Software-supported optimization of breeding plans

Dr. Alexandra Tomerius

Plant breeding consulting

Overview

- Situation of the breeder
- Model calculations
  - Hybrid winter rye (Tomerius, 2001 / 2008)
  - Winter wheat (Tomerius, unpublished)
- Breeding Simulation models
  - Parental selection in wheat (Wang et al., 2005)
- Conclusions and outlook

Situation of the breeder

"In planning a strategy aiming at the development of new, superior varieties, a breeder has to decide on the breeding scheme to be used as well as on the dimensioning of every breeding step involved."

F.W.Schnell, 1996
(Vortr. Pflanzenzuechtg. 33)

Situation of the breeder

Different possible breeding schemes exist

- **generally very complex**
- **may differ remarkably in efficiency**
- **often 'historical' reasons**

➔ **Breeder has to find the best possible scheme**

➔ **Experimental comparison hardly feasible**

➔ **Improvements based on experience and theoretical findings / expectations**

➔ **Judgement of efficiency often indirect**

AIB

Expected selection gain

Suitable criterion to judge the efficiency:

Expected selection gain $G = i\,\rho_{xy}\,\sigma_y$ *

per unit time and costs $G = (i\,\rho_{xy}\,\sigma_y) / (yr\ €)$

* i = selection intensity
ρ_{xy} = correlation between selection and gain criterion
σ_y = standard deviation in the gain criterion

G can be increased by reducing time and costs and / or increasing i, ρ_{xy}, σ_y

Components of the expected selection gain

Budget & restrictions

Number of candidates → **Selection intensity** i

Number of locations, years, replicates → **Heritability** (ρ pheno- / genotype) h^2

Genetic variances and covariances

- type(s) of candidates → **Heritability**; **Genetic correlation candidate / target unit** ρ_{ty}
- trait(s), base pop. → **Variance in gain criterion** σ^2_y

$$G = i\ h\ \rho_{ty}\ \sigma_y$$

Software supported optimization

- **Model calculations** and **breeding simulation studies** allow **a priori**-judgement by predicting

  the relative gain (= efficiency) and
  the respective optimum variant(s)
  of a virtual number of breeding schemes
  under various assumptions

→ MC/Sim are a **helpful tool for breeders** !

→ MC/Sim have different strenghts / limitations

Model calculations

Model calculations: General idea

Find

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

the combination of allocation parameters that maximizes the optimization criterion chosen.

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

Optimization criterion = expected selection gain <u>per year</u> under the restriction of a <u>fixed annual budget</u>

Model calculations: Requirements

1. Estimates of quantitative-genetic parameters
   - Genetic, genotype x environment interaction and error variances
   - Hybrids: Corr. line-testcross performance

-> derived from actual breeder's data

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

Model calculations: Requirements

2. Costs of individual breeding steps
- Development / multiplication of candidates (Crossing, selfing, DHL-production, ...)
- Field trials (rows, plots, special trials)

-> derived from **actual breeder's data**

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

Model calculations: Cost function

- Describes all breeding steps required to develop, propagate, and test the candidates at each selection stage
  -> detailed description of the scheme
- Candidate number at first (or last) selection stage is calculated from the cost function for each set of the other allocation parameters
  -> full use of the budget
- Allows reliable and meaningful comparisons of alternative breeding schemes

Model calculations: Possibilities

Model calculations

- **allow to optimize breeding schemes *per se***
- **and conclusively compare alternative <u>optimized</u> breeding schemes**

  for various genetic, economic, practical or even future situations ('what-if')
- **are highly relevant for practical breeding**

  -> estimates relate to concrete situation(s)
  -> practical requirements can be dealt with
- **are cost efficient and fast**

Model calculations: Limitations

Model calculations

- **require some simplifying assumptions**

  - genetic model (GPE, gene frequencies ~0.5)
  - limited number of traits, index selection

-> keep in mind when interpreting the results!

