



Optimization of breeding schemes with GS in R

Friedrich Longin

State Plant Breeding Institute, University of
Hohenheim, Stuttgart, Germany;
<https://lsa-weizen.uni-hohenheim.de>



- **3 faculties**
 - 1 – life sciences (biology, physics, biotechnology, food production and ingredients); ~ 14 institutes
 - 2 – agriculture (breeding, quantitative genetics, molecular genetics, agronomy, plant health, agroeconomics, animal sciences,...); ~ 15 institutes
 - 3 – Economy (management, economics, law, social sciences...); ~ 8 institutes
- ~ **4000 students** per year, all B.Sc. and M.Sc.
Master of crop science (english)
- **4 state institutes** incorporated into the University
- **Experimental station** for agriculture

Excellence Unit in Plant Breeding



Institute of Plant Breeding, Seed Science and Population Genetics

Subject Areas:

Applied Genetics and Plant Breeding

Prof. Dr. Albrecht E. Melchinger

Quantitative Genetics and Genomics

Prof. Dr. Scholten

Crop Biodiversity and Breeding Informatics

Prof Dr. Karl Schmid

Seed Science and Technology

Prof. Dr. Michael Kruse

State Plant Breeding Institute

Head: Prof. Dr. Tobias Würschum

Research groups:

Biotechnology and Mapping Strategies

Dr. Wilmar Leiser

Rye and Biotic Stress Resistance

Prof. Dr. Thomas Miedaner

Triticale and Breeding Methodology

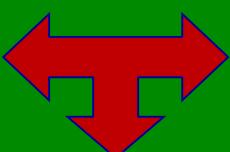
Dr. Hans P. Maurer

Wheat Breeding and Selection Theory

PD Dr. Friedrich Longin

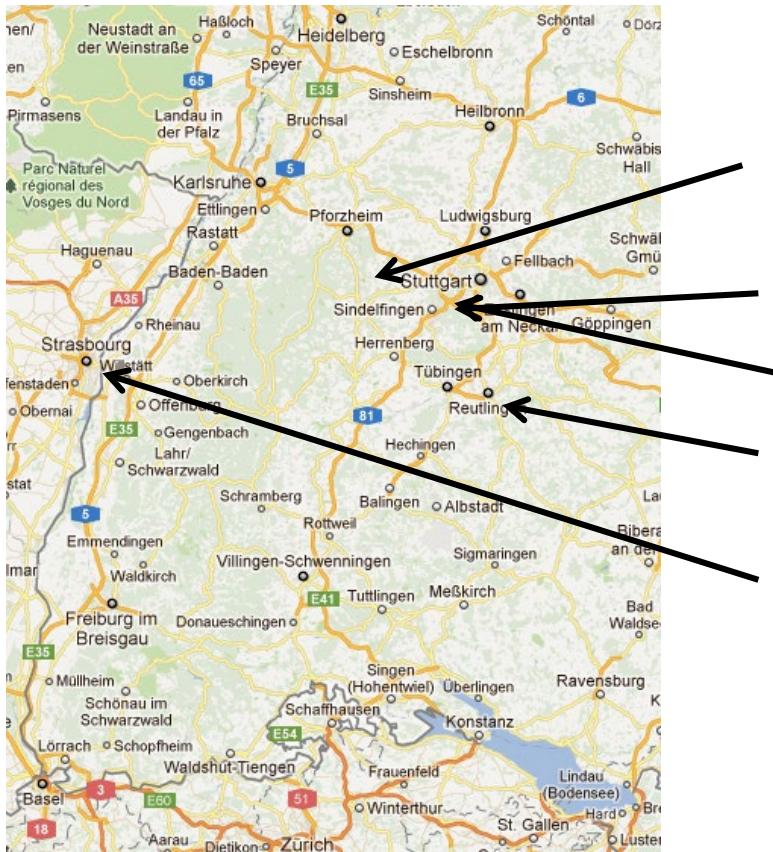
Sunflower and Soybean Breeding

Dr. Volker Hahn



Experimental Station

Experimental stations



Ihinger Hof
Heidfeldhof/Hohenheim
Kleinhohenheim/organic
Lindenhöfe
Eckartsweier

Five **experimental stations** ensure **high-quality phenotyping**
which is the basis for breeding and science

State Plant Breeding Institute (LSA)



- Six **scientists** with **broad** range of **expertise**,
- PhD and PostDocs (from agronomy, bioinformatics, mathematics)
- Experienced **technical staff** (>20) with phenotyping facilities, green houses, climate chambers, biotechlab,...
- ~ 20 peer-reviewed **publications** per year in leading international journals
- ~ 1.800.000€ extramural **grants** per year
- **Cooperations** with leading international **agricultural research centers** (e.g. CAAS, CIMMYT, INRA, ACPFG, NIAB)

Breeding activities at the LSA

Einkorn



Emmer



Spelt



Durum



Wheat



Rye



Triticale



Soybean



Sunflower



LSA has competitive breeding programs

Wheat group - aims

- Elite **breeding in rare wheat species**
- **Breeding programs in T. aestivum**
 - Pre-Breeding for grain yield: aim to deliver material to community for elite breeding starts
 - Breeding hybrid male parents: aim to deliver elite males to community
- **Research accompanying** these efforts
 - Genetic architecture of regarded traits
 - Optimized breeding schemes



Wheat group

- 1 scientist – Dr. F. Longin
- **6 technicians**
- 1 PostDocs, 1 PhD, students
- ~ 8 ha of nurseries
- > 25 different field locations
- Special nurseries for stress - **FHB**, virus, frost
- **Quality lab** for b- value, sds, falling, GI, protein content, vitreousness and dark points
- Machinery for threshing and dehulling



<https://lsa-weizen.uni-hohenheim.de/>

You are a breeder

- Congratulations → **fantastic job!**
- A breeder is the **head of product development**
 - You must be **innovative**
 - You must be able to **rapidly take decisions**
 - You must **define the strategies**
 - You must have **success**
 - You are **responsible for whole product chain**



You have to decide = you're responsible

- Marketing tells you your specific framework

? Strategies ?



You have to decide = you're responsible

- Marketing tells you your specific framework

? Strategies ?

Which breeding scheme ?

Which quality lab analyses ?

Molecular markers ?

Trial management ?

Priorisation of traits ?

Orga of phenotyping ?

Disease management ?

...

You have to decide = you're responsible

- Marketing tells you your specific framework

? Strategies ?

Which breeding scheme ?

Which quality lab analyses ?

Molecular markers ?

Trial management ?

Priorisation of traits ?

Orga of phenotyping ?

Disease management ?

...

Other lectures

Let's start – what we will do



- Introduction: Breeding categories → focus line breeding
- Theorectical background of selection gain package: formula of selection gain
- Variables influencing the selection gain
 - Important existing results
 - Realization in R package
- GS breeding schemes
 - Important results
 - Realization in R package
- Run your own first simulations

Breeding categories

- Line breeding

Wheat



- Hybrid breeding



Maize

- Clone breeding

Potato



- Population breeding



Rye

Line breeding based on per se performance

Line breeding

New breeding
lines from DH,
ssd,...



Field trials – per se
performance



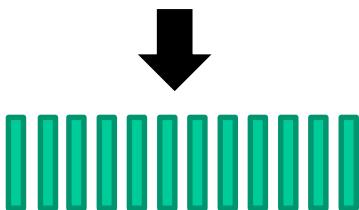
New line varieties

Hybrid breeding based on GCA



Line breeding

New breeding lines from DH, ssd, ...



Field trials – per se performance



New line varieties

Hybrid breeding

New breeding lines in heterotic group 1

GCA-Tester
GCA- Tester



Field trials –
combining ability



New lines **for hybrid production** in heterotic group 1

New lines **for hybrid production** in heterotic group 2

3 Phases in a breeding scheme

Example: phenotypic selection in hybrid breeding

Year 1 P1 x P2, ...

Year 2 DH-Production

Year 3 N_1 DH lines -
multiplication

Year 4

Hybrid
seed prod.
 $N_2 * T_2$

Year 5

Field test
 $N_2 * T_2 * L_2$

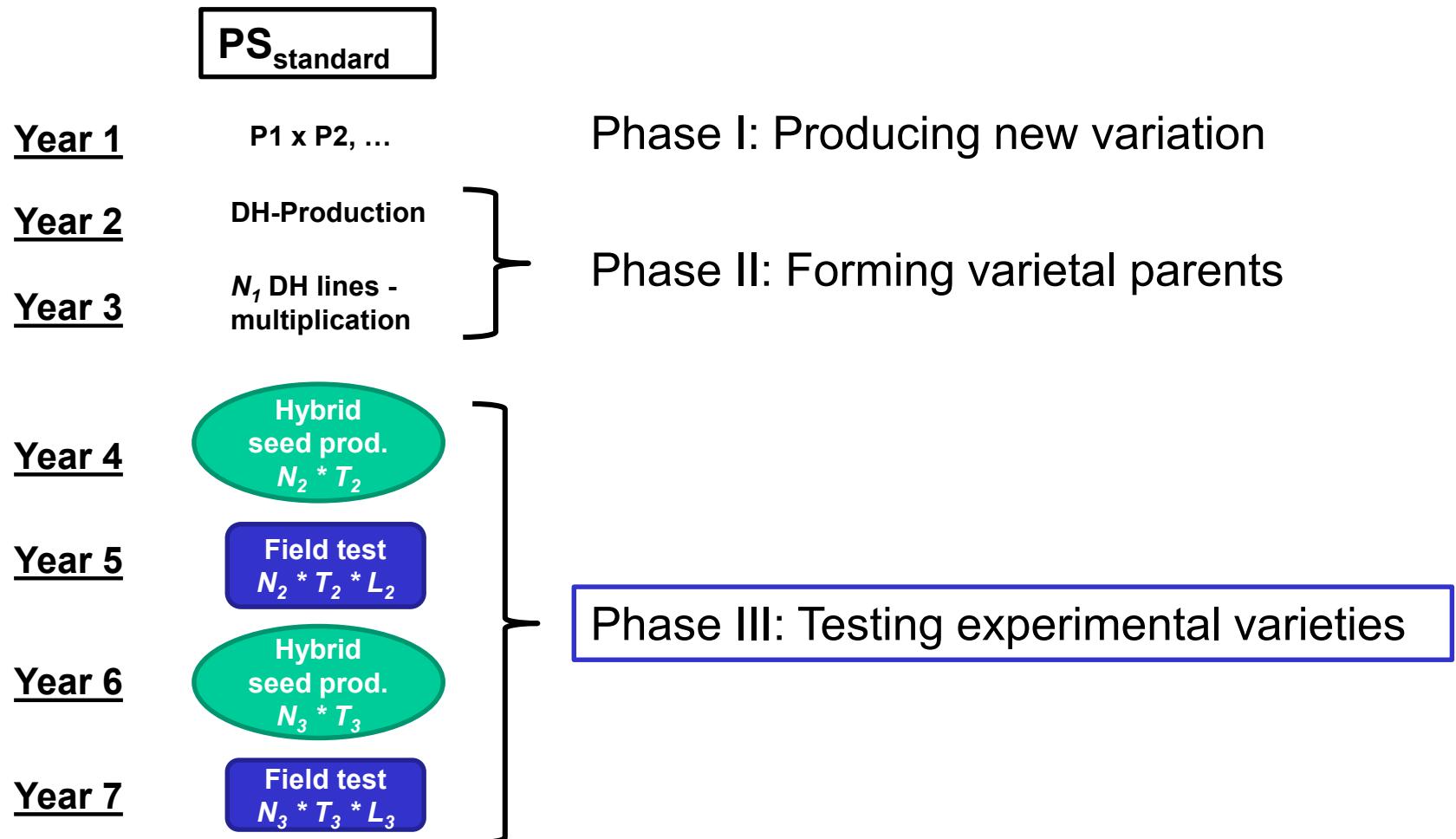
Year 6

Hybrid
seed prod.
 $N_3 * T_3$

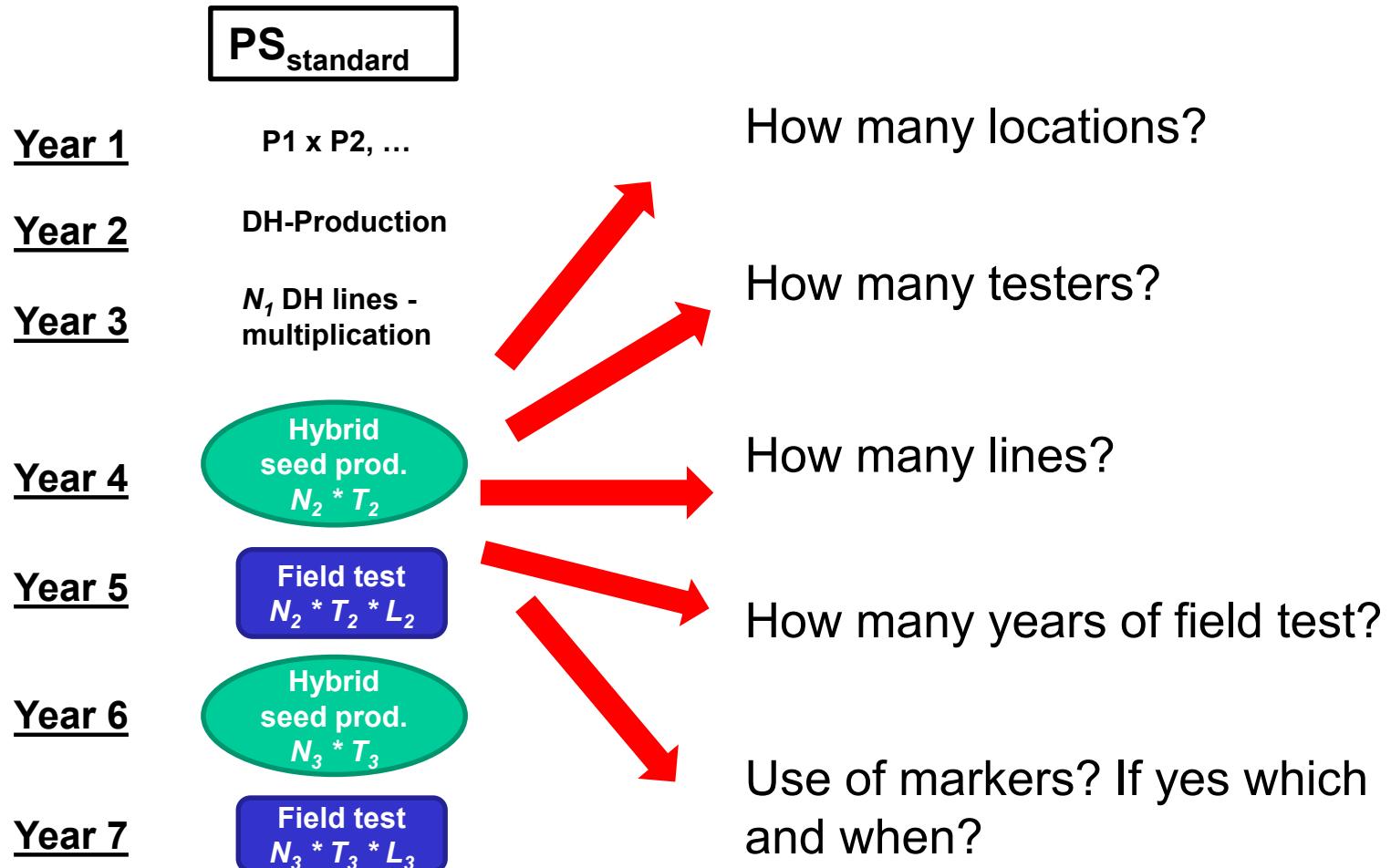
Year 7

Field test
 $N_3 * T_3 * L_3$

3 Phases in a breeding scheme



Many questions...



Option 1: Field trials

Year 1 P1 x P2, ...

Year 2 DH-Production

Year 3 N_1 DH lines -
multiplication

Year 4 Hybrid
seed prod.
 $N_2 * T_2$

Year 5 Field test
 $N_2 * T_2 * L_2$

Year 6 Hybrid
seed prod.
 $N_3 * T_3$

Year 7 Field test
 $N_3 * T_3 * L_3$

e.g. 1: maximum N
→ minimum L

Option 1: Field trials

Year 1

P1 x P2, ...

P1 x P2, ...

