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

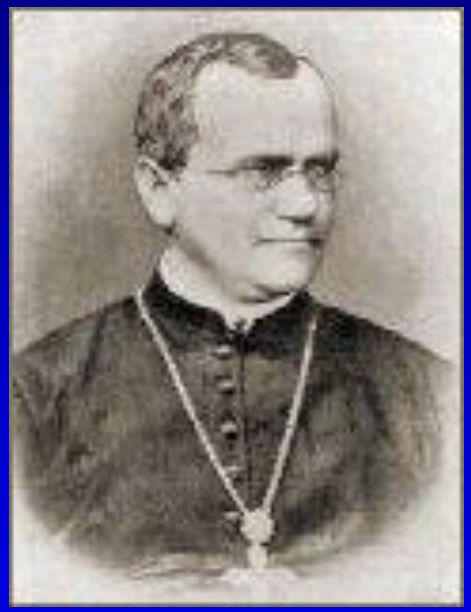
Outlines for presentation

- 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)

Qualitative traits in genetics

- Seed shape: round vs wrinkled
- Cotyledon color: yellow vs green
- Seed coat color: grey-brown vs white
- Pod shape: inflated vs constricted
- Unripe pod color: green vs yellow
- Flower position: axial vs terminal
- Stem length: long vs short

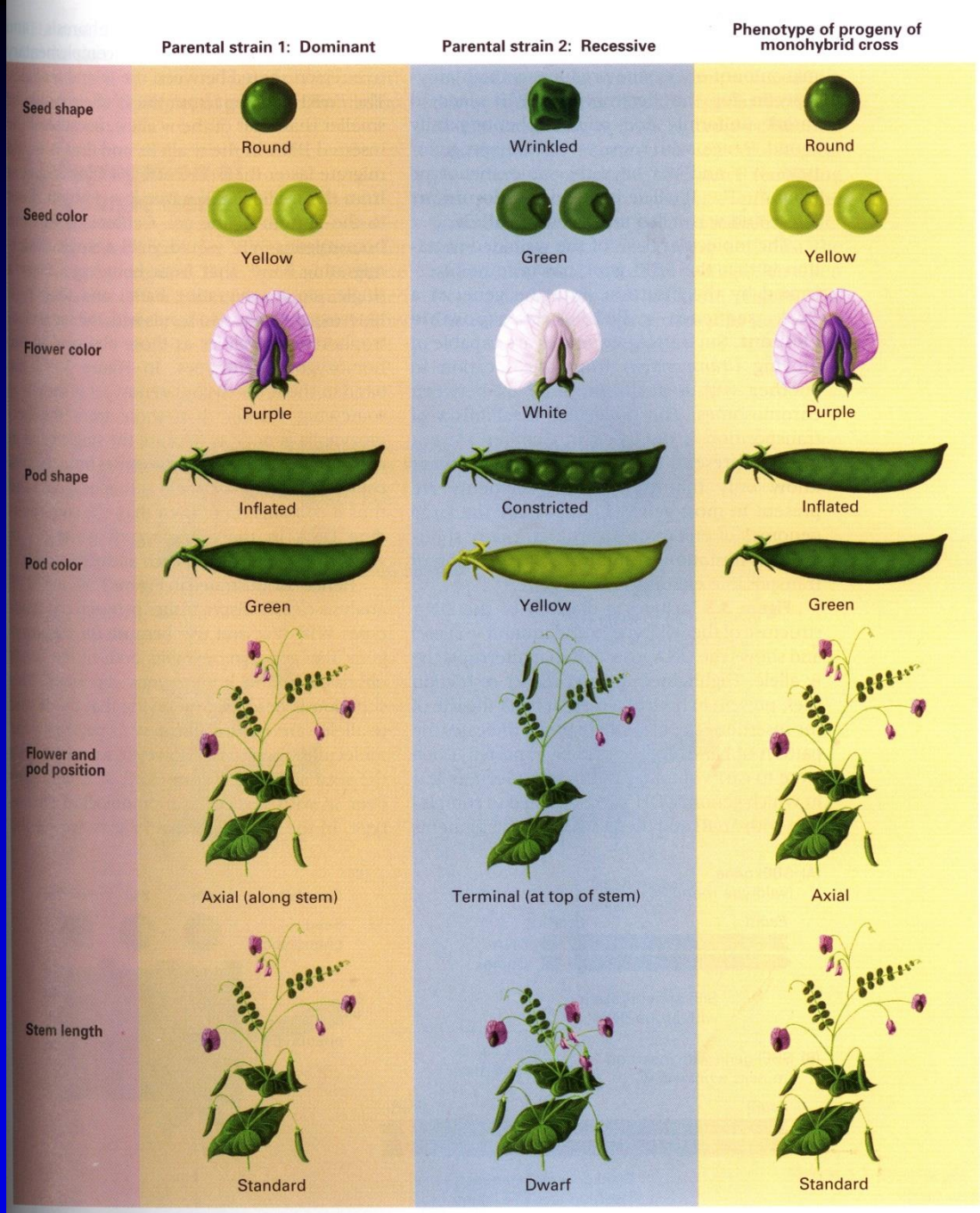
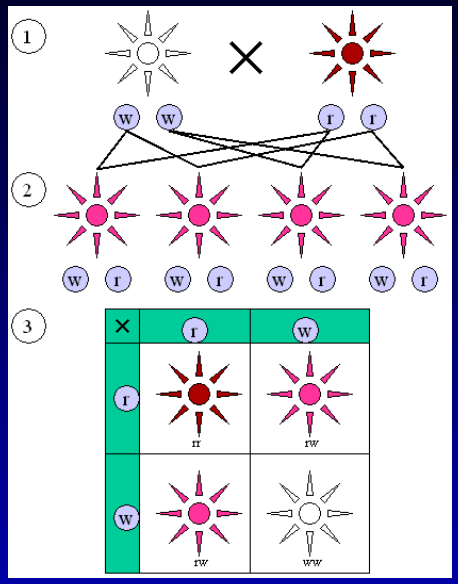


Figure 3.1 The seven different traits in peas studied by Mendel. The phenotype shown at the far right is the dominant trait, which appears in the hybrid produced by crossing.