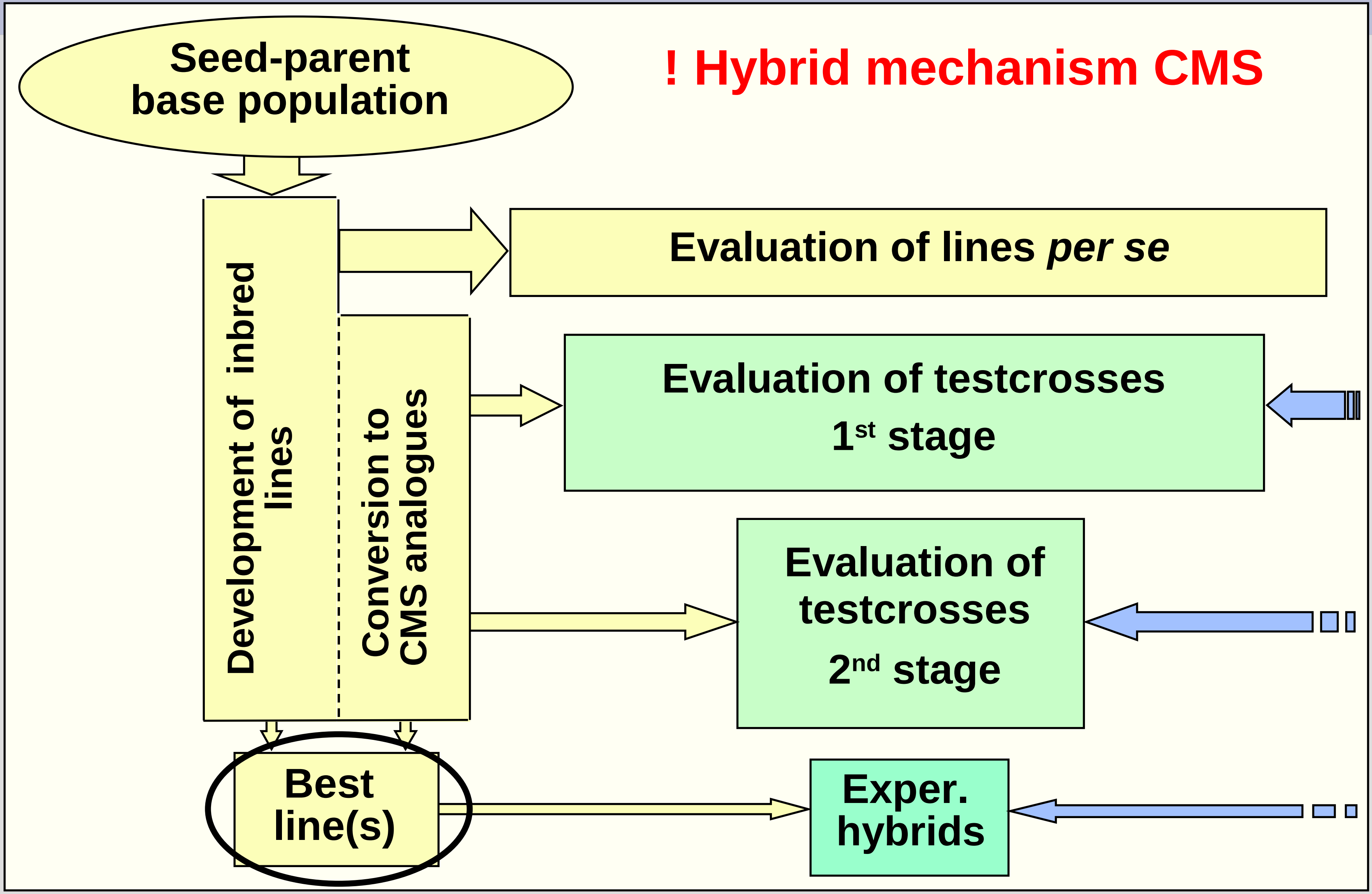
- **additional factors may be important in choice of scheme, e.g. simplicity, need for expensive technical facilities**

! MC results offer only <u>decision support</u> !

MC: Examples

- Hybrid Rye Breeding
- Wheat Breeding

Example 1: Hybrid rye breeding
Development of seed parent lines



Source: Tomerius, 2001

Example 1: Hybrid rye breeding
Development of seed parent lines

- Breeding scheme divided into 2 phases:
  - 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

Source: Tomerius, 2001

Hybrid rye: Standard scheme of seed parent development

						Field trial
1			S0-SP			
2			S1-SP			
3					S2L	PSP
4	CMS-Donor	x	S2-SP			
5	BC0	x	S3-SP			
6			S4-SP	BC1Lx T		
7	BC1L	x	S5L		BC1L•T	GCA1
8			S6L	BC2Lx T		
9					BC2L•T	GCA2
10-11			Multiplication Prod. of experimental hybrids			

