Generation
Cultivating Plant Diversity for the Resource-Poor

HarvestPlus
Breeding Crops for Better Nutrition

Simulation Approaches in Plant Breeding

Jiankang Wang
http://www.isbreeding.net

CIMMYT

THE UNIVERSITY OF QUEENSLAND

GRDC
Grains Research & Development Corporation

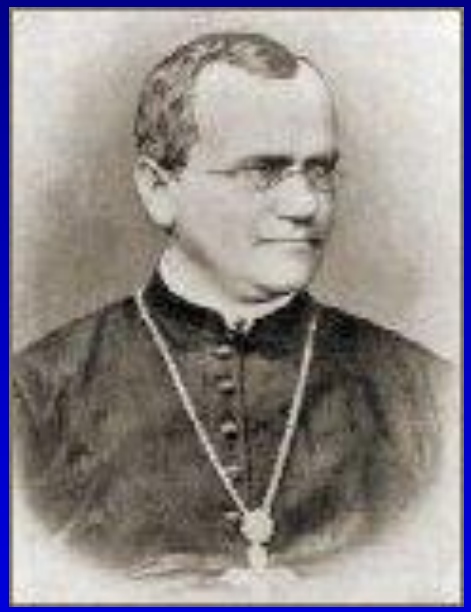
Outlines for presentation

- Conventional quantitative genetics
- Plant breeding and quantitative genetics
- QTL mapping
- Background of breeding simulation
- QuLine tool for genetics and breeding
- QuLine applications
- Demonstration and Hands-on

经典数量遗传学

孟德尔规律的重新发现

- 孟德尔 (1822-1884) 豌豆 (garden pea) 遗传试验 (1866年发表)：*Experiments with Plant Hybrids*
  - Seed shape: 5474 round vs 1850 wrinkled
  - Cotyledon color: 6022 yellow vs 2001 green
  - Seed coat color: 705 grey-brown vs 224 white
  - Pod shape: 882 inflated vs 299 constricted
  - Unripe pod color: 428 green vs 152 yellow
  - Flower position: 651 axial vs 207 terminal
  - Stem length: 787 long (185-230 cm) vs 277 short (20-50 cm)



- 孟德尔规律的重新发现 (1900年)

孟德尔学派和生物统计学派之争

- 生物统计学派 (K. Pearson)：认为连续变异是进化的重要原因，孟德尔规律不适用于连续变异。
- 孟德尔学派 (W. Bateson)：不连续变异是进化的重要因素，连续变异是不能遗传的。

- Yule (1906): There need be no conflict between Mendel's particulate inheritance and the inheritance of continuously varying traits, **provided many genes having similar small effects were responsible for continuously varying traits**.

孟德尔学派和生物统计学派的统一

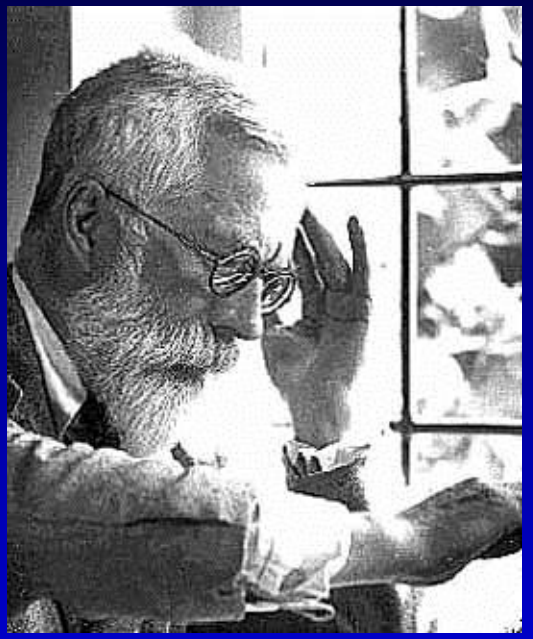
- W. L. Johannsen (1903) 的纯系学说将变异区分为遗传的变异与非遗传的变异，提出了基因型和表现型的概念，这为理解连续性变异也是遗传性状提供了依据 。
- Nilsson-Ehle (1909) 根据小麦粒色的遗传提出了数量性状的多因子假设，这一假设为E. M. East (1911) 玉米穗长和E. M. East (1913) 烟草花冠长度的遗传试验所证实。
- 通过多因子假设将数量性状的遗传纳入到孟德尔遗传的轨道，从而使两个学派的观点得到统一。

数量性状遗传的多基因假说

- R. A. Fisher (1918) “The correlation between relatives on the supposition of Mendelian inheritance”
- Multiple-factor hypothesis (polygene system)
  - A hypothesis to explain quantitative variation by assuming the interaction of a large number of genes (polygenes) each with a small additive effect on the character.
  - Number of genes, gene effects, environment

数量遗传基本理论体系的建立
(1920s-1940s)

- R. A. Fisher (1918) 发表了“The correlation between relatives on the supposition of Mendelian inheritance”。
- J. B. S. Haldane (1924-1927) 连续发表了“A mathematical theory of natural and artificial selection I-V”系列文章，用数学方法说明了数量性状在自然和人工选择下的遗传改变。
- S. Wright (1921) 的“Systems of mating”概括了群体的交配制度，并提出了近交系数的概念，将不同交配制度的群体通过近交系数相比较，并用其通径系数研究交配制度的遗传效果。
- J. L. Lush (1940) 在其“Animal breeding plan”一书中提出了遗传力以及广义遗传力和狭义遗传力的概念，用于研究选择效率和遗传进度。
- G. Malecot提出了亲本系数的概念，用来测量双亲的一对等位基因后裔同样的概率，并度量双亲间的亲缘程度以及该个体的近交程度，从而给出了亲属间协方差的通用表达公式。
- G. F. Sprague和L. A. Tatum (1942) 提出了杂种优势利用中亲本配合力的概念。

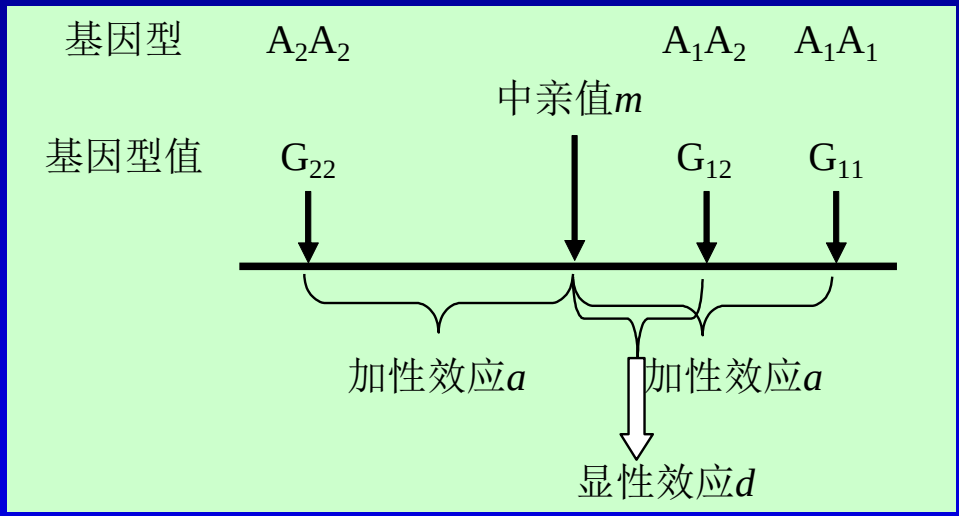


研究内容 — 群体遗传

- 群体遗传结构：基因和基因型频率
- 不同交配系统下基因和基因型频率的变化
- Hardy-Weinberg平衡的建立和平衡群体的性质
- 连锁对Hardy-Weinberg平衡的影响
- 迁移和突变
- 选择
- 突变和选择的联合效应
- 有限大小的理想群体
- 亲属关联和近交系数

研究内容 — 加性和显性效应模型

- $P=G+E+GE+\varepsilon$
- 单基因模型 (等位基因A_1和A_2，频率分别为p和q)
  - p^2: A_1A_1、2pq: A_1A_2、q^2: A_2A_2
  - 基因型A_iA_j的遗传值为G_{ij}
  - 中亲值：$m = (G_{11}+G_{22}) / 2$
  - 加性效应：$a = G_{11} - m$
  - 显性效应：$d = G_{12} - m$



  - $G_{11} = m+a$; $G_{12} = m+d$; $G_{22} = m-a$

研究内容 — 育种值和显性离差

- 单基因模型 (等位基因A_1和A_2)
  - 群体平均数: $\mu=m+(p-q)a+2pqd$
  - 等位基因效应: $\alpha_1=q[a+(q-p)d]$; $\alpha_2=-p[a+(q-p)d]$
  - 等位基因的替代效应$\alpha= \alpha_2 -\alpha_2=a+(q-p)d$
  - $G_{ij}= \mu + \alpha_i+ \alpha_j+ \delta_{ij}$
  - 育种值: $A_{ij}= \alpha_i+ \alpha_j$
  - 显性离差: $D_{ij}=\delta_{ij}$
- 容易推广到复等位基因

研究内容 — 上位性离差

- 2个基因位点模型（A_i、 A_j和B_k、 B_l）

$$G_{ijkl} = \mu + (\alpha_i + \alpha_j + \delta_{ij}) + (\alpha_k + \alpha_l + \delta_{kl}) + I_{ijkl}$$

- 育种值：A=Σ α
- 显性离差：D=Σ δ
- 上位性离差： I=Σ I
- P= G+ ε = μ +A+D+I+ ε
- V_P= V_G+ V_ε = V_A+V_D+V_I+ V_ε

研究内容 — 遗传方差的分解

效应的分解	方差的分解
表型值 (P)	表型方差 (V_P)
基因型值 (G)	遗传方差 (V_G)
育种值 (A)	加性方差 (V_A)
显性离差 (D)	显性方差 (V_D)
上位性离差 (I)	上位性方差 (V_I)
基因型X环境互作效应 (GE)	基因型X环境互作方差 (V_{GE})
随机误差 (ε)	随机误差方差 (V_ε)

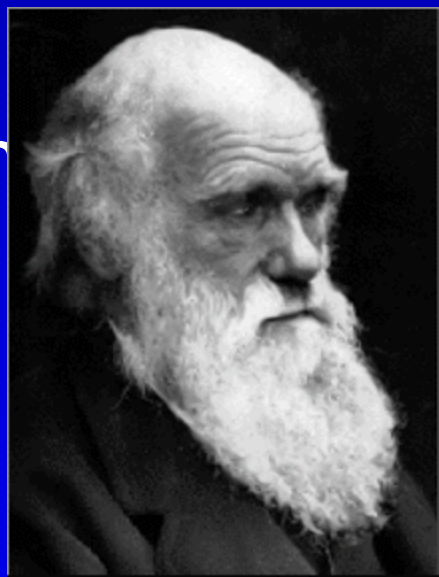
植物育种与数量遗传

植物育种的定义

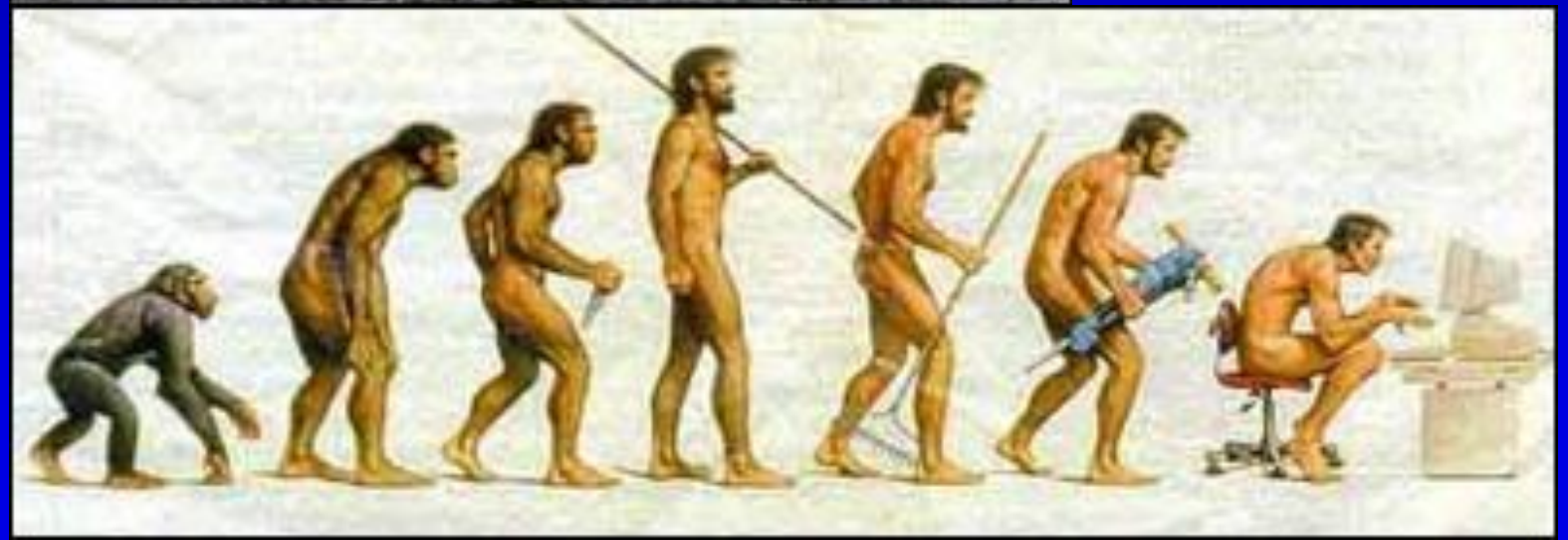
- 植物育种的主要任务是寻找控制目标性状的基因、研究这些基因在不同环境下的表型效应、挑选适当的亲本材料、设计杂交方案和分离群体种植方案、通过基因型/表型选择聚合存在于不同亲本材料中的有利基因、培育符合育种目标的基因型，为农业生产提供适宜的品种
- 植物育种中的科学
- 植物育种中的艺术
- 植物育种中的商业行为
- “机遇”或”运气”

达尔文(1809-1882)与进化论

- 《物种起源》(The Origin of Species) (1859)
  - On the Origin of Species by Means of Natural Selection
  - The Preservation of Favoured Races in the Struggle for Life
- “I have called this principle, by which each slight variation, if useful, is preserved, by the term Natural Selection.”



达尔文主义: 物竞天择, 适者生存



普通六倍体小麦的起源和驯化

Thank You

植物育种的一般过程

- 配：选择亲本、配制杂交组合
- 选：繁殖产生育种群体，根据育种目标选择理想基因型
- 比：多环境表型鉴定，通过比较确定适合特定环境的作物新品种

育种家关心的一些问题

- **配：** 选择谁做亲本进行杂交？采用什么杂交方式组合？如，
  - 单交：A×B
  - 回交：(A×B) ×B或 (A×B) ×A
  - 三交：(A×B) ×C
  - 双交：(A×B) × (C×D)
- **选：** 种植多大规模的群体？什么时候、选择什么性状、选择多少？如何提高育种效率？
- **比：** 如何评价和排除（或利用）基因型和环境互作？如何选择试验地点？田间如何布置基因型？

孟德尔遗传对现代育种的贡献

- 杂交育种：重组可以产生新的表型/基因型

杂交育种

假定有2个抗性基因R1和R2，其等位基因为r1和r2；3个高产基因Y1、Y2和Y2，其等位基因y1、y2和y3；一对未来病害基因R和r；不考虑连锁

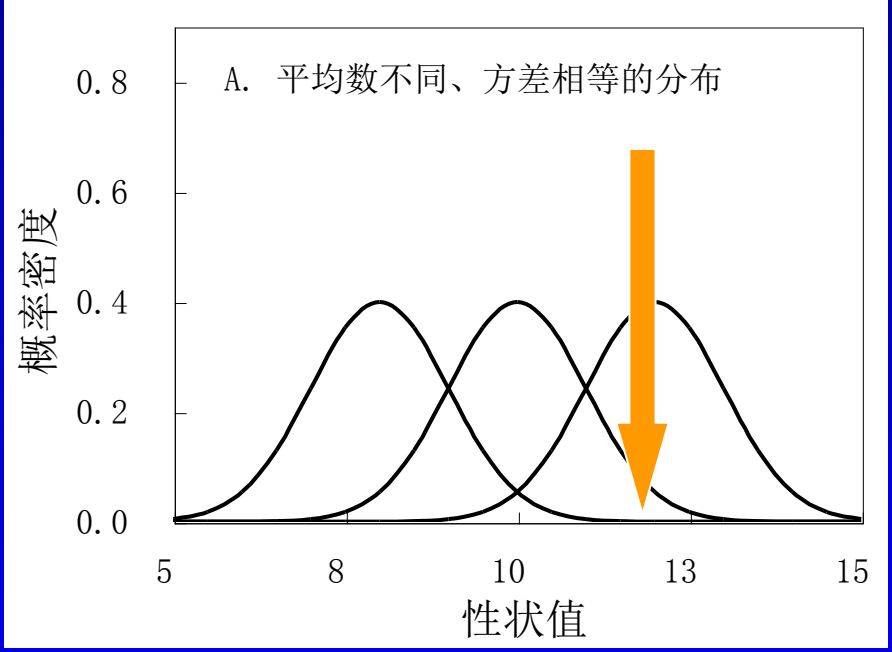
- 亲本基因型（2种）
  - P1：R1R1 r2r2 y1y1 Y2Y2 Y3Y3 rr：中抗、产量较高
  - P2：r1r1 R2R2 Y1Y1 y2y2 y3y3 RR：中抗、产量较低
- 目标基因型：R1R1 R2R2 Y1Y1 Y2Y2 Y3Y3 ??
- 后代基因型（2^5=32种）
  - 不抗病、产量低：r1r1 r2r2 y1y1 y2y2 y3y3 ??
  - 抗病、产量低：R1R1 R2R2 y1y1 y2y2 y3y3 ??
  - 等等
- 育种的目的是从32种可能的基因型中把理想的目标基因型选择出来

