Optimization and comparison of alternative breeding schemes

Pflanzengenetisches Kolloquium ws 2017/18
14th December 2017, University of Goettingen

Overview

- Introduction
- Expected selection gain
- Model calculations
  Hybrid winter rye (Tomerius 2001 / 2008)
- Simulating breeding programs
  Software package “SelectionTools”
- Conclusions & discussion

Introduction

Designing a breeding scheme means a **lot of choices:**

Which type of variety do you develop (line/OP/SYN/ Hybrid)?

Which / how many crosses do you produce?

How many progeny per cross?

Which methods are you using for line development?

If you develop hybrid varieties, what is your hybrid mechanism?

In which generation are you making test hybrids?

Which / how many testers do you use?

Introduction

How many selection stages do you use?

What type of trials do you use at the single stages?

Unreplicated observation -> multi-location replicated

How sharply do you select at each stage?

Fixed or variable selection intensity ?

How do you handle multi-trait selection?

Are you using markers / genomic selection ?

At which stages and for which traits?

Introduction

Many different possible breeding schemes exist

- often very complex
- efficiency may differ remarkably
- often used for 'historical' reasons
  - ➔ breeder should aim to find the best possible scheme

Problems:

- Practical comparison hardly feasible
- Improvements are based on experience / trial & error
- Judgement of efficiency is often indirect

Helpful tools:

- Model Calculations (MC)
- Breeding simulation studies (SIM)

Introduction

MC/SIM

allow an a priori-judgement between schemes

by predicting the relative efficiency and

giving the respective optimum variant(s)

of a virtual number of breeding schemes

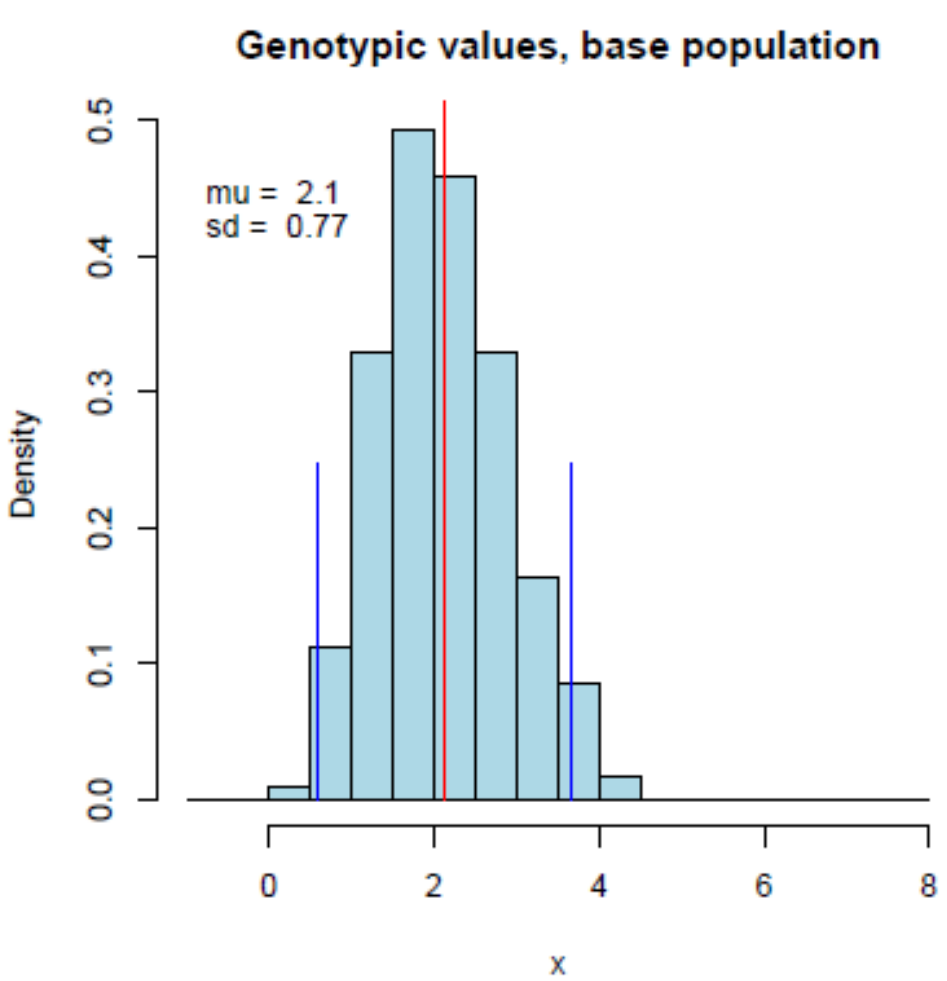
under various assumptions

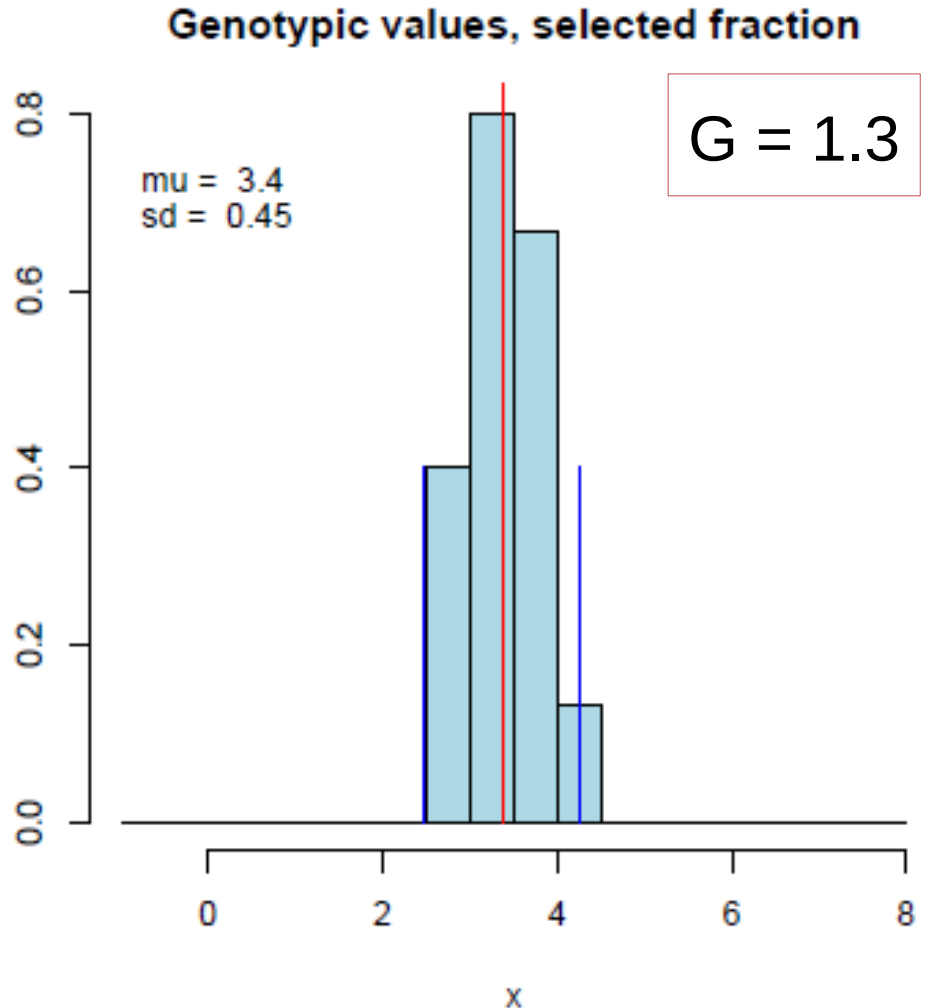
How to judge the efficiency of a breeding scheme?

