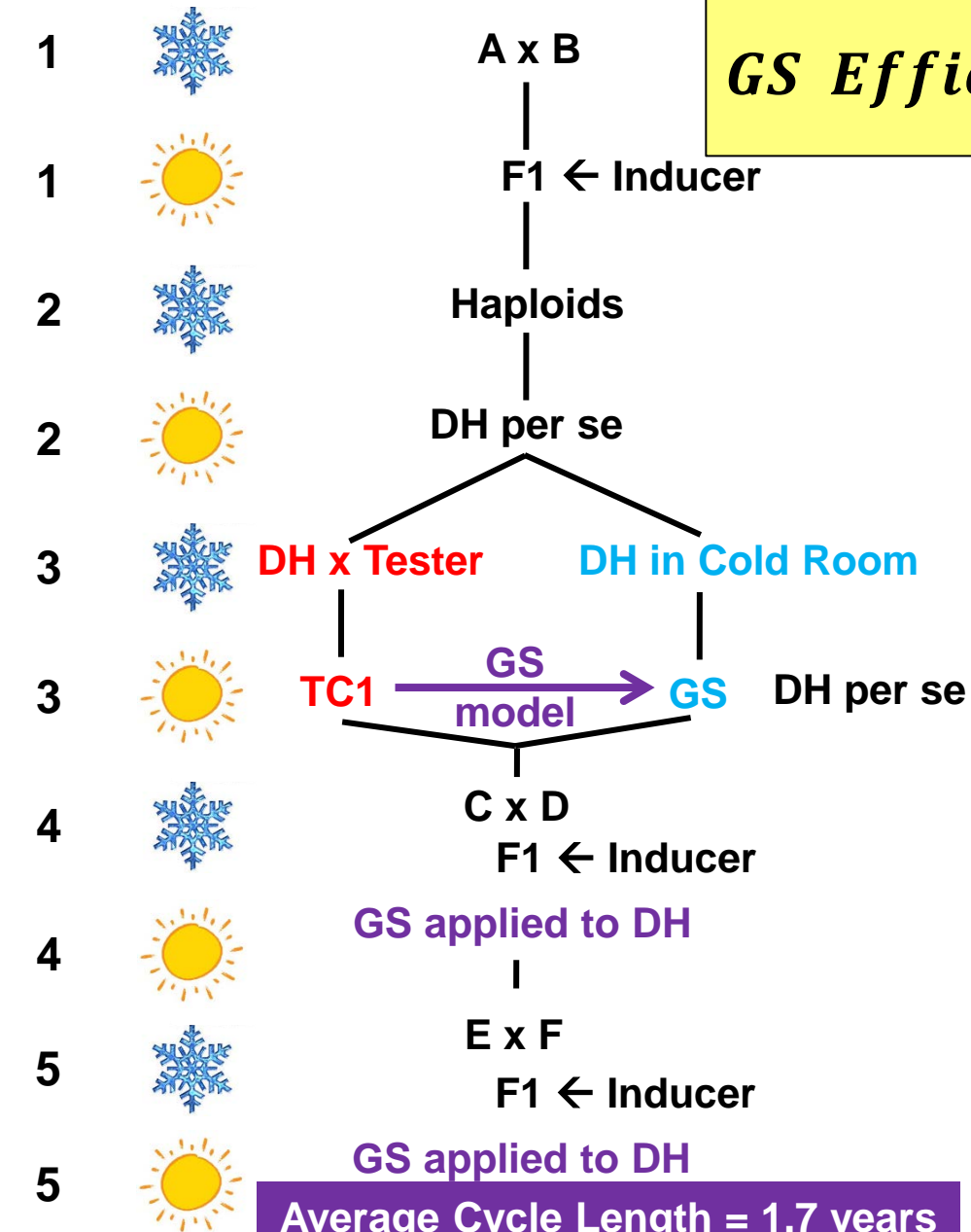
GS Efficiency: cycle length

Year Season

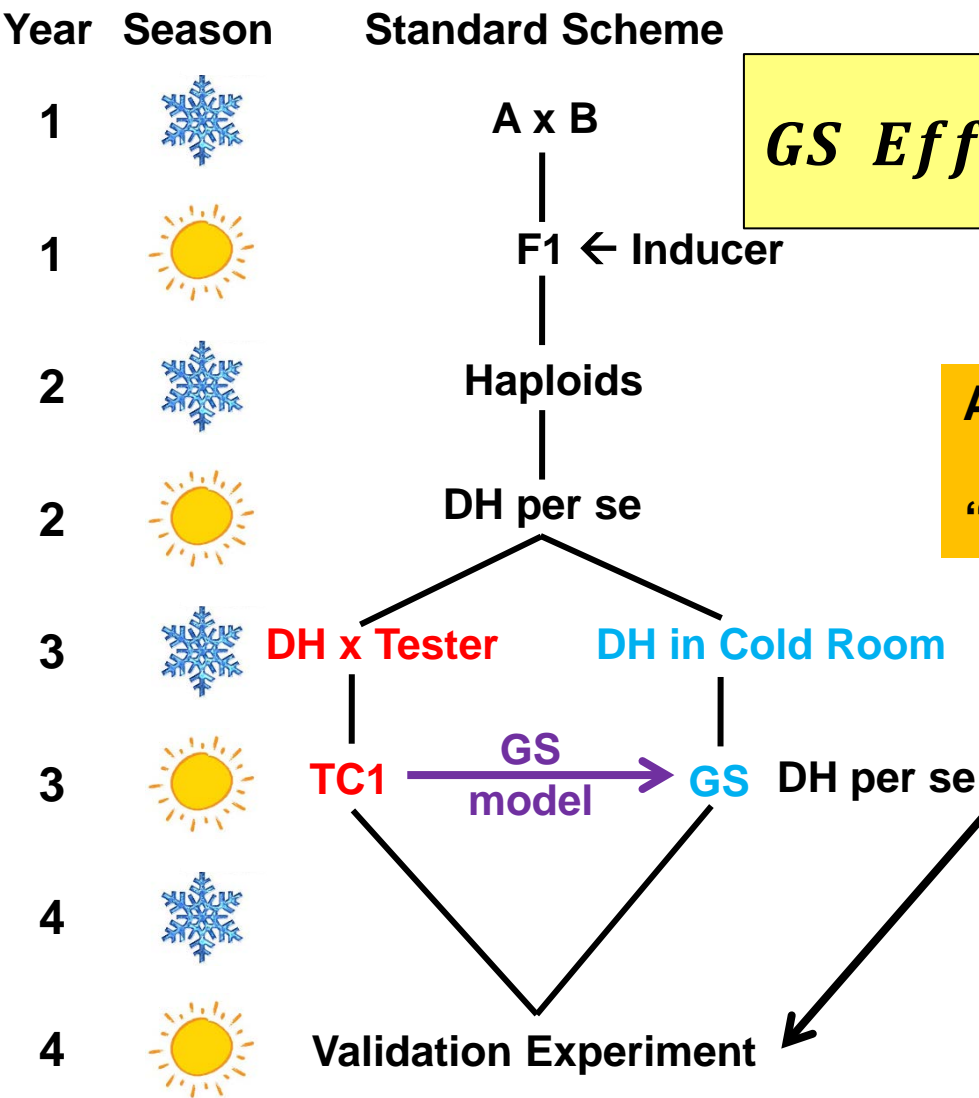
Phenotypic Selection & GS

$$GS \text{ Efficiency} = \frac{i_{GS} r_{M, G} / GS \text{ cycle length}}{i_{PS} r_{P, G} / PS \text{ cycle length}}$$

Length or "Speed"