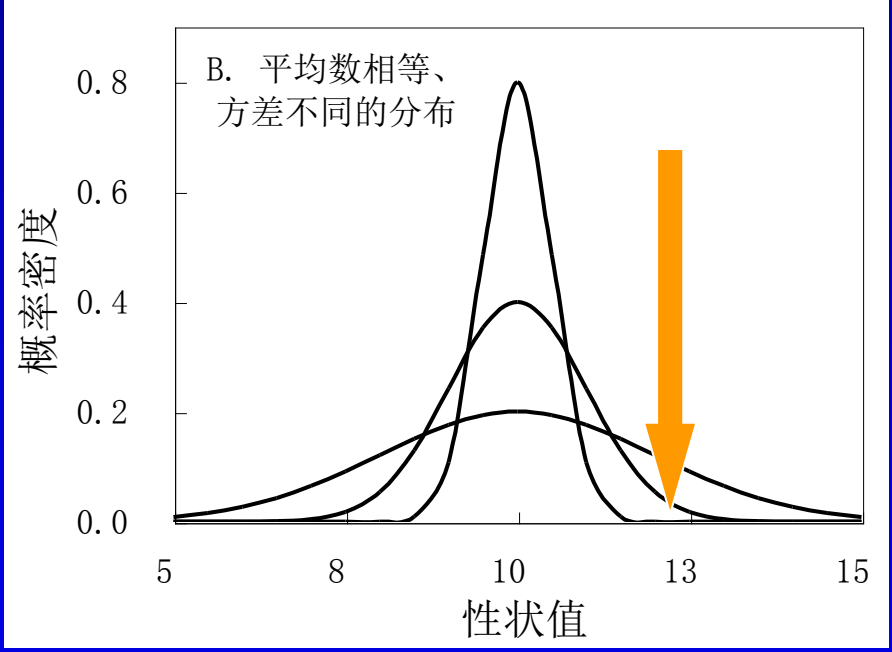
基因与基因型是两个不同的概念

- 当分离位点较多时，$2^{30}=1\times10^{9}$；$2^{50}=1\times10^{15}$
- 种质库的目的是保存尽可能多的基因；保存所有可能的基因是容易实现的；但是，保存所有可能的基因型几乎是不可能的
- 但是，也不是说越少越好，例如，
  - P1: R1R1 R2R2 Y1Y1 Y2Y2 Y3Y3 rr
  - P2: r1r1 r2r2 y1y1 y2y2 y3y3 RR
  - P3: R1R1 r2r2 Y1Y1 y2y2 Y3Y3 RR
  - 目标基因型: R1R1 R2R2 Y1Y1 Y2Y2 Y3Y3 RR
  - P1×P2育种群体中，有64种可能的基因型
  - P1×P3育种群体中，有8种可能的基因型
  - 因此，组配P1×P3，育种更容易取得成功

理想的育种群体

- 高群体平均数
- 大遗传变异





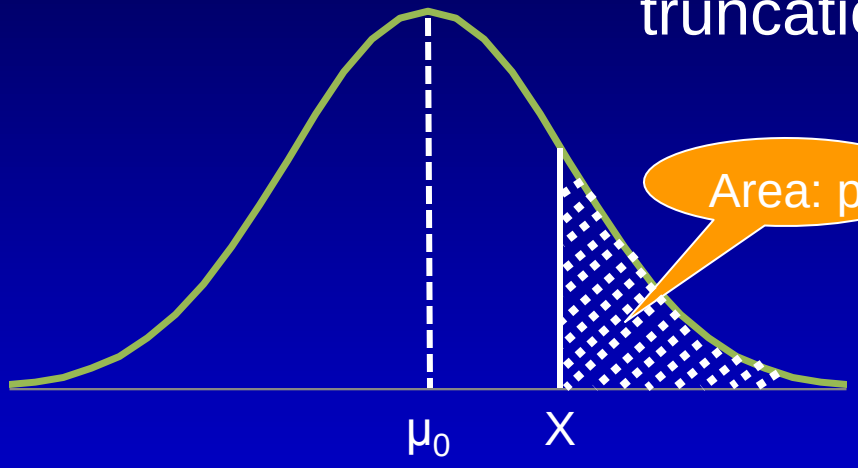
Genetic gain: response to selection

(or the change of population mean)

Criteria: tail selection (p) or truncation selection (X)

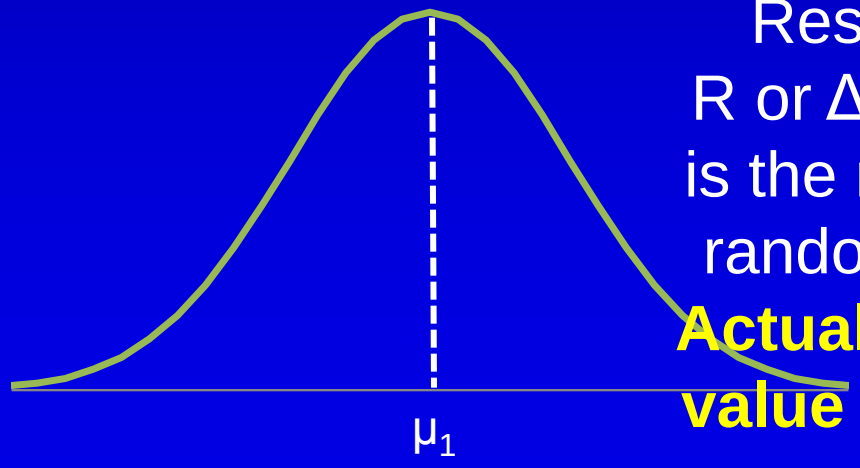
p ⇔ X

Distribution of parents



Selection differential: $S=\mu_s-\mu_0$, where μ_s is the mean performance of selected parents

Distribution of randomly-mated offspring



Response to selection: R or $\Delta G = \mu_1 - \mu_0$, where μ_1 is the mean performance of randomly-mated offspring.

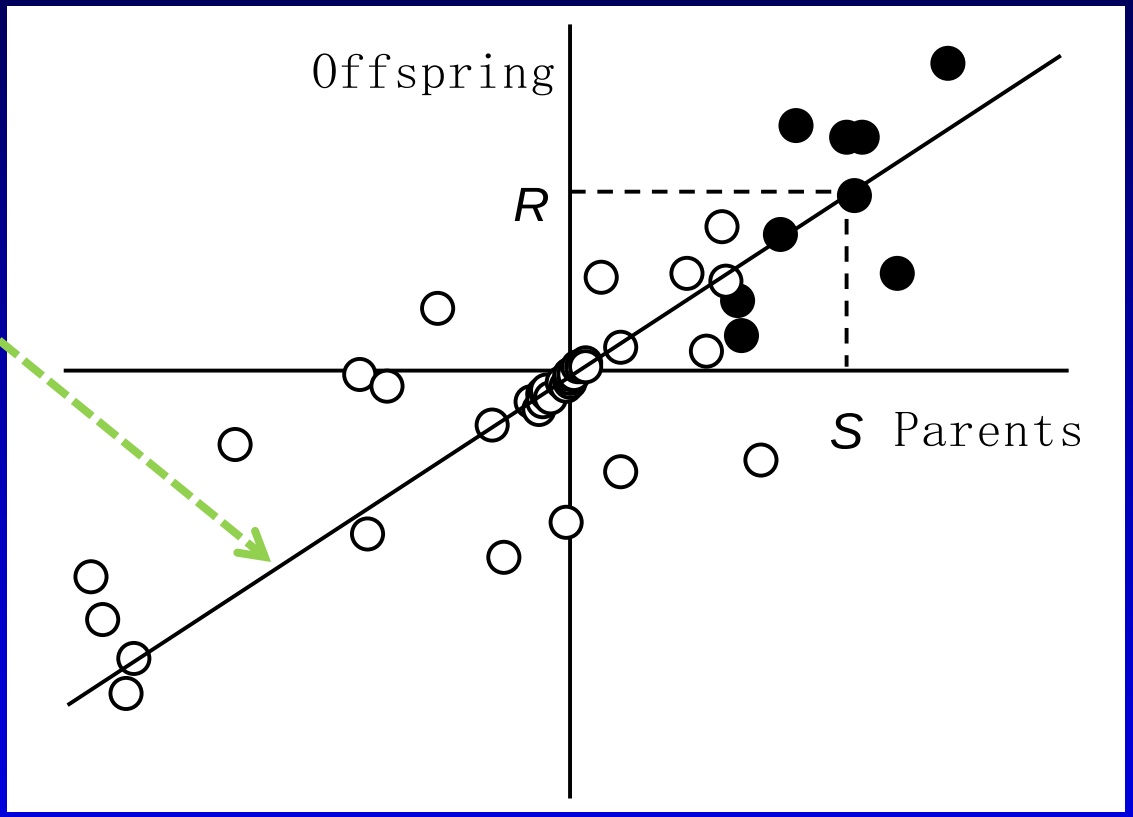
Actually, R is the breeding value of selected parents!

Estimation of R based on heritability

- $y = b\,x = h^2\,x$

- Response to selection or Genetic gain:

$R \text{ or } \Delta G = \mu_1 - \mu_0 = h^2\,S$



提高遗传进度的途径分析

$$R = k_p h \sqrt{V_A}$$

$$R = \frac{k_p V_A}{\sqrt{V_P}}$$

- 使用较小的选择比例，即提高选择强度 (k_p)
- 提高加性方差在遗传方差中所占的比例 (h^2)
- 提高加性方差本身 (V_A)
- 降低非遗传方差

QTL作图

定位数量性状基因; 估计单个数量性状基因的遗传效应

数量遗传研究主线

- P = G + E + GE + e
- G = A + D + I, GE互作分析
- G =主基因 + 微效多基因
- G =单个QTL效应之和

- 遗传效应的分解
- 遗传方差的分解

数量性状基因定位的目的

- 控制数量性状的基因
  - 有多少?
  - 在什么地方?
  - 表型效应有多大?
  - 和其它基因间有无互作?
  - 和环境间有无互作?

作图群体的分类

- 按基因型是否纯合
  - 暂时群体（Temporary population）
  - 永久群体（Permanent population）
  - 自然群体（Natural population）
- 按群体间的亲缘关系
  - 初级作图群体（Primary mapping population）
  - 次级作图群体（Secondary mapping population）

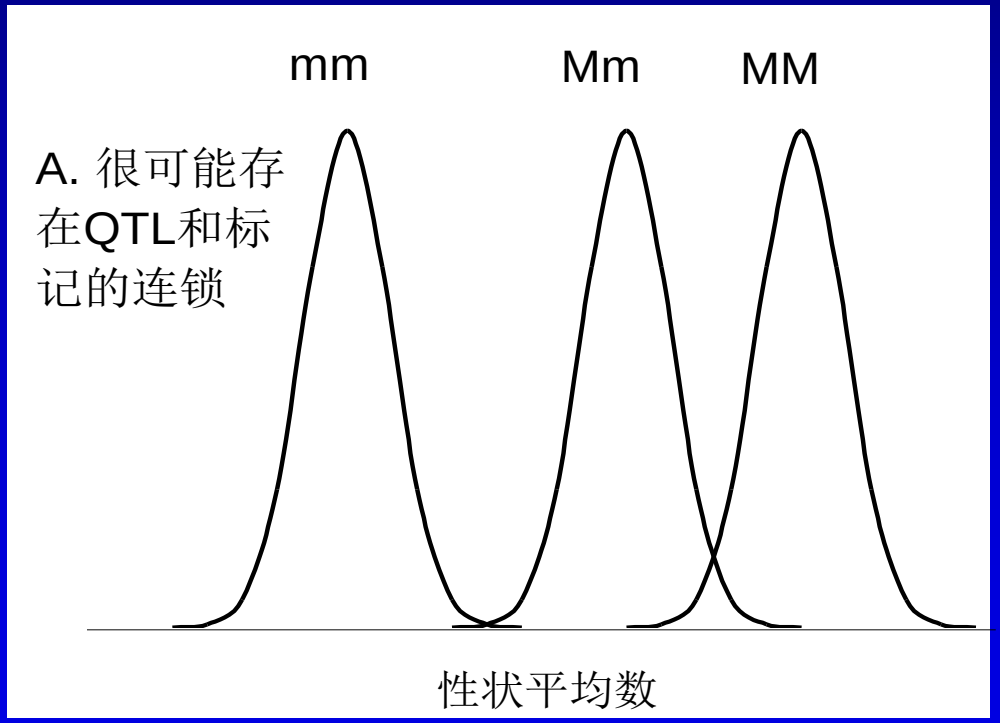
Example: 10 RILs of Rice (Linkage Map of Chr. 5)

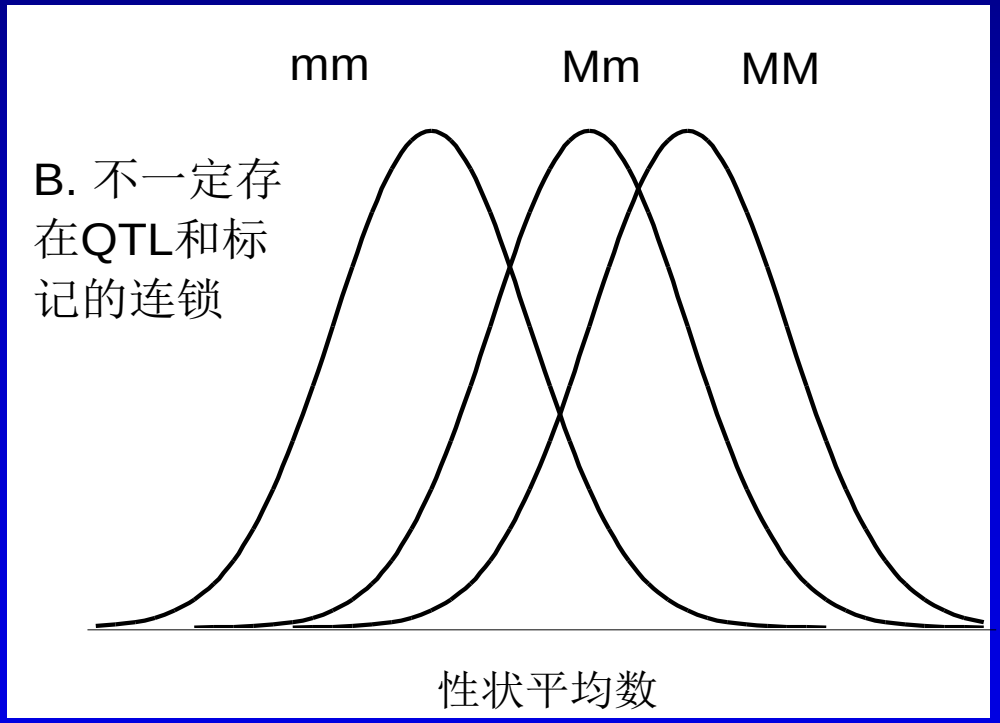
Marker	C263	R830	R3166	XNpb387	R569	R1553	C128	C1402	XNpb81	C246	R2953	C1447	Grain width (mm)
Position (cM)	0.0	3.5	8.5	19.5	32.0	66.6	74.1	78.6	81.8	91.9	92.7	96.8	
RIL1	1	1	1	1	1	1	1	1	1	1	1	1	2.33
RIL2	2	2	2	2	2	1	1	1	1	2	2	2	1.99
RIL3	1	2	2	2	2	2	2	2	2	2	2	2	2.24
RIL4	1	1	1	1	1	1	2	2	2	2	2	2	1.94
RIL5	1	1	1	1	1	2	2	1	1	1	1	1	2.76
RIL6	1	1	1	2	2	2	2	2	2	2	2	2	2.32
RIL7	1	1	1	1	1	1	1	1	1	1	1	1	2.32
RIL8	2	2	1	2	2	1	1	1	1	2	2	2	2.08
RIL9	1	1	1	1	2	2	1	1	1	1	1	1	2.24
RIL10	1	1	1	1	2	2	1	1	1	1	1	1	2.45

$$t = \frac{\hat{\mu}_{MM} - \hat{\mu}_{mm}}{\sqrt{\frac{s_e^2}{df_{MM}} + \frac{s_e^2}{df_{mm}}}}$$

QTL作图的基本原理

一个标记位点上3种基因型的性状平均数





表型对标记线性回归模型的性质

- 假定不同QTL间的效应是可加的，偏回归系数只依赖于两个相邻标记所标定区间上的QTL。
- 模型中加入非连锁标记，能有效控制剩余遗传方差，从而降低统计量的抽样方差，提高QTL的检测功效。
- 模型中的连锁标记可以降低连锁QTL对检验统计量的影响。
- 模型中的两个标记的偏回归系数是不相关的。

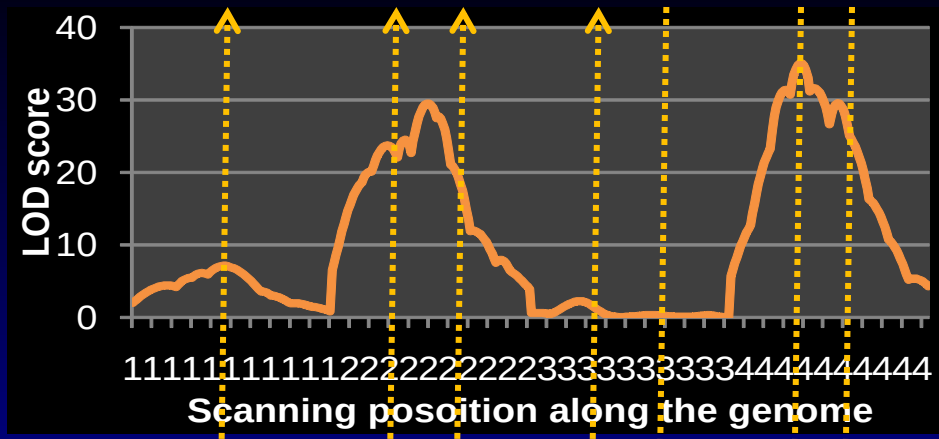
完备区间作图方法（ICIM）

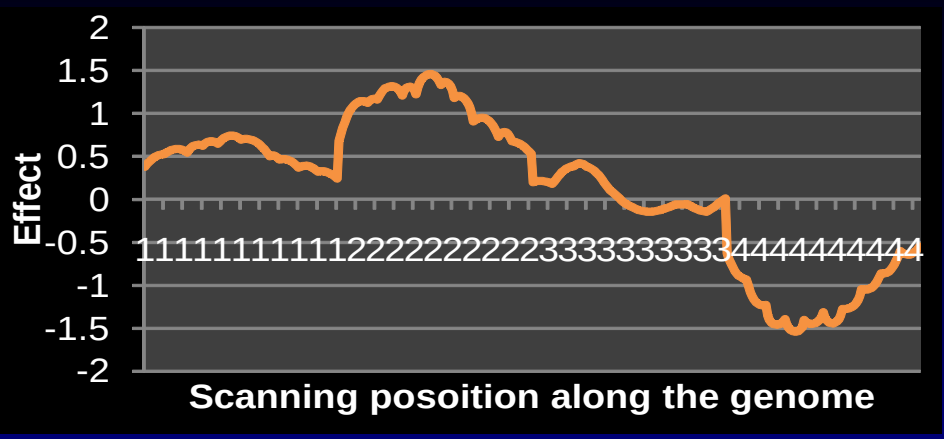
- One-dimensional scanning (interval mapping)

$$\Delta y_i = y_i - \sum_{j \neq k, k+1} \hat{b}_j x_{ij}$$

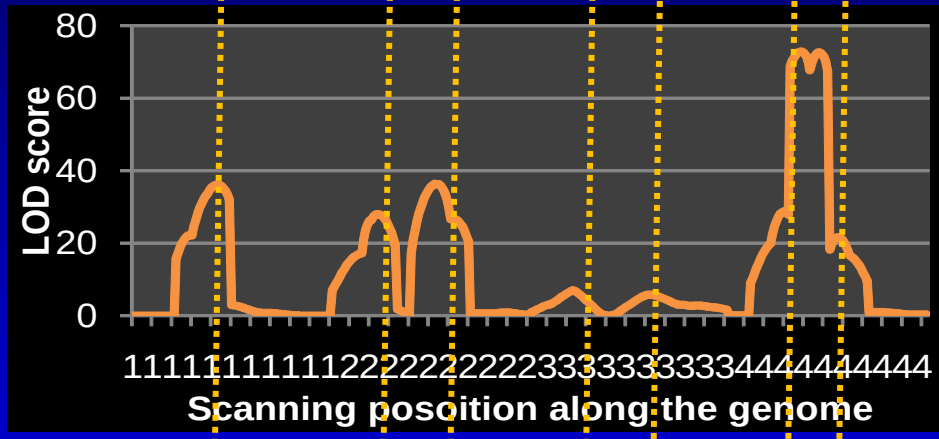
- Two-dimensional scanning (interval mapping)