Expected selection gain

Observed selection gain

Observed selection gain G is defined as difference between the means of the base population and the selected fraction





Expected selection gain

$$G = i \rho_{xy} \sigma_y$$

i = selection intensity (function of sel. fraction)

ρ_{xy} = correlation selection to gain criterion

σ_y = standard deviation in the gain criterion

$$G \uparrow \text{ if } i / \rho_{xy} / \sigma_y \uparrow$$

* Assumptions: Normal distribution of phenotypic values, single stage truncation selection

Expected selection gain: criteria

Selection criterion (x):

Performance mean of a candidate (T) across locations (L), years (Y) and replicates (R)

$\sigma^2_x = \sigma^2_t + \sigma^2_{tl}/L + \sigma^2_{ty}/Y + \sigma^2_{tly}/LY + \sigma^2_e/LYR$

-> Heritability $h^2_x = \sigma^2_t / \sigma^2_x$

Gain criterion (y):

- *Genetic* superiority of the target units for the trait(s) of interest
- Can relate to the total genotypic value (G) or the Additive genetic value (A)

Expected selection gain

$$G = i \rho_{xy} \sigma_y$$

ρ_{xy} can be expressed as the product of

- ρ_{xt} correlation between phenotypic and genotypic value of the test unit = h_x
- ρ_{ty} correlation between the genotypic value of the test and the target unit (eg F_∞-line) = r_A

$r_A = 1$ if test unit = target unit (e.g. DH lines)

$r_A < 1$ if test and target unit are related (eg Testcross)

Expected G: Efficiency

Breeding schemes also differ in costs and duration

-> Criterion to judge the value of a scheme:

G per unit time and costs -> Efficiency

$$Eff = (i\ h_x\ r_A\ \sigma_{A(y)}) / (yr\ €)$$

Eff ↑ if yr/ € ↓ and/or i / h_x / r_A / $\sigma_{A(y)}$ ↑

-> Suitable decision criterion

Model calculations

Model calculations: General idea

Find

for a given breeding scheme
assuming a set of quantitative-genetic
and a set of economic parameters

the combination of allocation parameters*

* = number of candidates, test locations, and replicates at each selection stage

that maximizes the optimization criterion*

* = efficiency

Model calculations: Requirements

- Flow Charts of Breeding scheme(s)
- Estimates of quantitative-genet parameters
- Costs of individual breeding steps

& Optimization software

Model calculations: Requirements

Flow Charts of Breeding scheme(s)

Detailed information on all breeding steps in each season (crossing, multiplication, tests...)

-> derived from breeder's data

-> used to develop the cost function

Model calculations: Requirements

Estimates of quantitative-genetic parameters

- Genetic, G x E and error variances
- Hybrids: Correlation between line and testcross performance

-> derived from actual breeder's data

-> used to calculate all genetic variances and covariances (among / between candidates, phenotypic variance, variance in selection and gain criterion)

Model calculations: Requirements

Costs of individual breeding steps

- Development / multiplication of candidates (Crossing, selfing, DHL-production, ...)
- Field trials (rows, plots, disease tests,...)
- Quality tests

-> derived from actual breeder's data

-> used to calculate the costs of a scheme in the cost function

Model calculations: Cost function

- „Heart“ of the optimization program
- Detailed description of the scheme in a single formula
- Candidate number at first (or last) selection stage is calculated for each set of the other allocation parameters

  -> make full use of the budget
- Allows reliable and meaningful comparisons of alternative breeding schemes

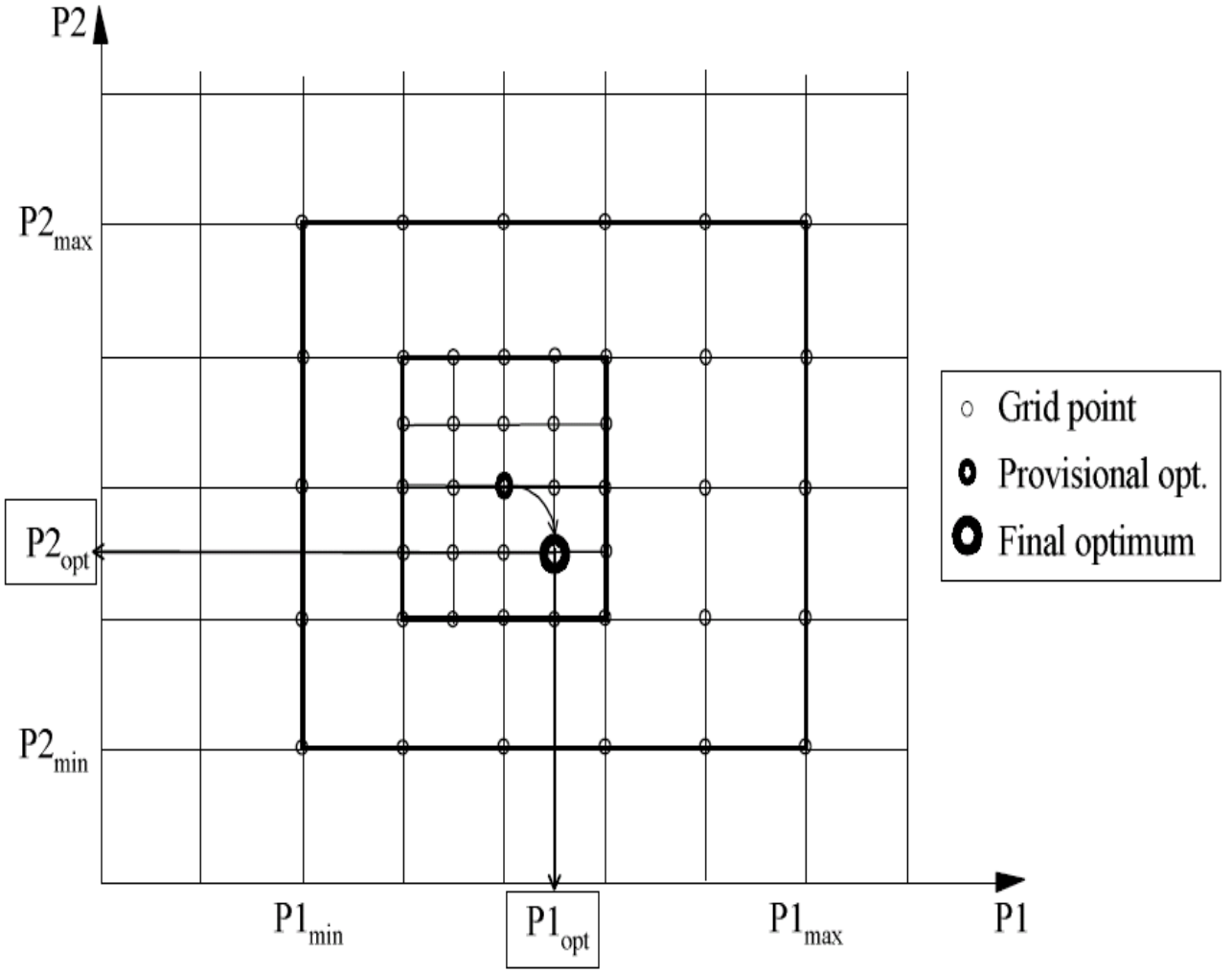
Model calculations: Optimization routine

Read input parameters
(genetic/economic; Min-Max N,(T),L,R; restrictions)

- Define first set of allocation parameter combinations
(covering Min-Max; only meaningful combinations)
- Calculate optimization criteria for 1st set
-> store provisional optimum
- Define new allocation parameter comb. set
(smaller range around prov. optimum for N,T,L,R)
- Calculate optimization criteria ....

... final optimum found -> store results

Optimization: grid search approach



Source: Tomerius, 2001

MC: Varying the parameters

- Find optimum under standard assumptions
- Then vary all / important parameters over
  a wide but meaningful range
- helps to identify parameters with large effect
  (crucial parameters for the breeder)
- gives an idea of the robustness of a scheme
  and identifies changes in the ranking
- measures the stability / reliability of the results
  (approximate measure of error of G -> impossible
  to compute in MC for multi-stage selection)

MC: Possibilities and limitations

☺ allow to optimize breeding schemes *per se* and compare alternative **optimized** breeding schemes

☺ investigate various genetic, economic, practical or even future situations ('what-if')

☺ are cost efficient and fast

BUT

- require some simplifying assumptions
- additional factors may be important in choice of scheme, e.g. simplicity, need for expensive technical facilities

! MC results offer only decision support !