Source: Tomerius, 2001

Hybrid rye: Assumptions & parameters

- Selection criterion: Selection index of five agronomic traits + grain yield (testcrosses)
- Optimization criterion: Selection gain per year in PSP and GCA (weighed 1:3)
- 3 best lines finally selected
- Budget: € 200,000 per year
- Estimates of genetic parameters from breeders' data (3 breeders) und official trials
- Cost parameters from breeders' calculations

Source: Tomerius, 2001

Hybrid rye: Standard set of quantitative-genetic parameters

Parameter	GY	PH	LR	TKW	FN	BR
	[dt ha^{-1}]	[cm]	[1 - 9]	[g]	[s]	[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

Source: Tomerius, 2001

Hybrid rye: Costs of individual breeding activities

Activity	Unit	€ p. unit
<u>Line development and seed multiplication</u>		
• 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
<u>Evaluation of test units</u>		
• Single row plots	1 row	5
• Large drilled plots	1 plot	20

Source: Tomerius, 2001

Optimization routine

- Read input parameters
  (genetic/economic; Min-Max N,(T),L,R; restrictions)
- Define first set of allocation p. 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

Calculation of the expected gain from multi-stage selection

- Formula $G = i\ \rho_{xy}\ \sigma_y$ 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)

- Multistage G depends on σ_y and
  - the fractions selected at the individual stages
  - the correlations between the selection criterion at each selection stage and the gain criterion
  - the correlations between the selection criteria used at the different selection stages

Optimum of the standard scheme under standard assumptions

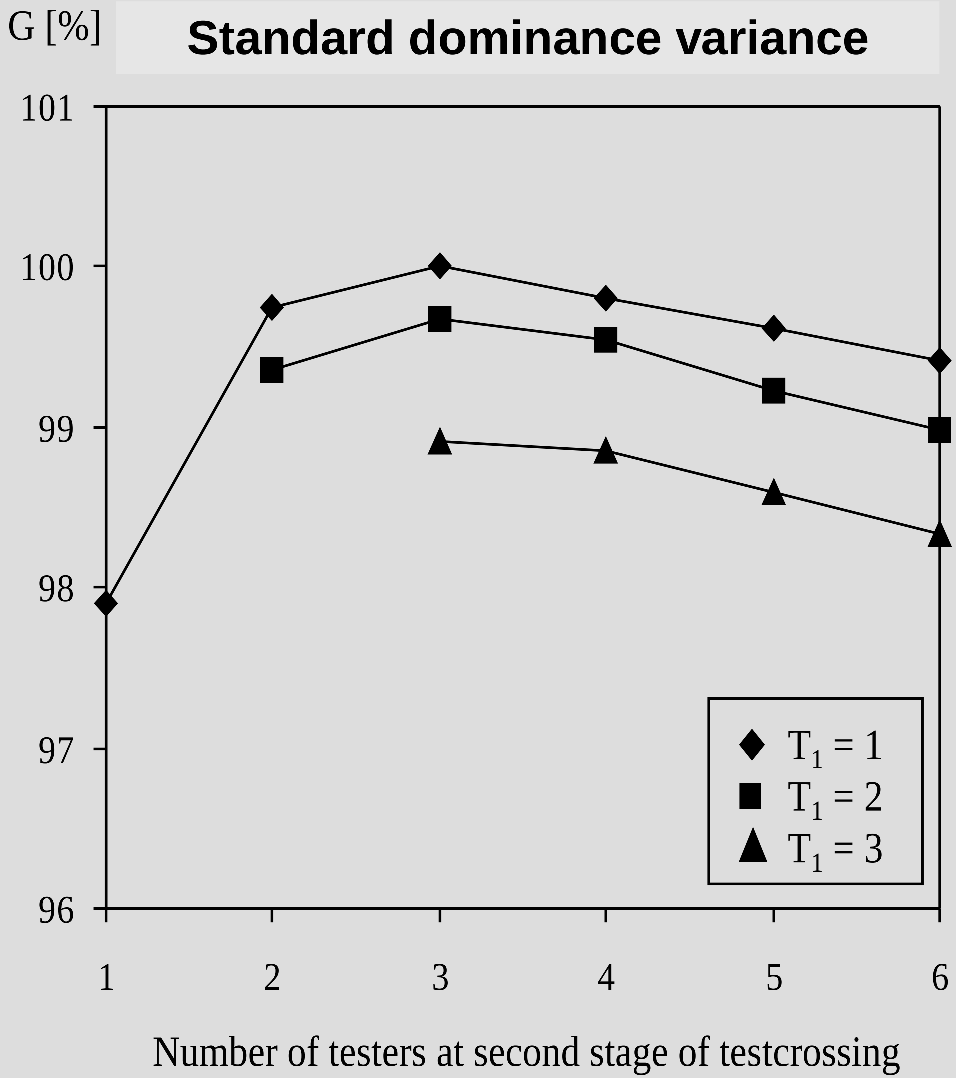
Trial	N	T	L	R	OptC (%)
PSP	2683	-	3[1]	1[2]	100.0
GCA1	188	1	4	2[2]	
GCA2	21*	3	11	2[2]	