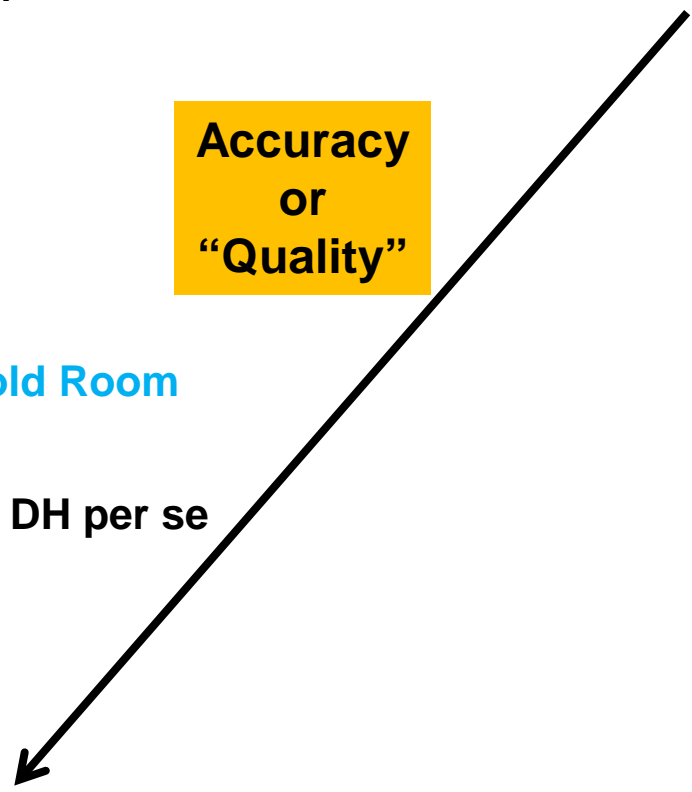


GS Efficiency: prediction ability

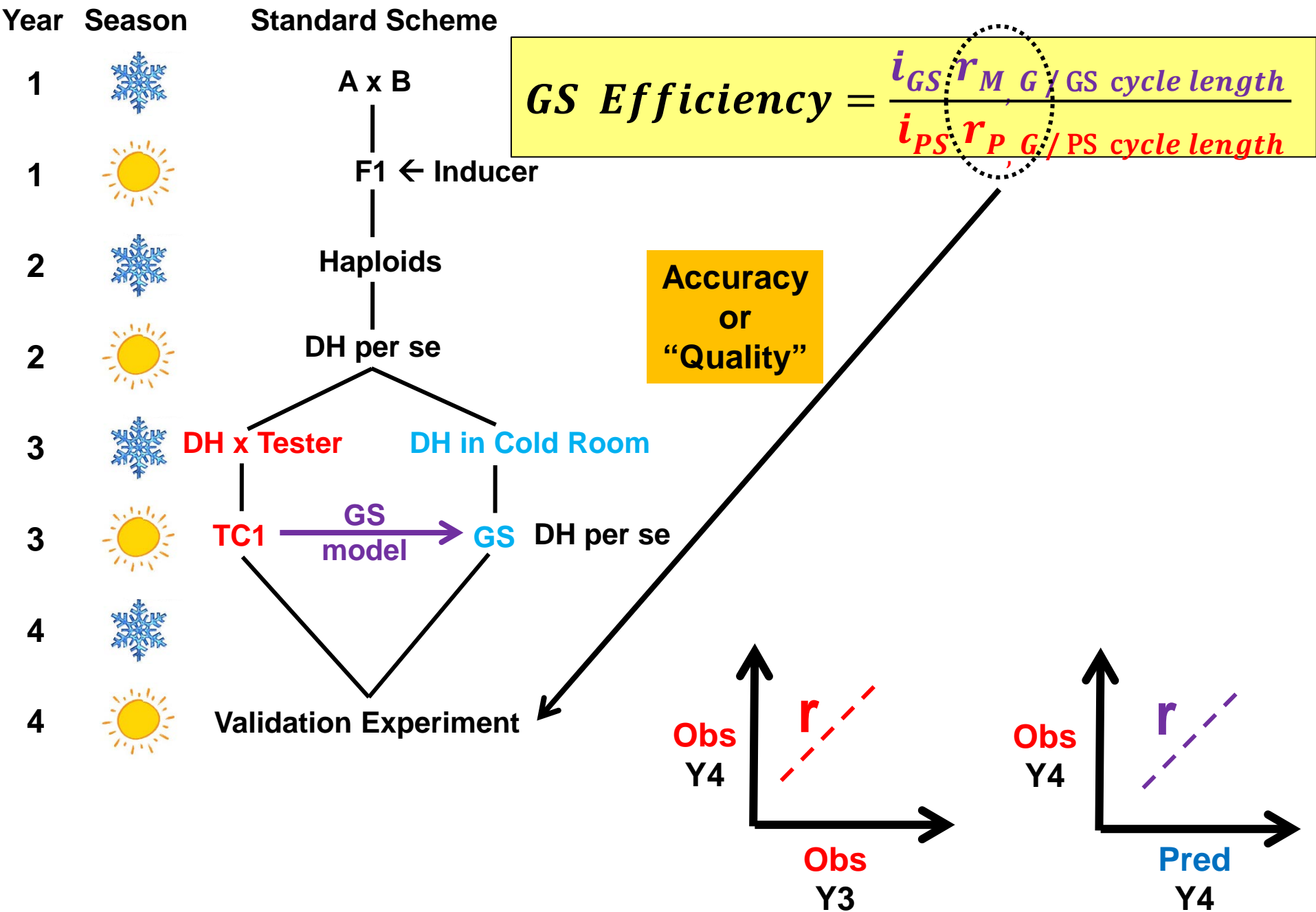


$$GS \text{ Efficiency} = \frac{i_{GS} r_{M, G} / GS \text{ cycle length}}{i_{PS} r_{P, G} / PS \text{ cycle length}}$$

Accuracy
or
"Quality"



GS Efficiency: prediction ability



Practical Considerations : data analysis

Phenotypic evaluation

- good trial network
- QC: field notes & outlier detection

Practical Considerations : data analysis

Phenotypic evaluation

- good trial network**
- QC: field notes & outlier detection**

Molecular data production

- good genotyping platform**
- QC: missing, heterozygous and error rates for individuals and markers**

Practical Considerations : data analysis

Phenotypic evaluation

- good trial network**
- QC: field notes & outlier detection**

Molecular data production

- good genotyping platform**
- QC: missing, heterozygous and error rates for individuals and markers**

Data storage into databases for traceability and efficiency

Practical Considerations : data analysis

Phenotypic evaluation

- good trial network
- QC: field notes & outlier detection

Molecular data production

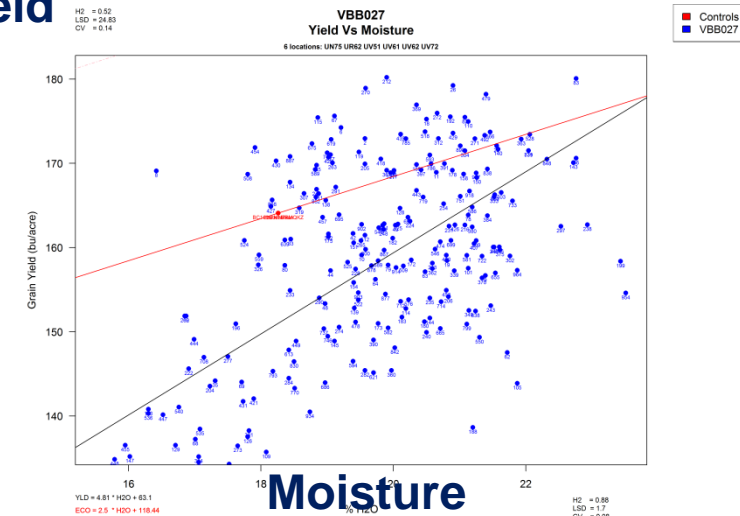
- good genotyping platform
- QC: missing, heterozygous and error rates for individuals and markers

Data storage into databases for traceability and efficiency

GS analysis