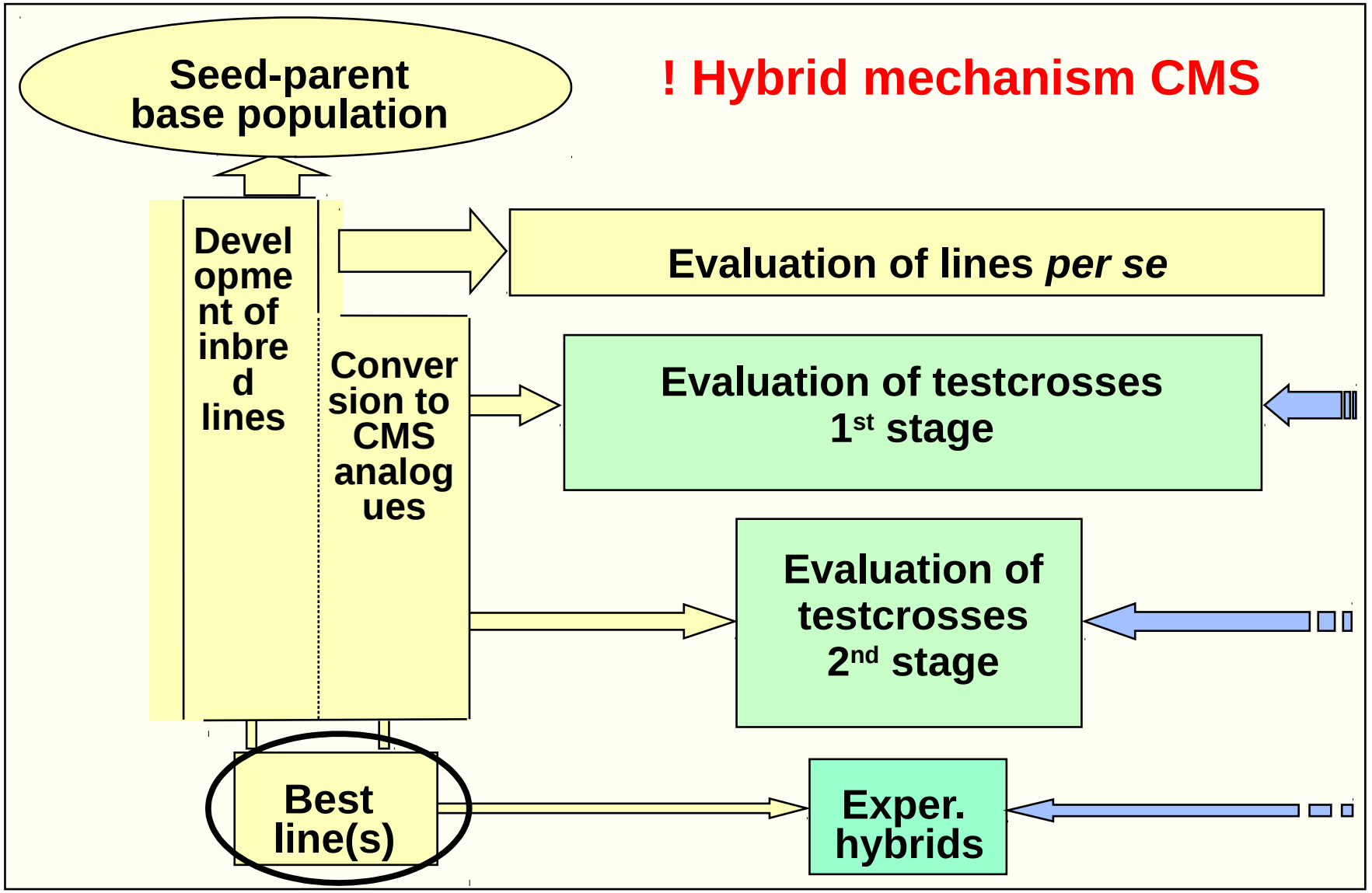
Model Calculations: Examples

Hybrid Rye Breeding

(Tomerius 2001 / 2008)

Hybrid rye breeding

Development of seed parent lines



Source: Tomerius, 2001

Hybrid rye breeding

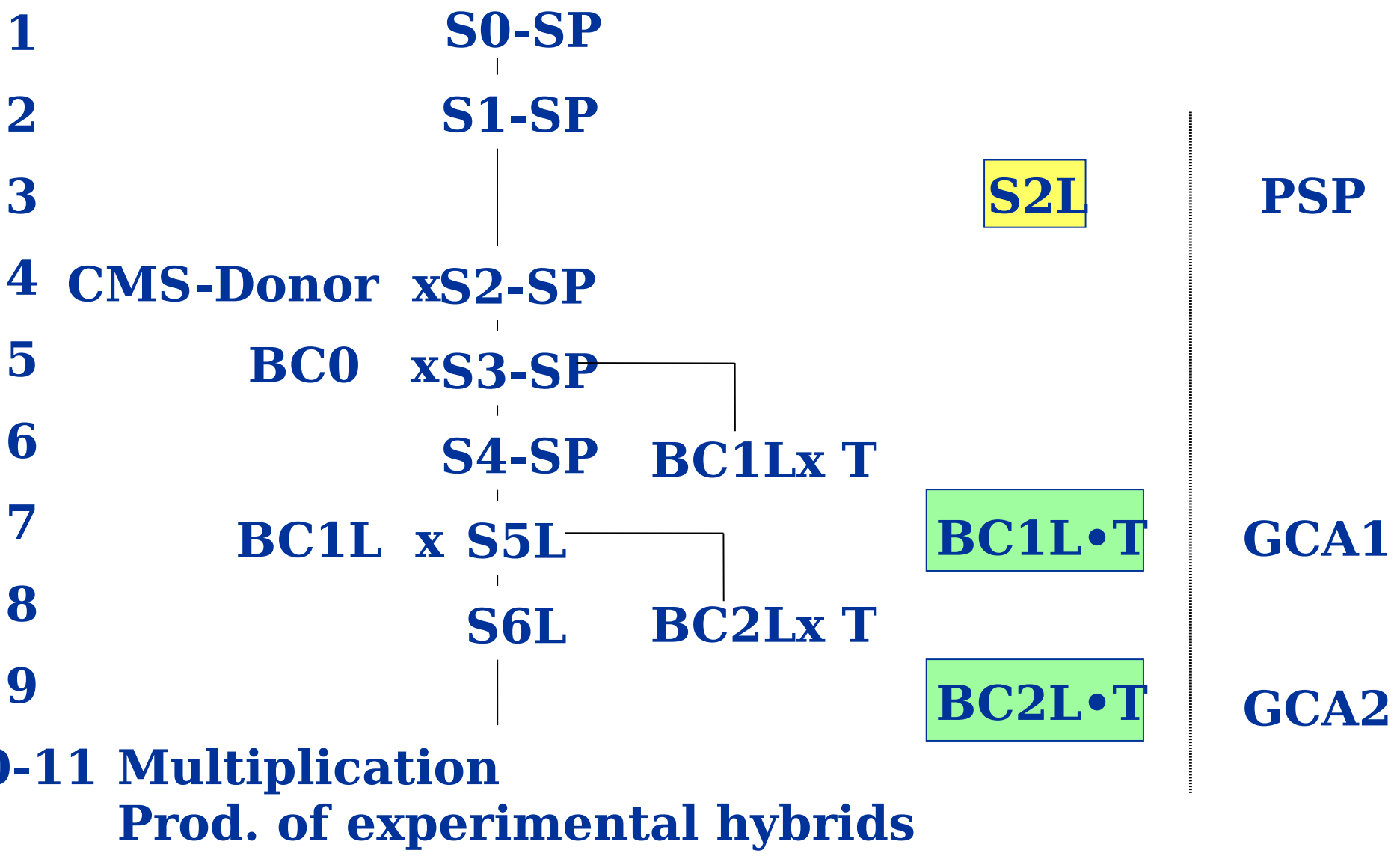
Development of seed parent lines

2 phases in breeding scheme:
- Preselection for per se performance (PSP)
- Selection for General Combining Ability (GCA) to pollinator gene pool

5 breeding schemes differing in
- basic material used
- type of test units
- number of selection stages
- length
- hybrid mechanism used

Hybrid rye breeding

Standard scheme of seed parent line development



Hybrid rye: Assumptions & parameters

- Selection criterion:

  PSP: Index of five agronomic traits

  GCA: same index + grain yield (most important)
- Optimization criterion: Selection gain per year

  in PSP and GCA (weighed 1:3) at a fixed budget
- 3 best lines finally selected
- Estimates of genetic parameters from breeders' data

  (3 breeders) und official trials
- Cost parameters from breeders' calculations (full costs)

Standard set of quantitative-genetic parameters

Parameter	GY [dt ha^{-1}]	PH [cm]	LR [1 - 9]	TKW [g]	FN [s]	BR [1 - 9]
Additive variance	24	46	1.5	7	900	0.9
Dominance var.	12	4	0.15	1	100	0.1
Error var. (PSP)	-	20	1.5	2.4	400	1.2
Error var. (GCA)	12	10	0.7	1.2	200	0.8
V_{GxL} (relative to V_G)	0.15	0.10	0.30	0.10	0.10	0.15
V_{GxY} (relative to V_G)	0.15	0.10	0.15	0.10	0.10	0.10
V_{GxLxY} (rel. to V_G)	1.00	0.30	0.90	0.40	0.40	0.60
Corr. Line -Testcr.	$-^{1}$	0.8	0.9	0.7	0.8	0.8

Costs of breeding activities

Activity	Unit	€ p. unit
Line development and seed multiplication		
• Production of selfed seed (Field / Greenhouse)	1 single plant	3 / 8.75
• Production of crosses (Field / Greenhouse)	1 pair of plants	4 / 17.5
• Production of Doubled Haploid Lines (DHL)	1 fertile DH-plant	22.5
• Male sterility checking	1 candidate	1.1
• Multiplication / crossing in plastic cabins	1 cabin	50
• Production of testcross seed (Topcross)	1 TC-plot	35
• Seed multiplication in small plastic house	1 plastic house	500
• Production of exp. hybrids in isolation plots	1 isolation plot	1000
Evaluation of test units		
• Single row plots	1 row	5
• Large drilled plots	1 plot	20