Year 2

DH-Production

DH-Production

Year 3

N_1 DH lines -
multiplication

N_1 DH lines -
multiplication

Year 4

Hybrid
seed prod.
 $N_2 * T_2$

Hybrid
seed prod.
 $N_2 * T_2$

Year 5

Field test
 $N_2 * T_2 * L_2$

Field test
 $N_2 * T_2 * L_2$

Year 6

Hybrid
seed prod.
 $N_3 * T_3$

Hybrid
seed prod.
 $N_3 * T_3$

Year 7

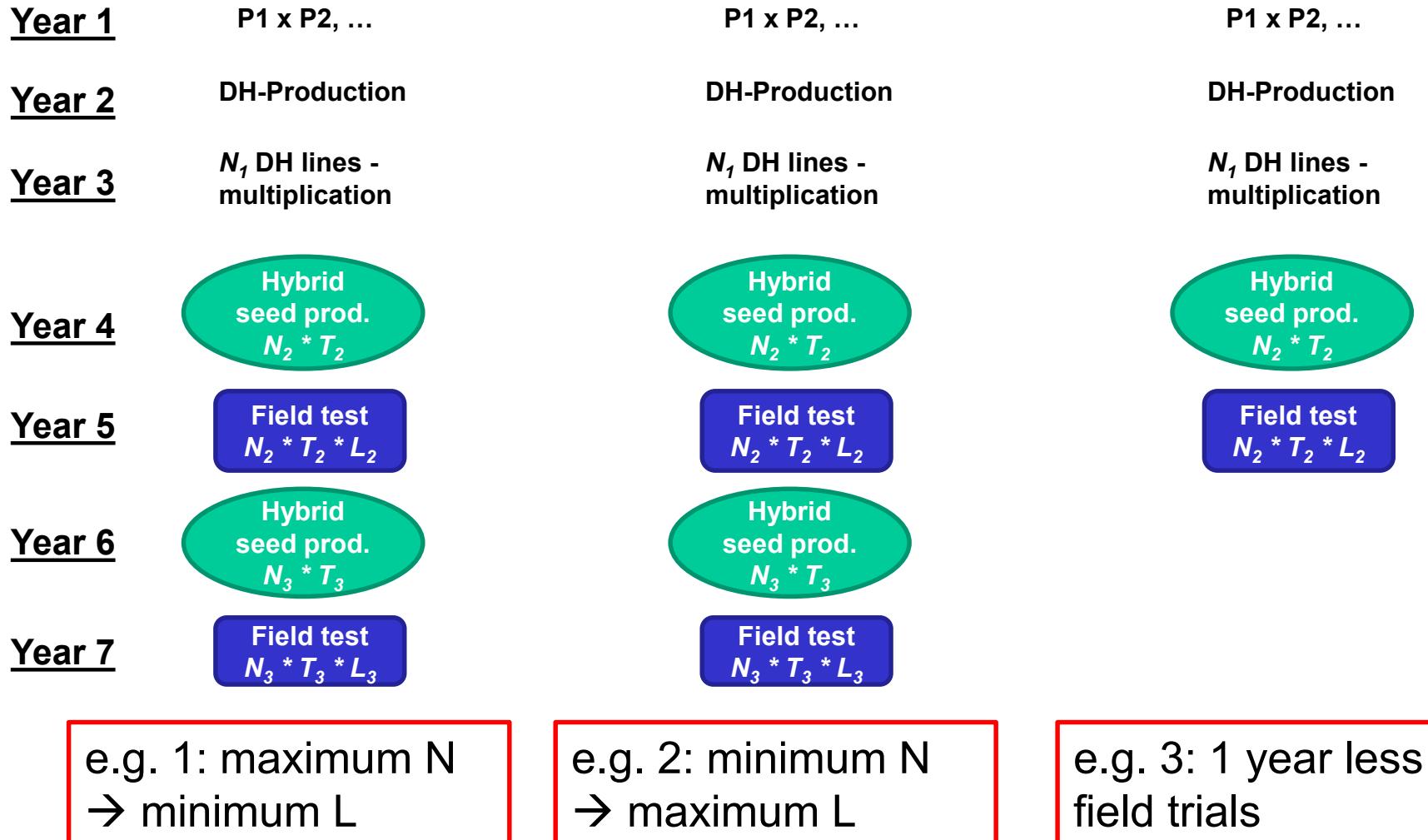
Field test
 $N_3 * T_3 * L_3$

Field test
 $N_3 * T_3 * L_3$

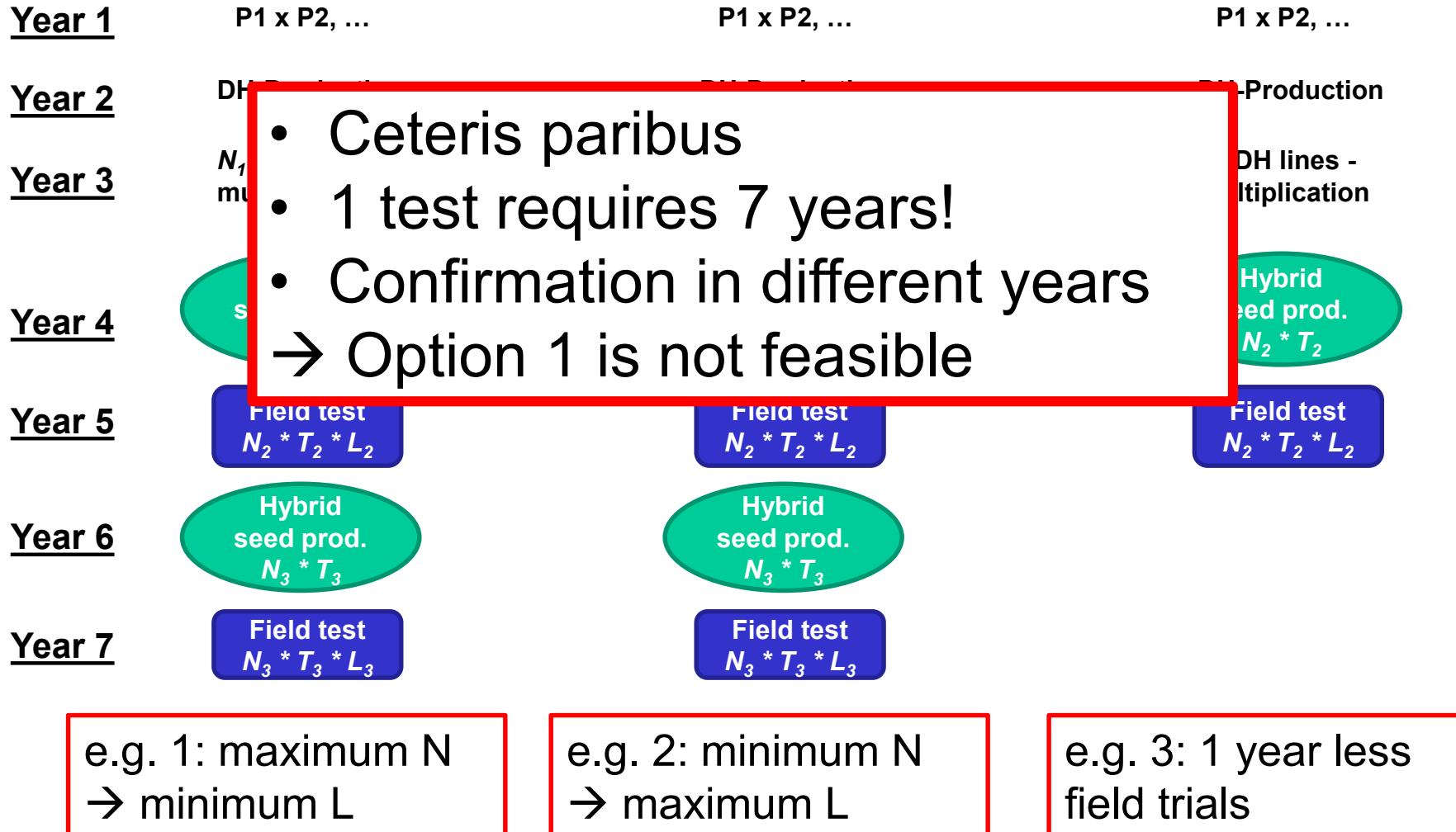
e.g. 1: maximum N
→ minimum L

e.g. 2: minimum N
→ maximum L

Option 1: Field trials



Option 1: Field trials



Option 2: Simulation of breeding methods



Year 1 P1 x P2, ...

Year 2 DH-Production

Year 3 N_1 DH lines -
multiplication

Year 4 Hybrid
seed prod.
 $N_2 * T_2$

Year 5 Field test
 $N_2 * T_2 * L_2$

Year 6 Hybrid
seed prod.
 $N_3 * T_3$

Year 7 Field test
 $N_3 * T_3 * L_3$

e.g. 1: maximum N
→ minimum L

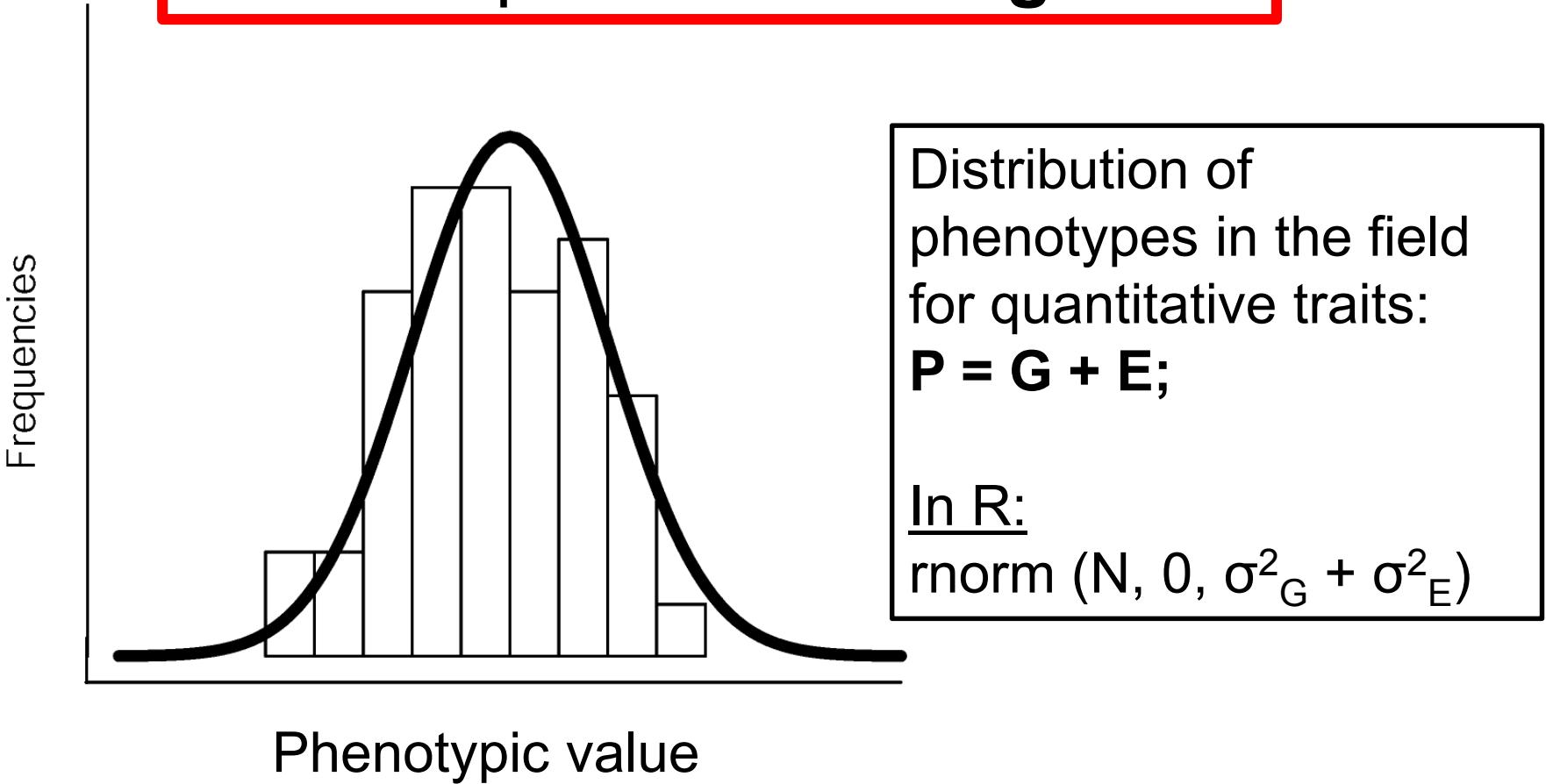
- Prediction the gain from selection in breeding schemes → Breeder's equation

- Simulation of different breeding schemes

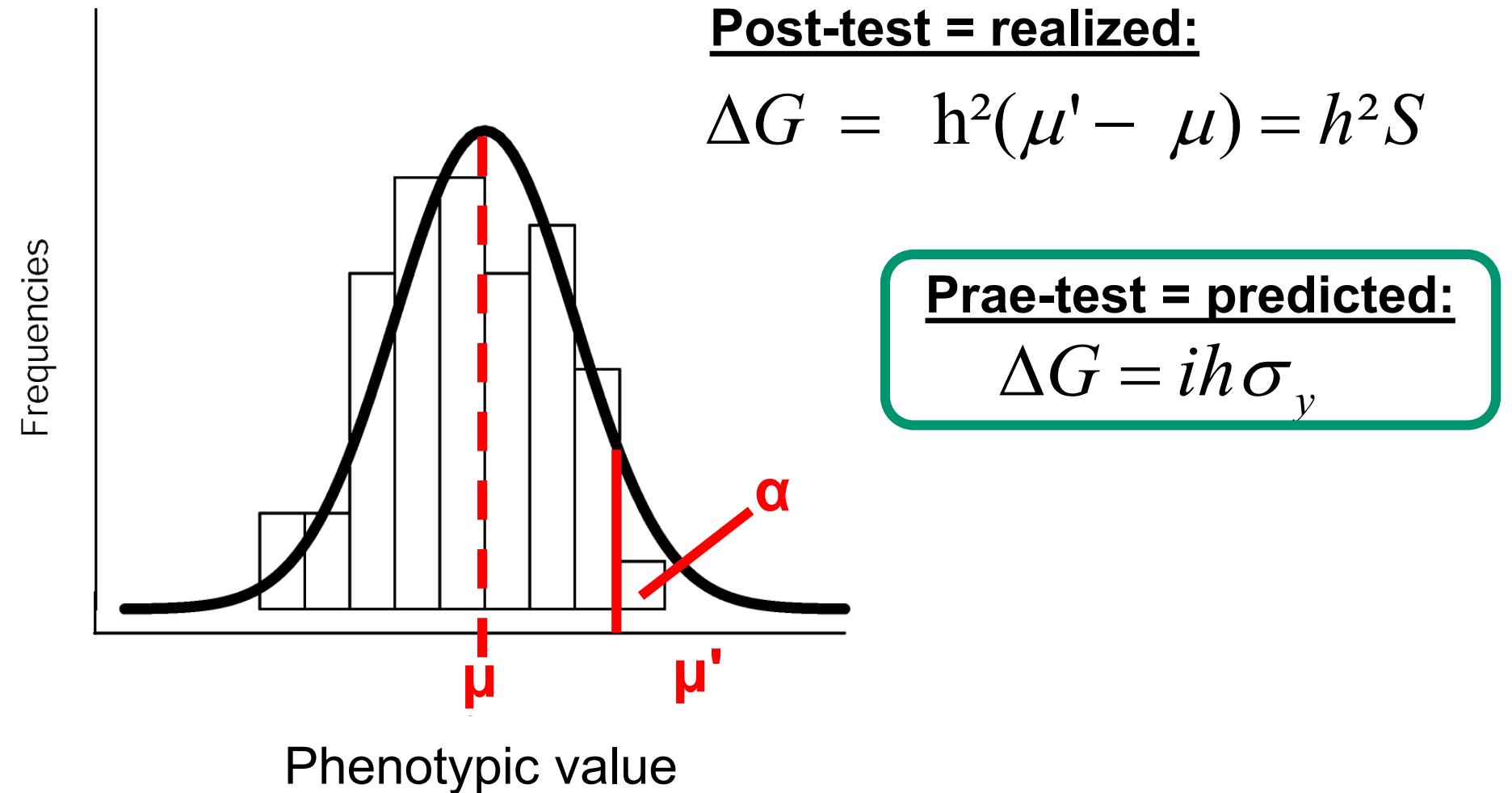
- R Package „selection gain“

Choice of breeding method

Concept of selection gain



Selection gain



Prediction of selection gain

Selection gain

$$\Delta G = ih\sigma_y$$

- i = selection intensity,
- h = square root of the heritability,
- σ_y = square root of the genetic variance of the target variable

Annual selection gain

$$\Delta G_a = ih\sigma_y / Y$$

- Y = no. of years required to finish one breeding cycle

Variables influencing selection gain

Annual selection gain

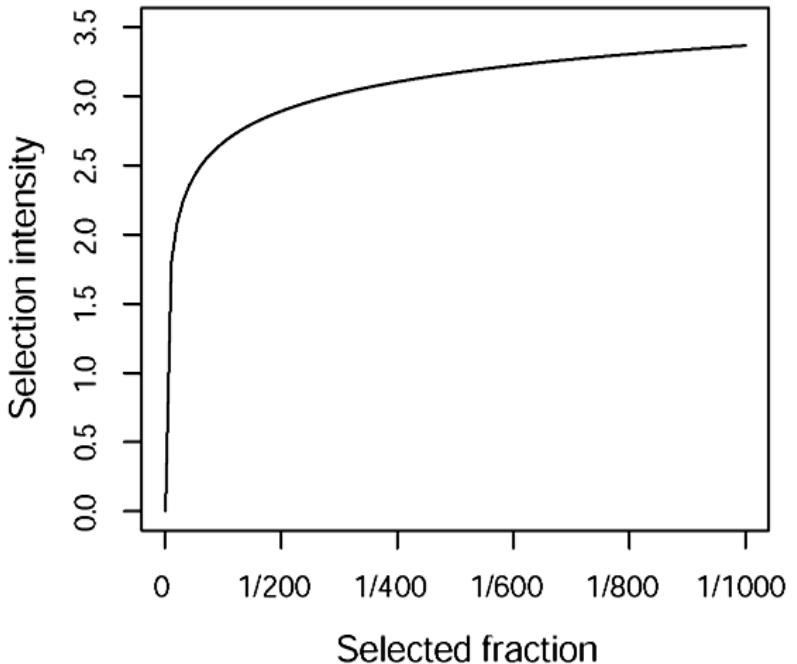
$$\Delta G_a = ih\sigma_y / Y$$

- i = selection intensity,
- h = square root of the heritability,
- σ_y = square root of the genetic variance of the target variable
- Y = no. of years required to finish one breeding cycle

Selection gain is maximized by an

- Increase of i → reduction of Y
- Increase of h
- Increase of σ_y

Increasing selection intensity



Selected fraction

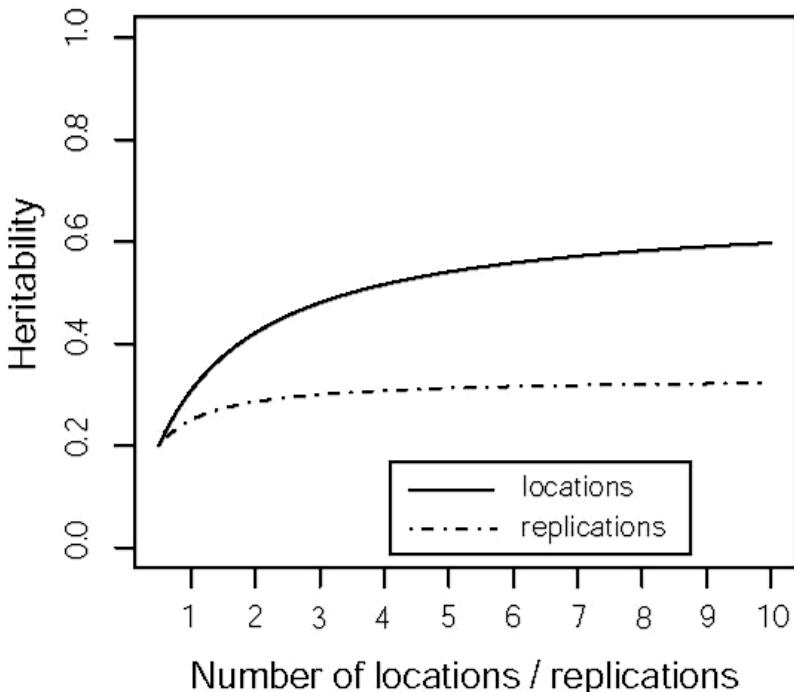
$$\alpha = \frac{\text{no. selected lines}}{\text{no. tested lines}}$$

Selection intensity is increased by

- Increasing the number of tested lines
- Decreasing the number of selected lines

BUT: Increase is not linear

Increasing heritability



$$h^2 = \sigma_G^2 / \sigma_P^2$$

$$\sigma_P^2 = \sigma_G^2 + \sigma_{GxE}^2 / L + \sigma_e^2 / (L * R)$$

L = Locs, R = Reps, σ_G^2 = genet. variance;
 σ_P^2 = phenotyp. variance, σ_{GxE}^2 = variance due to Genotyp-environment - interaction;
 σ_e^2 = error variance
 → estimated via ANOVA

Heritability is increased by

- Increasing the number of locations
- Increasing the number of replications, but less than for locations!