Mendelian Genetics

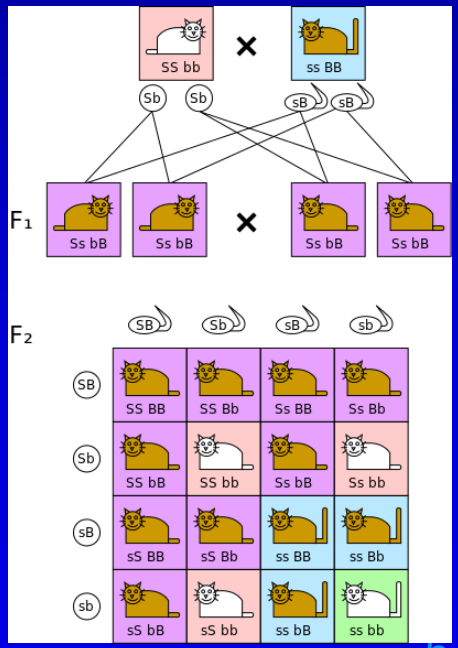
- The Principle of Segregation (The "First Law")
  - One gene (locus) Aa: ½A + ½ a (1:1)
  - Two independent genes (locus) AaBb:

    ¼AB + ¼Ab + ¼aB + ¼ab (1:1:1:1)