Calculation of expected gain from multi-stage selection

- "$G = i \rho_{xy} \sigma_y$ " is not valid for multistage G

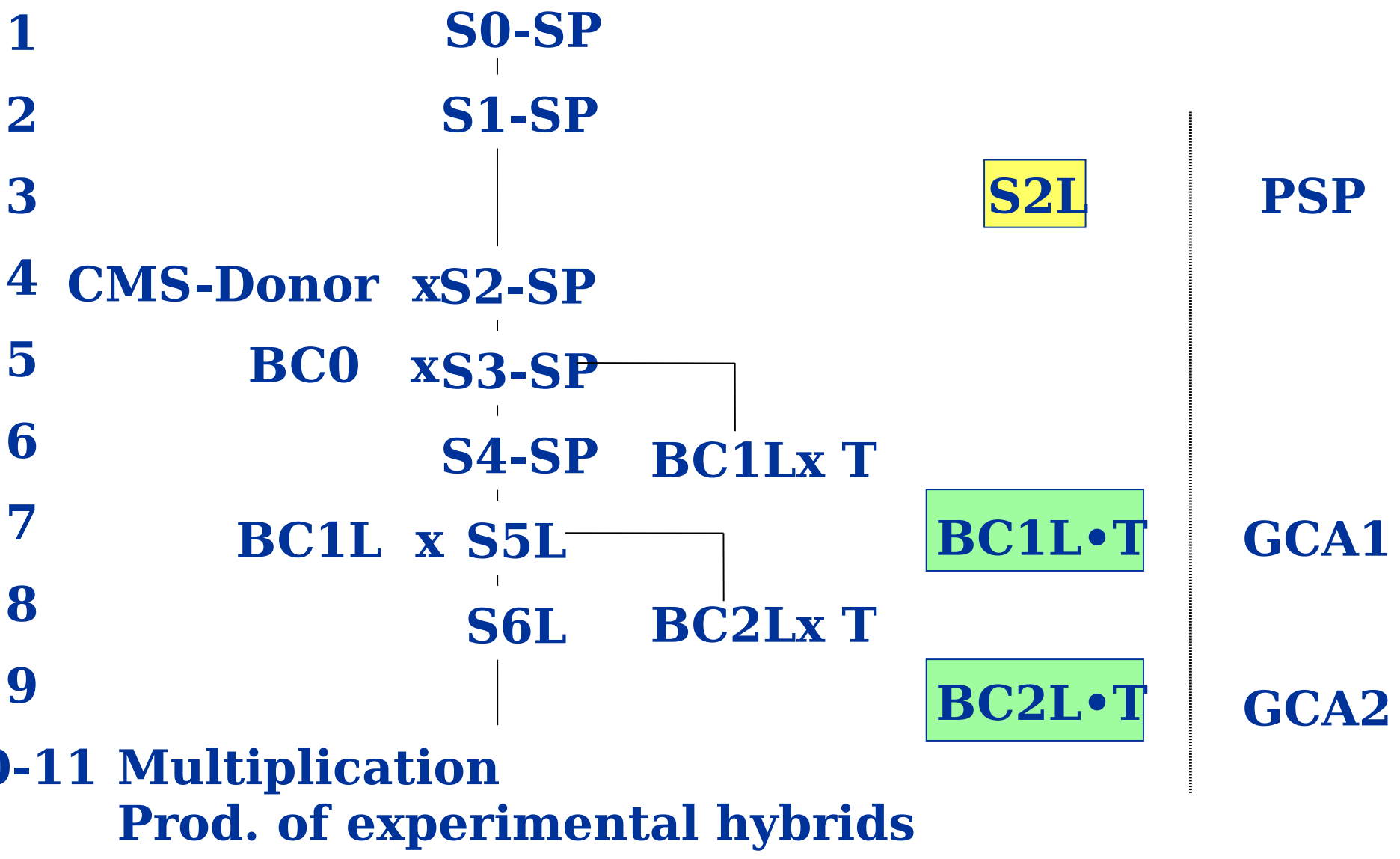
  -> each selection round diminishes genetic variance

  -> remaining candidates are not normally distributed

➔ Detailed formulae by Cochran (1951) resp. Utz (1969)

Hybrid rye breeding

Standard scheme of seed parent line development



Hybrid rye breeding

Optimum of standard scheme under std. assumptions

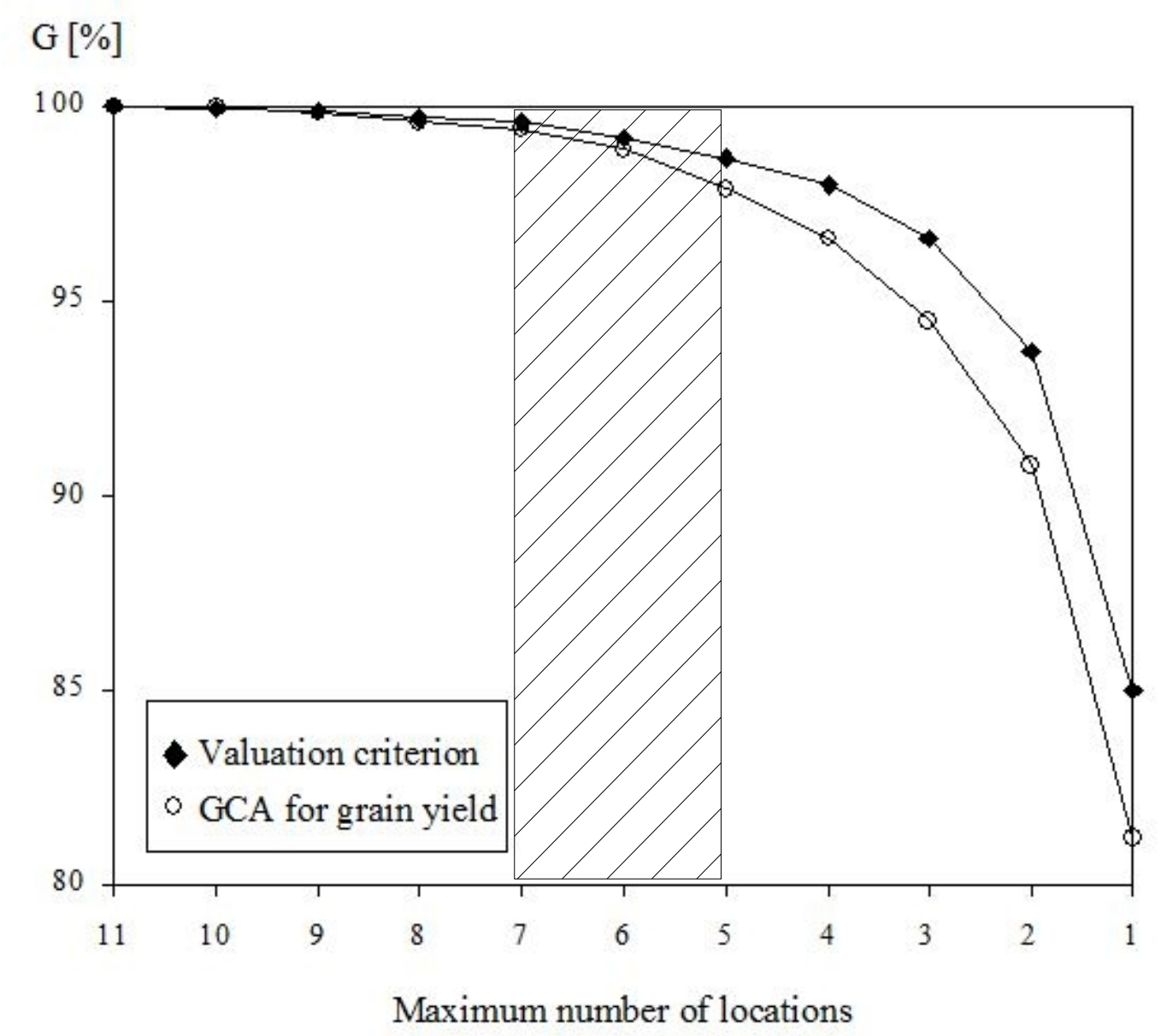
Trial	N	T	L	R	Effic. (%)
PSP	2683	-	3[1]	1[2]	100.0
GCA1	188	1	4	2[2]	
GCA2	21*	3	11	2[2]	

N, T, L, R = Number of candidates, testers, locations, replicates.
PSP, GCA = Selection for per se performance resp. GCA.

[1] maximum value due to limited seed availability [2] fixed values

* 3 finally selected.