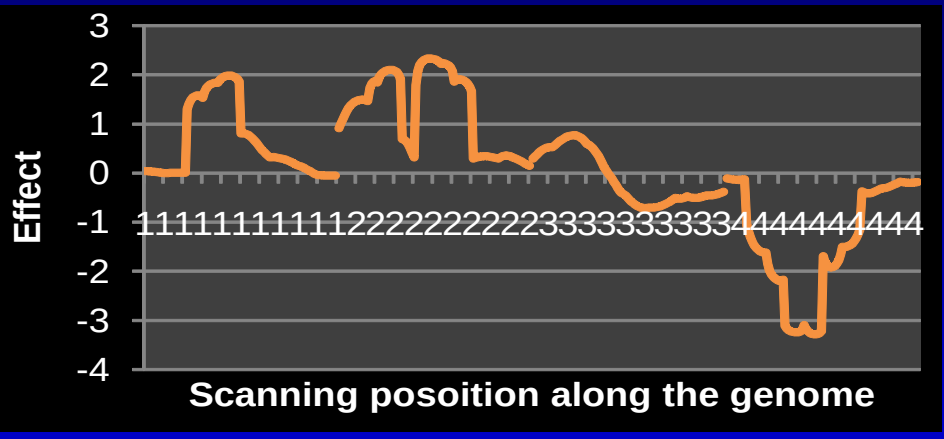
$$\Delta y_i = y_i - \sum_{r \neq j, j+1, k, k+1} \hat{b}_r x_{ir} - \sum_{\substack{r \neq j, j+1 \\ s \neq k, k+1}} \hat{b}_{rs} x_{ir} x_{is}$$



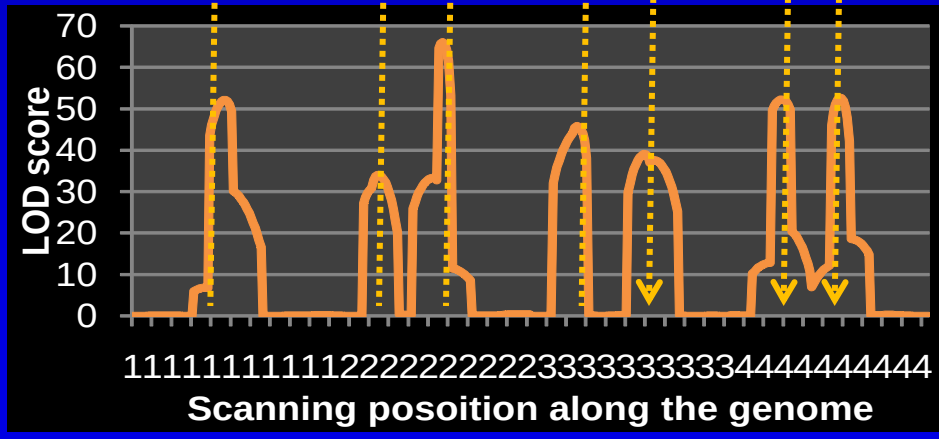


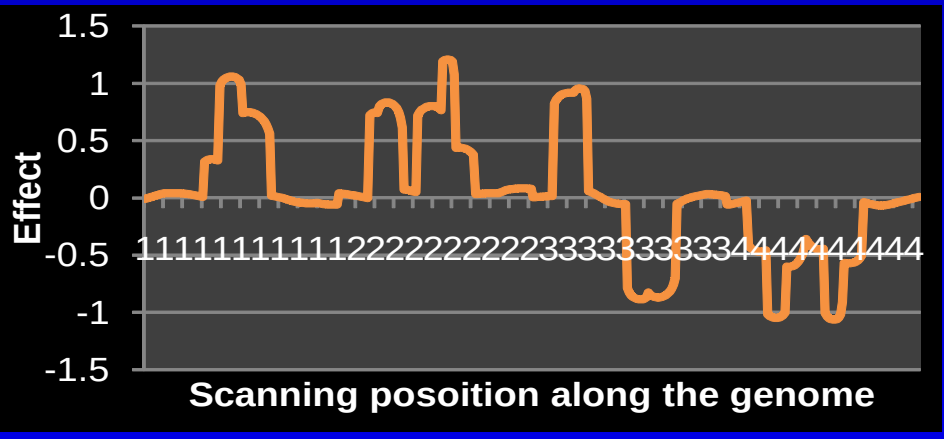
IM
1989





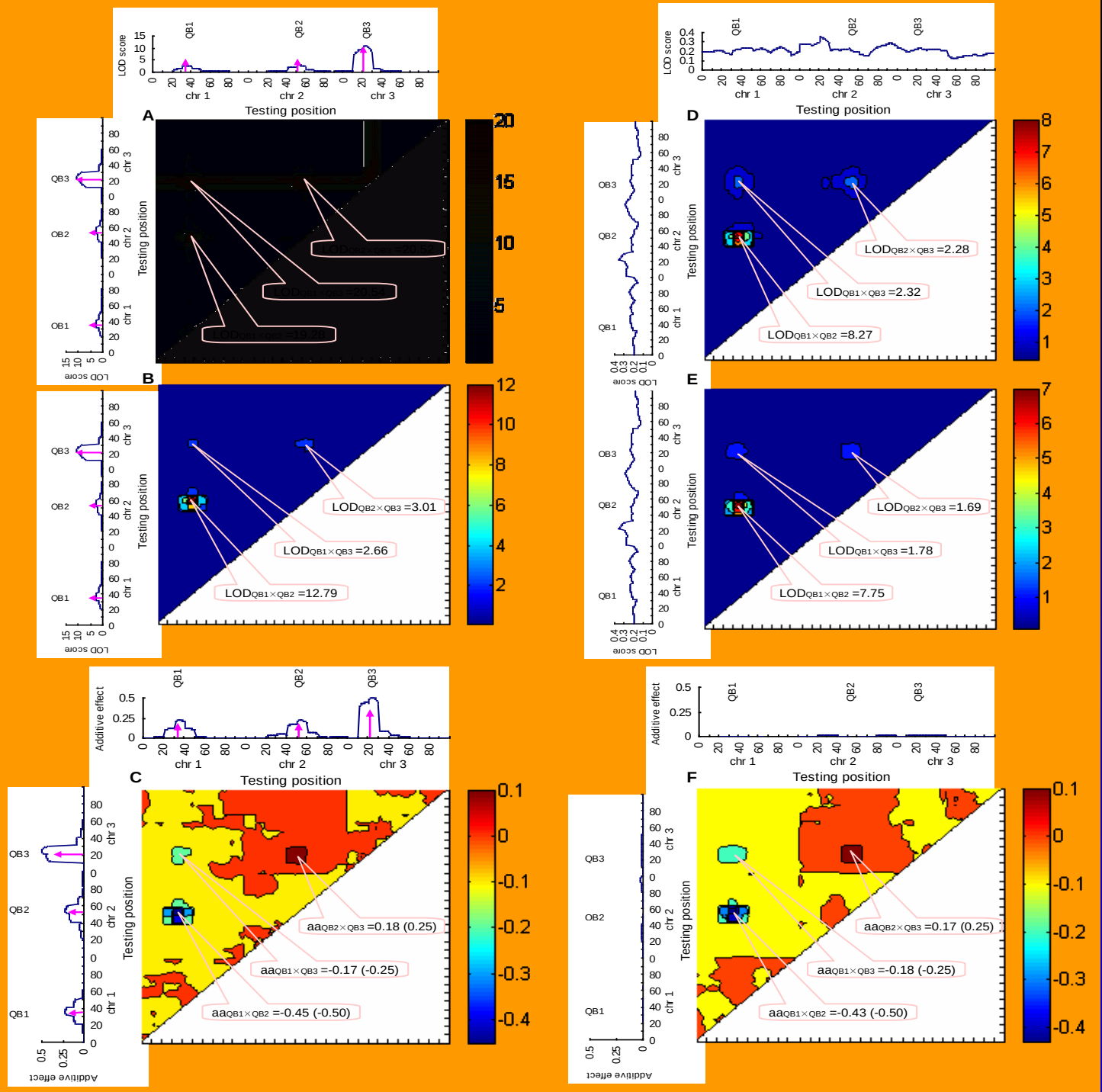
CIM
1994





ICIM
2006

- Detecting epistasis where the interacting QTL don't have significant additive effects



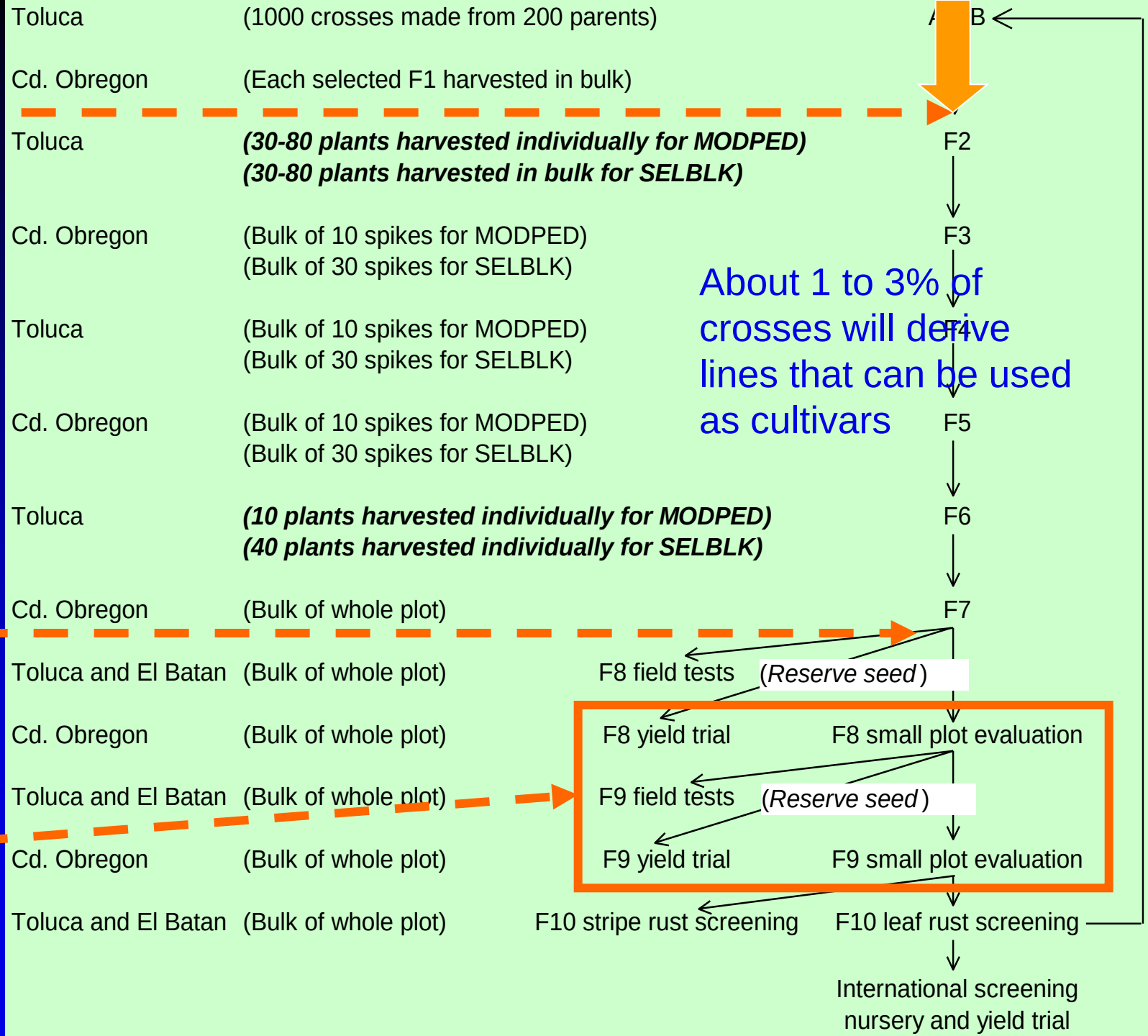
Background on breeding simulation

About 30% of the crosses are discarded in F1

PLANT BREEDING IS A COMPLEX PROCESS WITH MANY DECISIONS

About 40% of crosses retained after F7

About 20% of crosses retained after two rounds of yield trials



Why do we need tools in breeding?

- To improve the efficiency of traditional phenotypic selection through exploring various options
- To avoid the simplified assumptions made in classical quantitative genetic theory
- To better use the large amount of gene information available from
  - Genomics research
    - January 1992: 59,317 datasets, 77,805,556 bp
    - March 2005: 43,118,204 datasets, 47,009,081,750 bp
  - QTL mapping (CAB, April 2005)
    - 3497 publications on QTL mapping
    - 1581 publications in plants

Why do we need tools in breeding?

- To build a bridge between the biological data and breeders' requirements
- To combine all these sources of data into "knowledge" that breeders can use in their breeding programs

Questions that can be studied by QuLine: A genetic and breeding simulation tool

1. Comparison of breeding efficiencies from different selection strategies and their modifications. Which breeding method should be adopted?
2. Balance between the number of crosses and population size of segregating generations. What shall the breeder do if the available resources increase or reduce?
3. Evaluation of marker-assisted selection (MAS). When and how should MAS be used?
4. Comparative value of single, top, back, and double crosses in breeding?

Questions that can be studied by QuLine (mainly for inbred line development)

5. The correlation between parents and their offspring. Can F1 or F2 hybrids predict the performance of their advanced lines? For early generation selection, how early is too early?
6. When to use DH (doubled haploid)?
   - Early generation: many individuals need to be tested
   - Advanced line stage: lines will be good for other traits, but the desired genotype may be lost during selection due to population size, trait associations and genetic drift

QuLine tool for genetics and breeding

Available breeding simulation tools

- **QuLine**, a computer software that simulates breeding programs for developing inbred lines
- **QuHybrid**, a computer software that simulates breeding programs for developing hybrids
- **QuMARS**, a computer software that simulates marker-assisted recurrent selection and genome-wide selection

Test cross implemented in QuHybrid

Family	PopBefore(1)	PopBefore(2)	PopBefore(3)	…	Testers
Family size	5	3	6	…	3
	F11 F12 F13 F14 F15	F21 F22 F23	F31 F32 F33 F34 F35 F36	…	T1 T2 T3

Making testcrosses

Testcross 1	F11×**T1**	F12×**T1**	F13×**T1**	F14×**T1**	F15×**T1**	F21×**T1**	F22×**T1**	F23×**T1**	F31×**T1**	F32×**T1**	F33×**T1**	F34×**T1**	F35×**T1**	F36×**T1**
Testcross 2	F11×**T2**	F12×**T2**	F13×**T2**	F14×**T2**	F15×**T2**	F21×**T2**	F22×**T2**	F23×**T2**	F31×**T2**	F32×**T2**	F33×**T2**	F34×**T2**	F35×**T2**	F36×**T2**
Testcross 3	F11×**T3**	F12×**T3**	F13×**T3**	F14×**T3**	F15×**T3**	F21×**T3**	F22×**T3**	F23×**T3**	F31×**T3**	F32×**T3**	F33×**T3**	F34×**T3**	F35×**T3**	F36×**T3**

Among family selection

Testcross 1	F11×**T1**	F12×**T1**	F13×**T1**	F14×**T1**	F15×**T1**	F21×**T1**	F22×**T1**	F23×**T1**	F31×**T1**	F32×**T1**	F33×**T1**	F34×**T1**	F35×**T1**	F36×**T1**
Testcross 2	F11×**T2**	F12×**T2**	F13×**T2**	F14×**T2**	F15×**T2**	F21×**T2**	F22×**T2**	F23×**T2**	F31×**T2**	F32×**T2**	F33×**T2**	F34×**T2**	F35×**T2**	F36×**T2**
Testcross 3	F11×**T3**	F12×**T3**	F13×**T3**	F14×**T3**	F15×**T3**	F21×**T3**	F22×**T3**	F23×**T3**	F31×**T3**	F32×**T3**	F33×**T3**	F34×**T3**	F35×**T3**	F36×**T3**