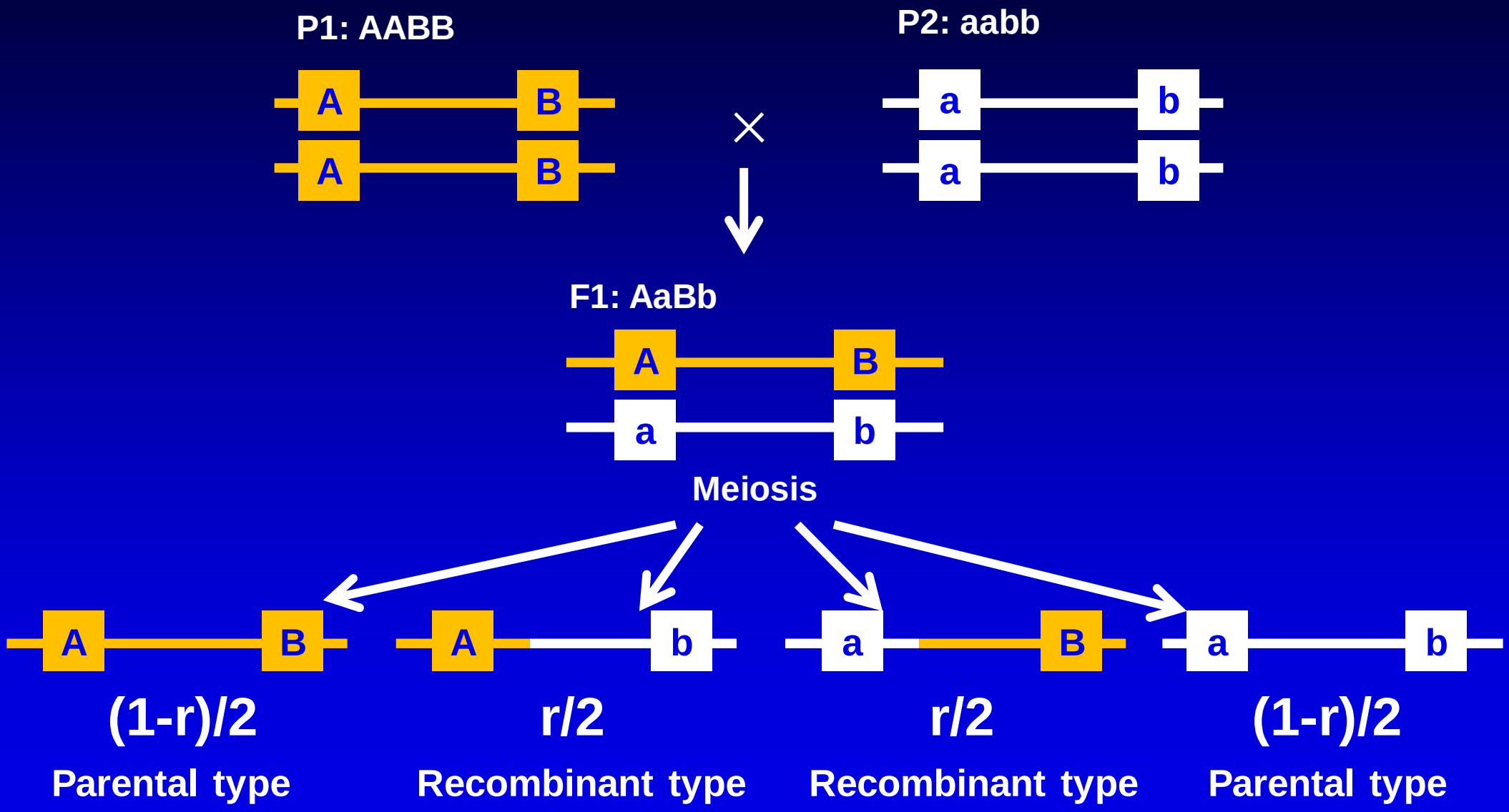


- The Principle of Independent Assortment (The "Second Law")
  - One gene (locus) Aa: $(½A + ½a)^2 = ¼AA + ½Aa + ¼aa$ (1:2:1)
  - Two independent genes (locus) AaBb:

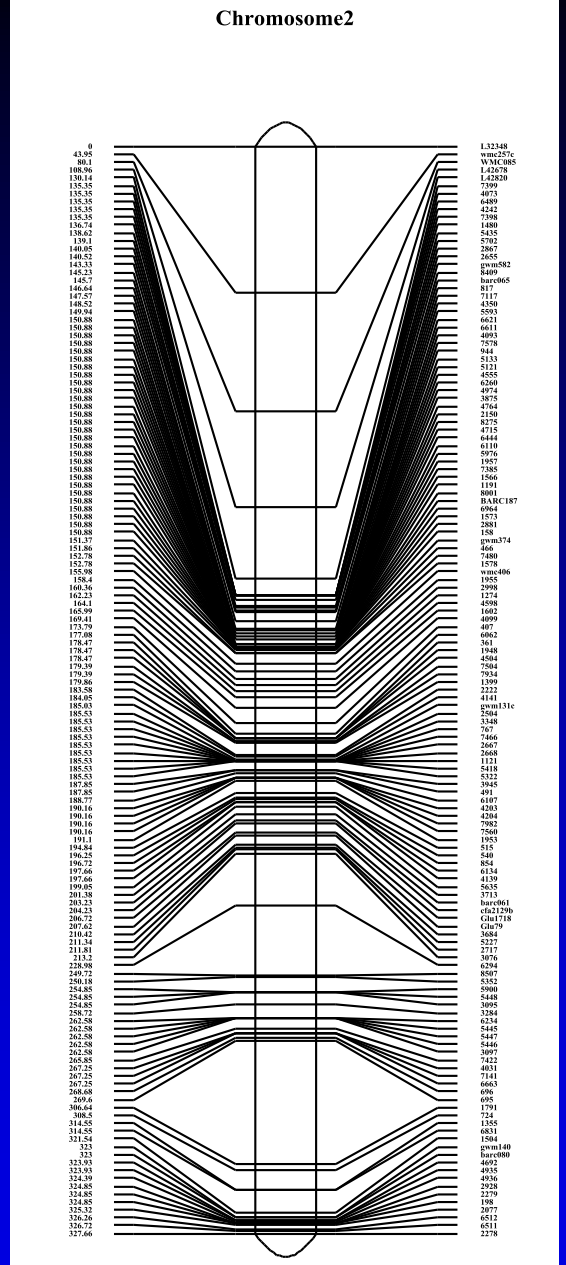
$(¼AB + ¼Ab + ¼aB + ¼ab)^2$ = (1/16)AABB + (2/16) AABb + (1/16) AAbb + (2/16) AaBB + (4/16) AaBb + (2/16) Aabb + (1/16) aaBB + (2/16) aaBb + (1/16) aabb or 1:2:1:2:4:2:1:2:1