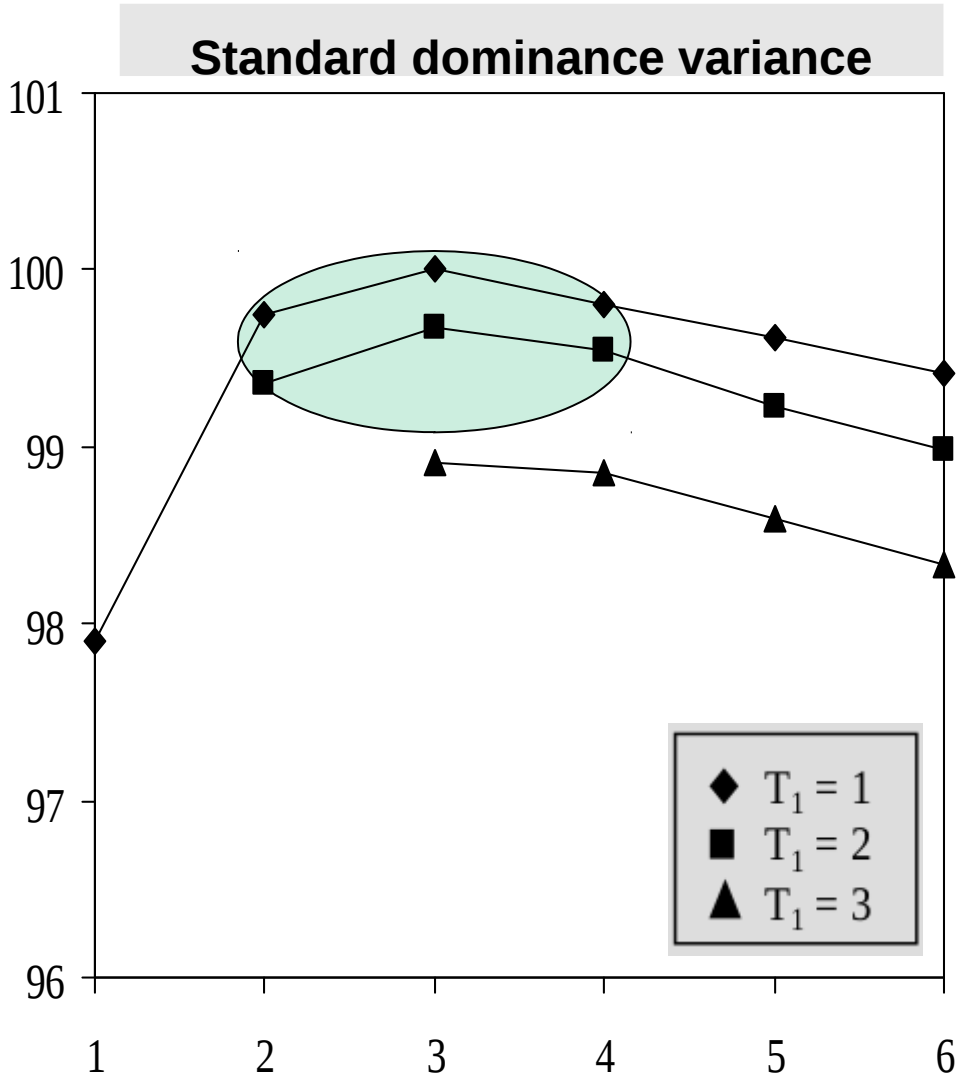
Deviating from the optimum: locations at the last selection stage

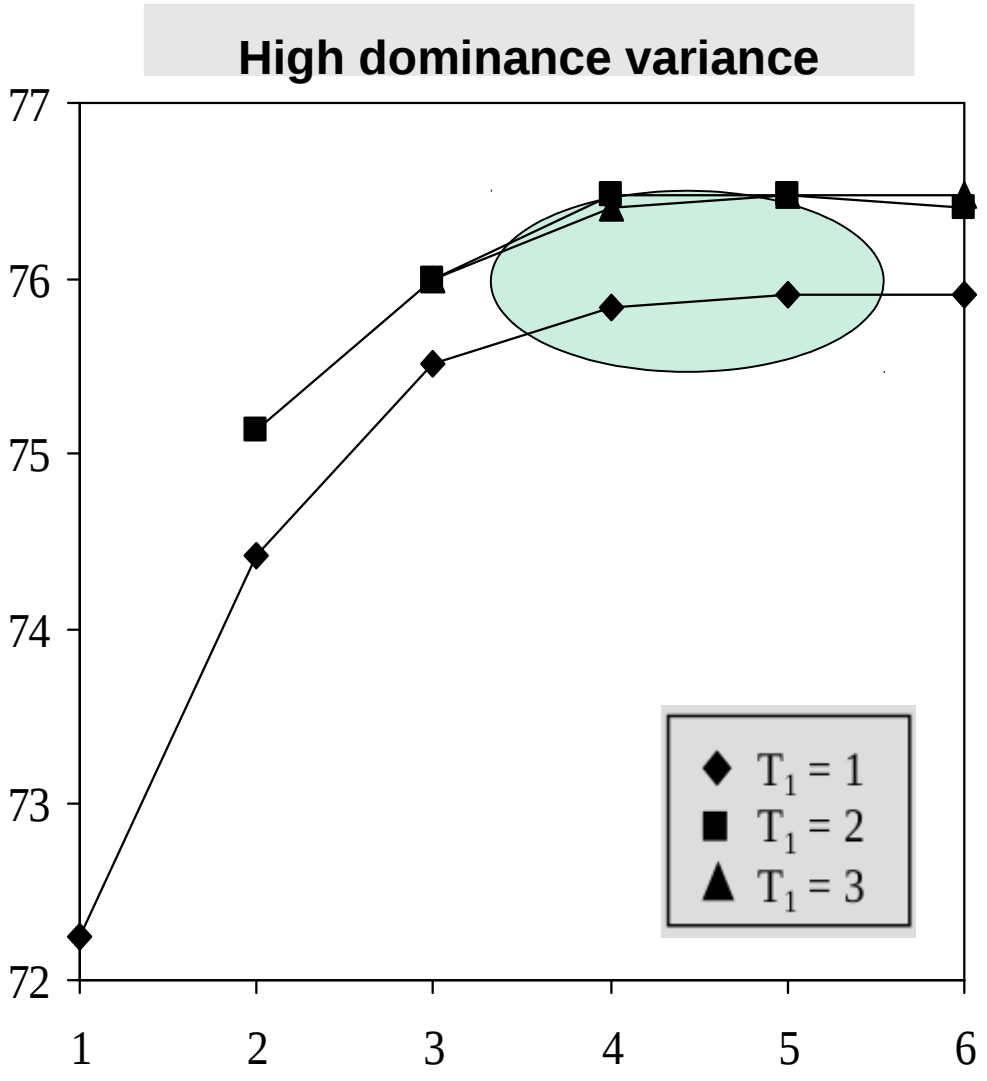


Influence of the dominance variance

Scheme	Stage	N	T	L	Effic. (%)
STD	PSP	2683	-	3	100.0
	GCA1	188	1	4	
	GCA2	21	3	11	
σ^2_D / 2	PSP	2798	-	3	112.2
	GCA1	198	1	4	
	GCA2	20	2	14	
σ^2_D * 2	PSP	2689	-	3	76.9
	GCA1	144	2	3	
	GCA2	19	5	9	

Optimum number of testers assuming standard / high dominance variance





Potential of shortening the breeding scheme by new technologies

Use of doubled haploids:

+ Shortens the scheme by one year
+ Full variance between candidates
- CMS-conversion remains necessary

Use of a gametozide:

+ Shortens the scheme by two years
+ Simplification of the scheme
+ Early testing on GCA possible

☹ both technologies not practicable to date

Potential of shortening the breeding scheme by new technologies

Scheme	Stage	N	T	L	Effic. (%)
STD	PSP	2683	-	3	100.0
11 years	GCA1	188	1	4	
	GCA2	21	3	11	
DHL	PSP	937	-	3	107.7
10 years	GCA1	125	1	5	
	GCA2	18	3	11	
GAM	PSP	2151	-	2	131.4
9 years	GCA1	281	1	4	
	GCA2	14	3	12	

Proportion of budget spent on different breeding operations

Breeding operation	Breeding scheme				
	CYC1_11	CYC1_21	POP2_11	DHL1_11	GAM1_11[1]
Inbred line product.[2]	26.7	25.8	30.3	56.8	18.0
Line *per se* evaluation	27.2	28.9	24.5	7.0	10.8
TCP production	7.8	8.2	7.5	3.2	7.5
TCP evaluation	30.8	31.1	30.2	26.0	34.0
EH prod.& line multipl.[3]	7.5	6.0	7.5	6.8	19.8

[1] Ten percent of the budget are spent on the gametocide. [2] Including production of CMS analogues. [3] Multiplication of finally selected seed-parent lines.