Within family selection

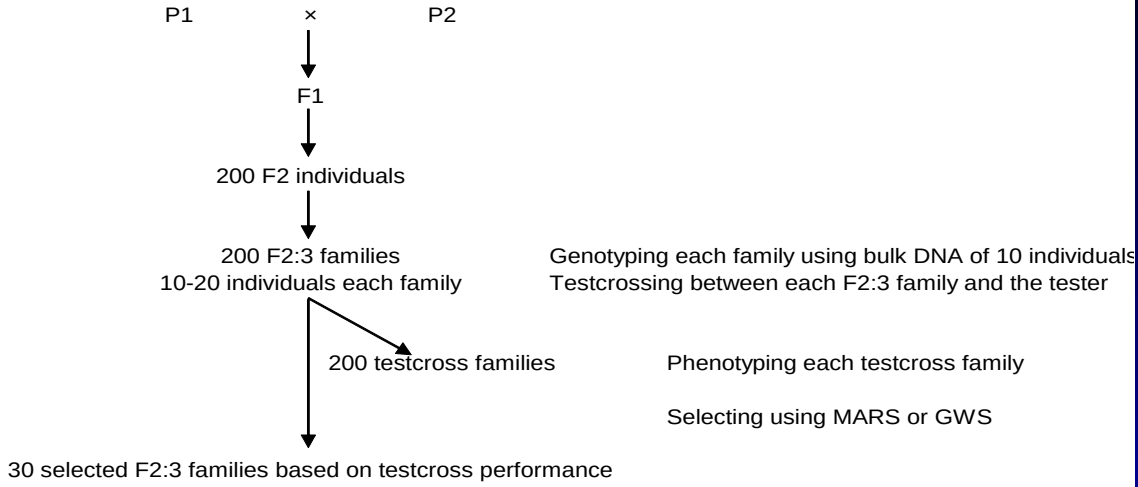
						F21×**T1**	F22×**T1**	F23×**T1**	F31×**T1**	F32×**T1**	F33×**T1**	F34×**T1**	F35×**T1**	F36×**T1**
						F21×**T2**	F22×**T2**	F23×**T2**	F31×**T2**	F32×**T2**	F33×**T2**	F34×**T2**	F35×**T2**	F36×**T2**
						F21×**T3**	F22×**T3**	F23×**T3**	F31×**T3**	F32×**T3**	F33×**T3**	F34×**T3**	F35×**T3**	F36×**T3**

Population after selection, this population will be used as the starting for the next generation

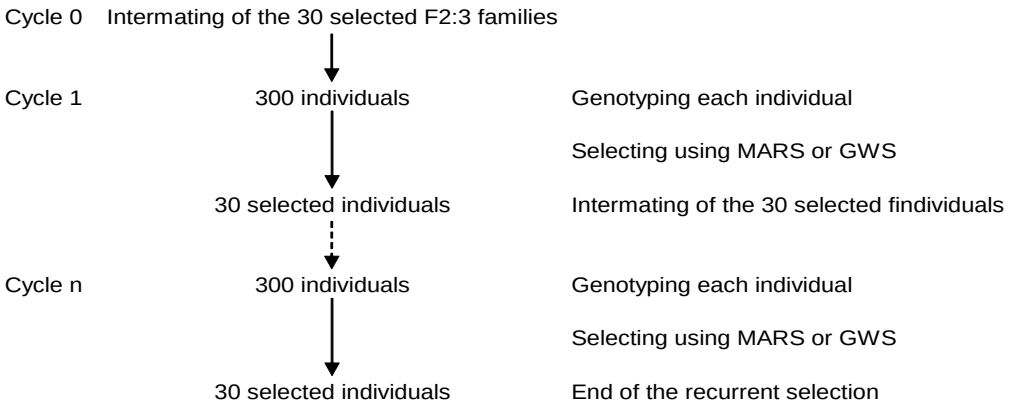
Family	PopAfter(1)	PopAfter(2)	PopAfter(3)
Family size	0	2	4
		F22 F23	F31 F33 F34 F36

The flowchart of QuMARS

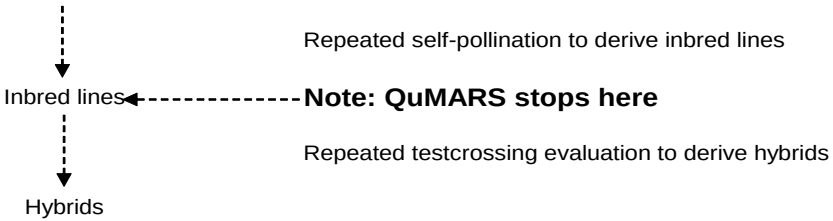
Stage I: Testcrossing: Starting from a single cross, To generate the initial population for recurrent selection; To build the prediction equation of breeding traits on markers



Stage II: Recurrent selection: To intermate the selected families (S0); To grow and genotype randomly mated progenies; To select individuals based on MARS or GWS



Stage III: Inbred development: To generate inbred lines through pedigree breeding



What can QuLine do?

- Comparison of genetic gains from different selection methods
  - Change in population mean
  - Change in gene frequency
  - Change in Hamming distance (distance of a selected genotype to the target genotype)
- Comparison of cross performance
  - Selection history
  - Rogers' genetic distance
  - Number of lines retained from each cross
- Comparison of cost efficiency
  - Number of families
  - Individual plants per generation
- Validation of theories

In genetics
(implemented by the QU-GENE engine)

- Most genetic phenomena, if not all, can be defined in the QU-GENE engine input file (QUG).
- Among them are:
  - Multiple alleles (e.g. Glutenin genes in wheat)
  - Linkage (between gene and marker, between genes, between markers)
  - Additive, dominance and epistasis
  - Pleiotropy (one gene effects multiple traits)
  - Genotype by environment interaction
  - Molecular markers (dominant, or co-dominant)

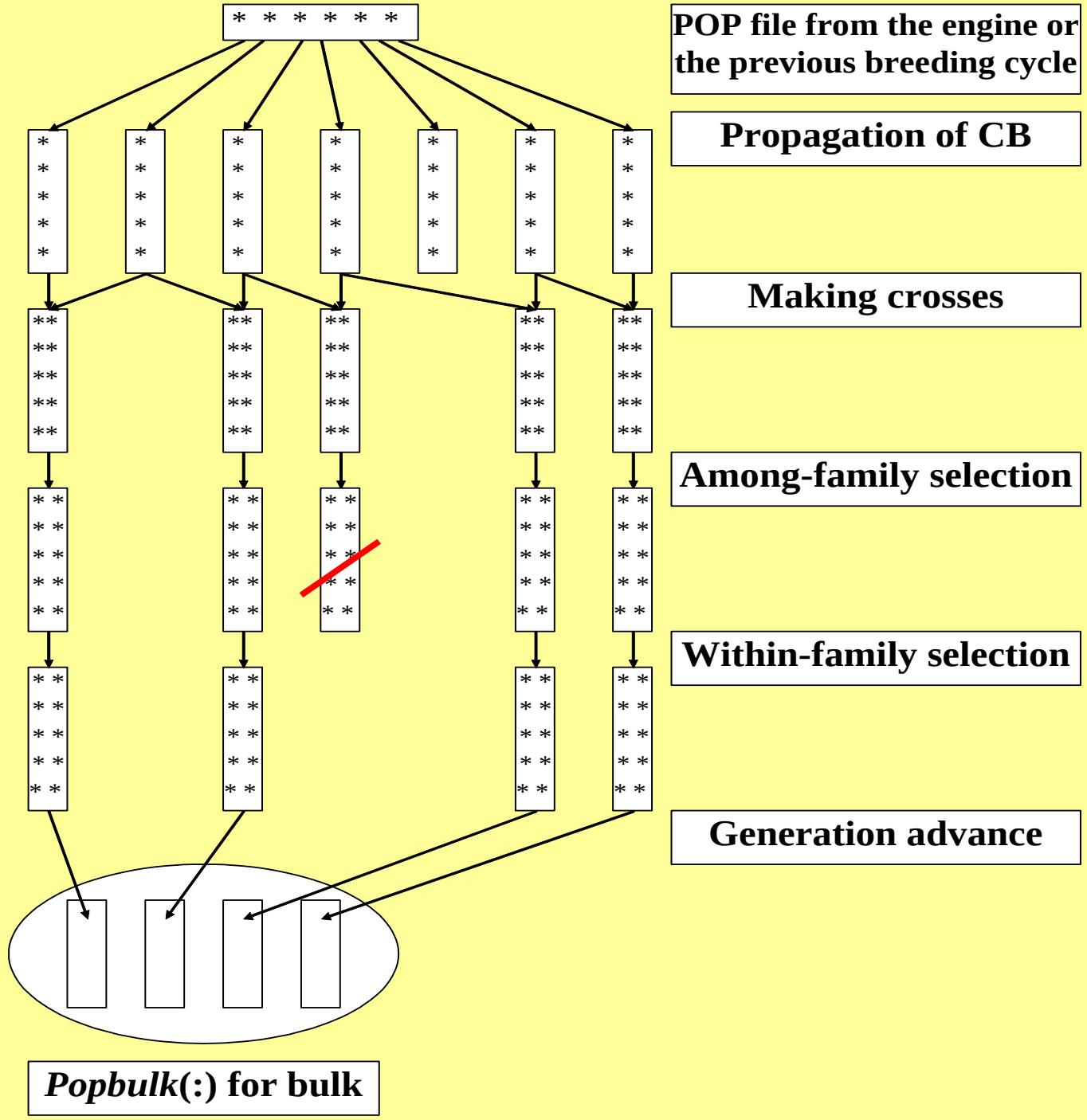
In breeding
(implemented by the QuLine module)

- Most, if not all, breeding methods for self-pollinated crops, can be defined and then simulated in QuLine.
- Among them are:
  - Pedigree system (including SSD)
  - Bulk-population
  - Doubled haploid
  - Marker-assisted selection (include marker-based selection)
  - Recurrent selection within one population
  - Many modifications and combinations

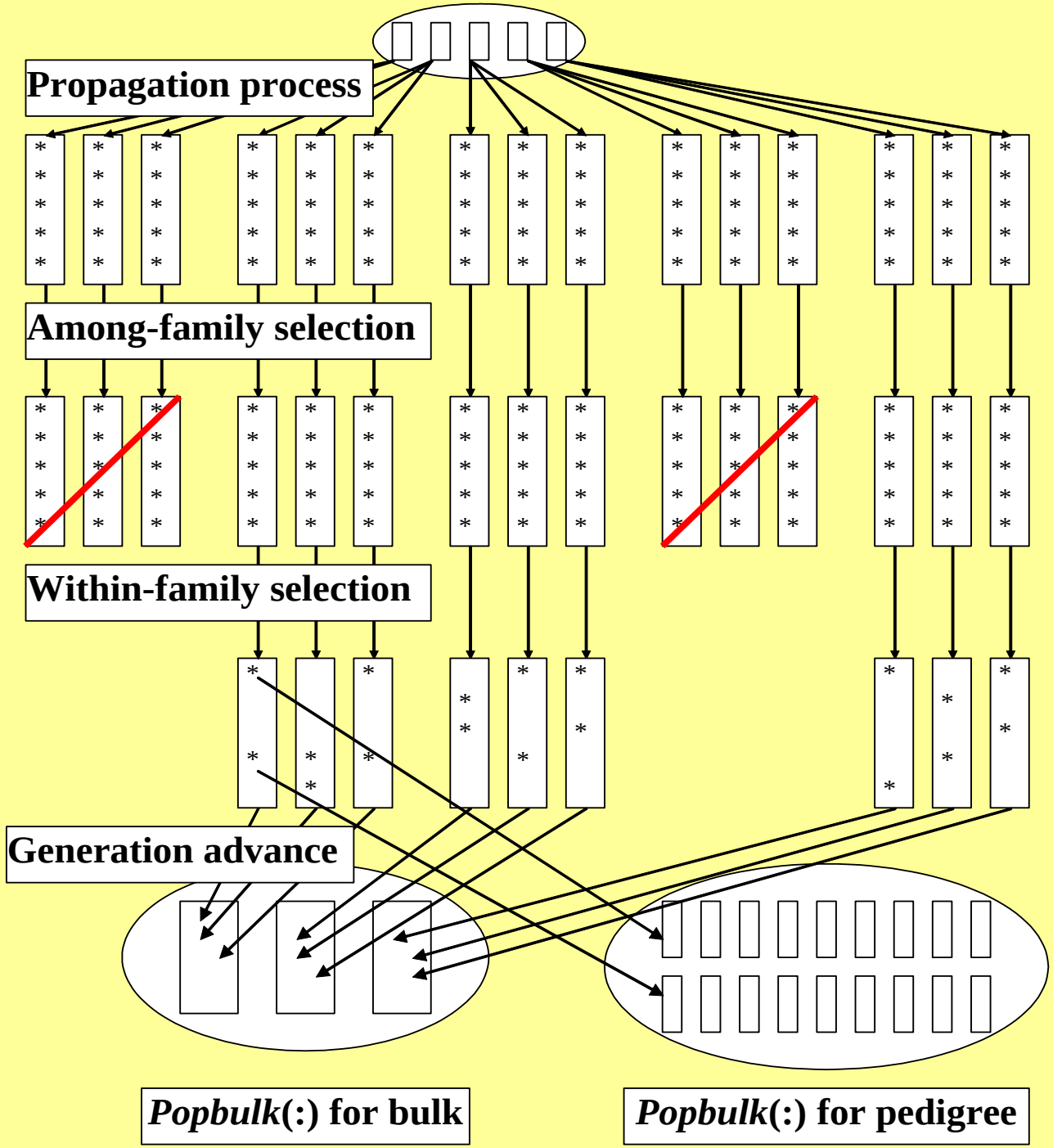
How does QuLine work?

- Two input files are needed
  - ***QUG file*** containing the necessary information for a genotype and environment (GE) system and initial population(s) of genotypes. It is the input for the QU-GENE engine. Two kinds of output files will be generated from the engine.
    - GES file for defining a GE system (= input for QuLine)
    - POP file for defining the initial population (= input for QuLine)
  - ***QMP file*** containing the necessary information for the breeding strategies to be simulated (e.g. pedigree, bulk, SSD, DH, etc.) (= input for QuLine)

General procedure for crossing and selection in F1



General procedure for propagation and selection in F2 and onwards



Define the QMP file for the selected bulk selection method: an example

Define the breeding method in a way that the computer can understand

General simulation parameters

- Number of runs: any integer
- Number of breeding cycles: any integer
- Number of crosses in F1 generation: any integer
- Indicator for crossing block update, 0 or 1.
  - 0: All individuals after a breeding cycle are used as the parents for the next cycle;
  - 1: The best individuals in the final selected population and the initial crossing block are selected as the parents for next cycle

General simulation parameters

- Indicator for outputting GE system details, 0 for no output, and 1 for output
- Indicator for outputting population details, 0 for no output, and 1 for output
- Indicator for outputting selection history, 0 for no output, and 1 for output
- Indicator for outputting Rogers distance for each cross and the lines retained from each cross, 0 for no output, and 1 for output
- Indicator for outputting correlation coefficient, 0 for no output, and 1 for output

The number of models in the GE system and the number of runs for breeding strategy

- Four nested loops in QuLine
  - Loop for all models

    All random effects in the GE system will be assigned values

    Determine the parents for all crosses for the first cycle
    - Loop for all runs
      - ✓Loop for all strategies
        - » Loop for all cycles
- All strategies start from the same point (same crossing block and same crosses), so that they can be properly compared.

Parameters to describe a set of breeding strategies

- Number of strategies
- For each strategy
  - Strategy name: any character
  - Number of generations in the strategy: any integer more than 0
  - Definition for each generation

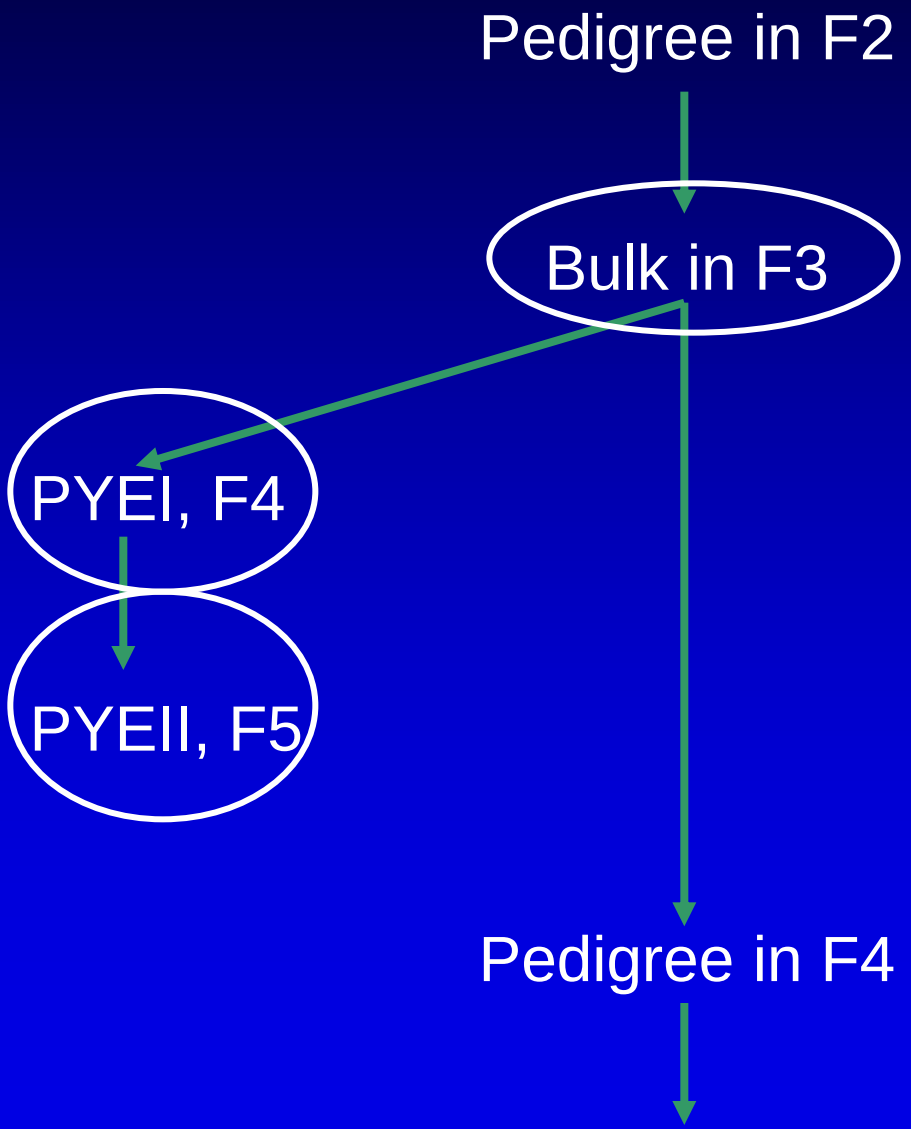
Definition of a generation

- Number of selection rounds in the generation: any integer more than 1
- Seed source indicator
  - 0: Seed for selection round *i* (*i* > 1) come from round 1
  - 1: Seed for selection round *i* (*i* > 1) come from round *i*-1
- Definition of each selection round