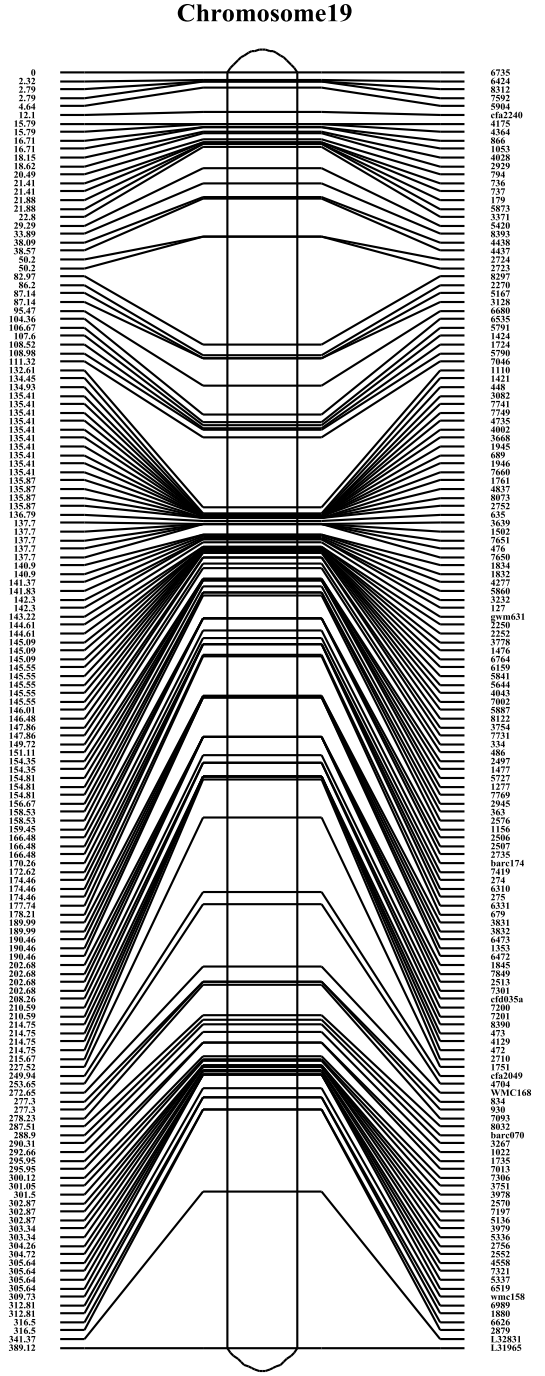


Genetic Linkage and Recombination

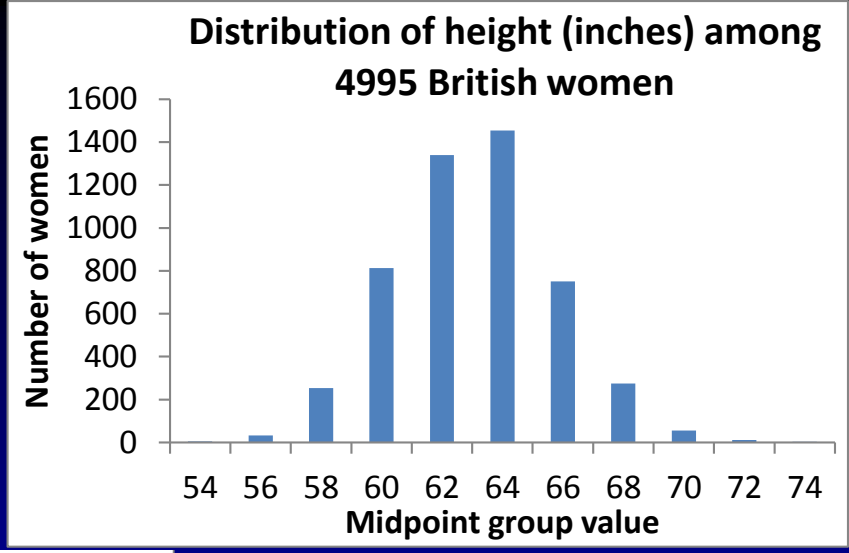


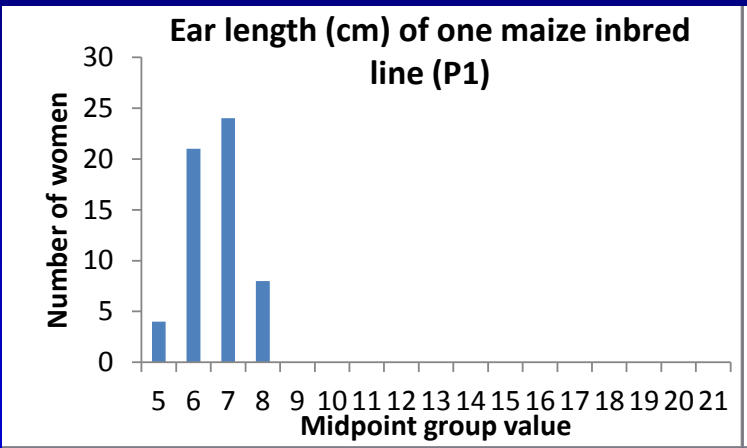
Genetic Linkage

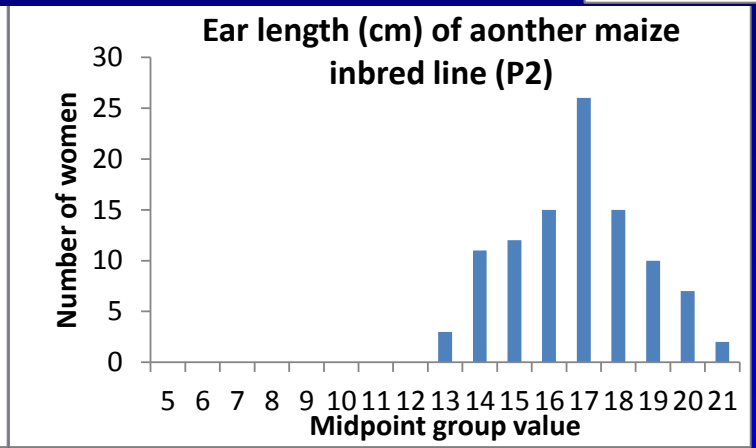


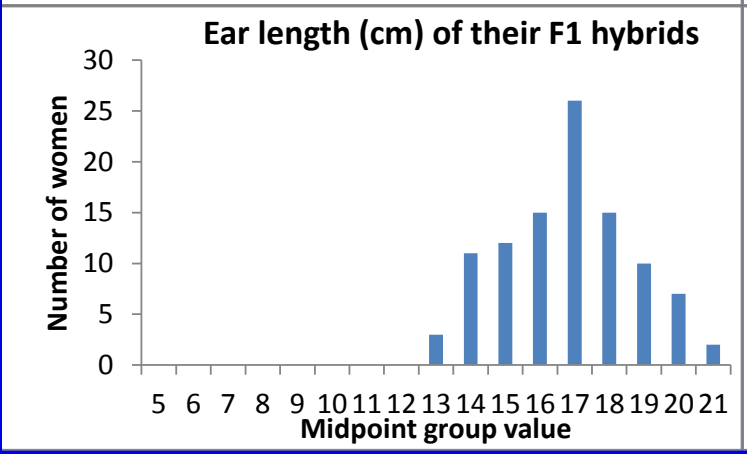


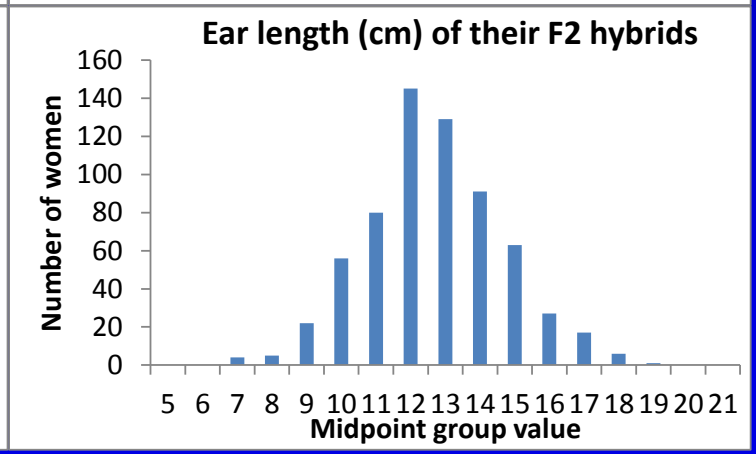
Quantitative traits in genetics