Conclusions from hybrid rye example

- Alternative breeding schemes differ in their efficiency
- Optimum dimensioning depends on genetic (and economic) parameters
- Small deviations from the optimum have no severe consequences (optima are flat)
- Shortening the breeding scheme increases gain -> new technologies, better organisation
- Increase of budget increases selection gain, but increase of gain is much lower *(not shown here)*
- Choice of more efficient scheme often much more effective than a budget increase *(not shown here)*

Simulation studies

Simulation studies: Possibilities

Breeding simulation studies
"provide a valuable tool for breeders to efficiently use the wide spectrum of genetic data and information available“

- allow definition of complicated genetic models (multiple alleles, pleiotropy, epistasis, GxE)
- allow to compare alternative breeding schemes
- allow to predict cross performance using known genetic information
- allow to optimize MAS / use of identified QTL

Simulation studies: Requirements

1. Information on the breeding scheme(s)

- seed propagation type (self, cross)
- selection stages and selection type
- virtual field design (L,R)
- selected fractions
- selection mode (top, bottom)

2. Information on the traits of interest

- Gene number and genetic values
- pleiotropic effects
- GxE-interaction effects
- Genetic model(s) investigated
- evtl. genetic map

➔ Obtained from real breeders' data, if possible

Simulation studies: Limitations

Require data and / or assumptions regarding the genetics of the traits under selection

(main problem: yield – not problematic with marker maps)

- Dimensioning (N, L, R) and selected fractions are often not optimized
- Costs are often not really accounted for

! Also SIM results offer only decision support !

Simulating breeding programs

Using the software package
“SelectionTools”

“SelectionTools” software

- is a collection of software from several research projects
- can be downloaded for free (incl. tutorial and examples)

  http://fb09-pg-s207.agrar.uni-giessen.de/~frisch-m/
- mostly based on R (with some C code)
- can be used for different topics
  - Genetic diversity analysis
  - Genetic simulation of breeding programs
  - Simulation of marker-assisted backcrossing
  - Genomic Selection

“SelectionTools” for simulating breding programs

Input is a data set with marker data, linkage map, and field (phenotypic) data

- Genetic architecture is estimated by genome wide prediction model
- Arbitrary genetic trait architecture possible (no assumption of normal distribution required)
- Model can be extended to plan number of locations, years, replications in field trials

“SelectionTools”: predicting selection gain from one generation of phenotypic selection

- consider (large) population of inbred lines
- phenotypic data, marker data and linkage map available
- inbred lines tested in field trial with $h^2 = 0.8$
- best 30 lines are selected
  - estimate single-stage gain from selection
  - later different values for h^2 / size of selected fraction

“SelectionTools”: predicting selection gain from one generation of phenotypic selection

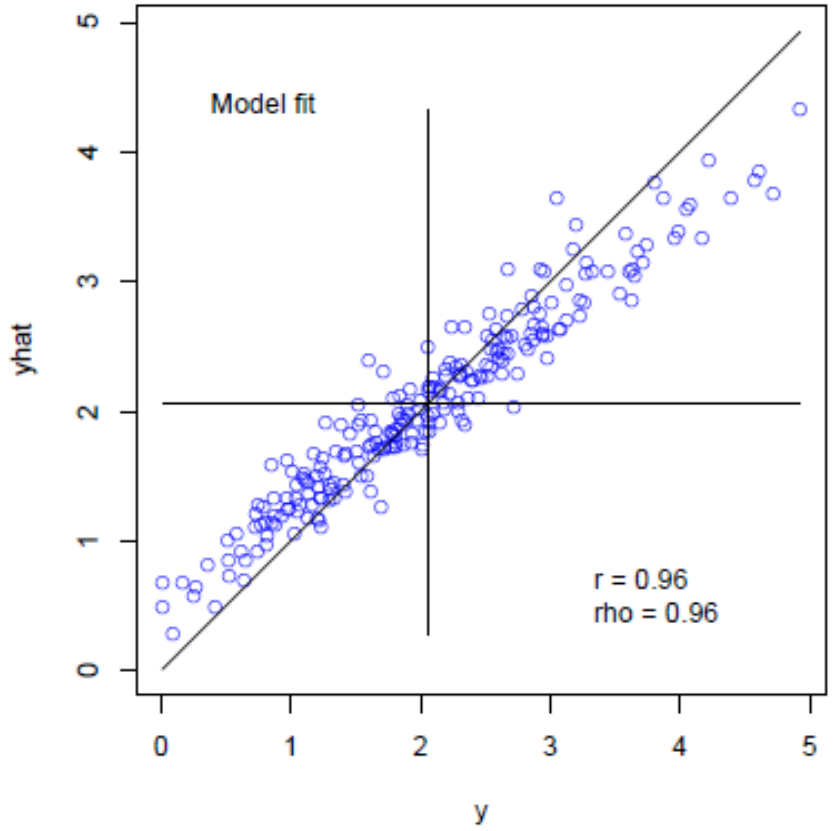
1. estimate marker effects (ridge regression) and check the model fit

“yhat” are estimated genotypic values of base population

-> used for simulations

- correlation r must be high

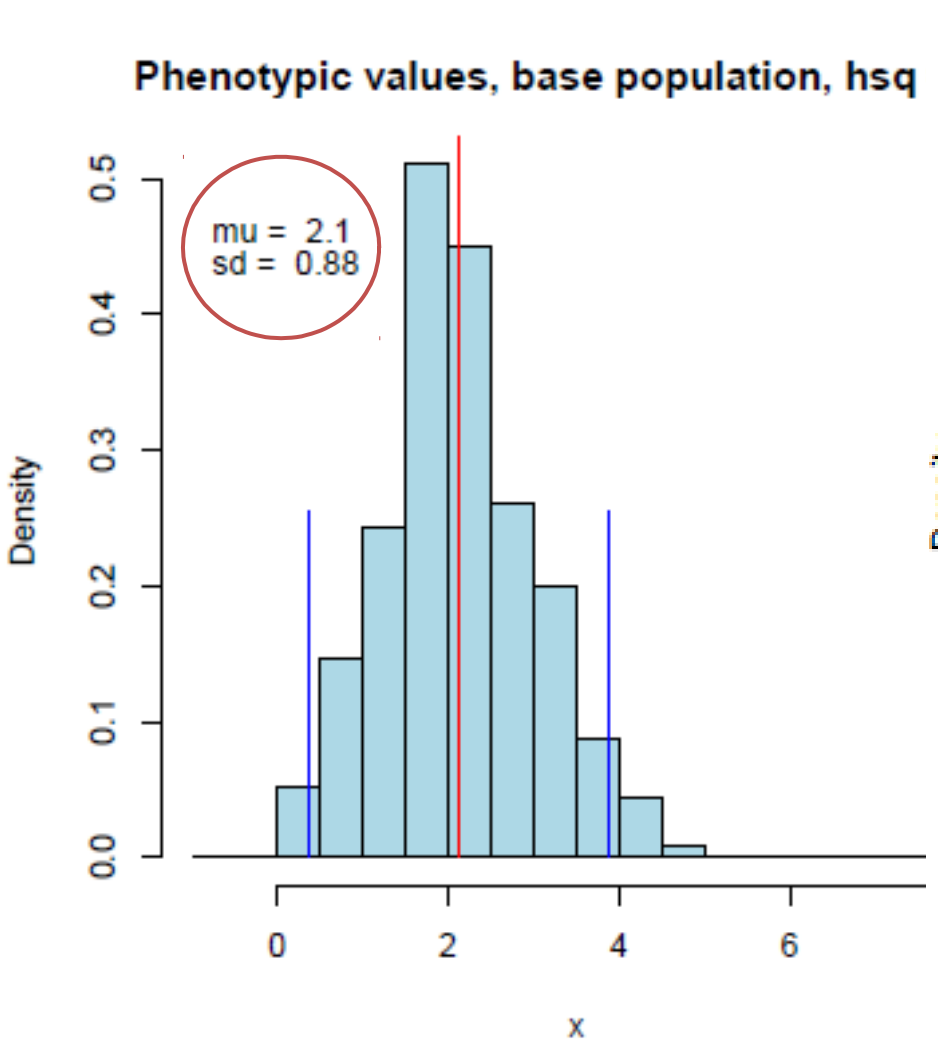
(if not, markers do not explain phenotypes well)