BUT: Increase is not linear

Allocation of resources

$$\Delta G = ih\sigma_y / Y$$

i: function of $\alpha = N_{sel} / N_1$

$$h^2 = \sigma_G^2 / \sigma_P^2 \rightarrow \sigma_P^2 = \sigma_G^2 + \sigma_{GxE}^2 + L + \sigma_e^2 / (L * R)$$

Allocation of resources

L = Locs, R = Reps, σ^2_G = genet. variance; σ^2_P = phenotyp. variance, σ^2_{GxE} = variance due to Genotyp-environment - interaction; σ^2_e = error variance

→ estimated via ANOVA

Optimize allocation of resources

- **Framework**
 - Target criteria: **maximize selection gain**
 - **Variables to optimize:**
 - No. of locations
 - No. of replications
 - No. of lines
 - No. of testers, type of tester
 - Splitting of lines on crosses and lines within crosses

→ **Fixed annual budget**

Budget of a breeding program = annual budget



Seed production		Seed stock		Breeding scheme					Special nurseries			
Resp.	Min trials	Ex of	Season	Prod.	Tests	Traits	Resp.	Allocation	Frost	WiSo	FHB	Virus
			Winter						HB	BY	BY	VL
BY			Summer	P1 x P2			BY	1 Loc				
BY			Winter									
BY			Summer	F1 self			BY	1 Loc				
BY			Winter									
BY			Summer	F2 self	Single plant	ear quality, kernel quality	BY	1 Loc				
BY		rest F1	Winter									
BY			Summer	F3 self	Ear to row	row obs, kernelquality, colour, sds	BY	1 Loc				
BY		rest F2	Winter									
BY			Summer	F4 self	Ear to row	row obs, kernelquality, colour, sds	BY	1 Loc	25 K			
BY		rest F3	Winter			heading, lodging, height, diseases,						
BY	1200g		Summer	F5 self	LP1	yield, colour, sds, protein, vitreousity	BH	4 Loc, 2 reps	25 K	1 row	3 rows	2 Loc, 2 rows
BY		rest F4	Winter			heading, lodging, height, diseases,						
BY	2500g		Summer	F6 self	LP2	yield, colour, sds, Protein, vitreousity	BH	10 Loc, 2 reps		1 row	3 rows	2 Loc, 2 rows

Official offer

	Breeding scheme											
Season	Prod.	Tests										
2013/14	P1 x P2		Breeding scheme									
2014/15			Prod.	Tests								
2015/16	F1 self		Breeding scheme									
2016/17			P1 x P2		Breeding scheme							
2017/18	F2 self	Single plant	Prod.		Breeding scheme							
2018/19			P1 x P2		Prod.		Tests		Breeding scheme			
2019/20	F3 self	Ear to row	Prod.		Tests							
			P1 x P2		Breeding scheme							
	F4 self	Ear to row	Prod.		Tests							
			P1 x P2		Breeding scheme							
	F5 self	LP1	Prod.		Tests							
			P1 x P2		Breeding scheme							
	F6 self	LP2	Prod.		Tests							
			P1 x P2		Breeding scheme							

Budget: horizontal and vertical!

Examples for budgets per program

- **Maize**
 - Large: >1.000.000€
 - Small: ~ 500.000€
- **Wheat**
 - Large: 600.000€
 - Small: 200.000€
- **Barley**
 - Large: 400.000€
 - Small: 100.000€

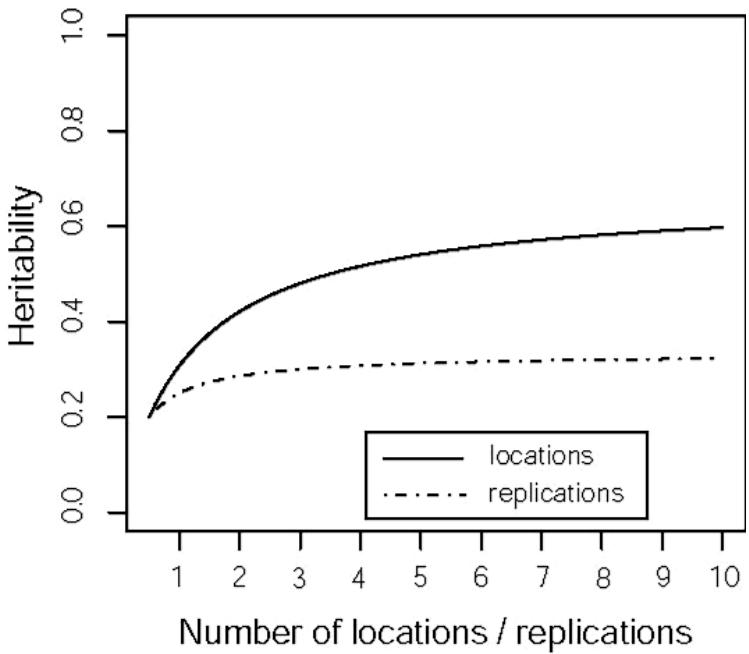
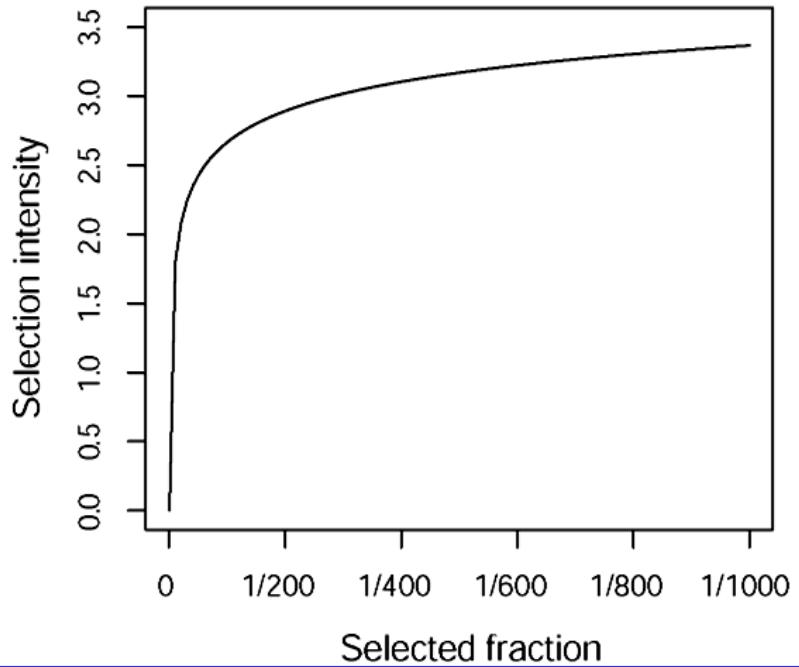


Simple budget formula

Budget = N*L*R (*T) + Production of N

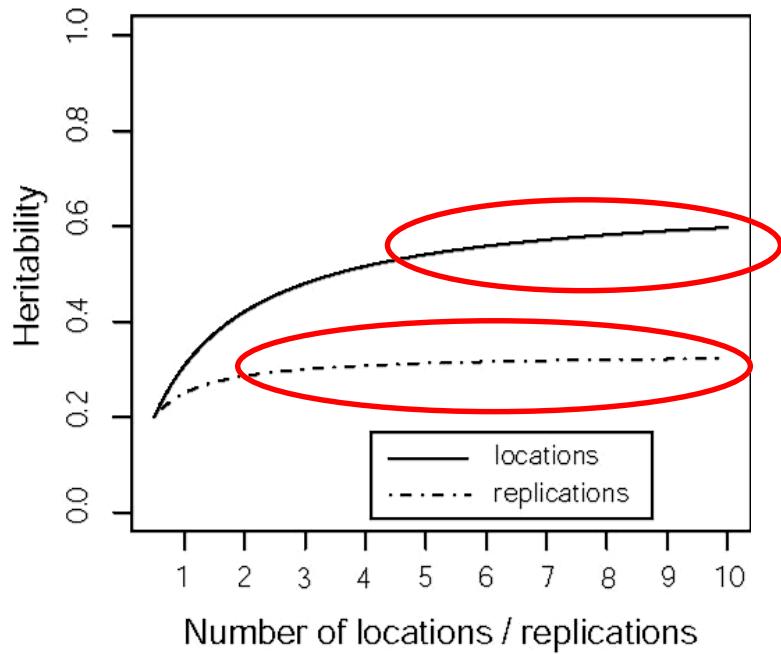
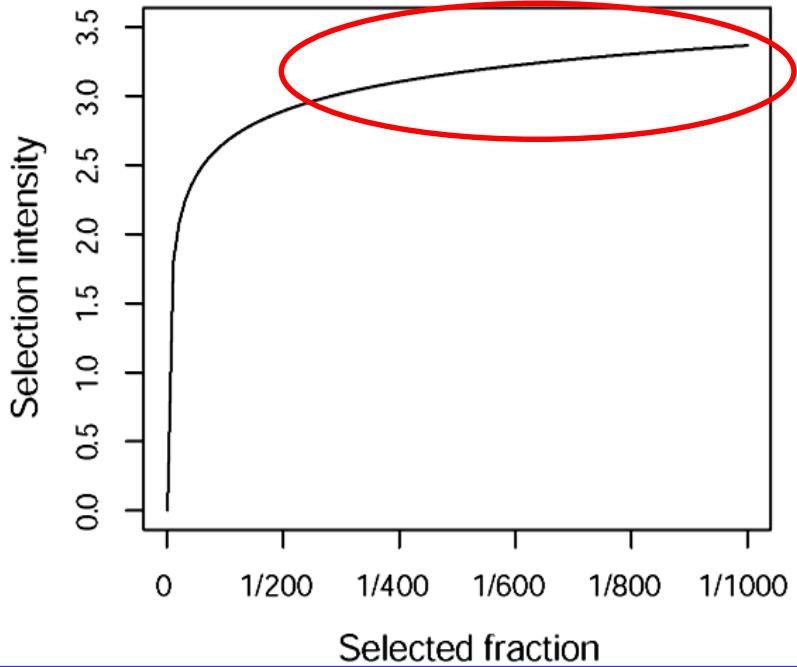
- N = no. of test candidates
- L = no. of test locations
- R = no. of replications
- T = no. of testers

Selection intensity vs. heritability



For a fixed budget, maximization of ΔG represents a compromise between a high number of test candidates and a high intensity of testing.

Golden rule



For a fixed budget, maximization of ΔG represents a compromise between a high number of test candidates and a high intensity of testing.

Golden rule: Curves of i and h^2 level off and increase by $L >$ than in R

Variables influencing selection gain

Annual selection gain

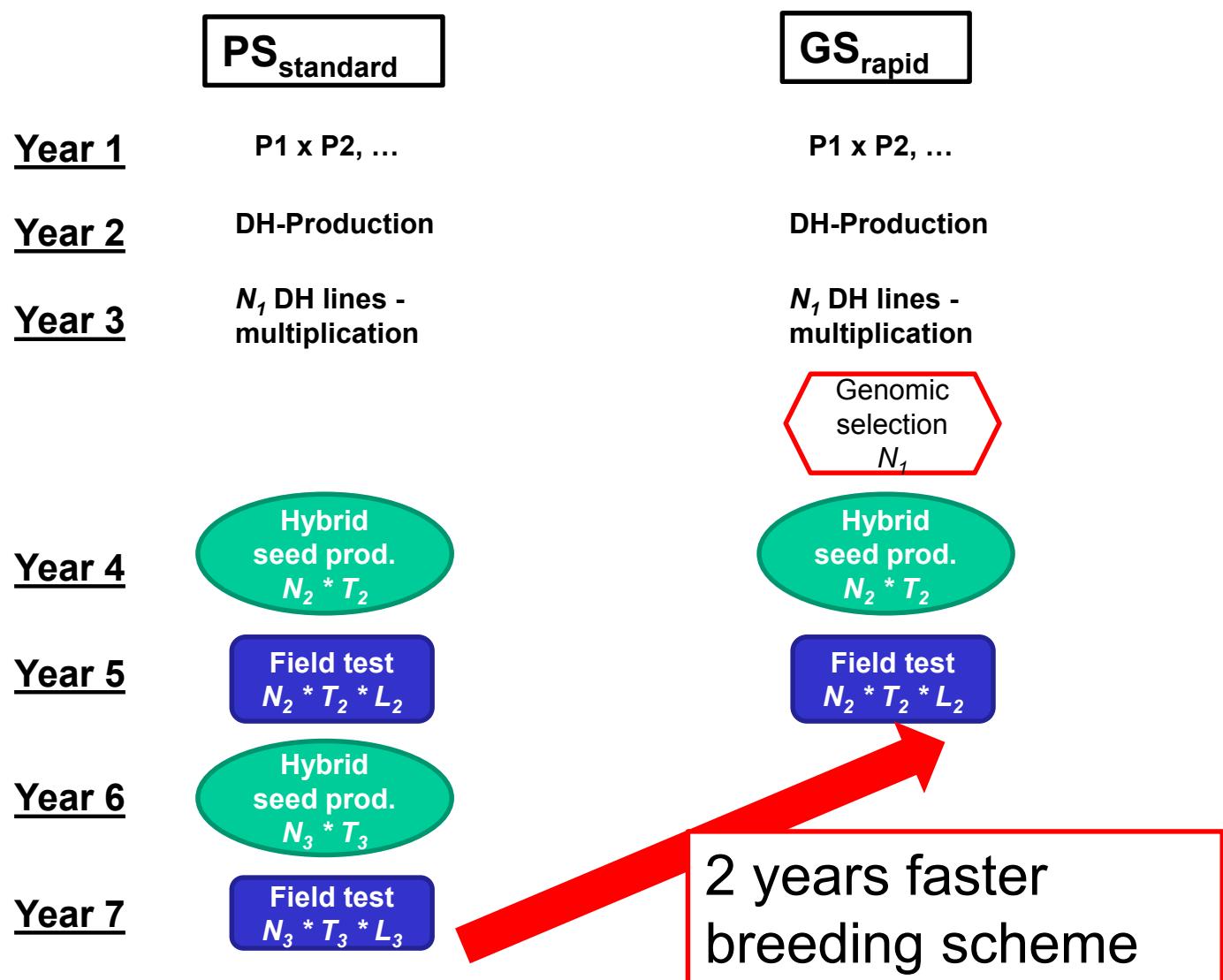
$$\Delta G_a = ih\sigma_y / Y$$

- i = selection intensity,
- h = square root of the heritability,
- σ_y = square root of the genetic variance of the target variable
- Y = no. of years required to finish one breeding cycle

Selection gain is maximized by an

- Increase of i → reduction of Y
- Increase of h
- Increase of σ_y

Reduce cycle length



Annual selection gain

$$\Delta G_a = ih\sigma_y / Y$$

Assumptions:

- $i = 2$
- $h = 0.7$
- $\sigma_y = 5 \text{ dt/ha}$

Example:

$$\Delta G = 2 * 0.7 * 5 / 7 = 1$$

$$\Delta G = 2 * 0.7 * 5 / 5 = 1.4$$

> 40 % higher annual selection gain

Cycle length has very strong effect on annual selection gain

→ You must be faster than your competitor

Variance components

Crop	Variance components due to				
	G	GxL	GxY	GxLxY	E
Winter wheat	14.2	2.4	2.4	9.2	14.4
Winter barley	6.1	1.6	1.6	6.7	13
Grain maize early	19.6	6.1	5.1	11.7	27.1
Forage maize early	38.7	15.6	8.6	17.8	80.8
Winter oil seed rape	3.3	1.7	2	3.8	9.7
Sugarbeet	33.6	9.1	1.7	4.8	35.2

Source: Laidig et al. 2008

Variance components

Crop	Variance components due to				
	G	GxL	GxY	GxLxY	E
Winter wheat	14.2	2.4	2.4	9.2	14.4
Winter barley	6.1	1.6	1.6	6.7	13
Grain maize early	19.6	6.1	5.1	11.7	27.1
Forage maize early	38.7	15.6	8.6	17.8	80.8
Winter oil seed rape	3.3	1.7	2	3.8	9.7
Sugarbeet	33.6	9.1	1.7	4.8	35.2

High variance due to

- genotype x year and
- genotype x year x location interaction

Year → large effect on genotype ranking

Genotyp	Rank 2015	Rank 2016	Yield 2016	Yield 2015
	21	1	58,71	77,10
Miradoux		2	58,66	
W-10066-217-316/14/3-512-2/1	8	3	58,57	82,33
W-10037-210-309/17/1-487-1/3	2	4	58,03	86,39
W-10021-204-307/4/3-468-2/1	1	5	57,89	90,33
W-10029-207-305/11/1-439-4/1	17	6	57,22	79,24
W-10066-217-316/9/2-511-3/1	6	7	57,02	83,49
W-10013-202-302/9/1-408-2/1	16	8	56,44	79,45
Lupidur	10	9	55,87	81,85
W-10058-214-313/21/2-501-1/3	9	10	55,68	82,07
W-10033-209-308/3/1-474-1/3	19	11	55,00	78,42
W-10064-216-315/10/3-506-2/1	13	12	54,97	80,86
W-10021-204-307/2/2-466-6/3	18	13	54,97	79,24
W-10058-214-313/11/1-499-1/3	15	14	54,89	80,46
W-10031-208-306/22/1-460-3/1	5	15	54,87	84,43
W-10066-217-316/23/3-514-6/3	4	16	54,65	84,81
W-10064-216-315/17/3-507-1/1	3	17	54,54	86,30
W-10043-211-310/19/2-494-3/3	12	18	54,34	81,28
W-10033-209-308/10/3-476-6/3	23	19	54,32	76,02

Further advantages of multiple year testing

- Disease resistance (natural occurring)
 - Frost
 - Drought, heat
- Speed of the program is also a compromise between a maximum annual selection gain and a security of the results

Compromise necessary

Annual selection gain

$$\Delta G_a = ih\sigma_y / Y$$

Annual selection gain is maximized by an

- increase of i → reduction of Y
- increase of h
- increase of σ_y

Compromise necessary between theory and practice!