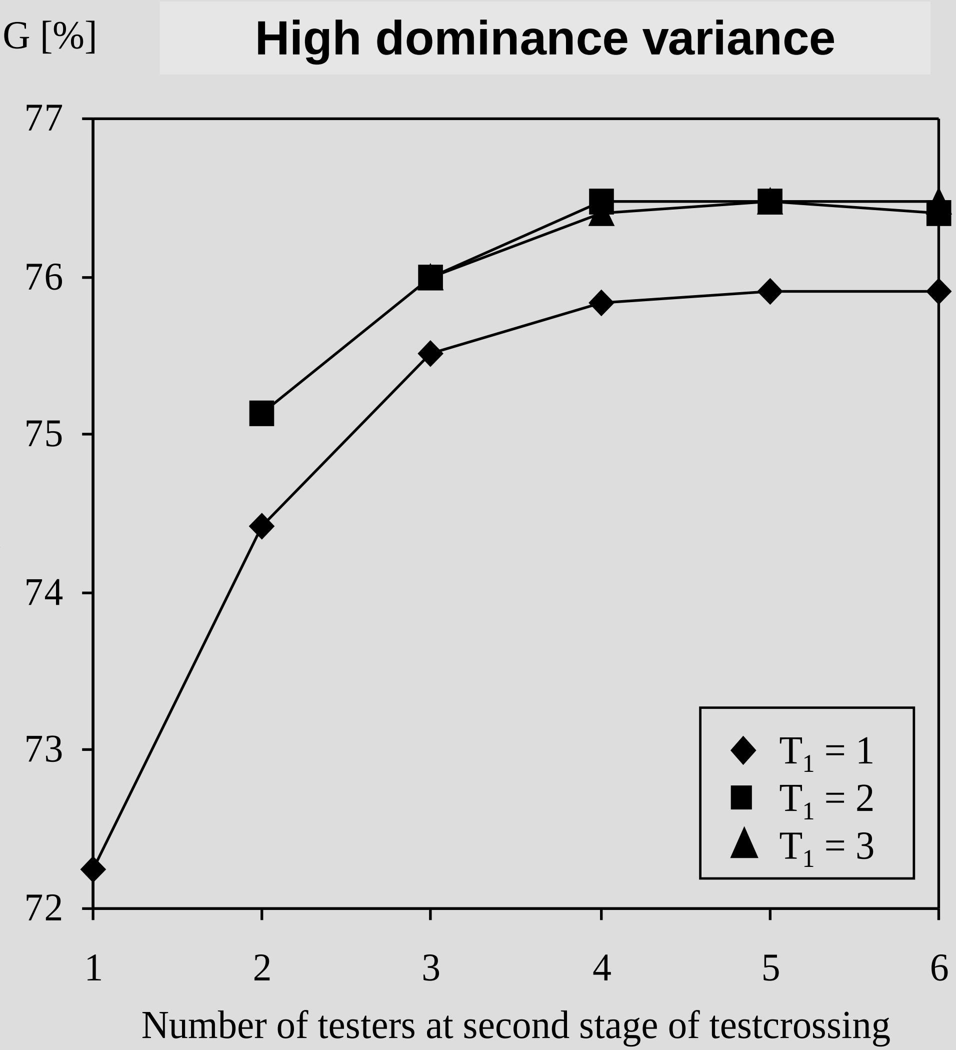
N, T, L, R = Number of candidates, testers, locations, replicates.
OptC (%) = Optimization criterion (relative) = Gain per year.
PSP, GCA = Selection for per se performance resp. GCA.