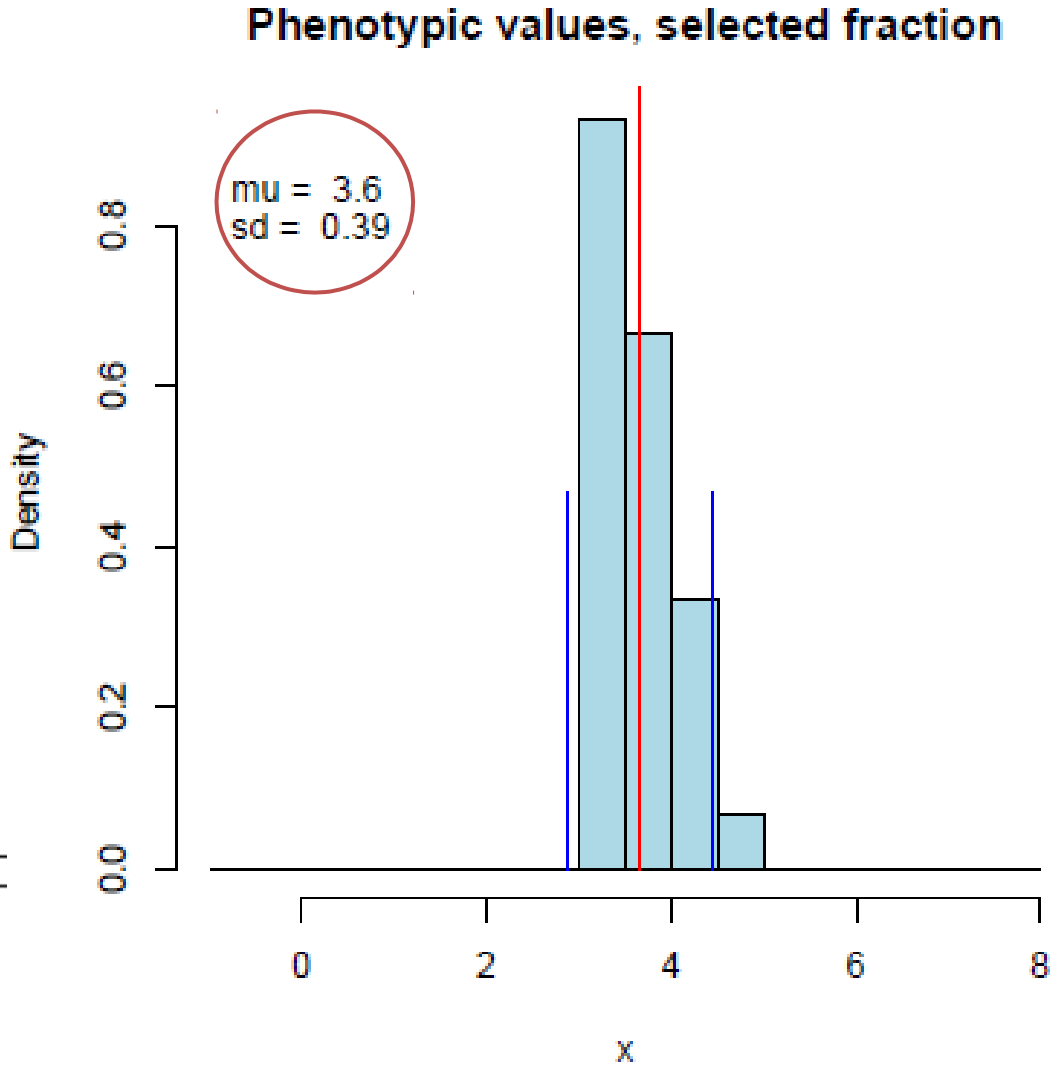


“SelectionTools”: predicting selection gain from one generation of phenotypic selection

2. Initialize simulation routines
3. Calculate genotypic value of all individuals
4. - simulation generates marker data matrix
5. - marker effects list used to calculate genotypic value
6. - return genotypic values
7. Calculate phenotypic values by adding a random realization of the masking variance
8. (deduced from h^2 as $s^2m = s^2g/h^2 – s^2g$)
9. Sort individuals by phenotypic values (desc/asc)
10. and plot phenotypes of all *vs.* selected individuals

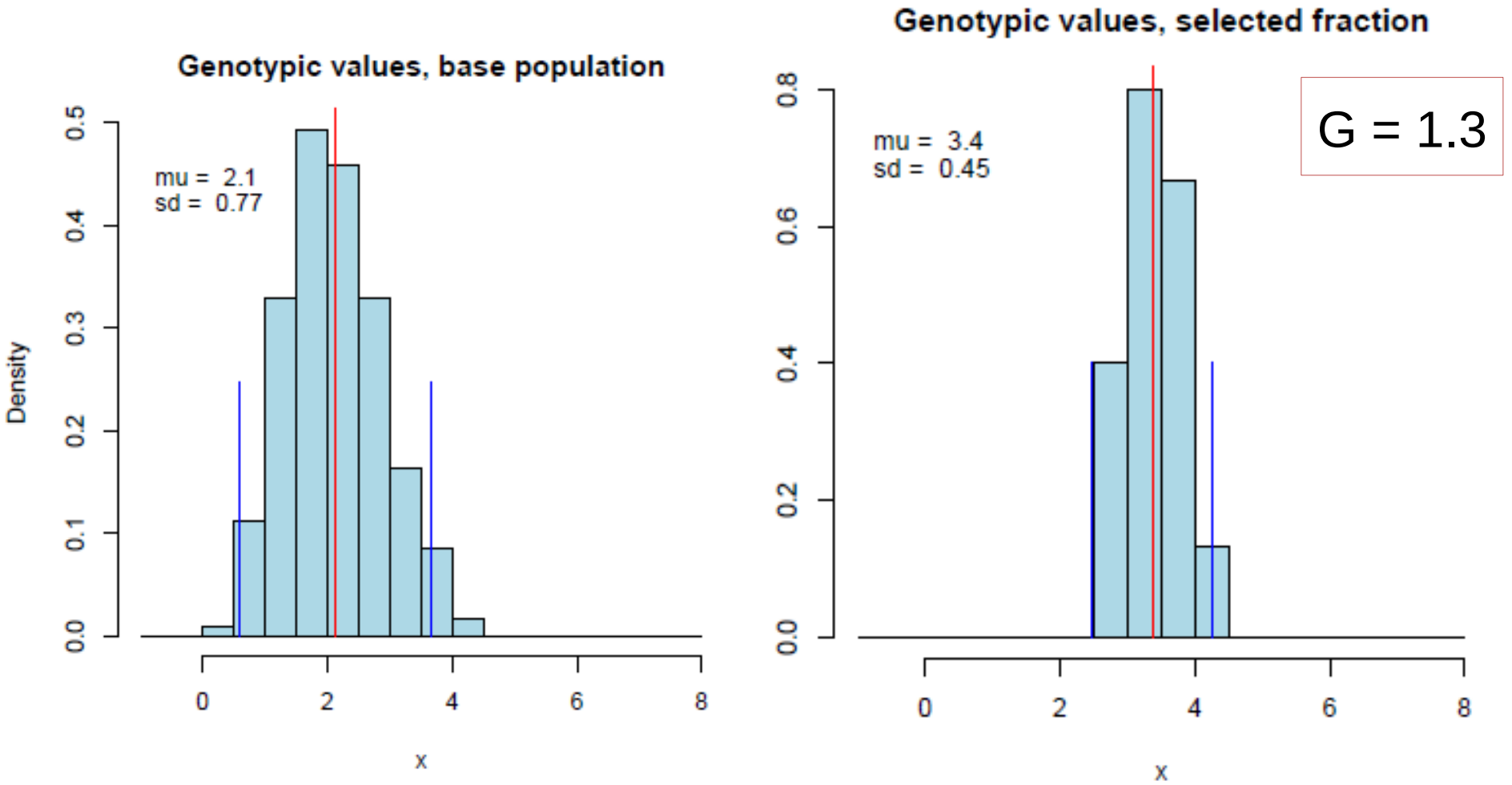
“SelectionTools”: predicting selection gain from one generation of phenotypic selection





“SelectionTools”: predicting selection gain from one generation of phenotypic selection

6. Get genotypic values of selected fraction
7. Calculate realized selection (G = mu_sel – mu_base)



“SelectionTools”: predicting selection gain from one generation of phenotypic selection