- accurate and robust method
- rapid for routine analysis
- standardization of the GS outputs
 - * predictions, h^2 , h , $r_{M,P}$ $r_{M,G}$
 - * graphical display

Yield



Efficient Computation of Ridge-Regression Best Linear Unbiased Prediction in Genomic Selection in Plant Breeding

H. P. Piepho,* J. O. Ogutu, T. Schulz-Streeck, B. Estaghvirou, A. Gordillo, and F. Technow

ABSTRACT

Computational efficiency of procedures for genomic selection is an important issue when cross-validation is used for model selection and evaluation. Moreover, limited computational resources may be a bottleneck when processing large datasets. This paper reviews several options for computing ridge-regression best linear unbiased prediction (RR-BLUP) in genomic selection and compares their computational efficiencies when using a mixed model package. Attention is also given to the problem of singular genetic variance-covariance. Annotated code

H.P. Piepho, J.O. Ogutu, T. Schulz-Streeck, and B. Estaghvirou, Bioinformatics Unit, Institute of Crop Science, University of Hohenheim, Fruwirthstrasse 23, 70599 Stuttgart, Germany; A. Gordillo, AgReliant Genetics, LLC, 4640 East State Road 32, Lebanon, IN 46052; F. Technow, Institute of Plant Breeding, University of Hohenheim, Fruwirthstrasse 21, 70599 Stuttgart, Germany. Received 11 Nov. 2011
*Corresponding author (piepho@uni-hohenheim.de).