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

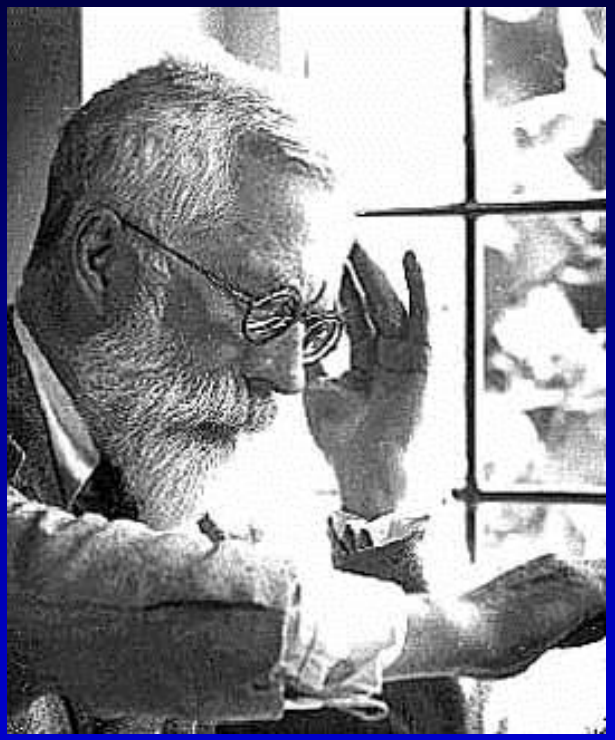
- W. L. Johannsen (1903) 的纯系学说将变异区分为遗传的变异与非遗传的变异，提出了基因型和表现型的概念，这为理解连续性变异也是遗传性状提供了依据 。
- Nilsson-Ehle (1909) 根据小麦粒色的遗传提出了数量性状的多因子假设，这一假设为E. M. East (1911) 玉米穗长和E. M. East (1913) 烟草花冠长度的遗传试验所证实。
- 通过多因子假设将数量性状的遗传纳入到孟德尔遗传的轨道，从而使两个学派的观点得到统一。