An example for seed source indicator 0

Practical breeding	Virtual breeding	Families selected
F6 field test at Toluca	F6	452 → 408 → 14,760
F7 field test at Cd. Obregon	F7	*Round 1 for F7* 14,760 → 3,868
F8 field test at Toluca and El Batan	F8 (T), F8 (B)	*Rounds 2 and 3* 3,868 → 2,163 → 1,974
F8 yield trial and F8 small plot evaluation at Cd. Obregon	F8 (YT) F8 small plot	*Round 4* 1,974 → 779 **779**

An example (LRC, Toowoomba, Australia) for seed source indicator 1



Families selected

Round 1 for F3

4000 → 1200

Round 2

1200 → 600

Round 3

600 → 100

100 → 1000

Definition of each selection round

- Title for the generation
- Seed propagation type (within a family)
  - *clone*, asexual reproduction
  - *DH*, doubled haploid
  - *self*, self-pollination
  - *backcross*, backcrossed to one parent
  - *topcross*, crossed with a third parent (three-way cross)
  - *doublecross*, crossed with another F1
  - *random*, random mating
  - *noself*, random mating but self-pollination is eliminated

Definition of each selection round

- Generation advance method (or harvest method): management of the selected individual plants in a family
  - *pedigree*: the selected plants in each family are harvested individually, resulting a few families in the next generation
  - *bulk*: the selected plants in each family are harvested in bulk, resulting one family in the next generation

Definition of each selection round

- Field experiment design
  - Number of replications for each family
  - Number of plants in each replication
  - Number of test locations
  - Environment type for each test location
    - defined in the GE system
    - if 0, randomly determined based on environment frequency

Definition of each selection round

- Among family and within family selection
  - Number of traits used for selection
  - Definition of each trait

Definition of each trait used in selection

- Trait number, for the trait in selection (0 when marker score is used in selection)
- Selection mode
  - **T for top, e.g. yield, tillering, grains per spike and 1000-kernel weight**
  - **B for bottom, e.g. lodging and rusts**
  - **M for middle, e.g. height and heading**
  - **R for random, for some special studies**
  - **TV for top value**
  - **BV for bottom value**
  - **TN for a number of individuals/families with top phenotypic values**
  - **BN for a number of individuals/families with bottom phenotypic values**
  - **RN for a number of individuals/families to be selected randomly**
- Selected proportion or value: the proportion or value of individual plants in a family (for within family selection) or of families (for among family selection) to be selected

Proportion selection

Threshold selection

Number selection

An example of generation definition

Rounds of selection	Seed source indicator	Generation title	Seed propagation type	Generation advance method	Replications	Plot size	Test locations	Environment type
1	0	F6	self	pedigree	1	750	1	2, Toluca
4	0	F7	self	bulk	1	70	1	1, Obregon
		F8(T)	self	bulk	1	70	1	2, Toluca
		F8(B)	self	bulk	1	70	1	3, El Batan
		F8(YT)	self	bulk	1	100	1	1, Obregon
1	0	F8(SP)	self	bulk	1	30	1	1, Obregon

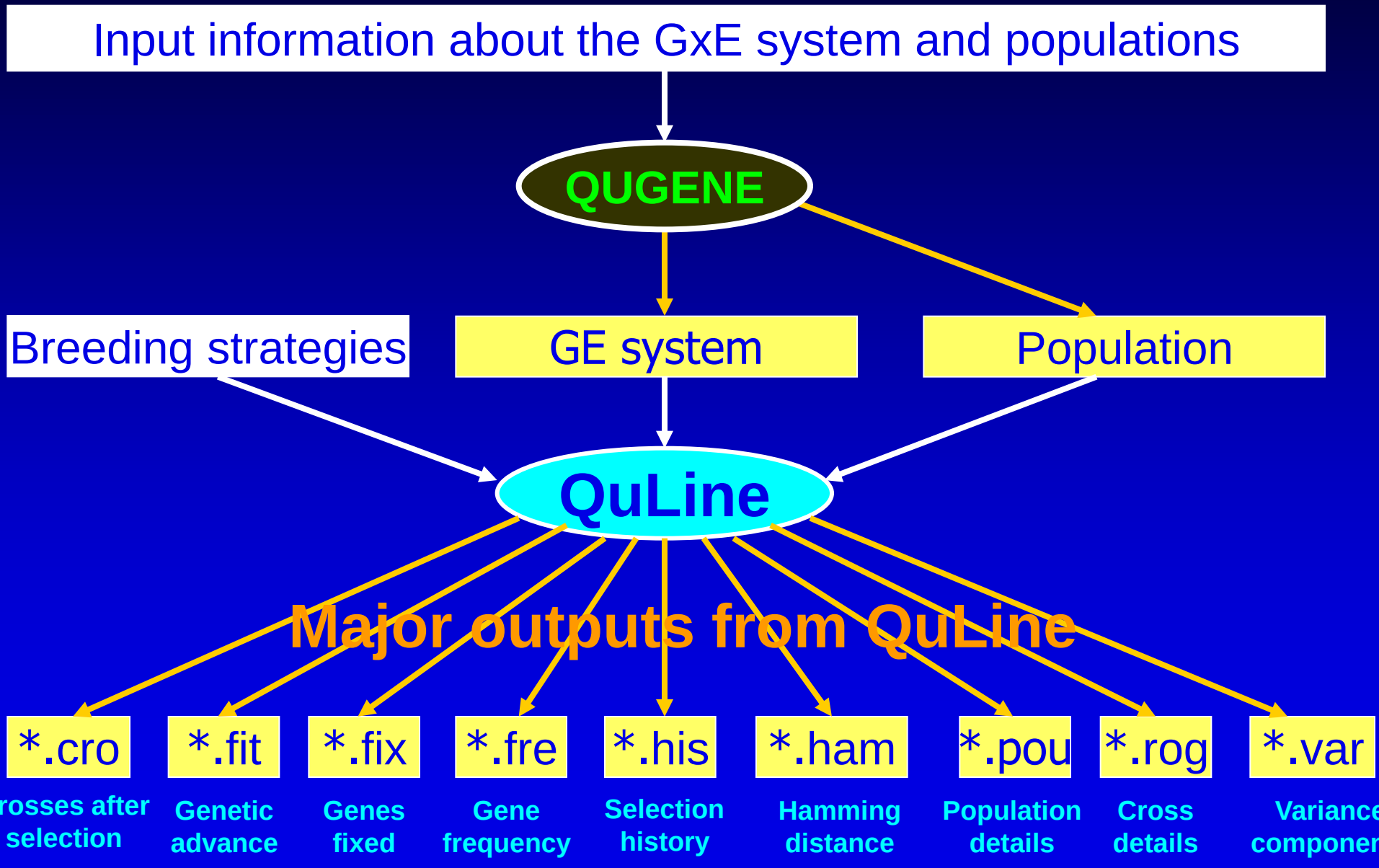
Traits, their selection modes and selected proportions

Trait	Yield	Lodg-ing	Stem rust	Leaf rust	Stripe rust	Height	Tillering	Heading	Grains per spike	1000 kernel weight	**Total**
Selection mode	**T**	**B**	**B**	**B**	**B**	**M**	**T**	**M**	**T**	**T**	
F6, among		**0.99**			**0.96**		**0.95**				**0.90**
F6, within		**0.90**			**0.70**	**0.90**	**0.95**	**0.98**	**0.10**		**0.05**
F7, among		**0.85**		**0.70**		**0.98**	**0.85**	**0.96**	**0.70**	**0.75**	**0.25**
F8(T), among		**0.55**			**0.70**	**0.99**	**0.98**	**0.99**	**0.90**		**0.55**
F8(B), among				**0.90**							**0.90**
F8(YT), among	**0.40**										**0.40**
F8(SP), among											**1.00**

In QuLine, a breeding program looks like …

```
!NR SS GT        PT         GA        RP PS    NL  ET...              Row 1
!                             AT (ID SP  SM)...                      Row 2
!                             WT (ID SP  SM)...                      Row 3
1   0  CB        self       bulk      1  10    1   2
                              0
                              0
1   0  F1     singlecross bulk       1  20    1   1
                              7   2 B 0.98  3 B 0.99  4 B 0.85  6 M 0.99  7 T 0.90  8 B 0.98  9 T 0.97
                              0
1   0  F2        self       pedigree  1  1000  1   2
                              7   2 B 0.99  3 B 0.99  5 B 0.90  6 M 0.99  7 T 0.99  8 B 0.99  9 T 0.99
                              8   2 B 0.95  4 B 0.99  5 B 0.40  6 M 0.85  7 T 0.60  8 B 0.90  9 T 0.50  10 T 0.60
1   0  F3        self       bulk      1  70    1   1
                              7   2 B 0.90  3 B 0.99  4 B 0.70  6 M 0.97  7 T 0.75  8 B 0.95  9 T 0.80
                              5   4 B 0.90  6 M 0.95  8 B 0.95  9 T 0.30 10 T 0.60
1   0  F4        self       bulk      1  70    1   2
                              6   2 B 0.90  5 B 0.65  6 M 0.97  7 T 0.85  8 B 0.97  9 T 0.85
                              5   5 B 0.90  6 M 0.95  8 B 0.95  9 T 0.30 10 T 0.60
1   0  F5        self       bulk      1  70    1   1
                              6   2 B 0.90  4 B 0.75  6 M 0.97  7 T 0.85  8 B 0.95  9 T 0.85
                              5   4 B 0.90  6 M 0.95  8 B 0.95  9 T 0.30 10 T 0.60
1   0  F6        self       pedigree  1  140   1   2
                              6   2 B 0.90  5 B 0.75  6 M 0.97  7 T 0.85  8 B 0.97  9 T 0.85
                              5   5 B 0.90  6 B 0.90  7 T 0.95  8 B 0.95  9 T 0.10
4   0  F7        self       bulk      1  70    1   1
                              7   2 B 0.90  4 B 0.75  6 M 0.97  7 T 0.90  8 B 0.95  9 T 0.85 10 T 0.75
                              0
       AL(T)     self       bulk      1  70    1   2
                              6   2 B 0.95  5 B 0.90  6 M 0.99  7 T 0.98  8 B 0.99  9 T 0.85
                              0
       AL(B)     self       bulk      1  70    1   3
                              1   4 B 0.90
                              0
       PYT       self       bulk      1  100   1   1
                              1   1 T 0.40
                              0
```

Steps to run QuLine



Comparison of two breeding strategies: modified pedigree (MODPED) and selected bulk (SELBLK)

Crop Science, 2003, 43: 1764-1773
Crop Science, 2004, 44: 2006-2018

Breeding methods with self-pollinated crops

Allard, R.W. 1960. Principles of plant breeding. John Wiley & Sons, Inc.

Stoskopf, N.C. 1993. Plant Breeding — Theory and Practice. Westview Press.

- Mass and pure-line selection
- The pedigree system
- The bulk population method
- The backcross breeding method
- Single seed descent (SSD), a special case of pedigree system
- Recurrent selection breeding method
- Mutation breeding
- Haploid breeding system (doubled haploid)

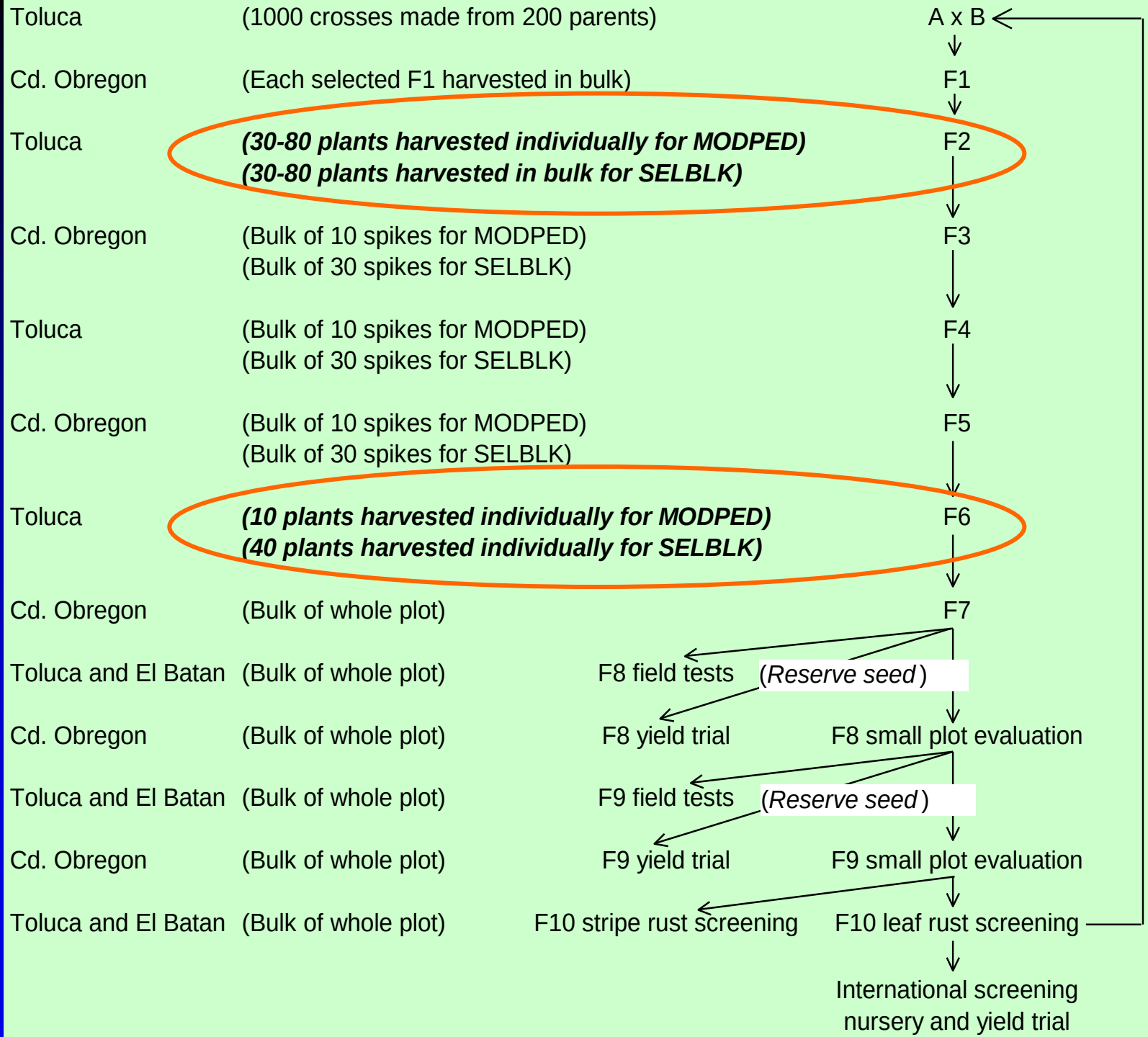
Breeding methods in CIMMYT's wheat breeding program

- Pedigree system: before 1984.
  - "Pedigree selection" is used from F2 to F6.
- Modified pedigree/bulk (MODPED): in 1985-1989/94.
  - "Pedigree selection" is used in F2 and F6, and "bulk selection" is used in other generations.
- Selected bulk (SELBLK): after 1995.
  - "Pedigree selection" is used only in F6, and "bulk selection" is used in other generations.

Germplasm flow for simple crosses made in Toluca and targeted to ME1

MODPED: modified pedigree/bulk

SELBLK: selected bulk methods



Trait, segregating gene number, gene effects and trait heritability

Trait	Genes	Gene effect type	AA	Aa	aa	Trait range	h_b^2 (Indiv. plant)
Yield	20, 40	E0, E1, E2	Random value from UD (0, 1)				0.05
Lodging	3	additive	0	5	10	0-30	0.10
Stem rust	5	additive	0	0.5	1	0-5	0.30
Leaf rust	5	additive	0	5	10	0-50	0.30
Yellow rust	5	additive	0	5	10	0-50	0.30
Height	3	additive	40	30	20	120-60	0.45
Tillers/plant	3	additive	5	3	1	15-3	0.35
Heading	5	additive	20	16	12	100-60	0.30
Grains/spike	5	additive	14	10	6	70-30	0.35
Seed weight	5	additive	12	8.5	5	60-25	0.35

Trait correlation and pleiotropy

Trait	Yield	Lodging	Stem rust	Leaf rust	Yellow rust	Height	Tillers/ plant	Heading	Grains/ spike	Seed weight
Yield		-0.50	-0.20	-0.10	-0.10	-0.50	0.40	0.30	0.50	0.40
Lodging	***-0.56***									
Stem rust	***-0.25***									
Leaf rust	***-0.05***									
Yellow rust	***-0.09***									
Height	***-0.62***									
Tillers/plant	***-0.08***								-0.20	-0.40
Heading	***0.60***									
Grains/spike	***0.09***						***-0.17***			-0.30
Seed weight	***-0.07***						***-0.30***		***-0.07***	