8. Study different h^2 and selected fractions in simulation

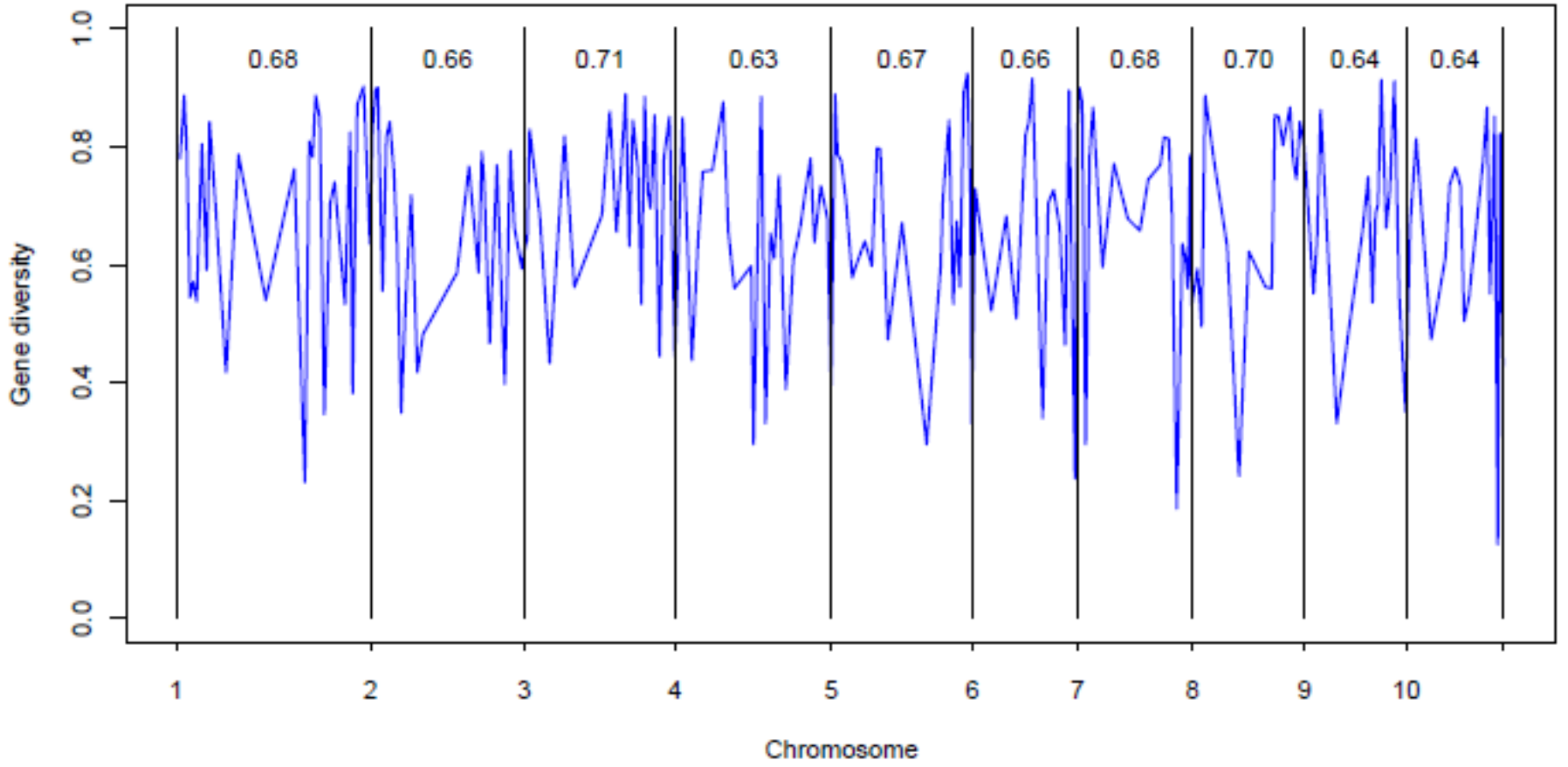
```
nsel <- c( 100, 50, 30, 10 )
h.sq <- c( 0.6, 0.7, 0.8, 0.9, 0.99 )
nrep <- 100
```

➢

```
> SG
          0.6       0.7       0.8       0.9      0.99
100 0.5511241 0.6132195 0.6430198 0.671797 0.7042547
50  0.8450615 0.9334265 0.9956299 1.040969 1.0878725
30  1.0577680 1.1767903 1.2070169 1.283798 1.3414516
10  1.4019869 1.5266215 1.5498781 1.656200 1.7221028
```

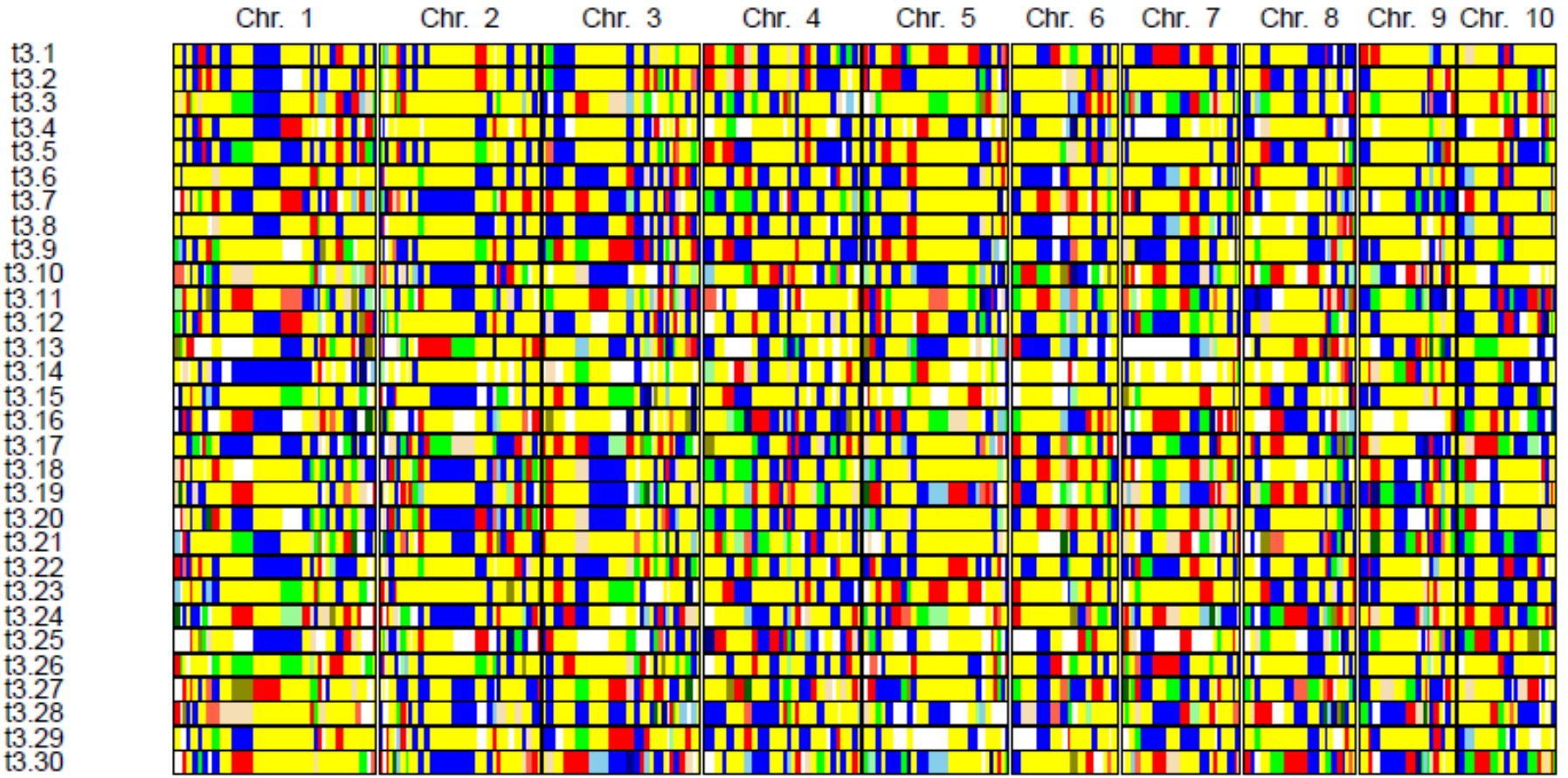
“SelectionTools”: further visualization of results

Plotting gene diversity along the chromosome



“SelectionTools”: further visualization of results

Plotting graphical genotypes of selected fraction



General Conclusions

For the optimization of breeding plans,

- MC / SIM are valuable support decision tools
- Optimization will become more important with increasing amount of genetic information
- Advances in genomics will help to build more realistic genetic models -> combination of MC/SIM interesting
- MC / SIM can not only confirm breeders' intuitive experience, but can also find out facts which breeders did not realize before

Discussion

- Gain from selection is only one parameter to judge a breeding scheme; strictly speaking applies better to recurrent population improvement

- Another suitable criterion:

  Probability of identifying superior genotypes [P(q)]

  - no reference to the mean of the selected group
  - depends on heritability and selection intensity, too
  -> Positively correlated with G

  Knapp 1998: Marker-assisted selection as a strategy for increasing the probability of selecting… Crop Sci 38:1164-1174

References

Tomerius, A.-M. 2001. Optimizing the development of seed-parent lines in hybrid rye breeding. Diss. Uni Hohenheim. Full text pdf available: opus.ub.uni-hohenheim.de/volltexte/2001/10/pdf/tomerius.pdf

Tomerius, A.-M., T. Miedaner, H.H. Geiger. 2008. A model calculation approach towards the optimization of a standard scheme of seed-parent line development in hybrid rye breeding. Plant Breeding 127(5):433–440.

Frisch, M. SelectionTools tutorial http://fb09-pg-s207.agrar.uni-giessen.de/~frisch-m/SelectionTools-tutorials.pdf

Thank you for listening!

AIB Dr. Alexandra Tomerius
Plant Breeding Consulting
Saffeweg 32
D-38304 Wolfenbuettel

aib@ing-tomerius.de
www.ing-tomerius.de