数量性状遗传的多基因假说

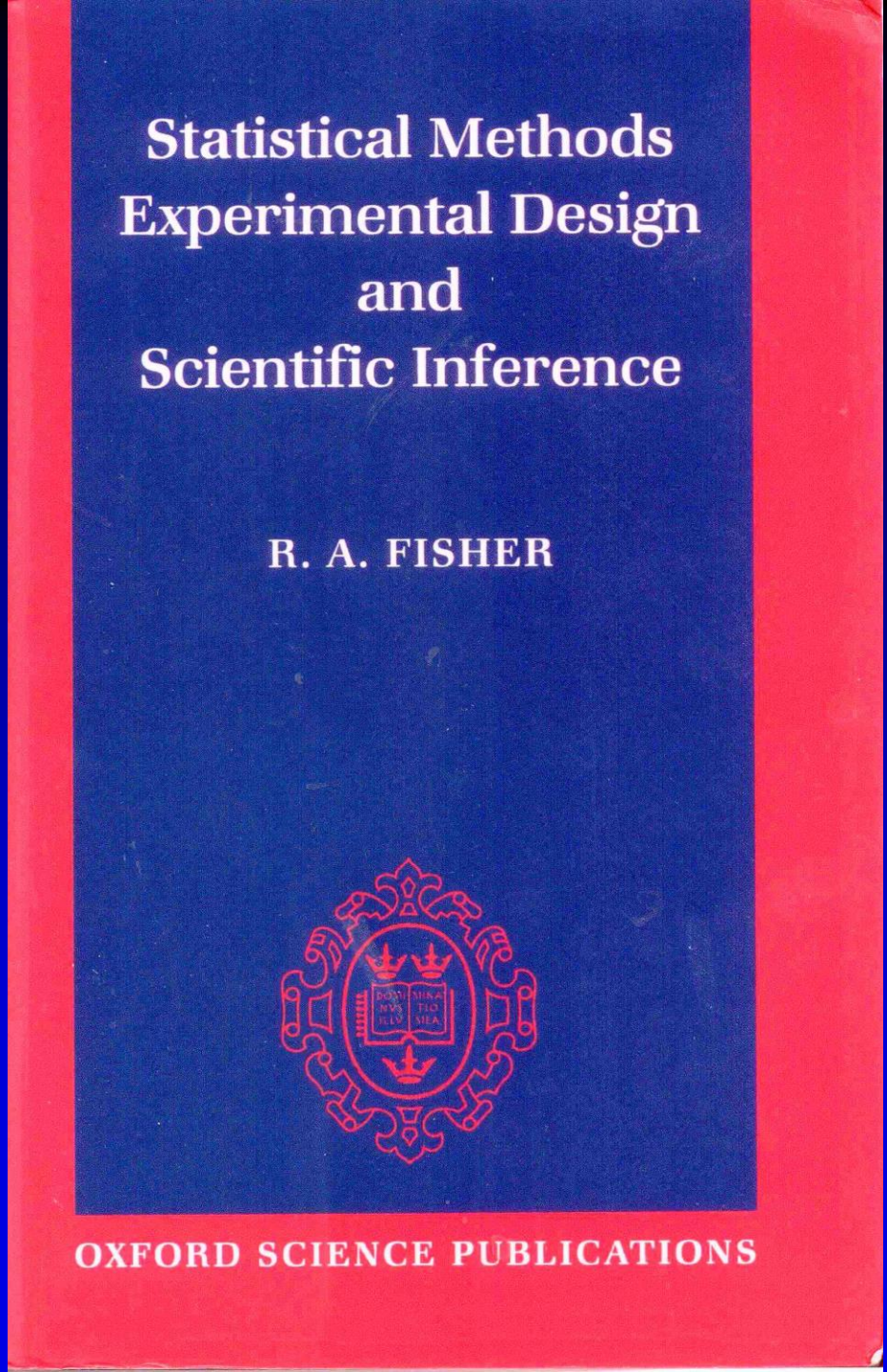
- R. A. Fisher (1918) “The correlation between relatives on the supposition of Mendelian inheritance”
- Multiple-factor hypothesis (polygene system)
  - A hypothesis to explain quantitative variation by assuming the interaction of a large number of genes (polygenes) each with a small additive effect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

数量遗传基本理论体系的建立
(1920s-1940s)

- 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。



Sir Ronald Fisher (1890-1962)



Mendel and Fisher

- Fisher(1936). Has Mendel's work been rediscovered? Annuals of Science 1:115-137. Mendel's data was so close to the values that Mendel expected under his theory that there must have been some manipulation, or omission, of data

- Dominant trait: 1/3 AA + 2/3 Aa
  - Family size: 10
  - Non-segregating (AA) : Segregating (Aa) = 1:2 (Mendel)
  - Fisher: Pro {Aa family classified as AA} = 0.75^10=0.0563
  - Pro {Non-segregating (AA)} =2/3*(1-0.0563)=0.6291
  - Non-segregating (AA) : Segregating (Aa) = 0.3709 : 0.6291 = 1 : 1.6961