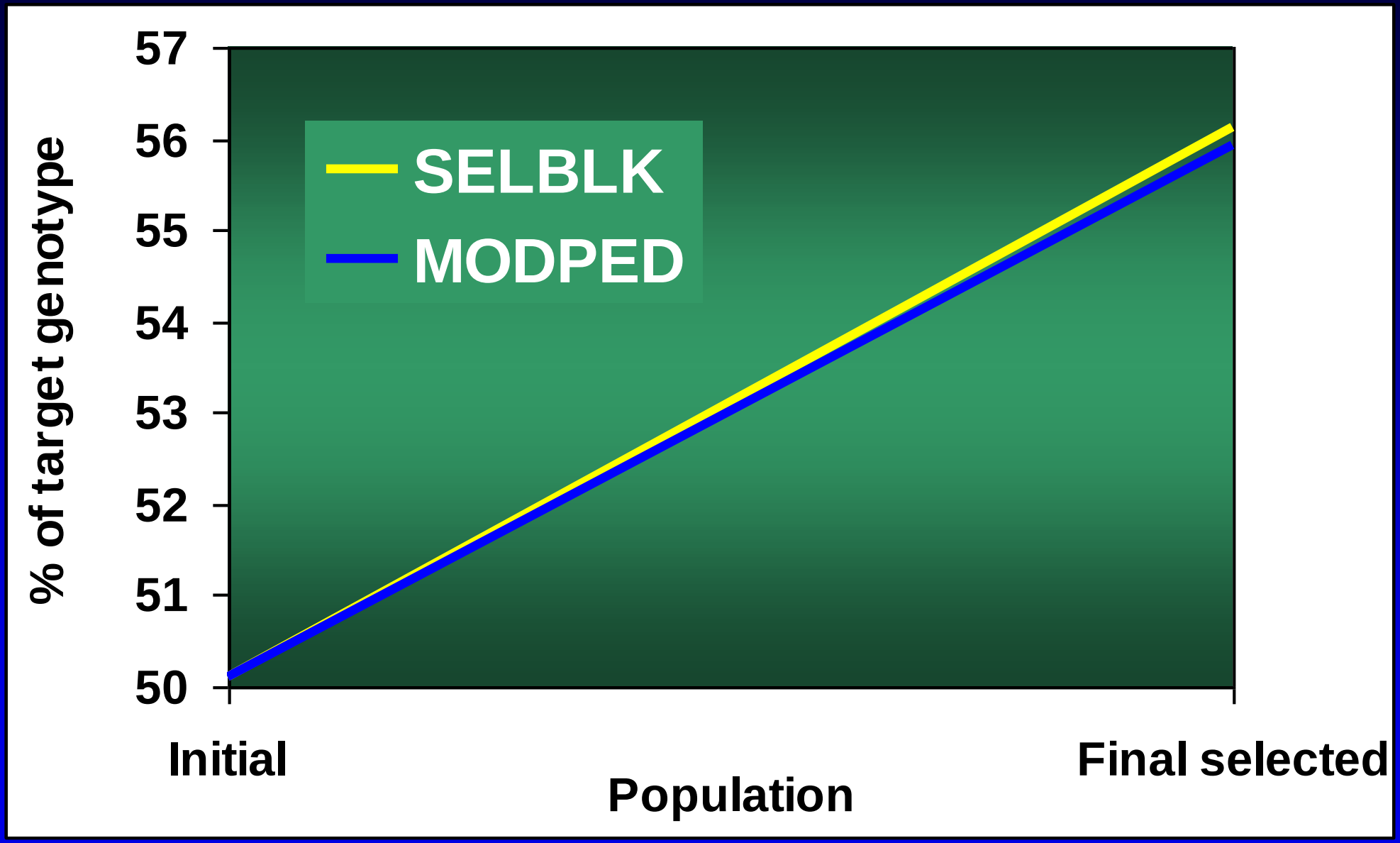
Estimated by CIMMYT breeders

Estimated from the defined genetic model

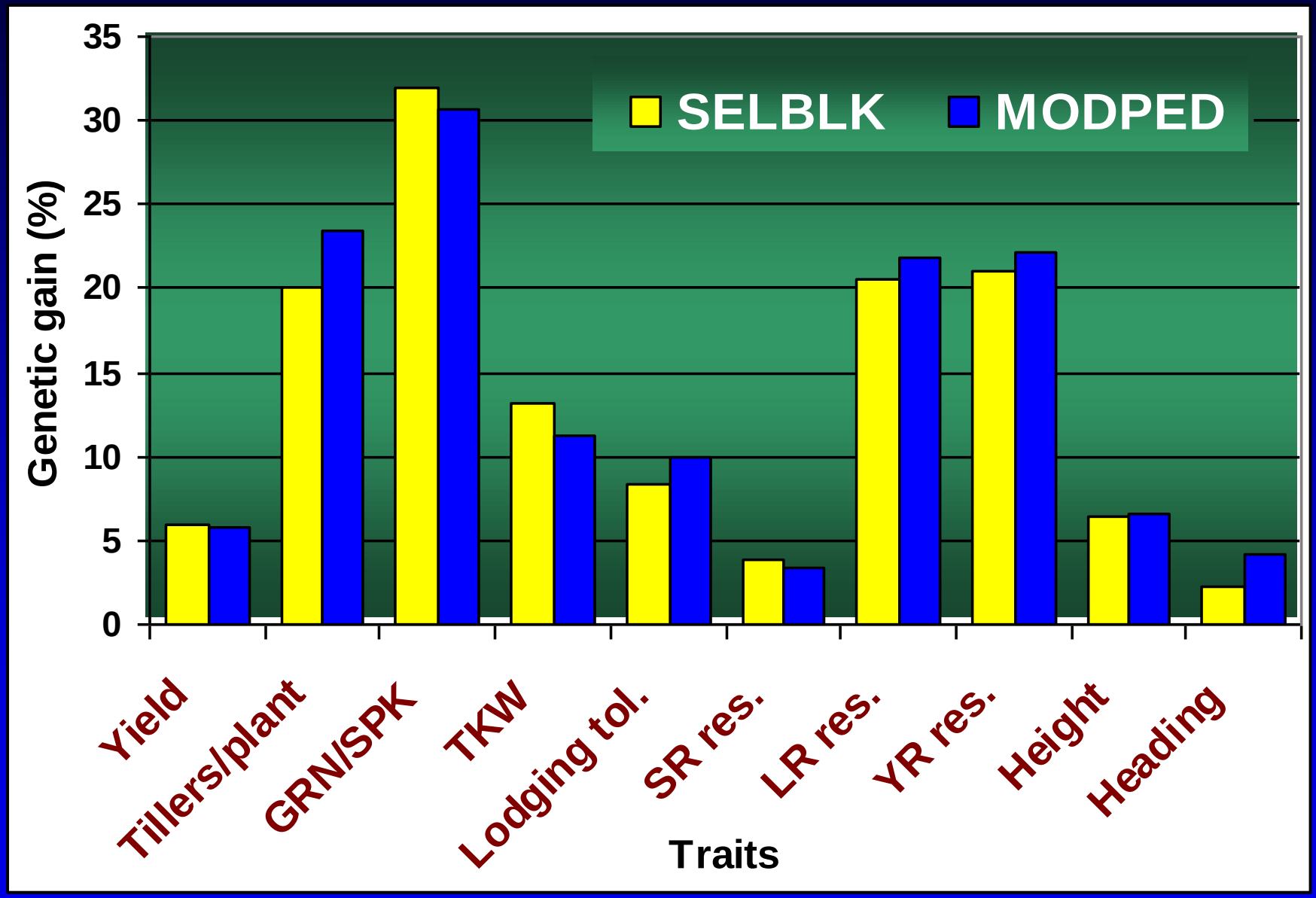
Experiment design

- 12 Genotype and environment (GE) systems
  - Two yield gene numbers: 20 and 40, two alleles for each gene
  - Pleiotropy (same gene effects various traits): absent and present
  - Epistasis (multiple gene interaction): no epistasis, digenic-epistasis, and tri-genic epistasis
  - Linkage: no linkage (independent gene segregation)
- Initial population
  - 200 parents, gene (allele) frequencies of 0.5 for all genes
- 1000 crosses were made
- 258 lines were selected after 10 generations of selection

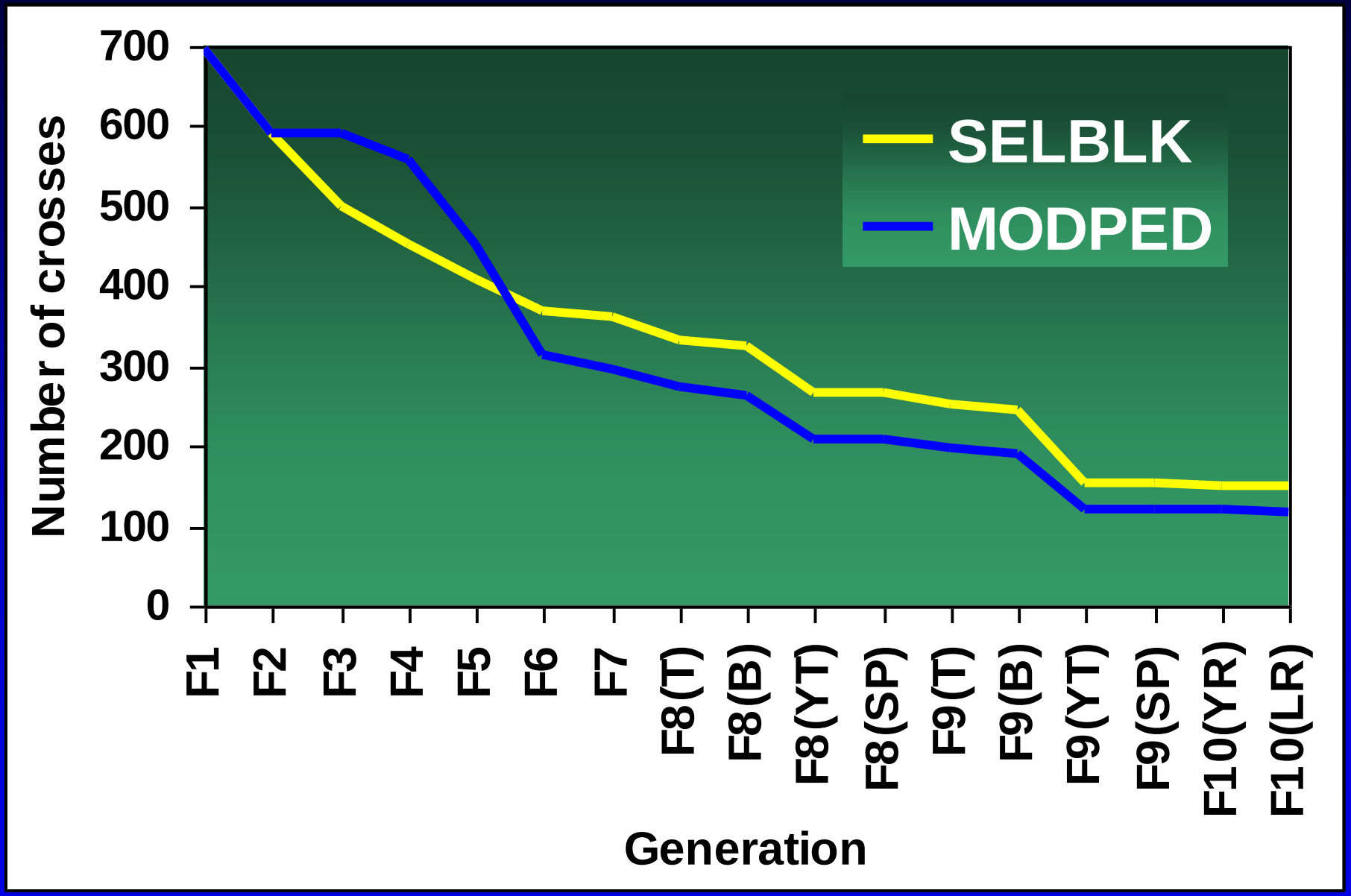
Result 1: Genetic gain in yield from SELBLK is 3.3% higher than MODPED. SELBLK is slightly more efficient.



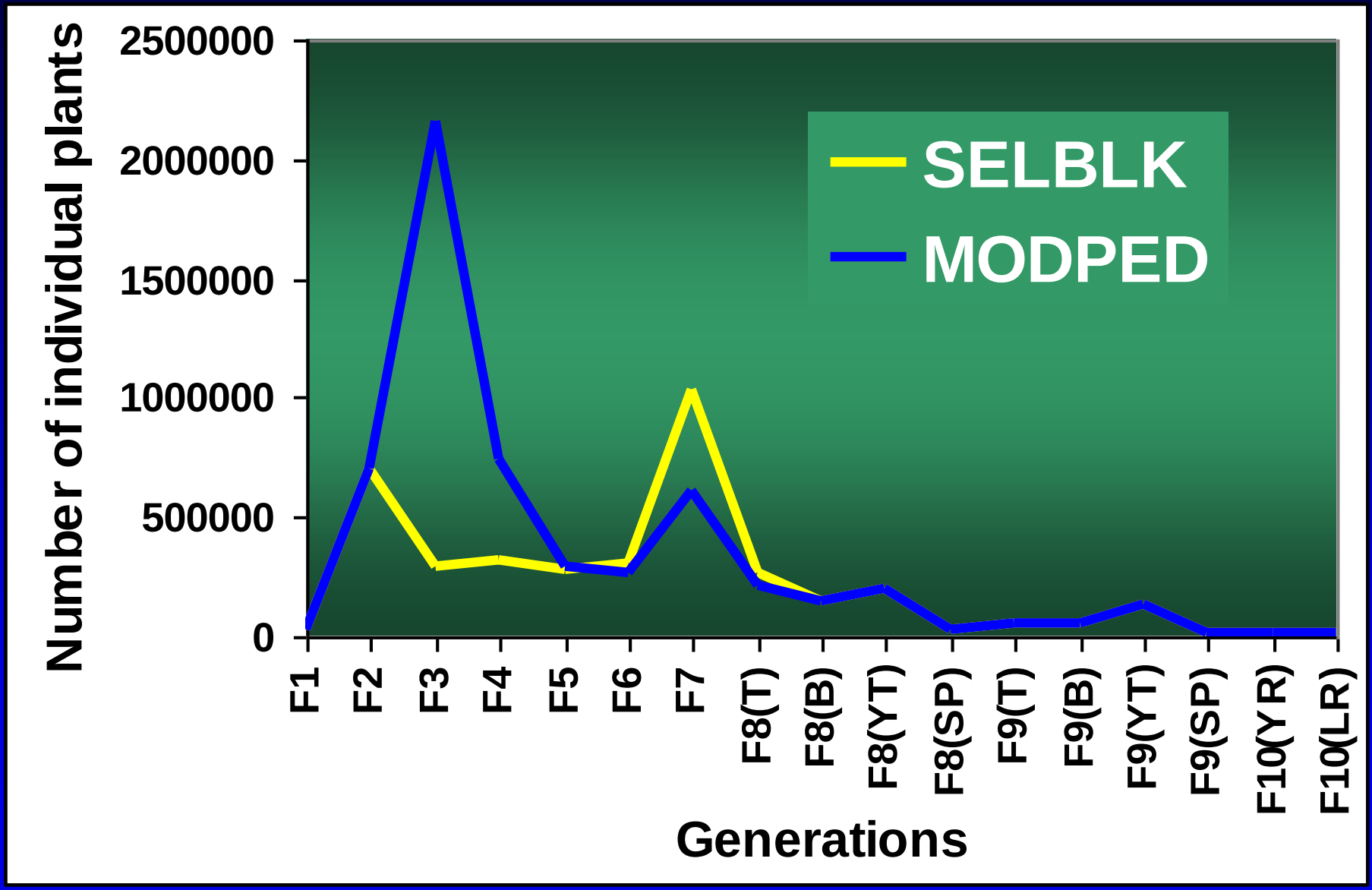
For gains per spike and 1000-kernel weight, SELBLK has a faster genetic gain. For tillers/plant, MODPED has a faster genetic gain.



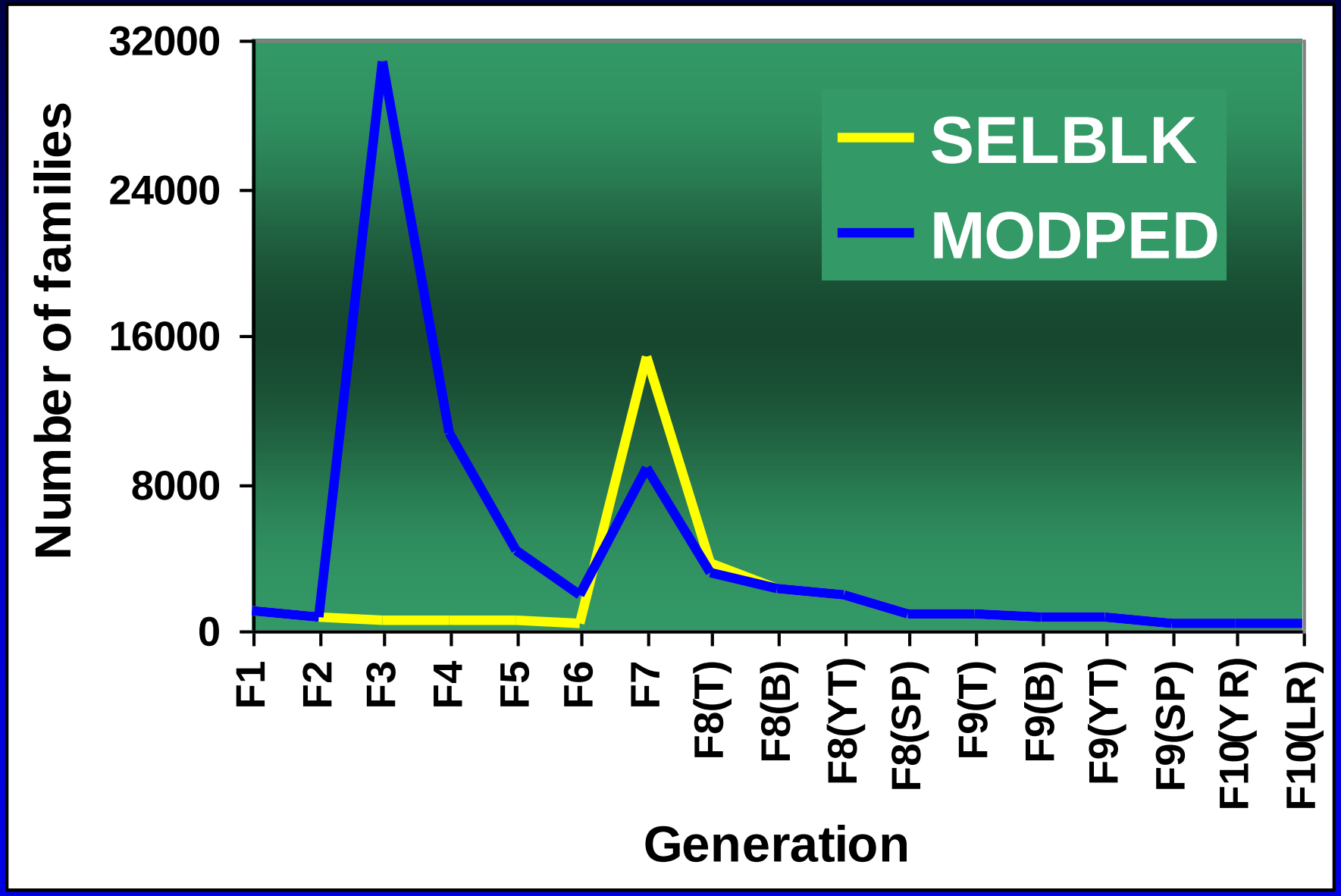
Result 2: SELBLK retained 25% more crosses in the final selected population (more genetic diversity retained)



Result 3: SELBLK required 1/3 less land from F1 to F8 than MODPED. SELBLK is more cost-effective.