R package „selection gain“

- Open source software package R (www.r-project.org)
- Package selectiongain
- <https://cran.r-project.org/web/packages/selectiongain/index.html>



Multi-stage selection is practiced in numerous fields of life and social sciences and particularly in breeding. A special characteristic of multi-stage selection is fraction of the superior candidates is selected and promoted to the next stage. For the optimum design of such selection programs, the selection gain plays a crucial role. While mathematical formulas for calculating the selection gain and the variance among selected candidates were developed long time ago, solutions for numerical optimization of selection programs for a given total budget and different costs of evaluating the candidates in each stage.

Version: 2.0.50.1
Depends: R ($\geq 3.0.0$), mvtnorm, parallel
Published: 2016-03-14
Author: Xuefei Mi, Jose Marulanda, H. Friedrich Utz, Albrecht E. Melchinger (Project contact person: Melchinger@uni-hohenheim.de)
Maintainer: Xuefei Mi <mi_xue_fei at hotmail.com>
License: GPL-2
NeedsCompilation: no
CRAN checks: [selectiongain results](#)

Downloads:

Reference manual: [selectiongain.pdf](#)
Package source: [selectiongain_2.0.50.1.tar.gz](#)
Windows binaries: r-devel: [selectiongain_2.0.50.1.zip](#), r-release: [selectiongain_2.0.50.1.zip](#), r-oldrel: [selectiongain_2.0.50.1.zip](#)
OS X Mavericks binaries: r-release: [selectiongain_2.0.50.1.tgz](#), r-oldrel: [selectiongain_2.0.50.1.tgz](#)
Old sources: [selectiongain archive](#)

Linking:

Please use the canonical form <https://CRAN.R-project.org/package=selectiongain> to link to this page.



Selectiongain - Manual

ctiongain.pdf

wetter

- + 90%

Package ‘selectiongain’

March 14, 2016

Type Package

Title A Tool for Calculation and Optimization of the Expected Gain from Multi-Stage Selection

Version 2.0.50.1

Date 2016-02-28

Author Xuefei Mi, Jose Marulanda, H. Friedrich Utz, Albrecht E. Melchinger (Project contact person: Melchinger@uni-hohenheim.de)

Maintainer Xuefei Mi <mi_xue_fei@hotmail.com>

Depends R (>= 3.0.0), mvtnorm, parallel

Description Multi-stage selection is practiced in numerous fields of life and social sciences and particularly in breeding. A special characteristic of multi-stage selection is that candidates are evaluated in successive stages with increasing intensity and effort, and only a fraction of the superior candidates is selected and promoted to the next stage. For the optimum design of such selection programs, the selection gain plays a crucial role. It can be calculated by integration of a truncated multivariate normal (MVN) distribution. While mathematical formulas for calculating the selection gain and the variance among selected candidates were developed long time ago, solutions for numerical calculation were not available. This package can also be used for optimizing multi-stage selection programs for a given total budget and different costs of evaluating the candidates in each stage.

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2016-03-14 12:58:02

R topics documented:

multistagecorr	2
multistagegain	4
multistagegain.each	6
multistageoptimum.grid	8
multistageoptimum.nlm	10
multistageoptimum.search	12

Selectiongain - download

Code View Plots Session Build Debug Tools Help

schemen_1versuch.R R_masterclass.R

Source on Save |   

```
library(selectiongain)
sessionInfo()
Budget=1000
VCGCAandError=c(5.7,
VCSCA=c(1.88,2.94,0,
# one-stage PS
res <- multistageopt
res
# two-stage PS
res <- multistageopt
```

Install Packages

Install from: [? Configuring Repositories](#)

Repository (CRAN, CRANextra)

Packages (separate multiple with space or comma):

selectiongain

Install to Library:

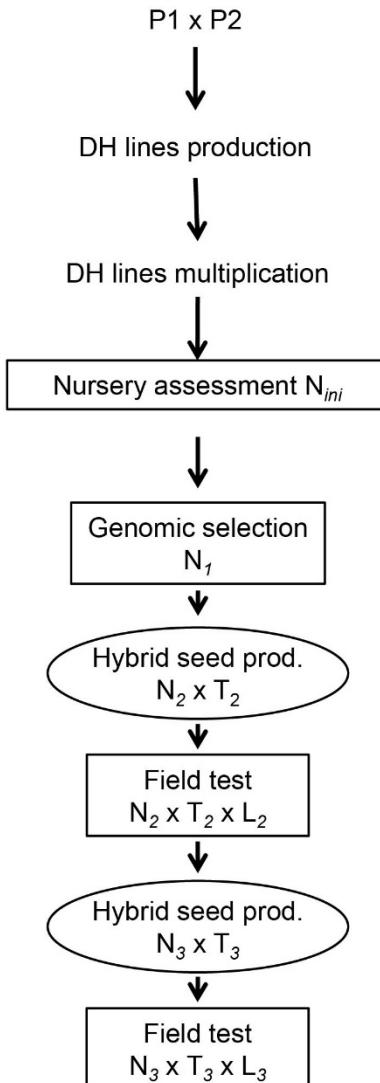
P:/WINNETZ/R/win-library/3.2 [Default]

Install dependencies

Install Cancel

```
VSCA=VCSCA, CostProd = c(0,4,4), CostTest = c
Nf = 5, Budget = Budget, N2grid = c(5, 511, 1
N3grid = c(5, 20, 1), L2grid=c(2,5,1), L3grid
Tgrid = c(1, 1, 1), Tgrid = c(1, 2, 1) R2=1, R3=1,
```

What is possible? – Breeding scheme



Breeding scheme:

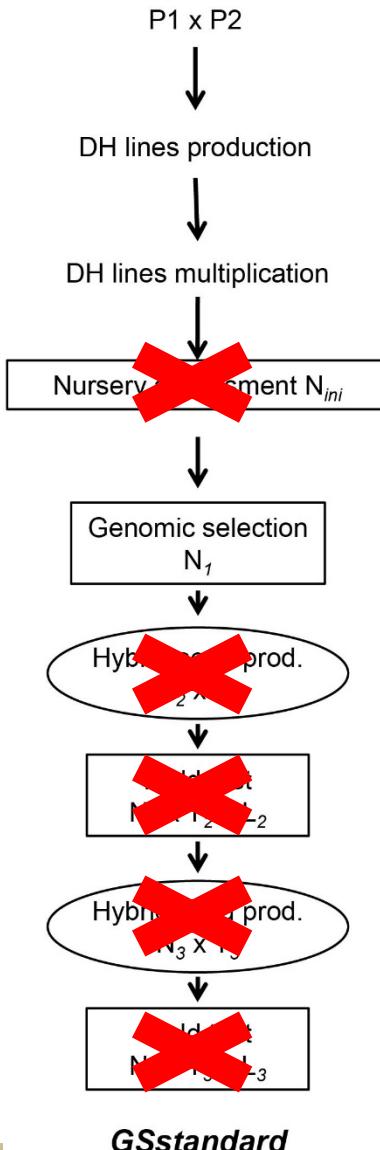
- DH production
- Nursery selection on traits not correlated to yield
- GS on yield
- 2 stage phenotypic selection on yield

N_i , L_i , R_i , T_i = number of lines, locations, replications and testers used in stage $i \rightarrow$ optimized

GSstandard



What is possible? – Breeding scheme



Breeding scheme:

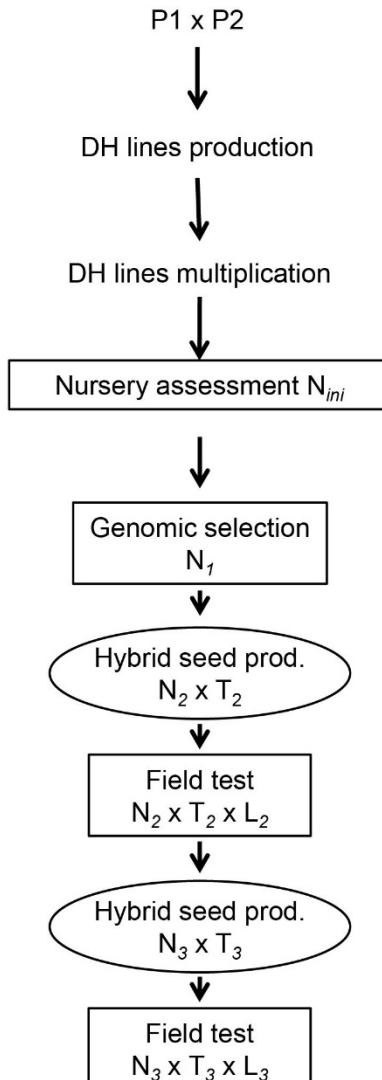
- DH production
- Nursery selection on traits not correlated to yield
- GS on yield
- 2 stage phenotypic selection on yield

Breeding scheme - modifications

- Each test stage can be switched off
- You can enter minimum/maximum numbers in each test stages of N, L, R, T
- GS → yes/no



Budget & Costs



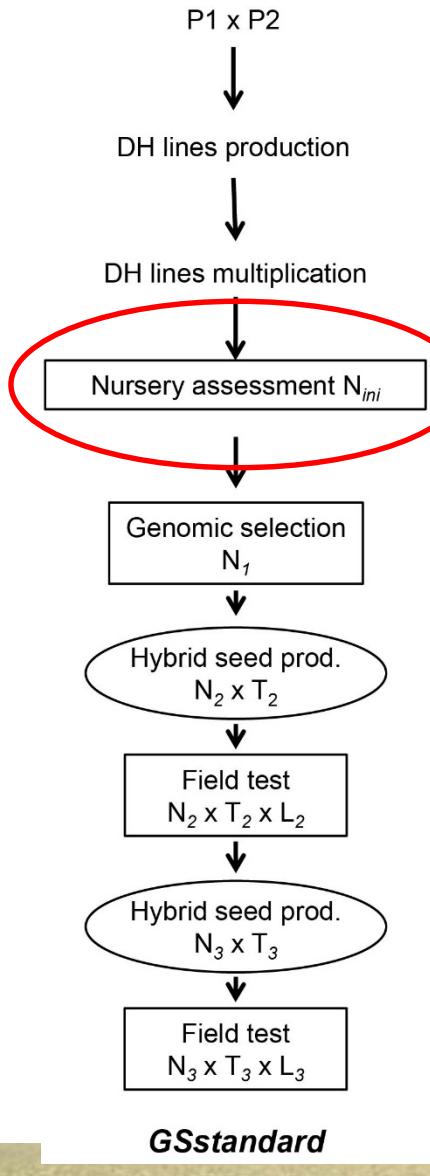
Budget =

$$\begin{aligned}
 & N_{ini}(Cost_{DH} + Cost_{nursery\ test}) \\
 & + N_1(Cost_{Genotyping}) + N_2 T_2 Cost_{Hybridseed} \\
 & + N_2 T_2 L_2 R_2 (Cost_{Test}) + N_3(T_3 - T_2) Cost_{Hybridseed} \\
 & + N_3 T_3 L_3 R_3 (Cost_{Test})
 \end{aligned}$$

Costs for:

- DH production
 - Nursery selection
 - GS
 - Hybrid seed production
 - Field plot,....
- All „redefined“ in field plot equivalents

Nursery selection

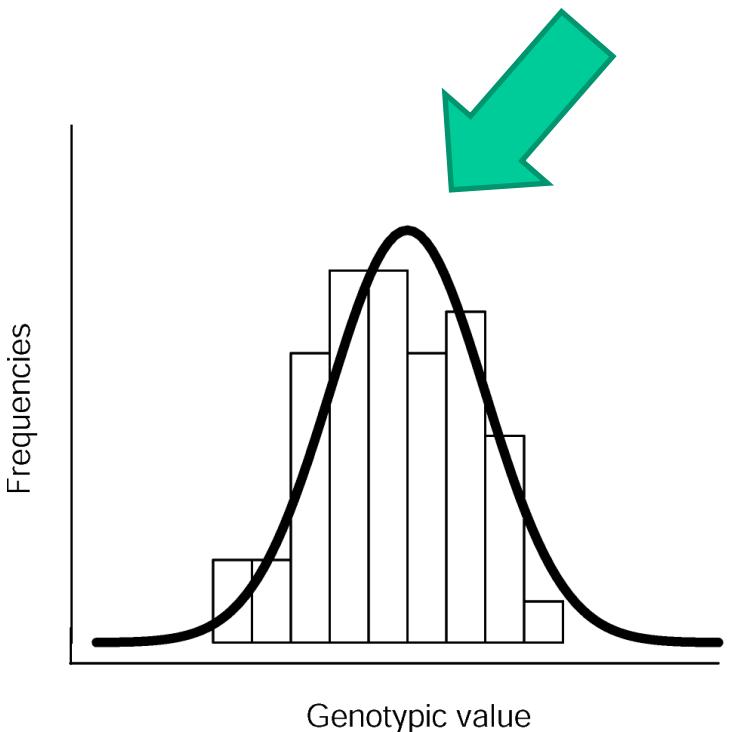
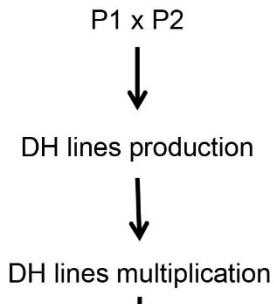


We assume:

- Nursery selection on traits not correlated to yield (e.g. lodging, leaf rust resistance, SDS,...)
- Nursery selection will not impact the selection gain formula
 - Nursery selection: costs money (budget impact)
 - Affects number of test lines
 - Nursery selection intensity is predefined (not optimized!) as $\alpha = \frac{N_{ini}}{N_1}$



First step of simulations



We define basics:

- Budget (for crossing, line development, GS, field tests) in field plot equivalents
- Crop & Trait to select for → variance components

$$\begin{aligned}\sigma_P^2 = & \sigma_{GCA}^2 + \sigma_{GCAxy}^2 + \sigma_{GCAXl}^2 / L + \sigma_{GCAXlxy}^2 / L + \\ & \sigma_{SCA}^2 / TM + \sigma_{SCAxy}^2 / TM + \sigma_{SCAXl}^2 / TML + \sigma_{SCAXlxy}^2 / TML \\ & \sigma_e^2 / (TLR)\end{aligned}$$

σ^2_{GCA} = GCA variance; σ^2_{SCA} = SCA variance ; σ^2_P = phenotyp. variance, $\sigma^2_{GCAX...}$ = variance due to Genotyp-environment - interaction; σ^2_e = error variance

→ estimated via ANOVA



Getting started with the code

library(selectiongain) → load the package

sessionInfo() → Info on the version you use

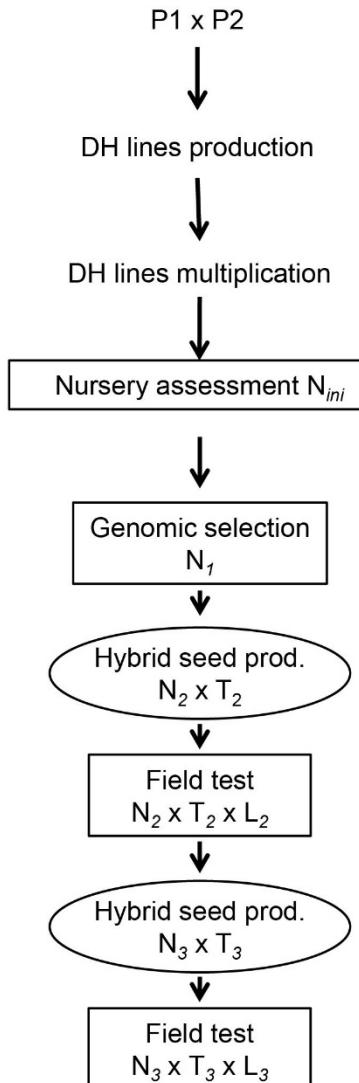
Important input parameters

- Budget = in field plot equivalents
- VCGCAandError =
 $c(GCA, GCA*loc, GCA*year, GCA*loc*year, error)$
- VCSCA = $c(SCA, SCA*loc, SCA*year, SCA*loc*year)$