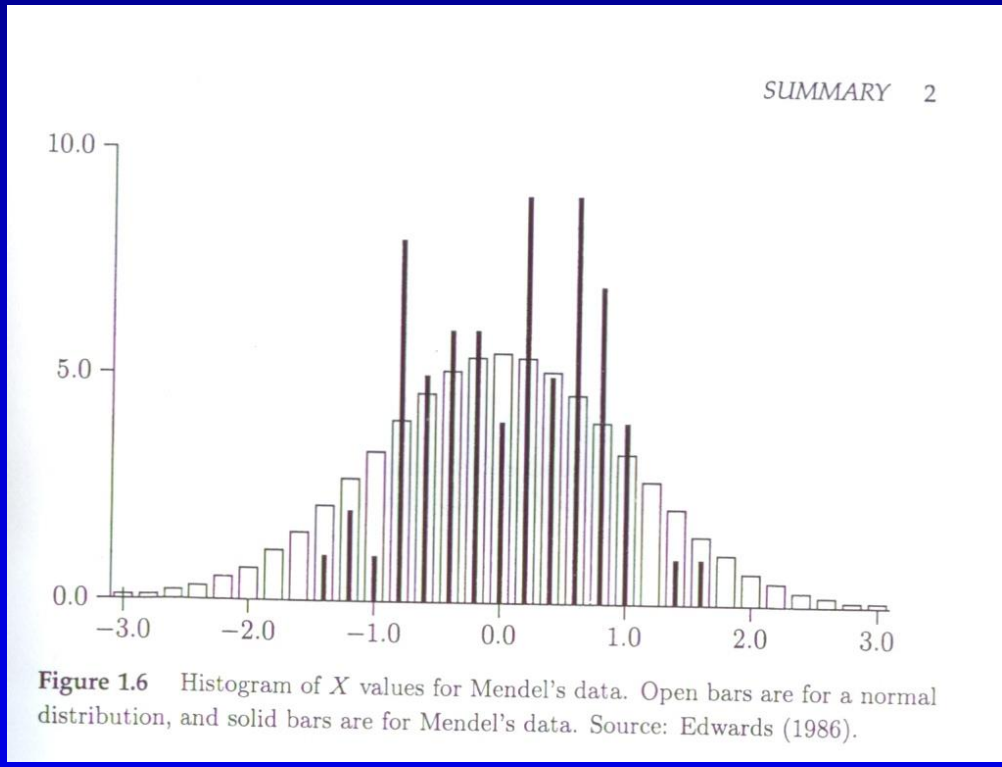


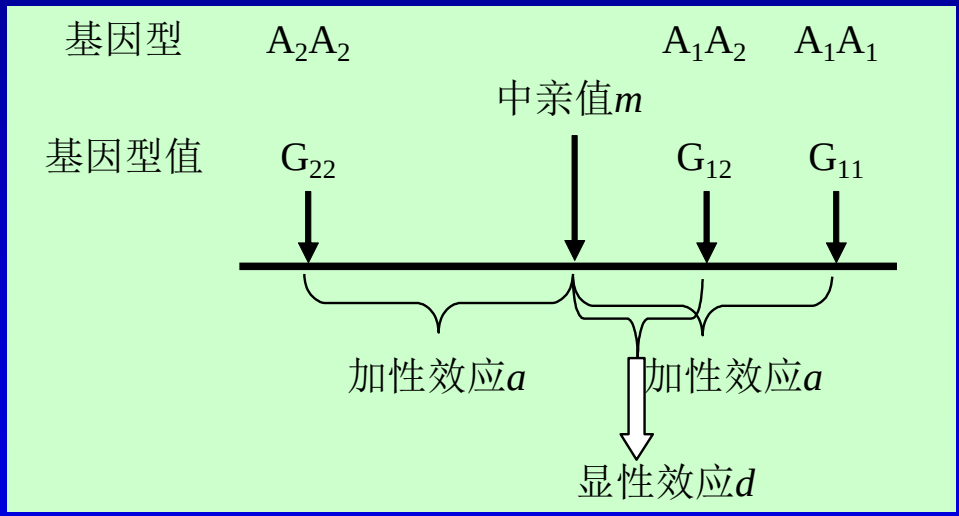
Figure 1.6 Histogram of X values for Mendel's data. Open bars are for a normal distribution, and solid bars are for Mendel's data. Source: Edwards (1986).

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+ ε
- 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)

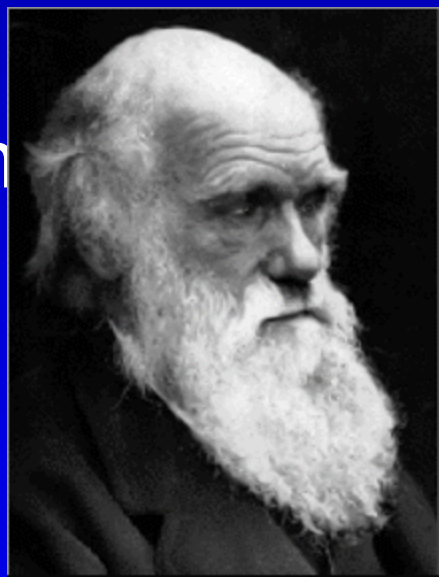
植物育种与数量遗传

植物育种的定义

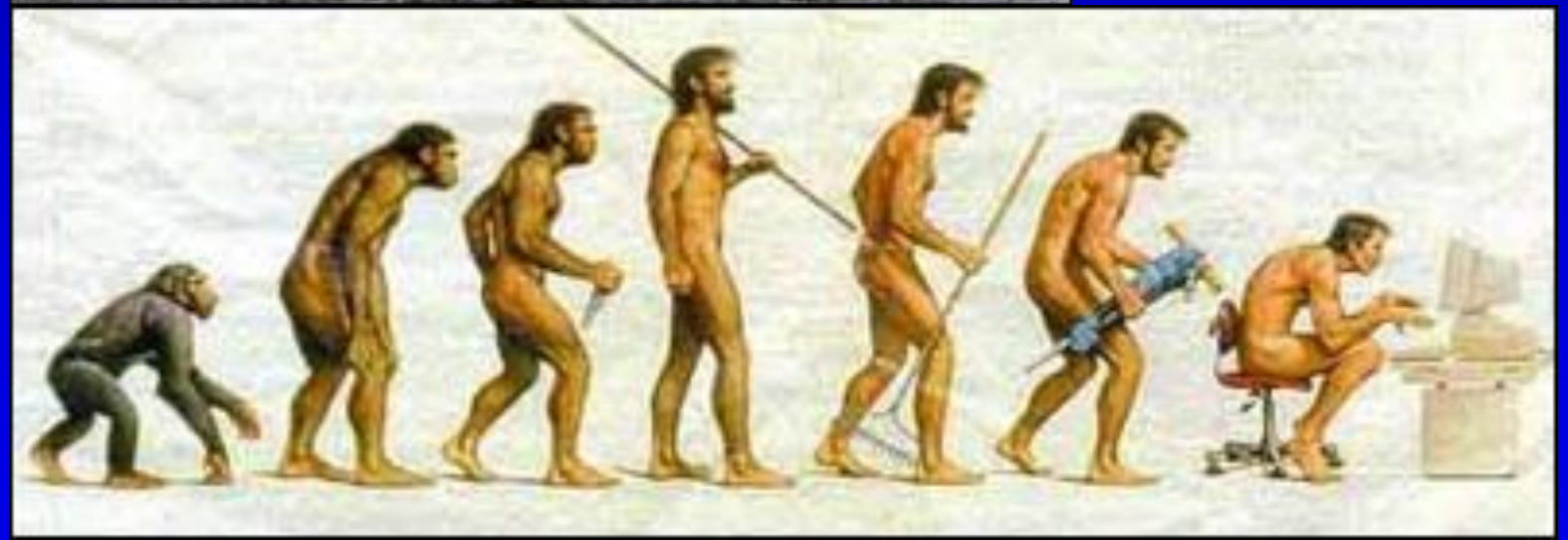
- 植物育种的主要任务是寻找控制目标性状的基因、研究这些基因在不同环境下的表型效应、挑选适当的亲本材料、设计杂交方案和分离群体种植方案、通过基因型/表型选择聚合存在于不同亲本材料中的有利基因、培育符合育种目标的基因型，为农业生产提供适宜的品种
- 植物育种中的科学
- 植物育种中的艺术
- 植物育种中的商业行为
- “机遇”或”运气”