Result 4: SELBLK produced 40% less families (plots) to be planted from F1 to F8 (less labor required)



Modeling of the Single Backcrossing Breeding Strategy (SBBS)

Theor. Appl. Genet., 2009, 118: 683-694

Estimated percentages of favourable alleles or gene combinations in different parental lines in wheat breeding at CIMMYT

Category	favorable genes (%)	Example	% total lines
Elite adapted lines (EAL)	80-85	Major released cultivars in targeted mega-environments (MEs) either developed by CIMMYT or by partners	10
Adapted lines (AL)	75-80	Elite advanced lines from CIMMYT's International Nursery and Yield Trials	60
Intermediate adapted lines (IAL)	65-75	Advanced lines from CIMMYT's Yield Trials in Ciudad Obregón and Toluca, Mexico	10
Un-adapted (or non-adapted) lines (UAL)	20-40	Land races	2
Second generation of re-synthesized wheat (SYNII)	40-60	Derived lines between the first generation of re-synthesized wheat derivatives and adapted lines	10
First generation of re-synthesized wheat (SYNI)	20-40	Derived lines between primary re-synthesized wheat and adapted lines	5
Primary re-synthesized wheat (SYN0)	0-30	Inter-specific crosses between *Triticum durum* and *Aegilops tauschii*	3

Two traits defined in QU-GENE

- Adaptation
  - 200 genes on the 21 wheat chromosomes
  - Lowest adaptation with no favorable alleles: 0
  - Highest adaptation with all favorable alleles: 100
  - Heritability: 0.5
- Donor traits (DT) to be transferred
  - 10 genes governing the donor traits
  - Lowest DT with no favorable alleles: 0
  - Highest DT with all favorable alleles: 10
  - Heritability: 0.5

Adapted parental groups in simulation

Gene frequency of favorable adaptation alleles fixed at 0.8

- A0: the frequency of favorable DT alleles is 0
- A2: the frequency of favorable DT alleles is 0.2
- A4: the frequency of favorable DT alleles is 0.4
- A6: the frequency of favorable DT alleles is 0.6
- A8: the frequency of favorable DT alleles is 0.8

Donor parental groups in simulation

Gene frequency of favorable DT alleles fixed at 1.0

- D0: the frequency of favorable adaptation alleles is 0
- D1: the frequency of favorable adaptation alleles is 0.1
- D2: the frequency of favorable adaptation alleles is 0.2
- D3: the frequency of favorable adaptation alleles is 0.3
- D4: the frequency of favorable adaptation alleles is 0.4
- D5: the frequency of favorable adaptation alleles is 0.5
- D6: the frequency of favorable adaptation alleles is 0.6
- D7: the frequency of favorable adaptation alleles is 0.7

Crosses made between different parental groups for wheat breeding at CIMMYT

Category	Percentage of total crosses	Similarity to defined parental groups
(EAL+AL) × (EAL+AL)	65	(A0+A2+A4+A6+A8) ×D7
(EAL+AL) × IAL	10	(A0+A2+A4+A6+A8) ×D5+D6)
(EAL+AL) × UAL	5	(A0+A2+A4+A6+A8) ×(D2+D3+D4)
(EAL+AL) × SYNII	10	(A0+A2+A4+A6+A8) ×(D6+D7)
(EAL+AL) × SYNI	7	(A0+A2+A4+A6+A8) ×(D4+D5)
(EAL+AL) × SYN0	3	(A0+A2+A4+A6+A8) ×(D0+D1+D2)

The single backcrossing breeding strategy (SBBS)

Generation	Seed propagation method	No. crosses or families grown	Individuals per cross or family	No. selected crosses or families	No. selected individuals in each cross or family
F_1	Hand pollination between adapted and donor lines	100	20	100	20
B_1F_1	Backcrossing to the adapted parents	100	400	100	50
B_1F_2	Selfing	100	1200	100	30
B_1F_3	Selfing	100	400	100	10
B_1F_4	Selfing	100	400	100	10
B_1F_5	Selfing	100	400	100	10
B_1F_6	Selfing	1000	200	30	200
Final selected advanced lined		10			

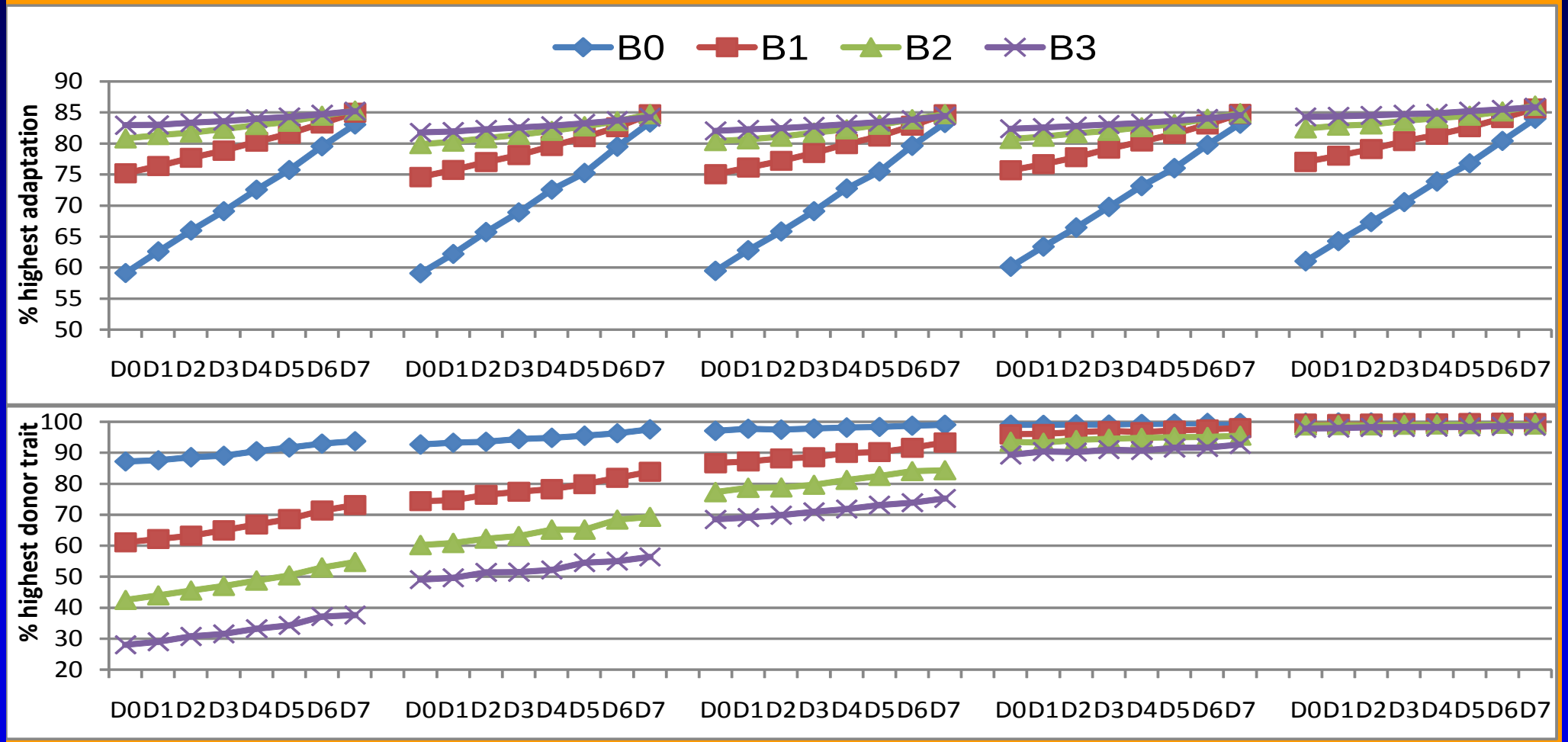
Four crossing strategies defined in QuLine

No backcross	One time of backcross	Two times of backcross	Three times of backcross	Generation advance method
B0	B1	B2	B3	
F_1	F_1	F_1	F_1	Bulk
F_2	BC_1F_1	BC_1F_1	BC_1F_1	Bulk
F_3	BC_1F_2	BC_2F_1	BC_2F_1	Bulk
F_4	BC_1F_3	BC_2F_2	BC_3F_1	Bulk
F_5	BC_1F_4	BC_2F_3	BC_3F_2	Bulk
F_6	BC_1F_5	BC_2F_4	BC_3F_3	Pedigree
F_7	BC_1F_6	BC_2F_5	BC_3F_4	Bulk

Six selection schemes

- **AD:** Adaptation is selected first, followed by the selection for DT
- **DA:** DT is selected first, followed by the selection for Adaptation
- **ADA:** Adaptation is selected first, followed by the selection for DT, and adaptation is selected again
- **DAD:** DT is selected first, followed by the selection for Adaptation, and DT is selected again
- **ADAD:** Adaptation and DT are selected two times in each generation, and adaptation is selected first
- **DADA:** Adaptation and DT are selected two times in each generation, and DT is selected first

Genetic advance of selection scheme AD



Conclusions

- We recommend the use of SBBS based on three assumptions:
  - multiple genes governing the phenotypic traits to be transferred from donor parents to adapted parents
  - donor parents still have some favorable genes that may contribute to the improvement of adaptation in the recipient parents even under low adaptation
  - the conventional phenotypic selection is applied or the individual genotypes cannot be precisely indentified

Breeding with known gene information

Aust. J. Agric. Res., 2005, 56: 465-473

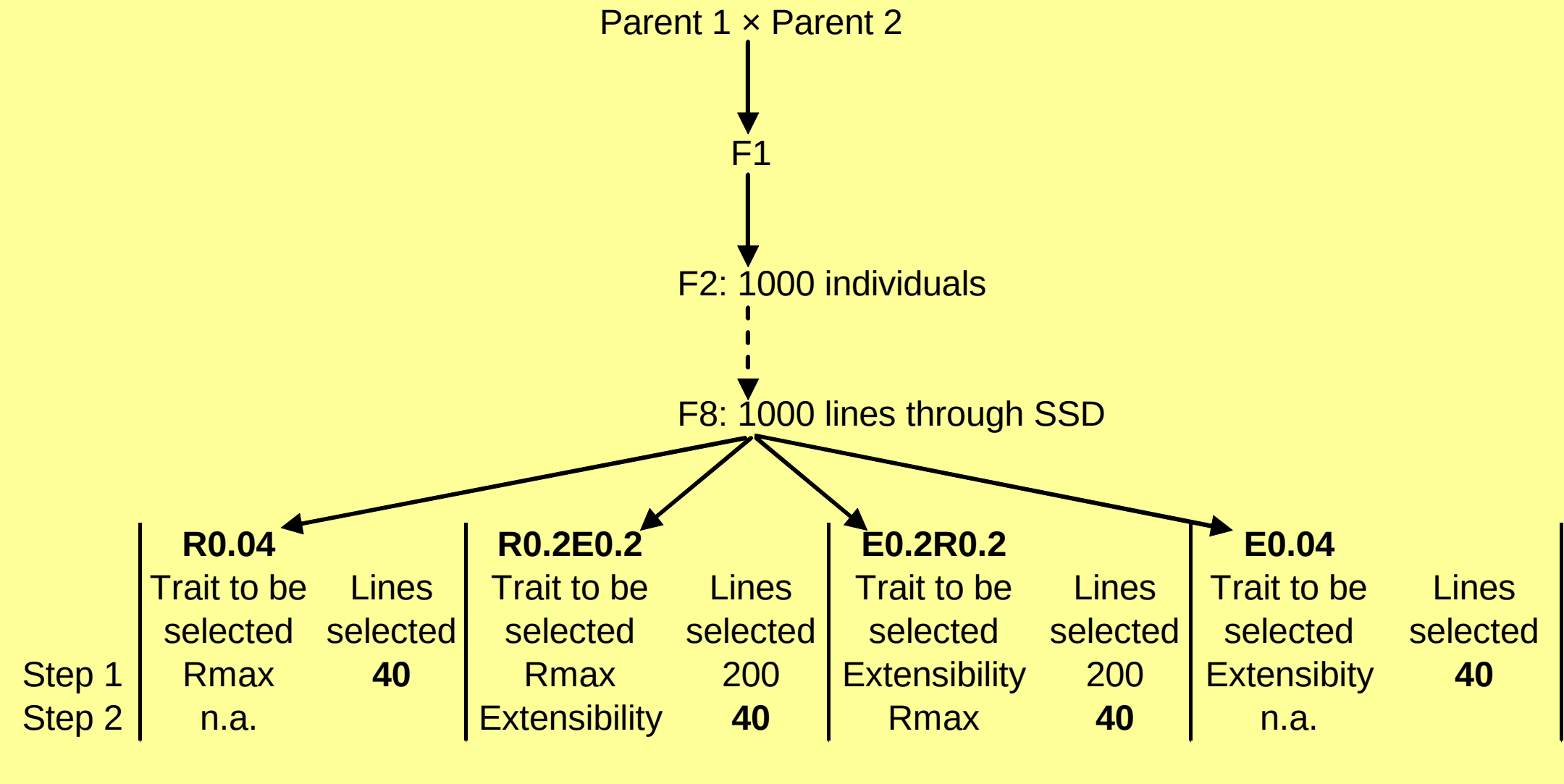
Glutenin genes and wheat quality

- Six glutenin genes
  - *Glu-A1* (1AL), *Glu-B1* (1BL), *Glu-D1* (1DL) for HMW
  - *Glu-A3* (1AS), *Glu-B3* (1BS), *Glu-D3* (1DS) for LMW
- Two end-use quality traits commonly used by wheat breeders
  - Rmax (BU), for maximum dough resistance
  - Extensibility (cm), for dough extensibility
- Multiple alleles on each gene locus
  - *Glu-A1*: *1*, *2*,* and *Null*
  - *Glu-B1*: *7*, *7+8*, *7+9*, *6+8*, *20*, *13+16*, *14+15*, *17+18*, and *23+24*
  - *Glu-D1*: *2+12*, *4+12*, *5+10*, and *2+T2*

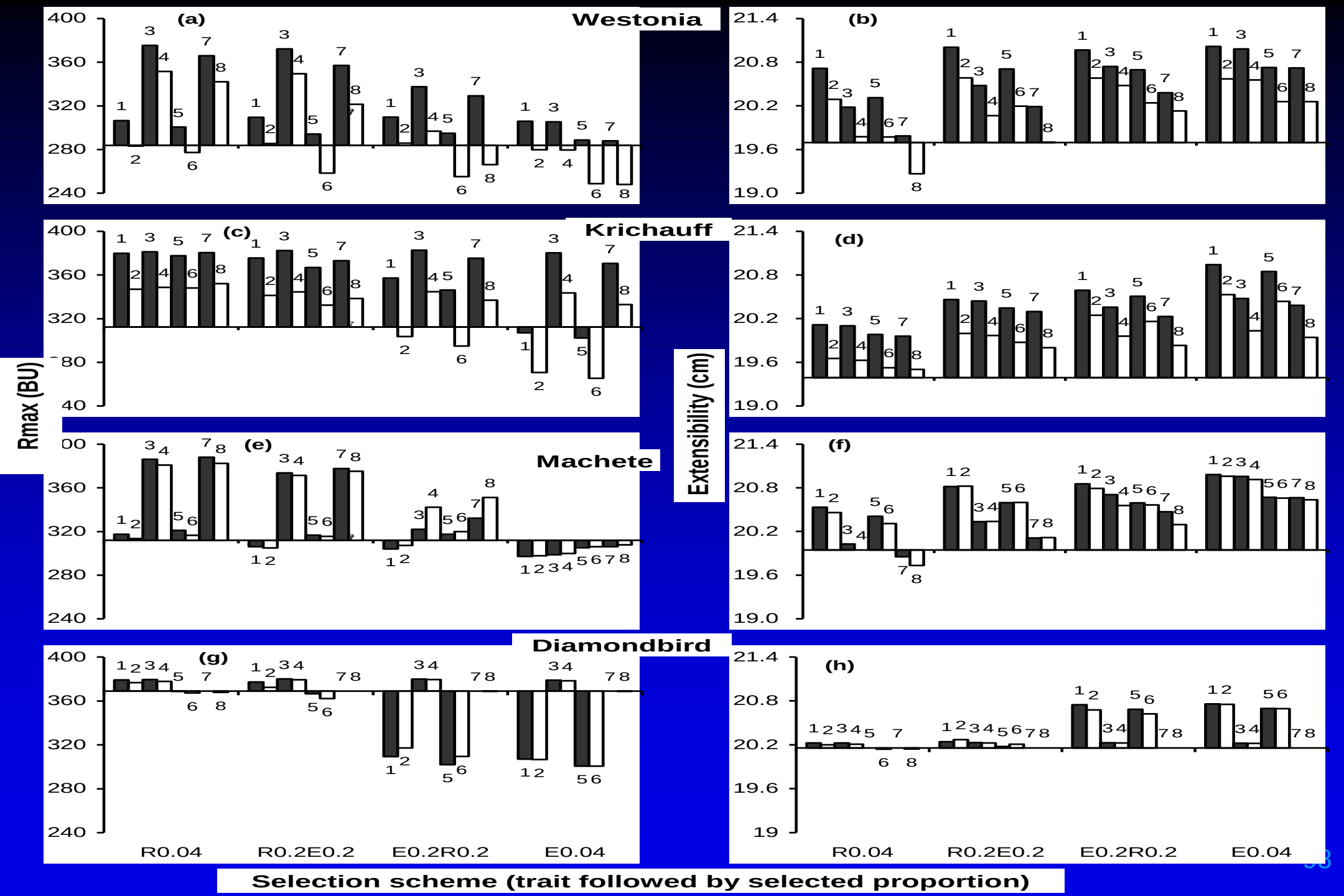
Selected parents

Parent	Rmax	Extensibility
Silverstar 1	309.80	20.78
Silverstar 2	270.20	20.31
…	…	…
Silverstar 7	368.90	20.15
Silverstar 8	329.30	19.68
Westonia	283.70	19.70
Krichauff	312.26	19.39
Machete	312.03	19.95
Diamondbird	368.88	20.16

Four selection schemes



	R0.04		**R0.2E0.2**		**E0.2R0.2**		**E0.04**	
	Trait to be selected	Lines selected	Trait to be selected	Lines selected	Trait to be selected	Lines selected	Trait to be selected	Lines selected
Step 1	Rmax	**40**	Rmax	200	Extensibility	200	Extensibity	**40**
Step 2	n.a.		Extensibility	**40**	Rmax	**40**	n.a.	