[1] maximum value due to limited seed availability [2] fixed values
* 3 finally selected.

Source: Tomerius, 2001

Influence of the number of testers for standard and high dominance variance





Source: Tomerius, 2001

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

Source: Tomerius, 2001

AIB

Standard scheme vs. schemes shortened by new technologies

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

Source: Tomerius, 2001

Conclusions from hybrid rye example

- Alternative breeding schemes differ clearly 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 the selection gain markedly
  -> new technologies, better organisation

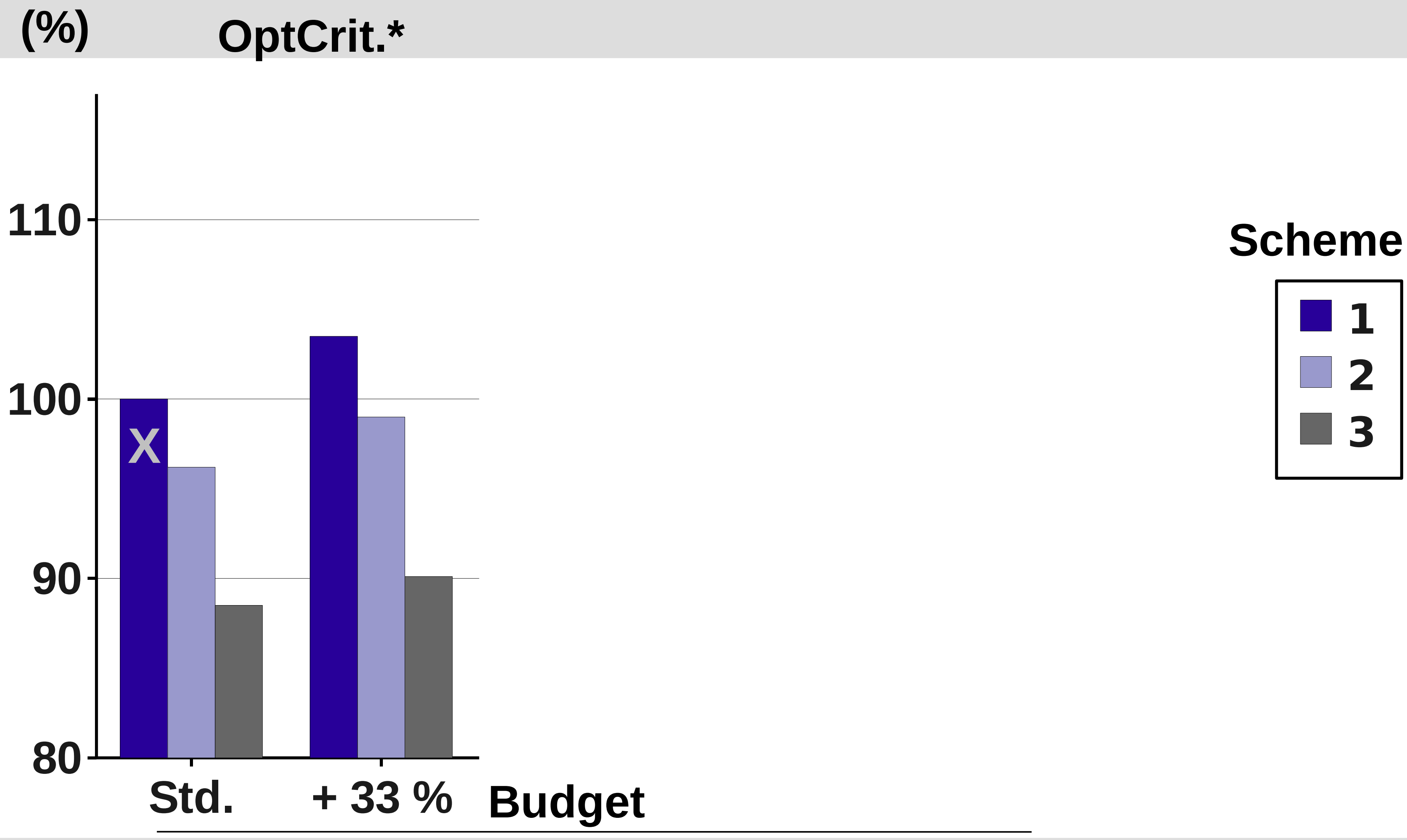
Example 2: Winter wheat

- <u>Goal</u>: Optimize variety (= line) development -> gain criterion is always line performance
- <u>3 breeding schemes:</u>
  - Single Seed Descent (SSD)
  - F2-Pedigree
  - F3-Pedigree
- <u>2 phases</u>:
  - Preselection for 3 'Non-yield-traits' (NYT)
  - Field trials for NYT + grain yield (GY)
- <u>Optimization criterion</u>: selection gain per year in NYT and GY (weighed)