Example for hybrid wheat:

- Budget = 10 000
- VCGCAandError = $c(5.7, 5.19, 0, 0, 24.37)$
- VCSCA = $c(1.88, 2.94, 0, 0)$

Second step of simulations



We define breeding operations:

- Intensity of nursery selection
- GS: yes/no; predictive ability
- Number of test stages in field
- Maximum numbers of testers, locations,...
- Costs for each operation

GSstandard



Most important code

Budget = 10000

VCGCAandError = c(5.7,5.19,0,0,24.37)

VCSCA = c(1.88,2.94,0,0)

multistageoptimum.search (

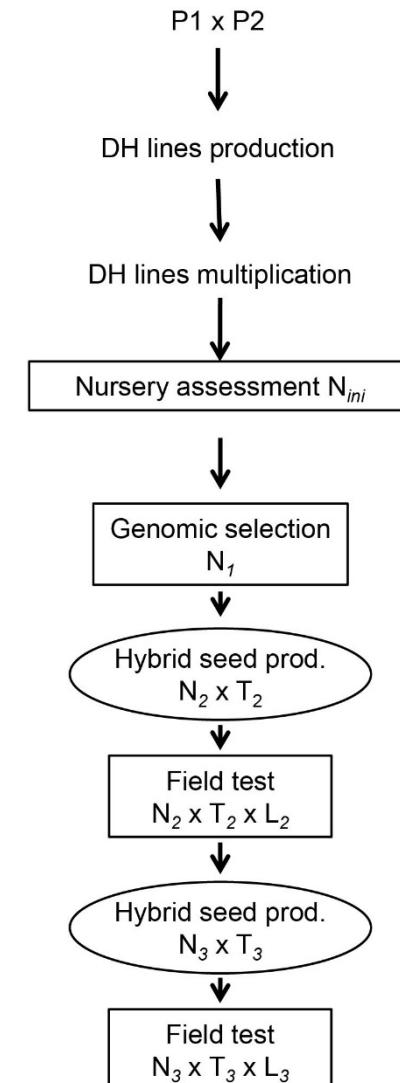
```
maseff=NA, alpha.nursery = 1,
VGCAandE=VCGCAandError, VSCA=VCSCA,
cost.nursery = c(1,0.3), CostProd = c(0,4,4),
CostTest = c(2,1,1), t2free = T,
Nf = 5, Budget = Budget,
N2grid = c(Nf, 511, 10), N3grid = c(Nf, 20, 1),
L2grid=c(2,5,1), L3grid=c(1,5,1),
T2grid=c(1,1,1), T3grid=c(1,5,1),
R2=1, R3=1, alg = Miwa(),
detail=FALSE, fig=FALSE)
```

Nf	Nini	N1	N2	N3	L2	L3	T2	T3	R2
5.000	204.000	0.000	51.000	9.800	5.000	10.000	1.000	2.000	1.000

R3 Bini B1 B2 B3 Budget Gain

R3	Bini	B1	B2	B3	Budget	Gain
1.000	265.200	0.000	459.000	274.900	1000.000	3.278

3.278



GSstandard

ΔG

Most important code - explained

multistageoptimum.search (

maseff = GS pred. ability,

alpha.nursery = selected fraction in disease nursery,

VGCAandE=**VCGCAandError**, **VSCA**=**VCSCA**,

cost.nursery = c(line prod., test in nursery),

CostProd = c(0, hybrid seed prod., hybrid seed prod.),

CostTest = c(GS, yield plot, yield plot), **t2free** = T,

Nf = no. finally selected lines, **Budget** = Budget,

N2grid = c(Nf, 511, 10), **N3grid** = c(Nf, 5, 1),

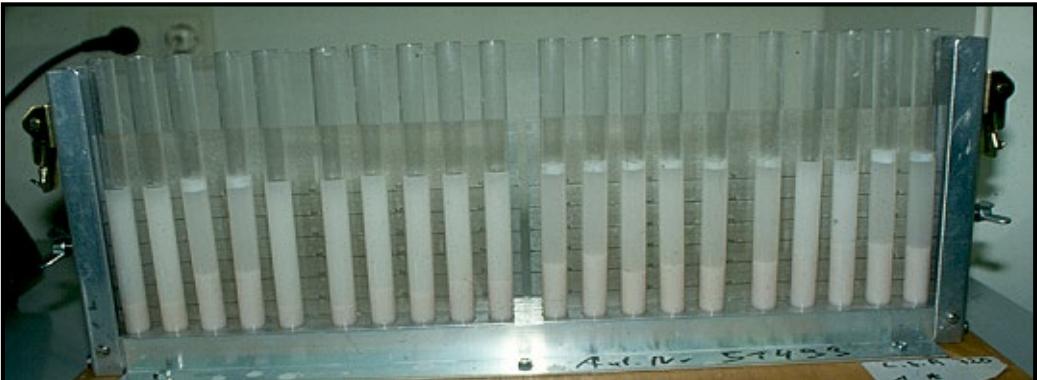
L2grid=c(1,5,1), **L3grid**=c(1,5,1),

T2grid=c(1,1,1), **T3grid**=c(1,3,1),

R2=1, **R3**=1, **alg** = Miwa(),

detail=FALSE, **fig**=FALSE)

Let's go into breeding world



Modeling the optimum allocation

1. Basic level

- Target criterion
- Trait

2. Breeding level

- Scheme
- Scenario

3. Optimization level

- Test resources

Modeling the optimum allocation

1. Basic level

- Target criterion = selection gain
- Trait = grain yield in wheat

2. Breeding level

- Scheme = PS standard
- Scenario = variance components, budget, selected fraction, technical requirements,...

3. Optimization level

- Test resources = number of test locations, testers, replications, DH lines

Determining the opt. allocation within a given model framework



1. Basic level

- Selection gain
- Maize grain yield

2. Breeding level

- PSstandard
- Given scenario

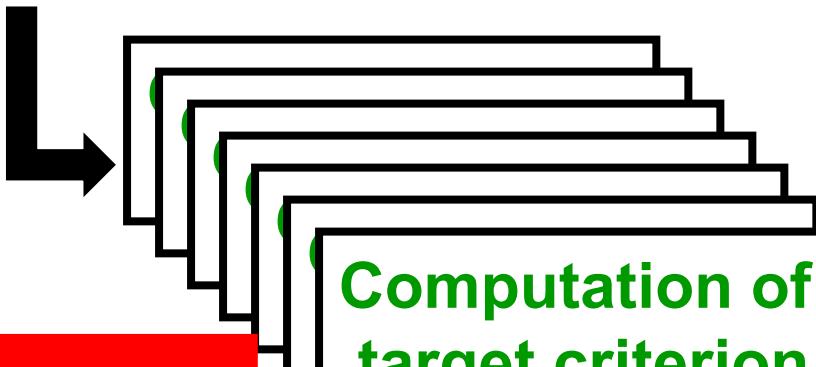
3. Optimization level

specific allocation of the number of testers, test locations, DH lines, replications

**AIM: Find the allocation
maximizing selection gain for
that specific def. of level 1
and 2**

= optimum allocation

**Computation of
target criterion**



Use of molecular markers

Nothing else than indirect selection

$$\Delta G = i\rho \sigma_y / Y$$

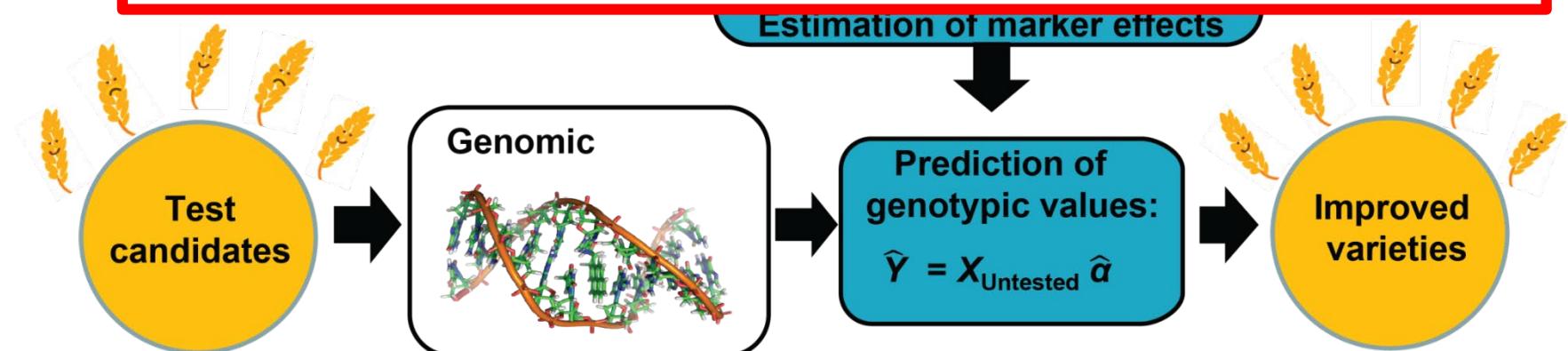
- $0 < \rho < 1$: **selection gain is only increased if** the use of the test criteria enables
 - Increase of $i \rightarrow$ **high throughput: N_1**
 - Increase of h
 - Increase of σ_y
 - Decrease of $Y \rightarrow$ **fast recycling**

Genomic selection

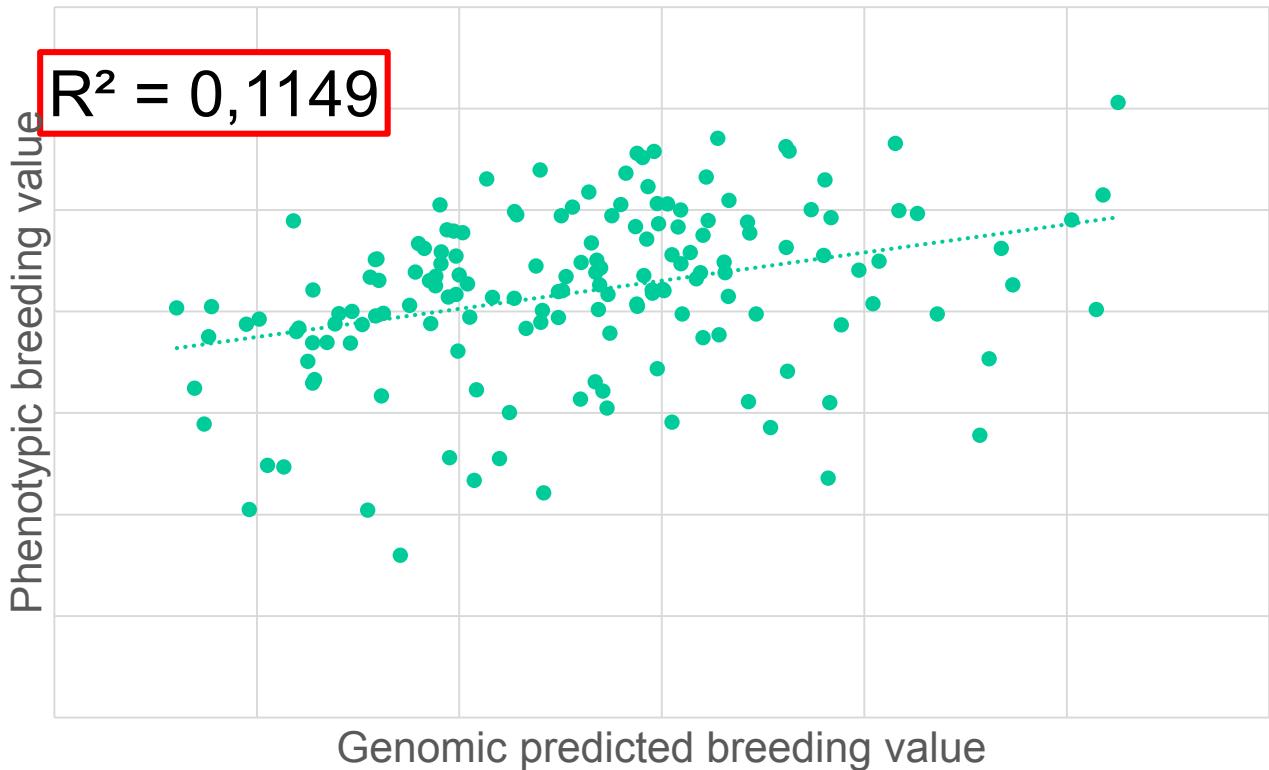


We assume, that

- prediction models have been developed in previous projects → no budget required for it now!
- Retraining of the model is done with routine field trials → no additional budget required for it!



We need prediction ability



Correlation between observed and predicted breeding value = prediction ability
→ that's what we need for our simulations

$$\text{Prediction accuracy} = \frac{\text{prediction ability}}{h}$$

Modifications of the framework

Besides budget, variance components, costs for line production and phenotyping we need further data:

- Correlation GS with GCA: $\rho(GS, GCA) = 0.3$ (for T0 scenario; Zhao et al. 2014)
- Costs GS = high density genotyping of 1 line costs as much as 2 field plots

- (data is shown for wheat with framework based on papers below)

Breeding schemes

PS_{standard}

Year 1 P1 x P2, ...

Year 2 DH-Production

Year 3 N_1 DH lines -
multiplication

Year 4 Hybrid
seed prod.

$$N_2 * T_2$$

Year 5 Field test

$$N_2 * T_2 * L_2$$

Year 6 Hybrid
seed prod.