Westonia
Krichauff
Machete
Diamondbird
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Rmax (BU)
Extensibility (cm)
R0.04
R0.2E0.2
E0.2R0.2
E0.04
Selection scheme (trait followed by selected proportion)

The best sister lines under each breeding objective and selection scheme

Parent	Objective	R0.04	R0.2E0.2	E0.2R0.2	E0.04
Westonia	High Rmax	3, 7	3, 7	3, 7	1, 3
	High Ext.	1	1, 5	1, 3, 5	1,3,5,7
Krichauff	High Rmax	1,3,5,7	1,3,5,7	3, 7	3, 7
	High Ext.	1,3,5,7	1,3,5,7	1, 5	1, 5
Machete	High Rmax	3,4,7,8	3,4,7,8	4, 8	None
	High Ext.	1,2,5,6	1,2,5,6	1, 2, 3	1,2,3,4
Diamondbird	High Rmax	1,2,3,4	1, 3, 4	3, 4	3, 4
	High Ext.	None	None	1,2,5,6	1,2,5,6

Efficient selection of multiple genes via marker-assisted selection, an example in wheat

Wang, J.,* S.C. Chapman, D.B. Bonnett, G.J. Rebetzke, and J. Crouch. 2007. Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. **Crop Science 47**: 580-588.

Nine major genes to be pyramided in wheat

Gene	***Rht-B1***	***Rht-D1***	***Rht8***	***Sr2***	***Cre1***	***VPM***	***Glu-B1***	***Glu-A3***	***tin***
Chr.	**4BS**	**4DS**	**2DL**	**3BS**	**2BL**	**7DL**	**1BL**	**1AS**	**1AS**
Marker	Codom	Codom	Codom	Codom	Dom	Dom	Codom	Codom	Codom
MK-gene distance	0	0	0.6	1.1	0	0	0	0	0.8
HM14BS	***Rht-B1a***	***Rht-D1a***	***Rht8***	*sr2*	*cre1*	*vpm*	*Glu-B1a*	*Glu-A3e*	*Tin*
Sunstate	***Rht-B1a***	*Rht-D1b*	*rht8*	***Sr2***	*cre1*	***VPM***	***Glu-B1i***	***Glu-A3b***	*Tin*
Silverstar+ tin	*Rht-B1b*	***Rht-D1a***	*rht8*	*sr2*	***Cre1***	*vpm*	***Glu-B1i***	*Glu-A3c*	***tin***
Target	***Rht-B1a***	***Rht-D1a***	***Rht8***	***Sr2***	***Cre1***	***VPM***	***Glu-B1i***	***Glu-A3b***	***tin***

One strategy identified by QuLine to combine the nine genes from topcross

- Selection of **Sunstate as the final parent** (having largest number of favorable alleles) in the topcross

- Stage I: Selection for *Rht-B1a* and *Glu-B1i* homozygotes, and enrichment of *rht8*, *Cre1*, and *tin* in TCF1
- Stage II: Selection of homozygotes for one target allele, e.g. *Rht8*, and enrich remaining target alleles in TCF2
- Stage III: Selection of the target genotype in DHs/RILs

Comparison with other strategies

- For this strategy, one target genotype can be selected by screening **< 600 individuals/lines**

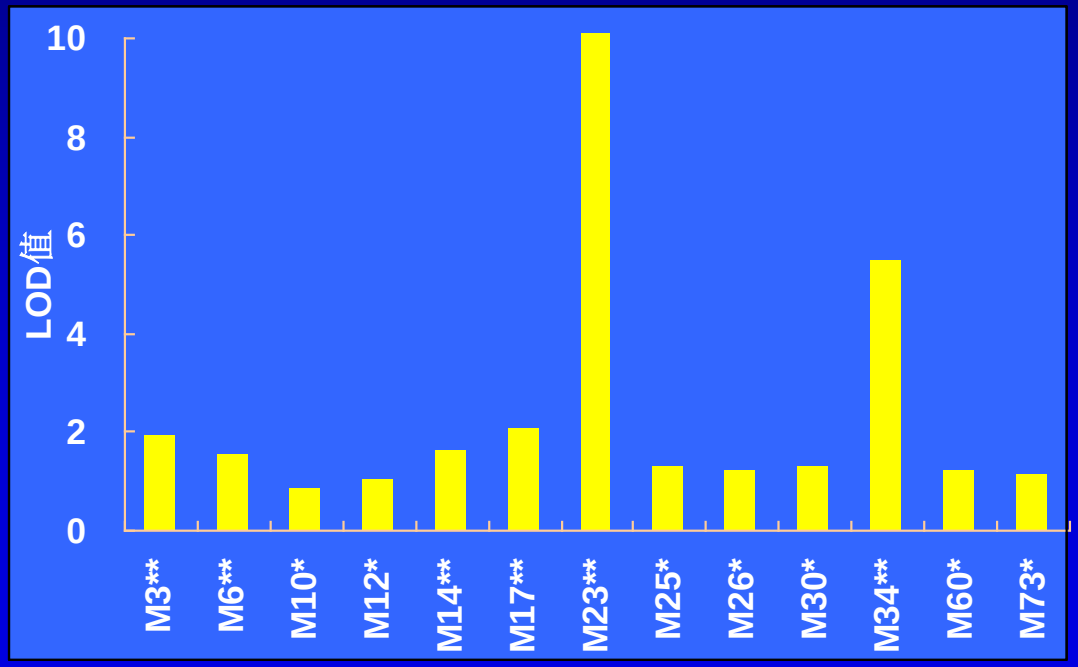
- For one-stage selection in advanced generations, one target genotype can be selected be screening **> 3500 lines**
- For one-stage selection in early generations, say TCF2, one target genotype can be selected be screening **millions of individuals**

Using QTL mapping results to design the breeding program

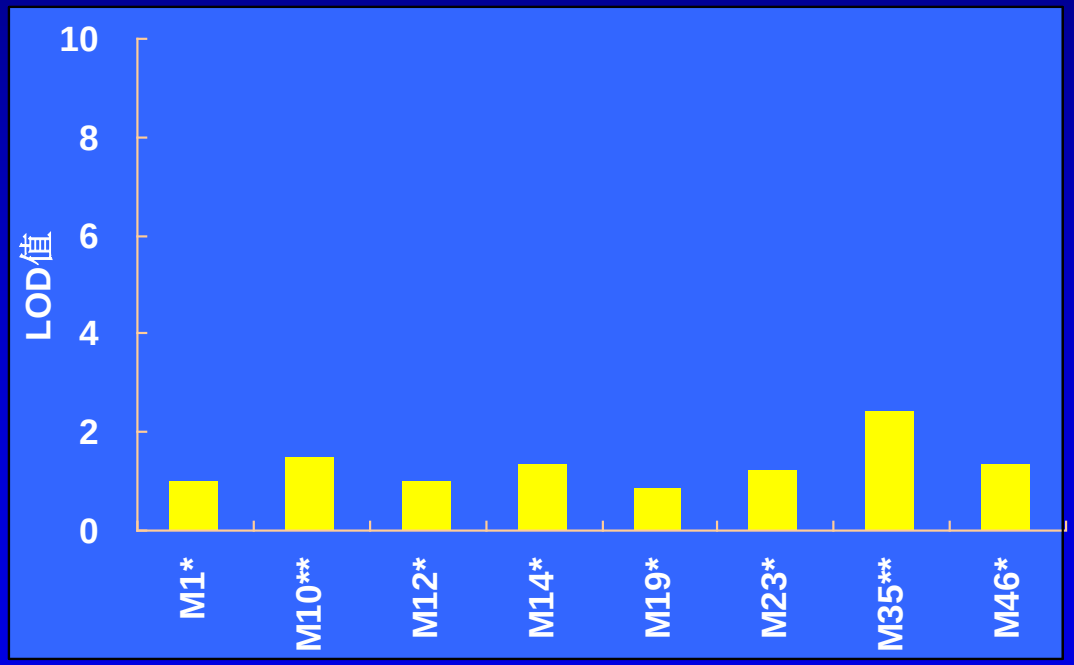
Genetical Research, 2006, 88: 93-104
Theor. Appl. Genet., 2007, 115: 87-100

QTL for grain length and grain width in rice using 65 CSS lines

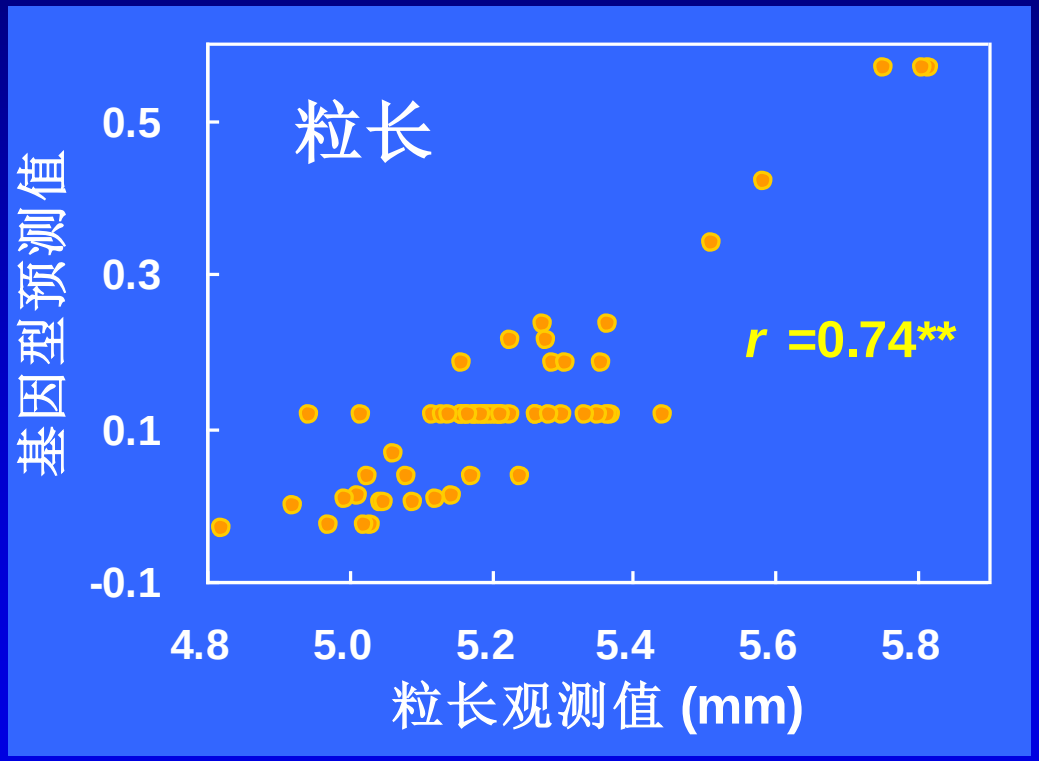
Grain length

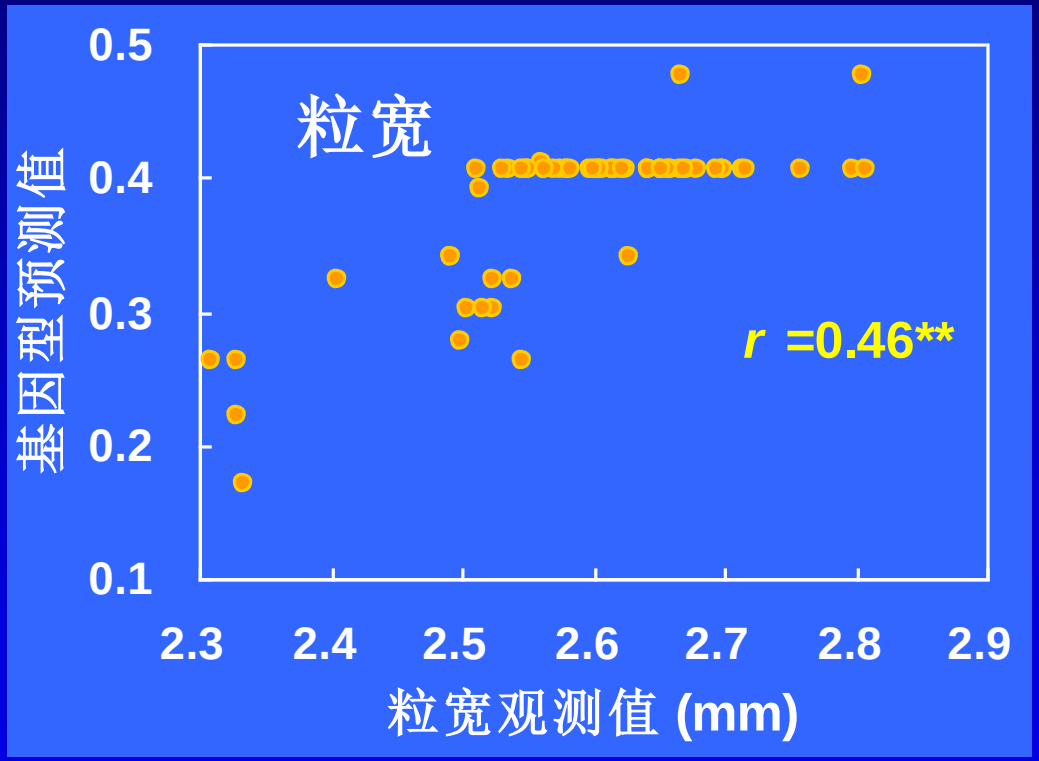


Grain width

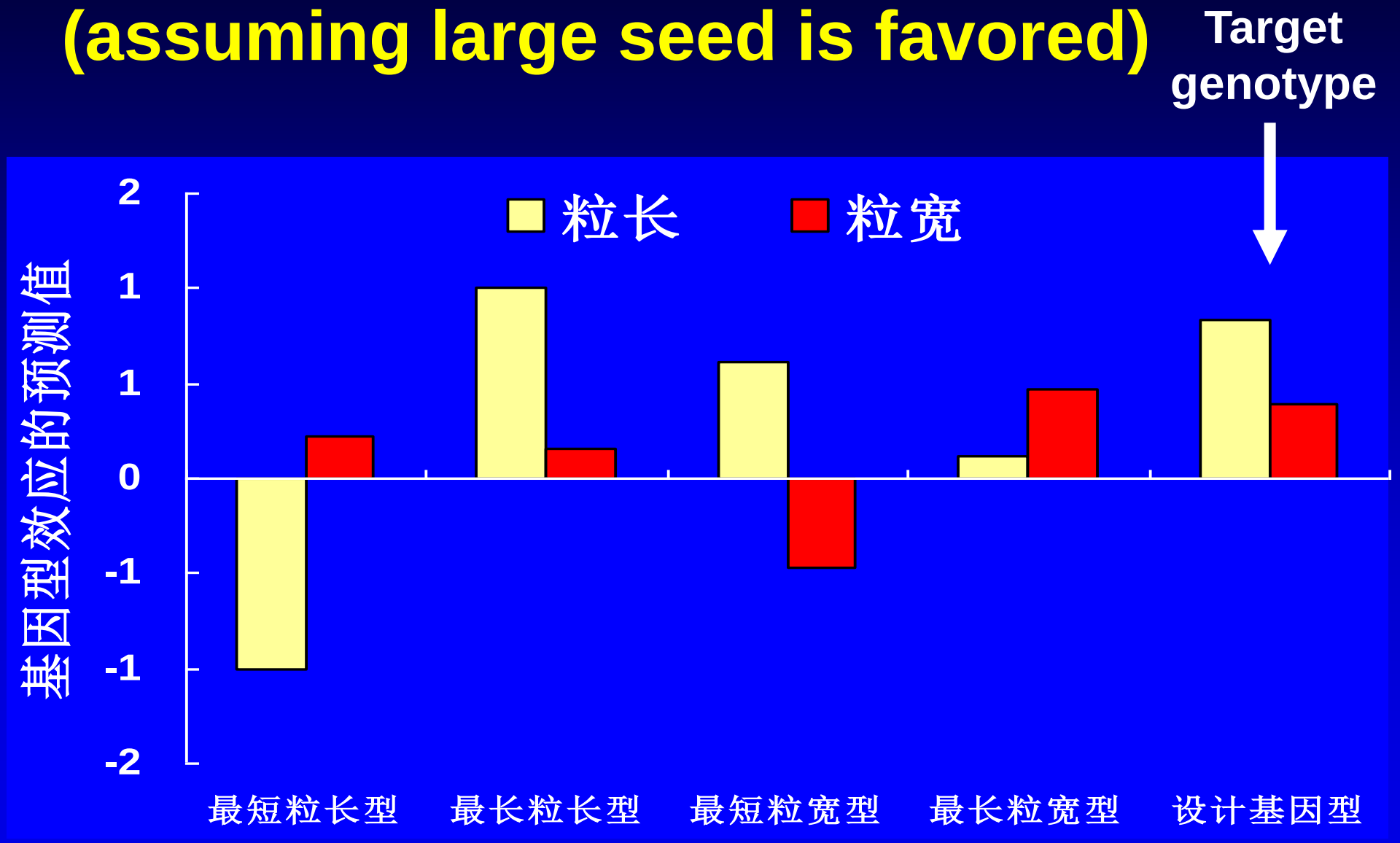


Prediction using identified QTL





Designing the target genotype
(assuming large seed is favored)



Choosing parental lines

Genotype	M1	M6	M12	M23	M25	Prediction in	
						GL (mm)	GW (mm)
Longest grain	1	2	2	2	2	6.21	2.74
Widest grain	2	1	1	1	1	5.32	3.07
Designed TG	2	2	1	2	2	6.05	2.98
CSSL5	1	2	1	1	1	5.44	3.00
CSSL16	2	1	1	2	1	5.77	2.98
CSSL19	1	1	2	1	2	5.54	2.93

Achieving the designed TG

- Three option for crossing
  - TC1：(CSSL5×CSSL16)×CSSL19
  - TC2：(CSSL5×CSSL19)×CSSL16
  - TC3：CSSL5×(CSSL16×CSSL19)
- Two MAS schemes (just for example)
  - Scheme1：MAS in F8 only
  - Scheme2：MAS in F2 and F8

Choosing the best crossing and selection schemes

TC	MAS	Selected Ind. In F2	F8 families before MAS	F8 families after MAS (S.E.)	DNA samples	DNA samples per TG
TC1	Scheme1	**100**	**3000**	**7.6 (3.27)**	**3000**	**395**
	Scheme2	**12.0**	**359**	**7.6 (3.37)**	**459**	**60**
TC2	Scheme1	**100**	**3000**	**24.3 (7.06)**	**3000**	**123**
	Scheme2	**24.8**	**745**	**23.3 (7.16)**	**845**	**36**
TC3	Scheme1	**100**	**3000**	**11.2 (5.45)**	**3000**	**268**
	Scheme2	**7.5**	**226**	**12.3 (5.14)**	**326**	**26**

Demonstration of QuLine