Winter wheat: Parameters varied

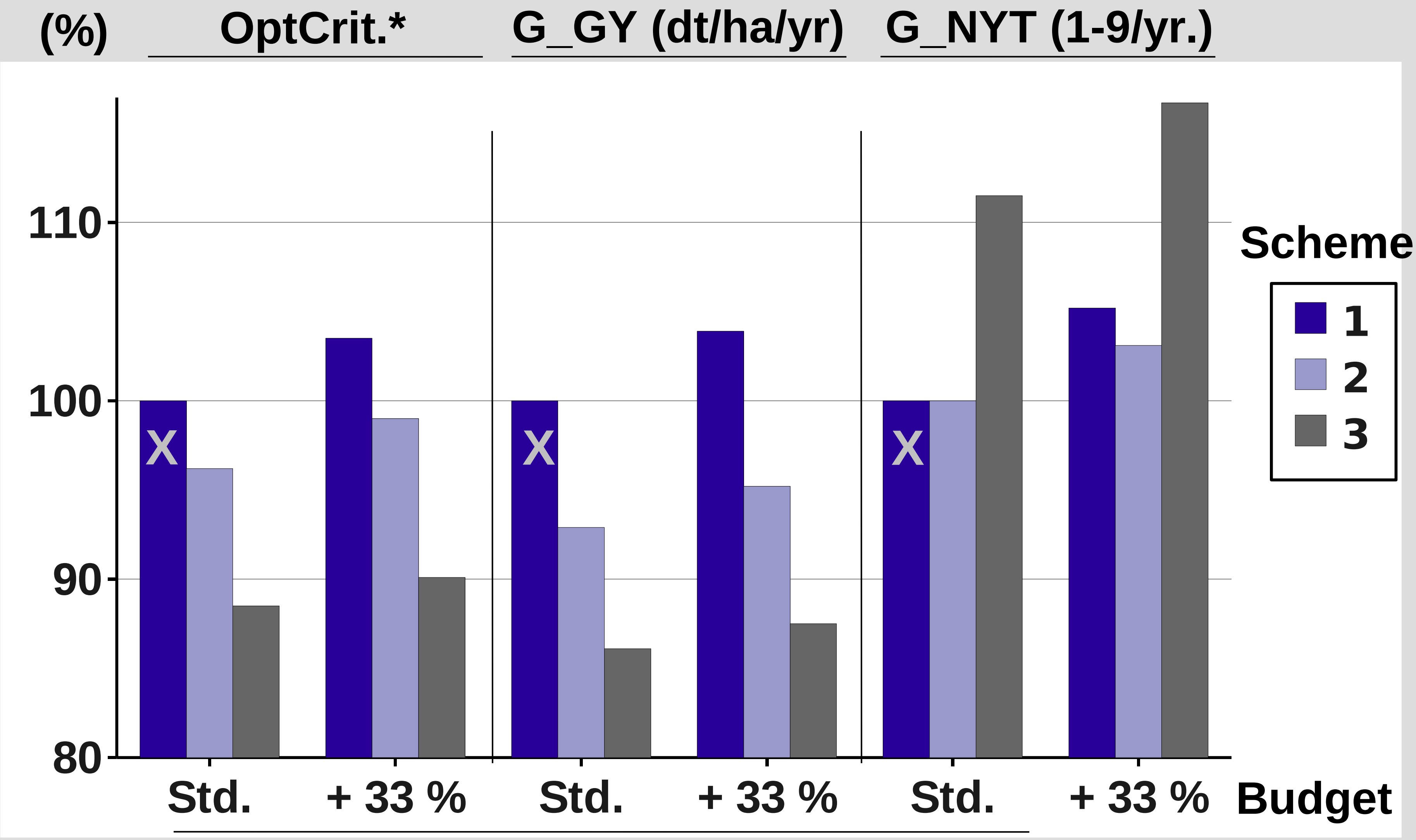
- Amount of total variance available
- Variance between / within crosses
- Genotype x environment interaction variance
- Experimental error
- Weighing of NYT / GY
- Number of NYT selected for
- Budget
- Costs of individual breeding steps
- Number of finally selected lines
- ...