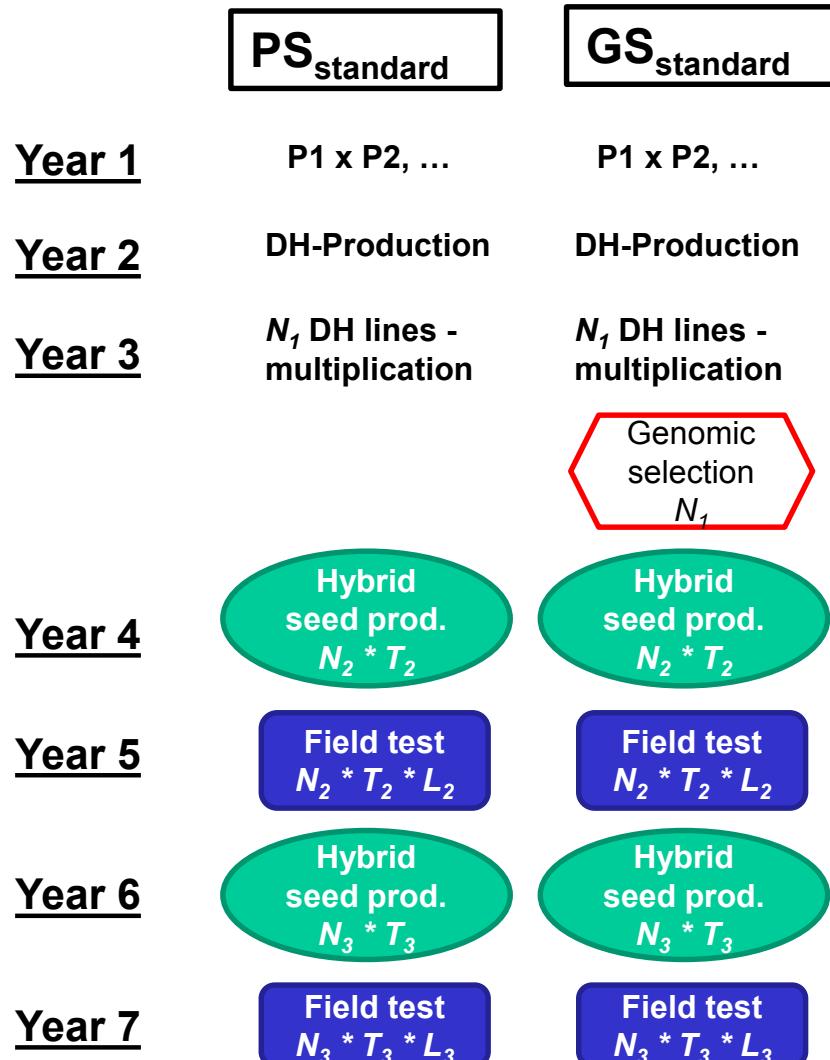
$$N_3 * T_3$$

Year 7 Field test

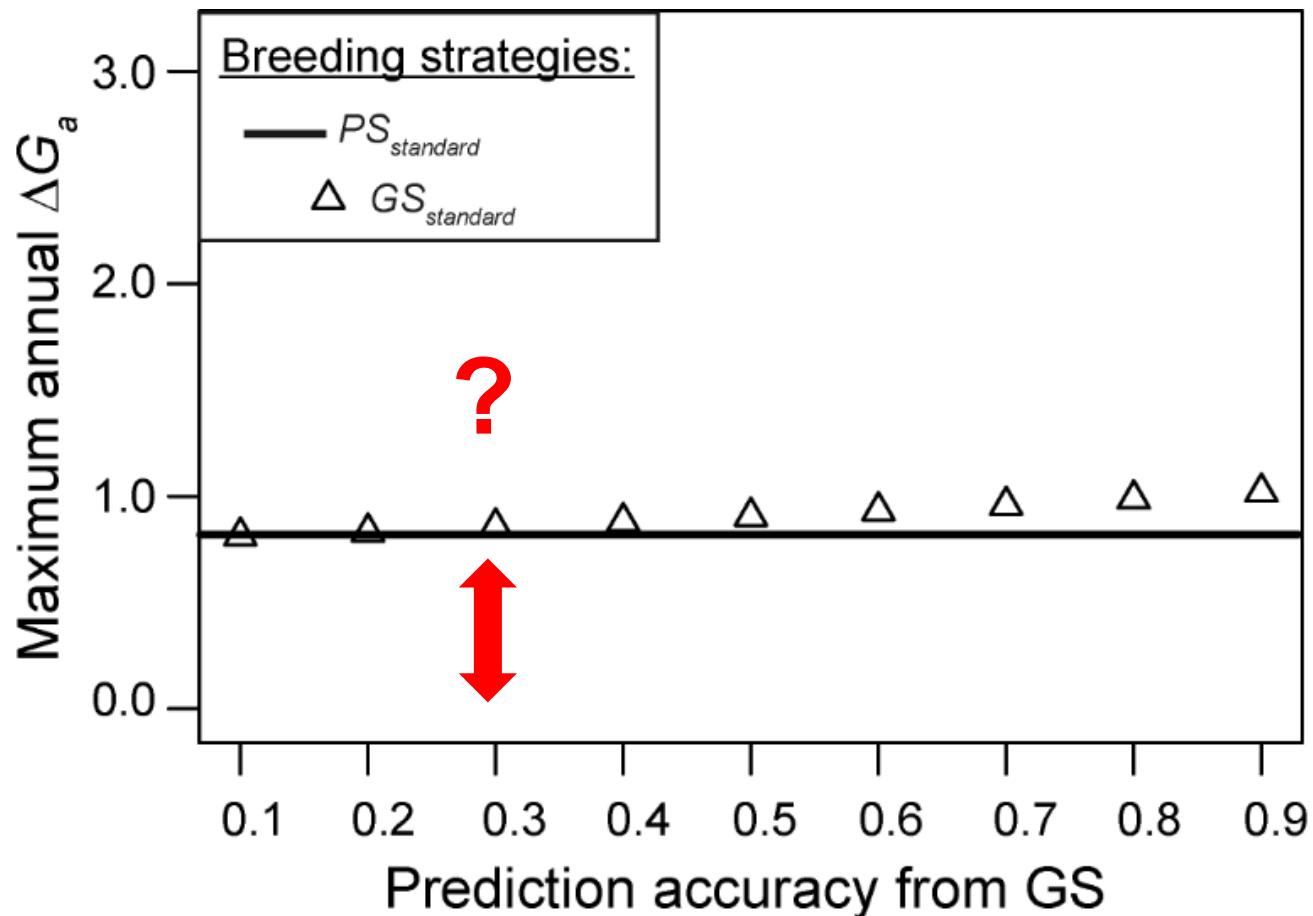
$$N_3 * T_3 * L_3$$

Year 8, 9,... Pre-registration trials

Breeding schemes



Increase in selection gain with GS

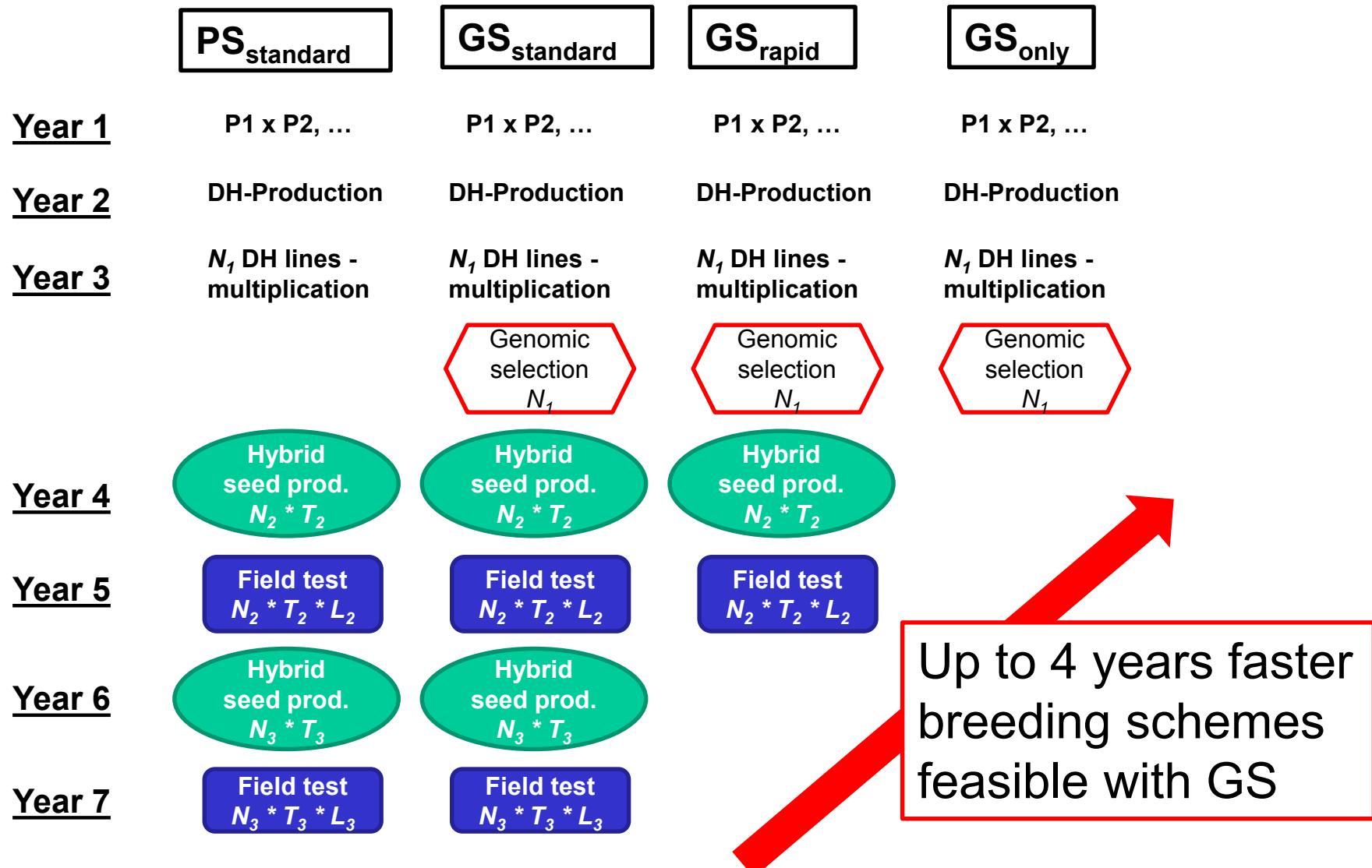


The higher the prediction accuracy the larger is the advantage of GS schemes

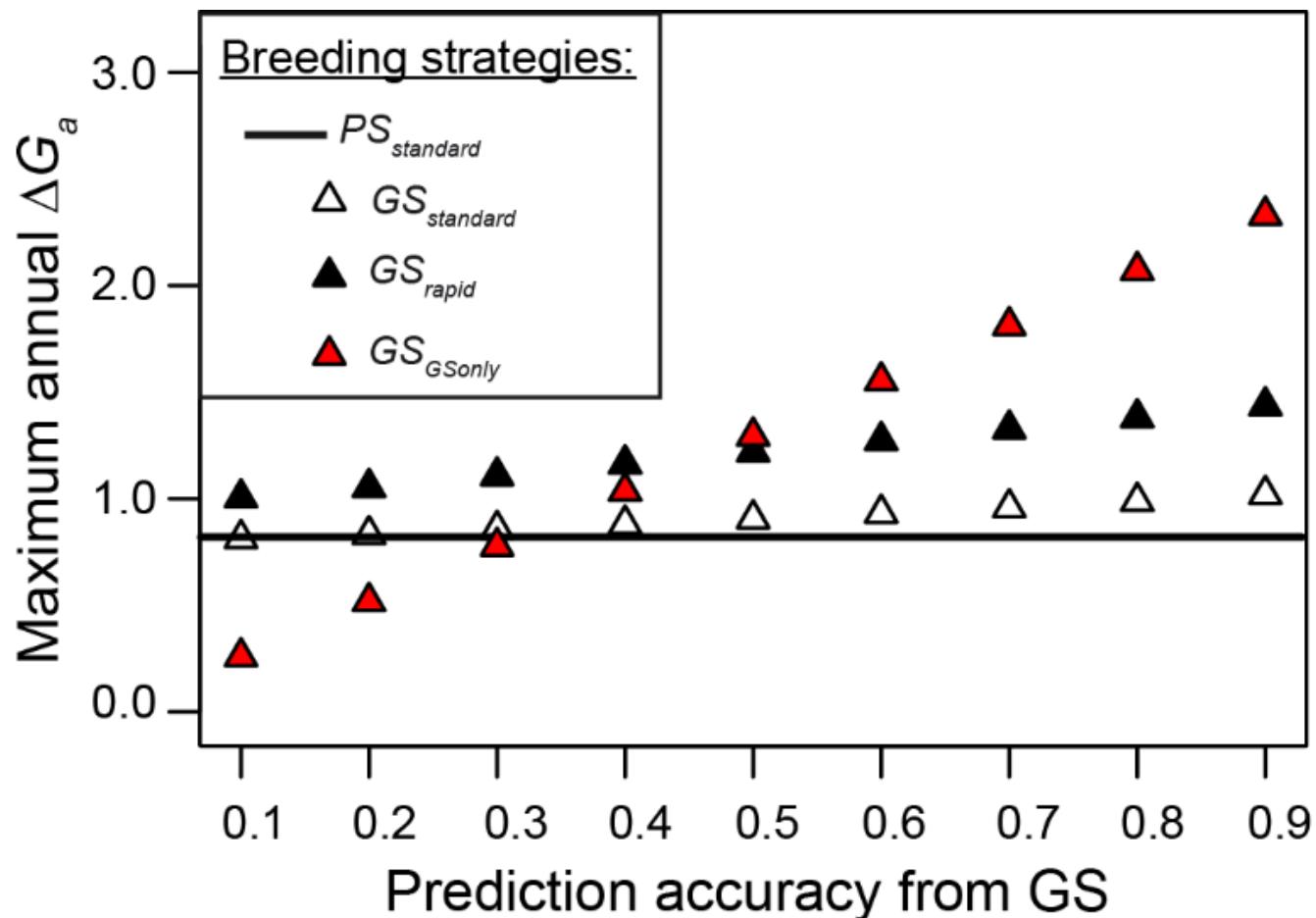
Reduce cycle length with GS

	PS _{standard}	GS _{standard}	GS _{rapid}	GS _{only}
<u>Year 1</u>	P1 x P2, ...	P1 x P2, ...	P1 x P2, ...	P1 x P2, ...
<u>Year 2</u>	DH-Production	DH-Production	DH-Production	DH-Production
<u>Year 3</u>	N_1 DH lines - multiplication	N_1 DH lines - multiplication	N_1 DH lines - multiplication	N_1 DH lines - multiplication
		Genomic selection N_1	Genomic selection N_1	Genomic selection N_1
<u>Year 4</u>	Hybrid seed prod. $N_2 * T_2$	Hybrid seed prod. $N_2 * T_2$	Hybrid seed prod. $N_2 * T_2$	
<u>Year 5</u>	Field test $N_2 * T_2 * L_2$	Field test $N_2 * T_2 * L_2$	Field test $N_2 * T_2 * L_2$	
<u>Year 6</u>	Hybrid seed prod. $N_3 * T_3$	Hybrid seed prod. $N_3 * T_3$		
<u>Year 7</u>	Field test $N_3 * T_3 * L_3$	Field test $N_3 * T_3 * L_3$		

Reduce cycle length with GS

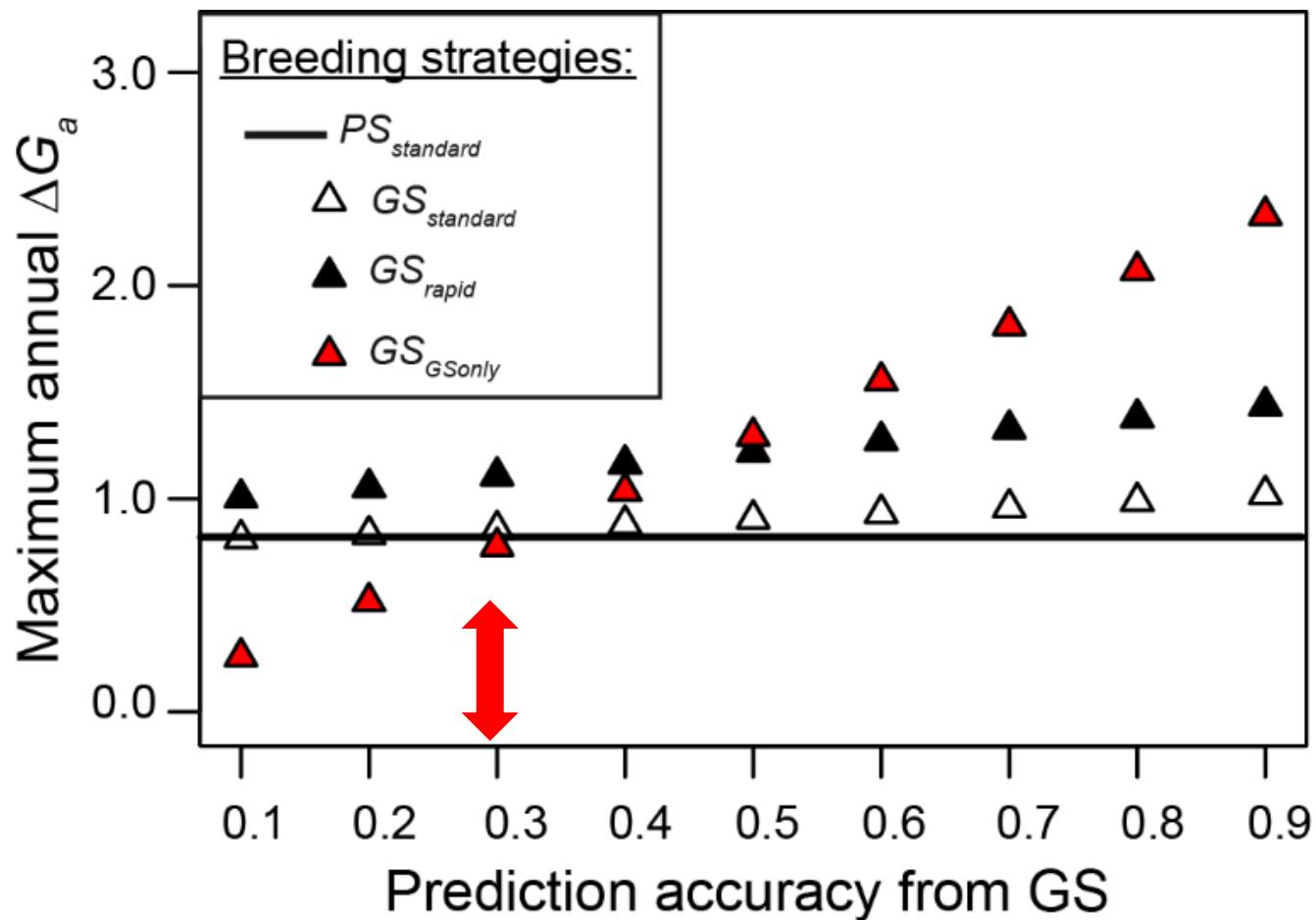


GS for yield is interesting



Genomic selection is promising for grain yield especially when used to shorten breeding cycle length

GS for yield is interesting



With recent GS accuracy breeding scheme GS_{rapid} seems most promising: + 35% in annual selection gain

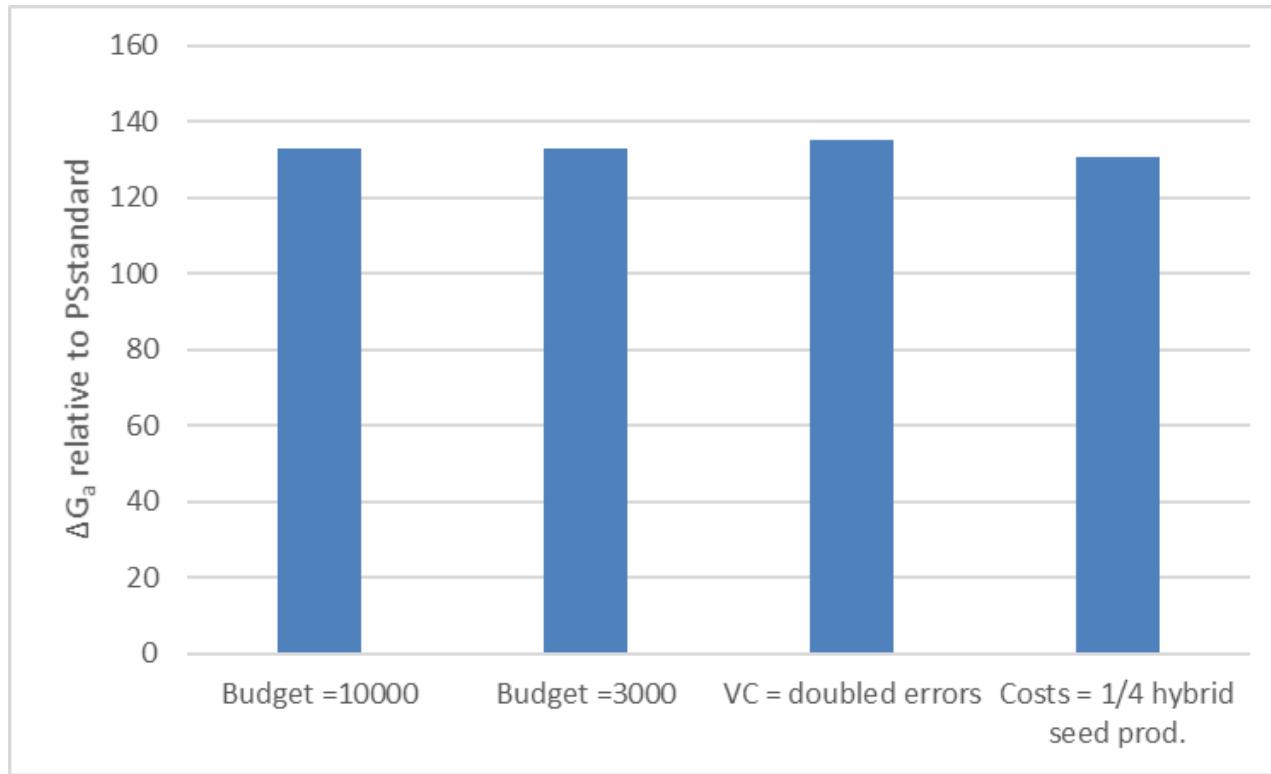
Generalisation of results

We state that a breeding scheme using GS, namely GSrapid, is top and should be used; but is it also the truth for

- small budgets ?
- different variance components ?
- reduced hybrid seed production costs ?