Abbreviations: BLUP, best linear unbiased prediction; GS, genomic selection or genome-wide selection; MET, multienvironment trials; MME, mixed model equations; REML, restricted maximum likelihood; RR-BLUP, ridge-regression best linear unbiased prediction; SNP, single nucleotide polymorphism.

Practical Considerations : GS implementation

GS Scheme Class:

Same breeding cycle

Different breeding cycles

Practical Considerations : GS implementation

GS Scheme Class:

Same breeding cycle

Different breeding cycles

Number of individuals and markers:

Pilot study: GS accuracy = fct(# individuals or markers)

Start with a simple GS scheme

Practical Considerations : GS implementation

GS Scheme Class:

Same breeding cycle

Different breeding cycles

Number of individuals and markers:

Pilot study: GS accuracy = fct(# individuals or markers)

Start with a simple GS scheme

Prediction accuracy assessment:

Validation experiment

$r_{M,G} = r_{M,P} ?$

$$GS \text{ Efficiency} = \frac{i_{GS} r_{M,G} / GS \text{ cycle length}}{i_{PS} r_{P,G} / PS \text{ cycle length}}$$

Practical Considerations : GS implementation

GS Scheme Class:

Same or Different breeding cycles
Match training and prediction data sets

Number of individuals and markers:

Pilot study: GS accuracy = fct(# individuals or markers)
Start with a simple GS scheme

Prediction accuracy assessment:

Validation experiment

$$r_{M,G} = r_{M,P} ?$$

$$GS \text{ Efficiency} = \frac{i_{GS} r_{M,G} / GS \text{ cycle length}}{i_{PS} r_{P,G} / PS \text{ cycle length}}$$

Customize GS scheme:

Low scale implementation
Then scale up

Conclusions: Pros & Cons

Phenotypic selection:

- + assess additive and non-additive model
- does not account for relatedness between plants

Genomic selection:

- + Reduced cycle length → higher genetic gain per time unit
- + Increase selection intensity → higher genetic gain, reduced cost
- + Prediction accuracy comparable to that of phenotypic selection
- + accounts for relatedness between individuals
- additive model assumed (most of the time)

Conclusions: success of GS

GS is very **popular** in plant breeding

- more breeders involved
- part of the routine work
- increasing proportion of the GS breeding schemes

Is GS **successful** ?

- GS derived lines coded and in advanced stages of evaluation

What is a successful variety?

- “ It is planted by the farmer “

Conclusions: success of GS

GS is very **popular** in plant breeding

- more breeders involved
- part of the routine work
- increasing proportion of the GS breeding schemes

Is GS **successful** ?

- GS derived lines coded and in advanced stages of evaluation

What is a successful variety?

- ~~“ It is planted by the farmer “~~
- “ It is planted by the farmer **again** ”

Harry Brokish, director of production, AgReliant

Conclusions: success of GS

GS is very **popular** in plant breeding

- more breeders involved
- part of the routine work
- increasing proportion of the GS breeding schemes

Is GS **successful** ?

- GS derived lines coded and in advanced stages of evaluation

What is a successful variety?

- ~~“ It is planted by the farmer “~~
- “ It is planted by the farmer **again** ”

Harry Brokish, director of production, AgReliant

→ It is anticipated that the proportion of lines deriving from GS will increase

Perspectives

GS can be integrated in hybrid maize breeding schemes

It can be designed novel types of GS analysis schemes:

- combine testers**
- use at different selection stages**
- combine years**
- etc...**

Perspectives

GS can be integrated in hybrid maize breeding schemes

It can be designed novel types of GS analysis schemes:

- combine testers**
- use at different selection stages**
- combine years**
- etc...**

→ GS will **never replace the breeder**

→ GS **depends on the breeder**

Thank You !