WW: Relative selection gain as a function of the available budget



* = G_NYT + G_GY (weighed) per year

WW: Relative selection gain as a function of the available budget



* = G_NYT + G_GY (weighed) per year

AIB Tomerius 2010

Conclusions from wheat example: Influence of the budget

- Increase of budget increases selection gain, but increase of gain is much lower
- Clearly diminishing return on investment
- Alternative schemes benefit differently
- Choice of more efficient scheme is often much more effective than a budget increase!
- Budget increases should primarily be used to increase the number of candidates at all selection stages (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 gene information**
- **allow to optimize MAS / use of identified QTL**

Simulation studies: Requirements

1. Information on the breeding scheme(s)
   - selection stages, seed propagation type (self, cross), selection type (pedigree / bulk), virtual field design (L,R), selected fractions, selection mode (top, bottom) etc.
2. Information on the traits of interest
   - Gene number and genetic values, pleiotropic effects, GxE-interaction
   - Genetic model(s) investigated

➔ Obtained from breeders' data, if possible

Simulation studies: Limitations

Breeding simulation studies

- **require data and / or assumptions regarding the genetics of the traits under selection (main problem: yield!)**
- **Dimensioning (N, L, R) and selected fractions are usually not optimized**
- **Costs are ususally not really accounted for**

! SIM results offer only decision support !

SIM: Example

- Parental Selection in Wheat Quality Breeding

Parental selection

- Parental selection is the first and an essential step in breeding
- Promising crosses show a high progeny mean and a large genetic variance
- Breeders' information sources:
  - parental performance
  - genetic diversity among crossing partners
  - progeny performance

Source: Wang et al. 2005

Predicting cross performance

- **Mean** can be predicted fairly good in an additive genetic model by midparent value
- **Variance** prediction is inaccurate / difficult

=> so far, no method has given a precise prediction of cross performance

- Yet, cross performance can be accurately predicted when **all gene information** about the traits of interest exists

Source: Wang et al. 2005

Predicting cross performance for quality traits in wheat

- For certain aspects of wheat quality,such genetic information is available
- Idea: predict best crosses for grain quality between 8 Silverstar lines (differ in R_{max} and Extensibility) and 4 other cultivars
- 1000 F8-lines 'developed' from 1000 individual F2 plants by SSD, 40 F8-lines selected
- Two different breeding objectives, 4 different selection schemes (one-step, two-step)

Source: Wang et al. 2005

Predicting cross performance for quality traits: Results

Parent	Objective	Best Silverstar Lines	
		Two step§	One step§
Westonia	High Rmax	3,7	1,3
	High extensib.	1,3,5	1,3,5,7
Krichauff	High Rmax	3,7	3,7
	High extensib.	1,5	1,5
Machete	High Rmax	4,8	-
	High extensib.	1,2,3	1,2,3,4
Diamondbird	High Rmax	3,4	3,4
	High extensib.	1,2,5,6	1,2,5,6

§ Two step: select 200 lines on basis of ext., then 40 on basis of Rmax
One step: select 40 lines on basis of extensibilty only

Source: Wang et al. 2005

Predicting cross performance for quality traits: Results

- **Parental selection depends on the breeding objective and the definition of the selection scheme (e.g. order of trait selection)**

Source: Wang et al. 2005

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**
- **MC / SIM can not only confirm breeders' intuitive experience, but can also find out facts which breeders did not realize before**

AIB

Outlook

- **Advances in genomics will help to build more realistic genetic models**
- **Future goals:**
  - **Develop a combined MC/SIM approach to combine strengths of both methods**
  - **Find optimal ways to integrate multitude of genetic information into breeding practice**

Thank you for listening!

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

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

References

- **Tomerius, A.-M. 2001. Optimizing the development of seed-parent lines in hybrid rye breeding. Diss. Uni Hohenheim. pdf-Dokument unter: 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.**
- **Wang, J., H.A. Eagles, R Trethowan, M. van Ginkel. 2005. Using computer simulation of the selection process and known gene information to assist in parental selection in wheat quality breeding. Aust J Agr Res 56:465-473.**