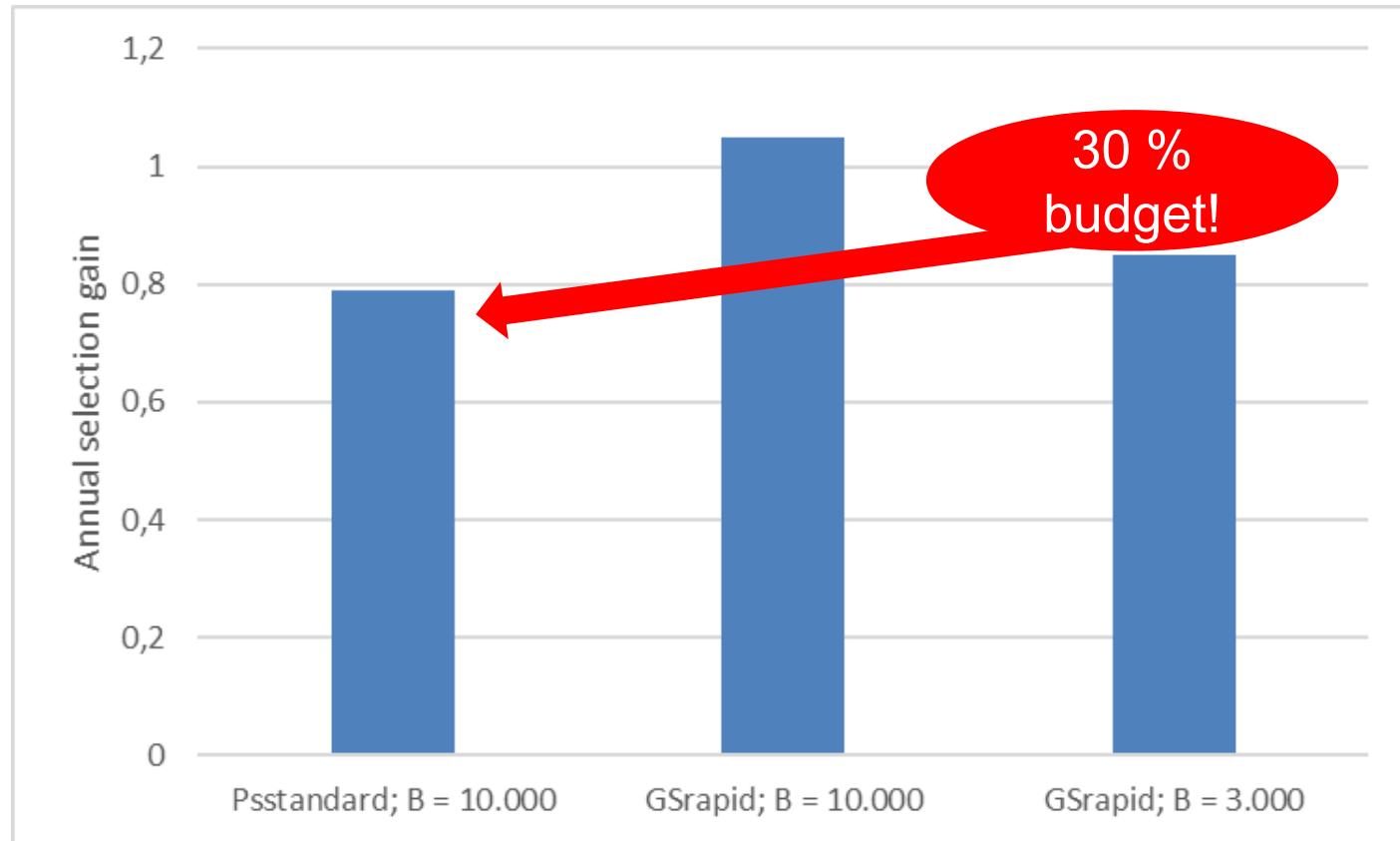


Broad advantage of GSrapid



The use of GS in elite breeding is recommended for a broad range of scenarios; also for small breeding programs!

Think about....



Using **GSrapid** with a budget of 3.000 field plots has a 7.6 % higher annual selection gain than PSstandard with a budget of 10.000 field plots!

Realization of GSrapid

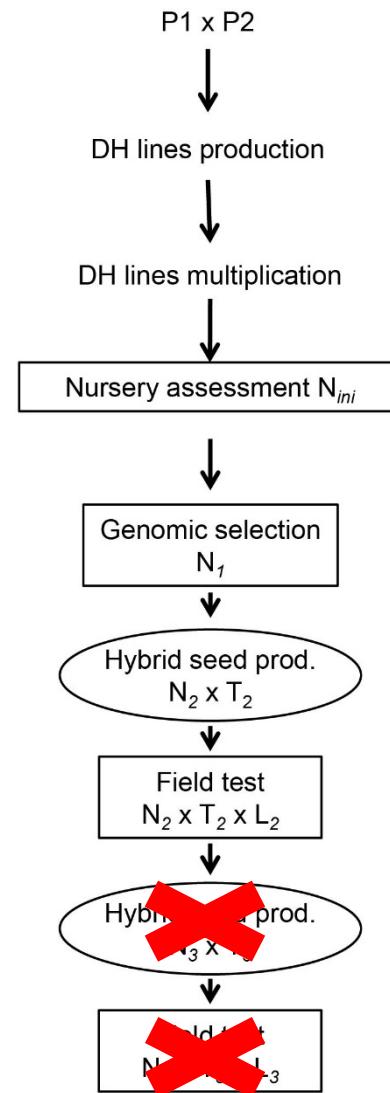
Budget = 10000

VCGCAandError = c(5.7,5.19,0,0,24.37)

VCSCA = c(1.88,2.94,0,0)

multistageoptimum.search (

 maseff=0.3, alpha.nursery = 0.25,
 VGCAandE=VCGCAandError, VSCA=VCSCA,
 cost.nursery = c(1,0.3), CostProd = c(0,4,0),
 CostTest = c(2,1,0), t2free = T,
 Nf = 5, Budget = Budget,
 N2grid = c(Nf, 511, 10), N3grid = c(Nf, 5, 1),
 L2grid=c(1,5,1), L3grid=c(0,0,1),
 T2grid=c(1,3,1), T3grid=c(0,0,1),
 R2=1, R3=1, alg = Miwa(),
 detail=FALSE, fig=FALSE)



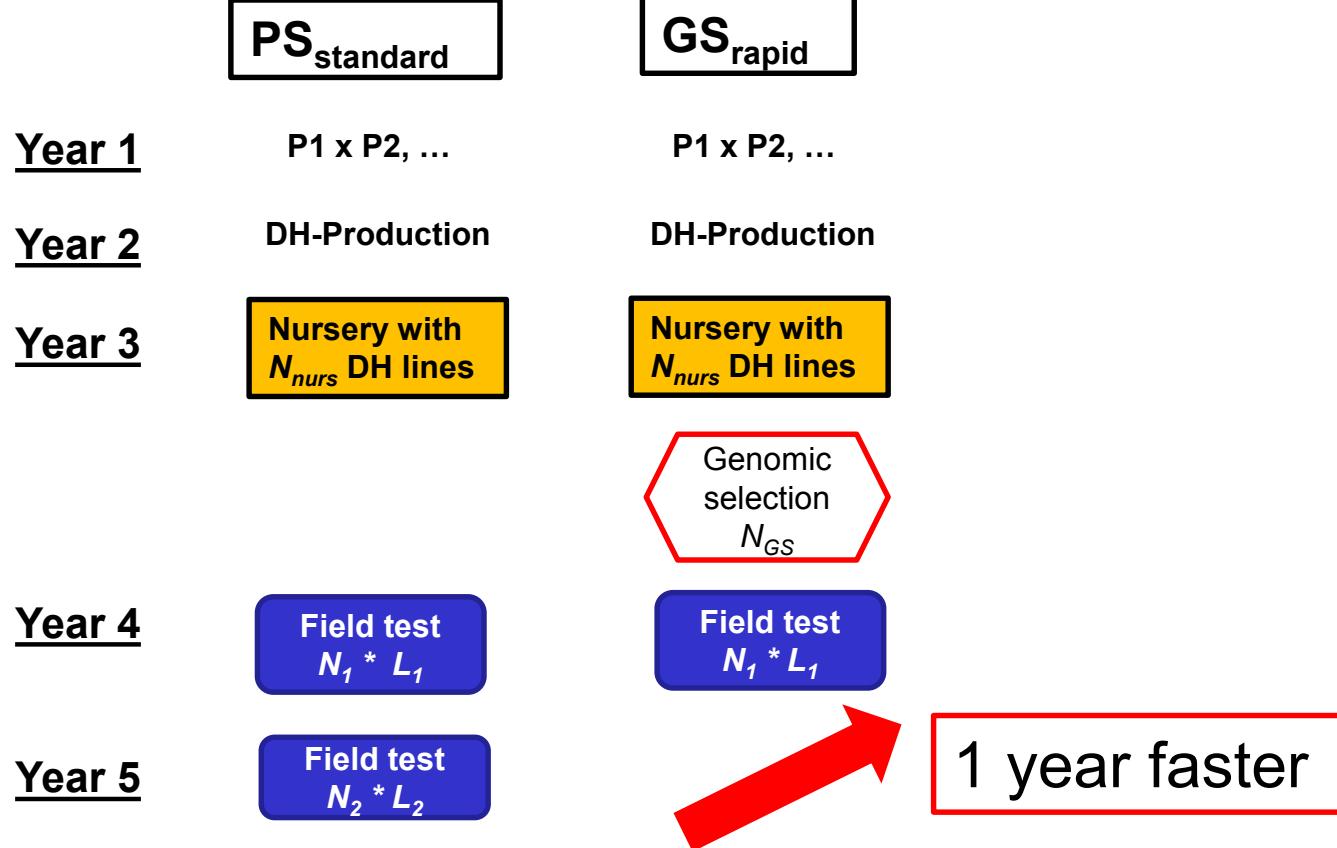
GSstandard

Student simulations

?



Line breeding



Questions – durum breeding

- Is GSrapid better than PSstandard also for line breeding?
 - Speed up in DH method worthwhile also if very expensive?
 - What about breeding schemes for durum breeding in developing countries?
- Look on annual and absolute selection gain and elaborate potential differences in the allocation of resources



Questions – durum breeding

- Is GSrapid better than PSstandard also for line breeding?
- Speed up in DH method worthwhile also if very expensive?
- What about breeding schemes for durum breeding in developing countries?

→ Look on annual and absolute selection gain and elaborate potential differences in the allocation of resources

- Three student groups
 1. Optimize both classical breeding schemes for different budgets, variance components, GS pred. abilities
 2. Optimize both breeding schemes with fast DH for different budgets, variance components, ...
 3. Optimize both breeding schemes in a context of very low budgets and high GS costs

How to realize

Excel files

- Breeding schemes defined
- All necessary input parameters (variance components, costs,...)
- Just let the package work for you:
 - For each defined scenario, you will get the maximum possible absolute selection gain and the allocation of resources → put it all to an Excel table
 - Run next scenario
 - Finally compare the results and find the best scheme with its best allocation of resources

Example table

Breeding strategy	Alpha in Nursery	Optimum allocation of test resources									ΔG_a
		<i>Nini</i>	<i>N1</i>	<i>N2</i>	<i>N3</i>	<i>L2</i>	<i>L3</i>	<i>T2</i>	<i>T3</i>		
PSstandard	0.1	3170	-	317	30	7	10	1	6	0.75	
PSstandard	0.2	2190	-	438	35	6	10	1	6	0.79	
PSstandard	0.3	1680	-	504	35	6	10	1	6	0.80	
PSstandard	0.4	1395	-	558	40	5	10	1	6	0.81	
PSstandard	0.5	1178	-	589	40	5	10	1	6	0.81	
GSrapid	0.1	3440	344	115	-	10	-	3	-	0.99	
GSrapid	0.2	2790	558	125	-	10	-	3	-	1.04	
GSrapid	0.3	2500	750	125	-	10	-	3	-	1.06	
GSrapid	0.4	2060	824	135	-	10	-	3	-	1.07	
GSrapid	0.5	1882	941	135	-	10	-	3	-	1.08	

Changes in code

Budget = 10000

VCGCAandError = c(14.06,22.27,0,0,24.37)

VCSCA = c(**0,0,0,0**)

multistageoptimum.search (

maseff=NA, alpha.nursery = 0.25,

VGCAandE=VCGCAandError,

VSCA=VCSCA, cost.nursery = c(1,0.3),

CostProd = c(0,**0,0**),

CostTest = c(2,1,1), t2free = T,

Nf = 5, Budget = Budget,

N2grid = c(Nf, 6011, 40), N3grid = c(Nf, 1511,

5), L2grid=c(1,5,1), L3grid=c(2,10,1),

T2grid=c(**1,1,1**), T3grid=c(**1,1,1**),

R2=1, R3=1, alg = Miwa(),

detail=FALSE, fig=FALSE)

P1 x P2



DH lines production



DH lines multiplication



Nursery assessment N_{ini}



Genomic selection



Hybrid seed prod.



Field test



$N_2 \times \text{red X} \times L_2$



Hybrid seed prod.



$N_3 \times \text{red X} \times L_3$



Field test

$N_3 \times \text{red X} \times L_3$

GSstandard



Contact

PD Dr. Friedrich Longin

State Plant Breeding Institute, University of Hohenheim,

Fruwirthstrasse 21

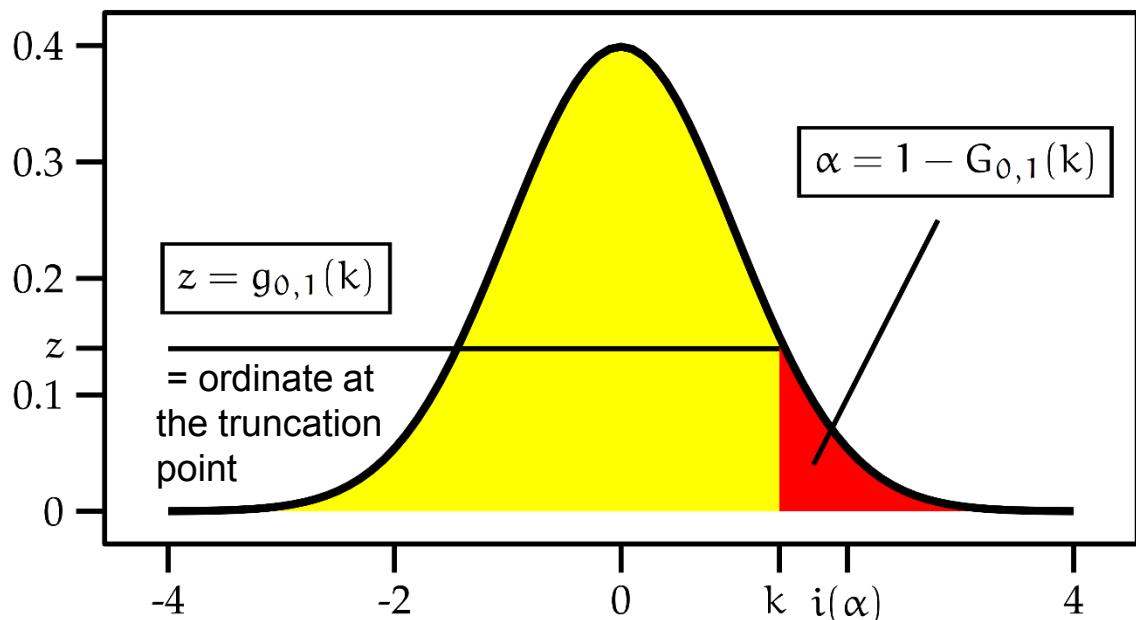
70593 Stuttgart, Germany

Phone: (++49) 0711 459 23846

friedrich.longin@uni-hohenheim.de



Prediction of selection intensity



$$S = i(\alpha)\sigma_P$$

$$\alpha = \frac{\text{no. selected lines}}{\text{no. tested lines}}$$

= truncation point calculated by the cumulative density function

$$i(\alpha) = \frac{z(\alpha)}{\alpha} = \frac{g_{0,1}(k)}{\alpha} = \frac{g_{0,1}[G_{0,1}^{-1}(1 - \alpha)]}{\alpha}$$

Realisation in R

$$i(\alpha) = \frac{z(\alpha)}{\alpha} = \frac{g_{0,1}(k)}{\alpha} = \frac{g_{0,1}[G_{0,1}^{-1}(1-\alpha)]}{\alpha}$$

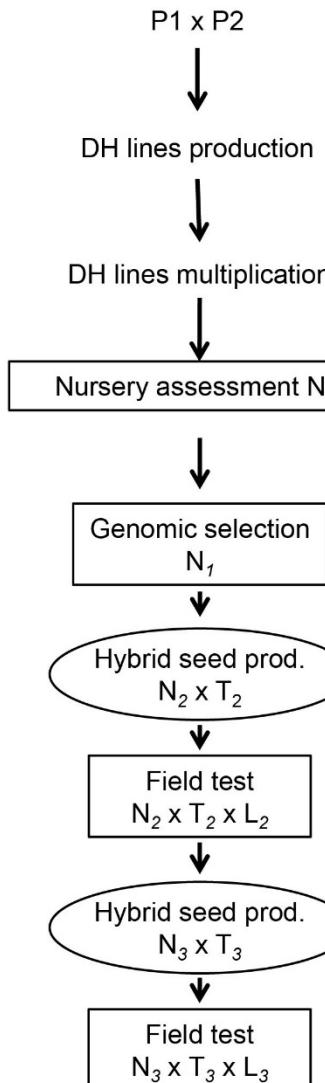
Numerical calculation:

	R	Excel
$g_{\mu, \sigma^2}(x)$	<code>dnorm(x, mu, sigma)</code>	<code>=NORMVERT(x; mu; sigma; 0)</code>
$G_{\mu, \sigma^2}(x)$	<code>pnorm(x, mu, sigma)</code>	<code>=NORMVERT(x; mu; sigma; 1)</code>
$G_{\mu, \sigma^2}^{-1}(1 - \alpha)$	<code>qnorm(1-alpha, mu, sigma)</code>	<code>=NORMINV(1-alpha; mu; sigma)</code>

Keep in mind:

- these formula assume infinite number of candidates
- for exact calculation → order statistics but difficult to realize for multistage selection
- no difference in allocation of resources and selection gain only slightly overestimated by infinitesimal model

Multistage selection = multivariate distributions



Approximation from Dickerson and Hazel 1944

$$\Delta G = i_1 \rho_1 \sigma_y + i_2 \rho'_2 \sigma'_y$$

- ρ_1 = heritability
- σ'_y = genetic variance after first selection =

$$\sqrt{\sigma_y^2 (1 - \rho_1^2 (i_1(i_1 - k_1)))} \rightarrow > \text{for decreased } \alpha$$
- $\rho'_2 = \frac{\rho_2 - \rho_1 \rho_{12} i_1 (i_1 - k_1)}{\sqrt{(1 - \rho_1^2 i_1 (i_1 - k_1))(1 - \rho_{12}^2 i_1 (i_1 - k_1))}}$

GSstandard





Interested in more details?

Vortr. Pflanzenzüchtg. 7, 30-40 (1984)

- 31 -

Calculating and maximizing the gain from selection

H.F. Utz

Institute of Plant Breeding, Seed Science, and Population Genetics, University of Hohenheim

1. Introduction

The breeder must often predict the gain due to selection. As early as 1934, STUDENT used the gain to interpret a selection experiment. With the aid of the expected gain, SMITH (1936) derived the optimum weighting of characters, i.e. the selection index. In the meantime, this equation has become a standard instrument in deciding which alternative selection procedure is better or in answering questions such as how many replications or candidates would be needed to reach a maximum of gain. KEMPTHORNE (1977) succinctly stated that the expected gain is "the 'work horse' formula of quantitative selection".

Whoever calculates expected gains will be confronted with certain numerical problems. It will therefore be useful to discuss the calculation and maximization of such gains. In the following paper, the basic formulae will be presented and some helpful approximations given for the cases of one-stage and multi-stage selection.

2. Predicting gain from selection

where σ_y is the standard deviation of the genetic values y of the candidates,
 ρ_{yn} is the correlation coefficient between y and the phenotypic measurement n , which can be a single observation, a mean of values, or a selection index criterion, and
 $i_{(\alpha)}$ is the selection intensity or the standardized selection differential.

The determination of σ_y and ρ_{yn} is rather a problem of estimation, since the parameters of the population undergoing selection are usually unknown. The size of ρ_{yn} depends on the test situation, the number of replications, of locations, and also which parts of the genetic variance can be exploited. A detailed discussion of these topics can be found in the textbook of HALLAUER and MIRANDA (1981) or in SCHNELL (1982).

Assuming a normal distribution for y and a great population of candidates, then the selection intensity can be obtained as

$$i_{(\alpha)} = z_{(\alpha)} / \alpha \quad (2)$$

where $z_{(\alpha)}$ is the ordinate of the standardized normal distribution at the point of truncation, and α is the selected fraction.

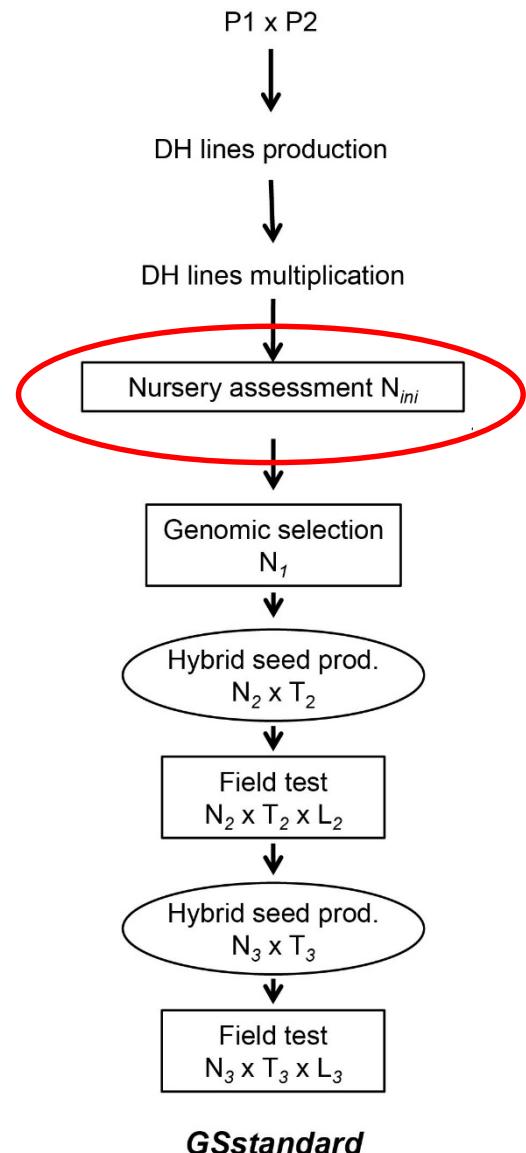
The ordinate $z_{(\alpha)}$ for certain α is tabulated in standard statistical tables (cf. PEARSON and HARTLEY, 1970). The direct use of tables of $i_{(\alpha)}$ is more convenient. KONDO and ELDERTON (1931) tabulated $i_{(\alpha)}$ for $\alpha = 0.001(0.001)1$ with 10 decimals. BECKER (1975) gives tables for the same steps

- For exact formulas: Mi Papers, Cochran (1951),...
- For finite sample size: MC simulations or order statistics

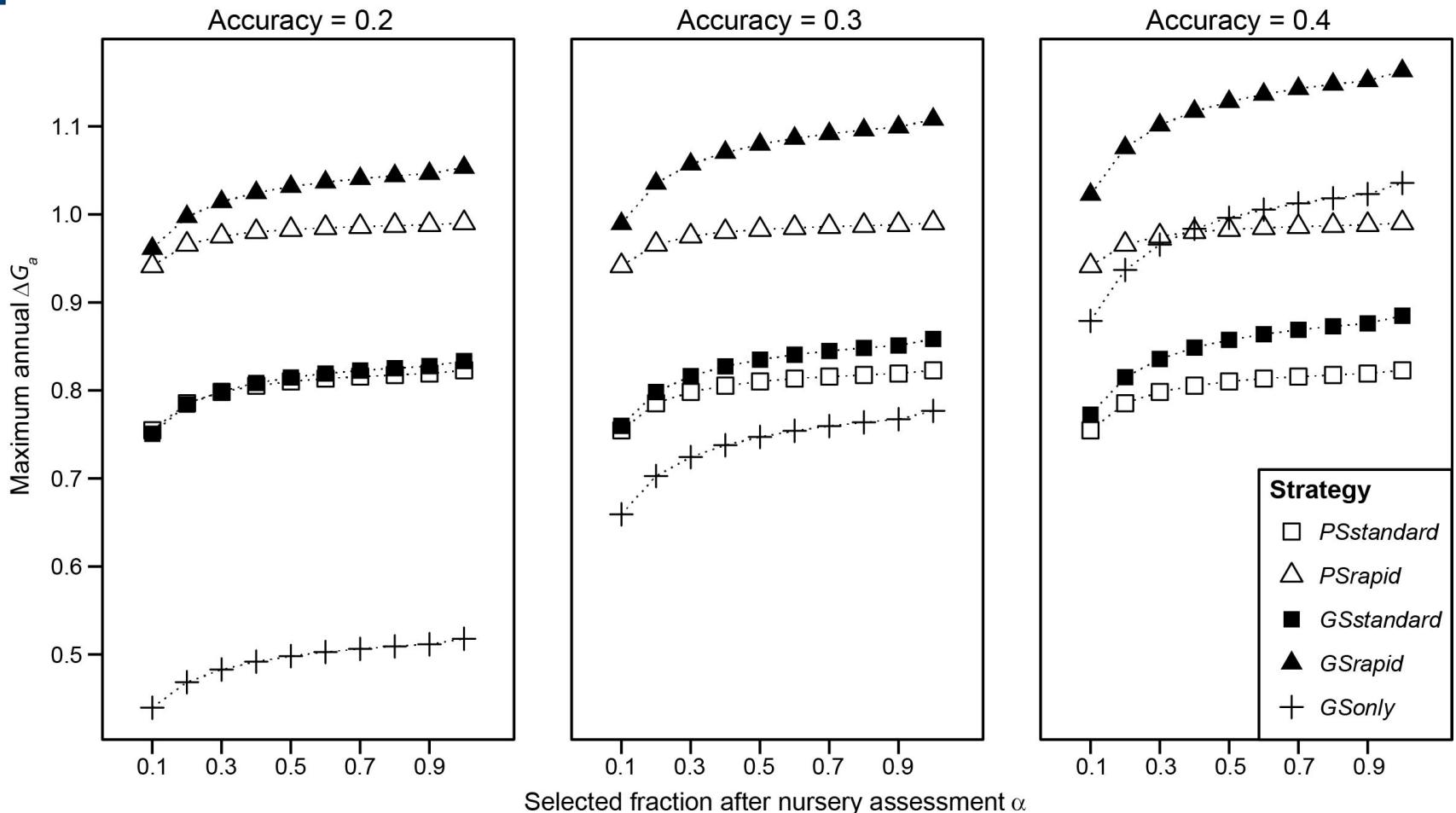
Nursery selection

Further data required:

- Breeders usually select lines first based on high heritable traits (diseases, qualities,...)
- Assumption: traits observed in nursery do **not correlate with target trait, eg. GCA for grain yield**
- Cost for nursery selection = 0.3 yield plots

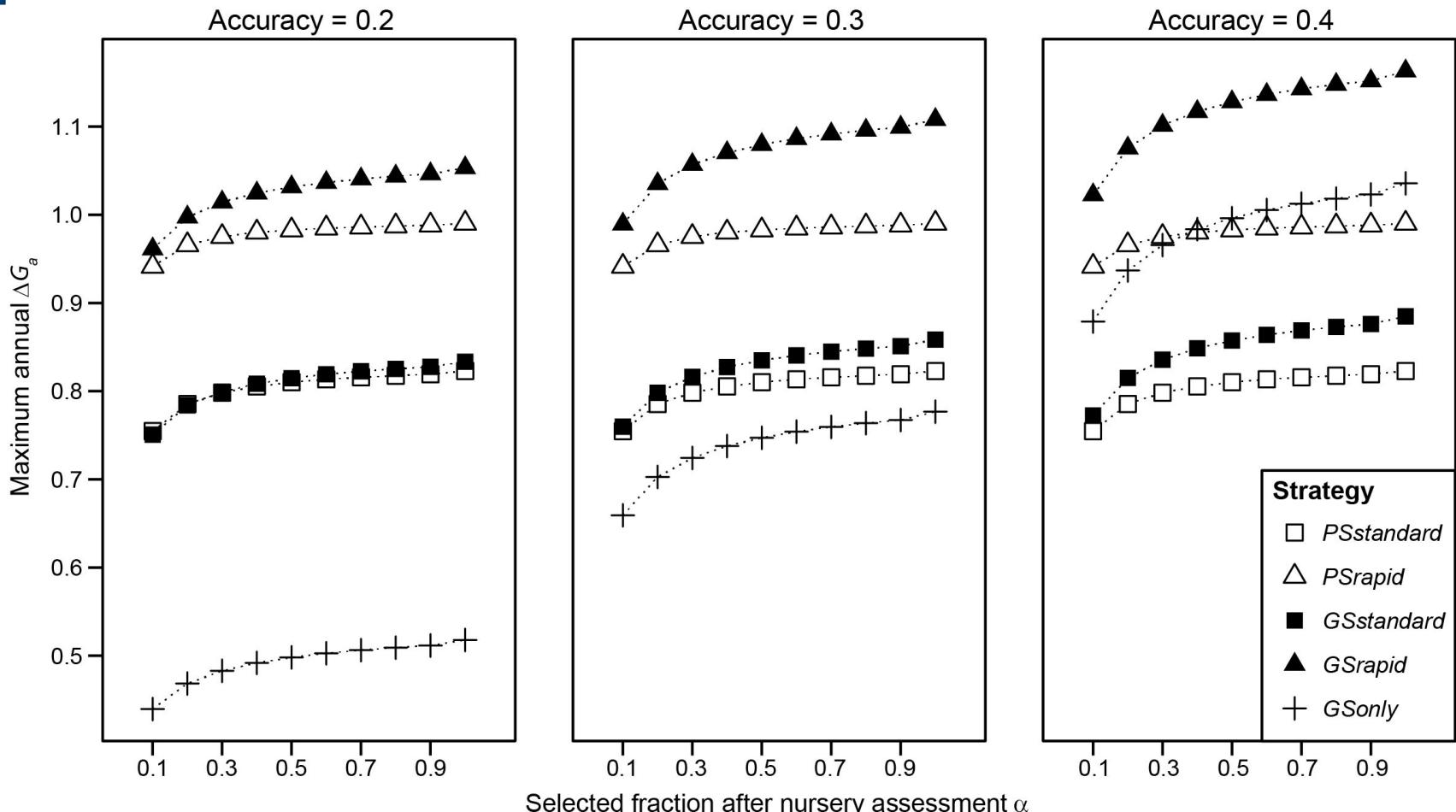


Impact of nursery selection



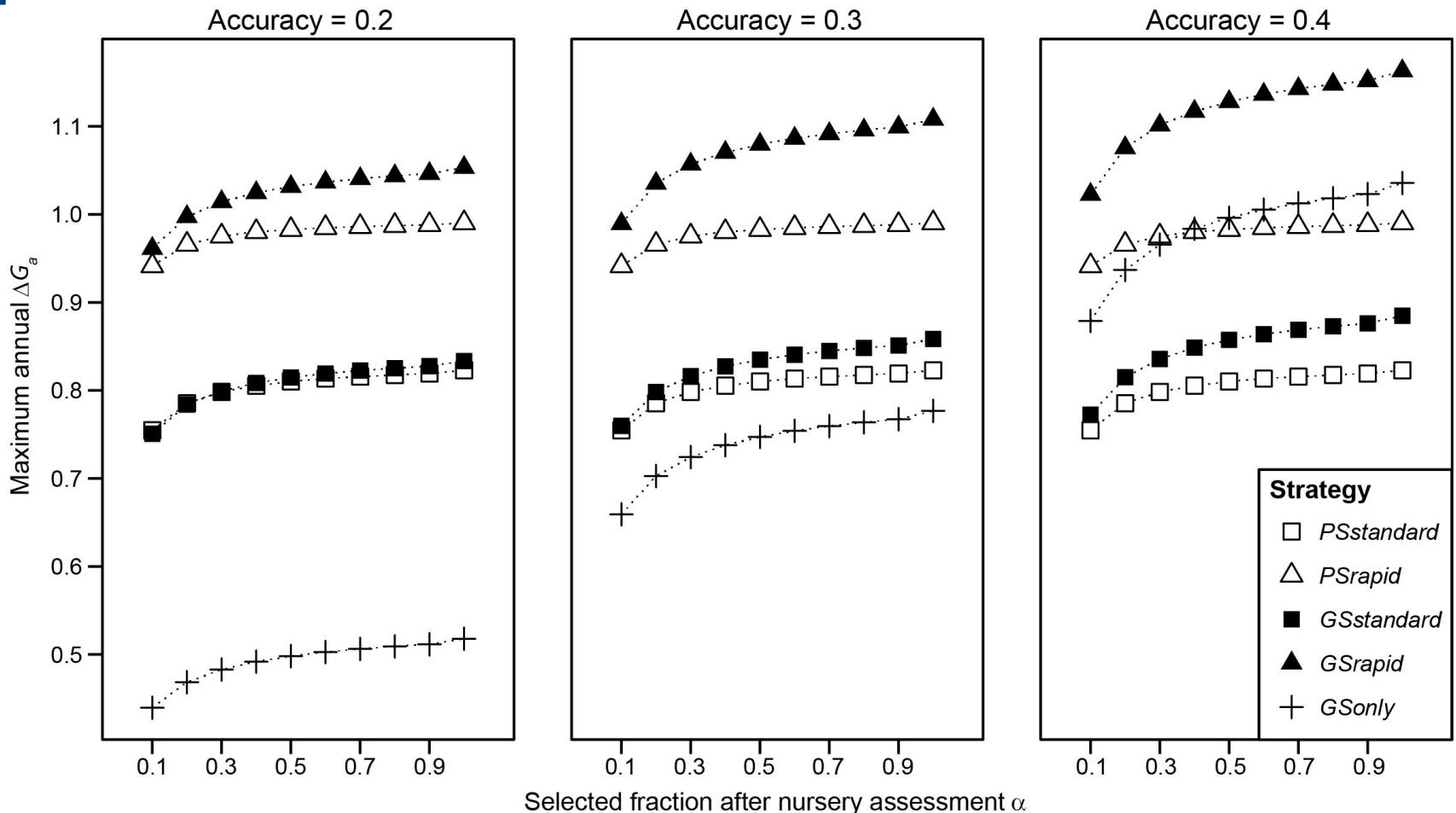
1. Ranking of schemes nearly not affected by nursery selection → **GSrapid is top!**

Impact of nursery selection



2. With increasing nursery selection (smaller α), ΔG_a for grain yield is reduced

Impact of nursery selection



3. This reduction is larger for GS schemes and for increased GS prediction accuracy