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

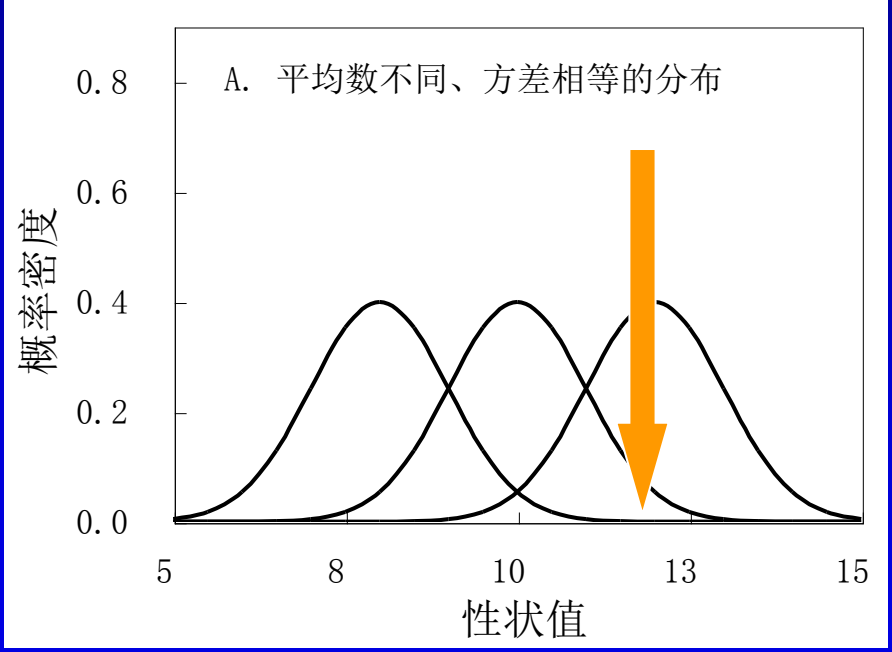
- 亲本基因型（2种）
  - P1：R1R1 r2r2 y1y1 Y2Y2 Y3Y3 rr：中抗、产量较高
  - P2：r1r1 R2R2 Y1Y1 y2y2 y3y3 RR：中抗、产量较低
- 目标基因型：R1R1 R2R2 Y1Y1 Y2Y2 Y3Y3 ??
- 后代基因型（2^5=32种）
  - 不抗病、产量低：r1r1 r2r2 y1y1 y2y2 y3y3 ??
  - 抗病、产量低：R1R1 R2R2 y1y1 y2y2 y3y3 ??
  - 等等
- 育种的目的是从32种可能的基因型中把理想的目标基因型选择出来

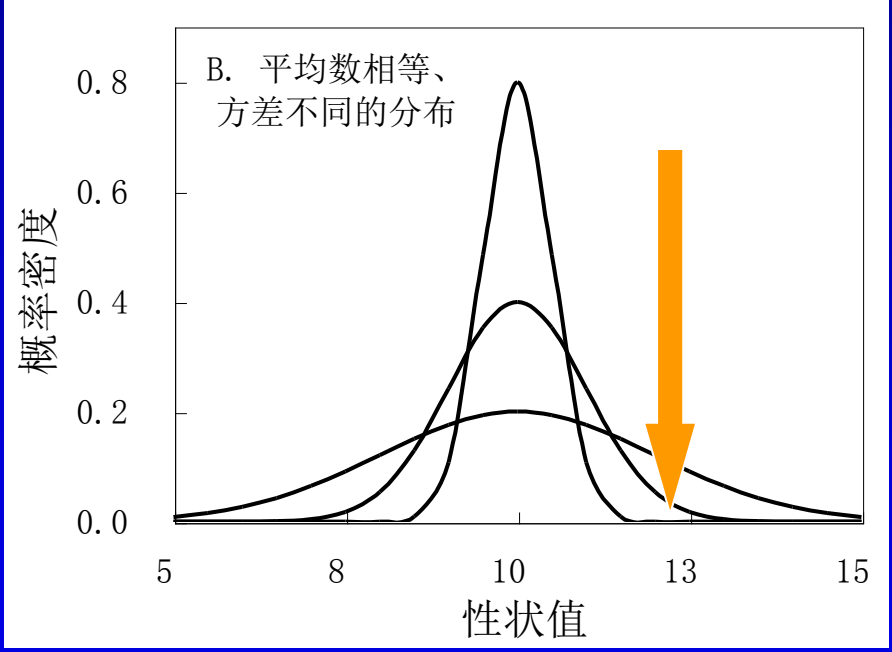
基因与基因型是两个不同的概念

- 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





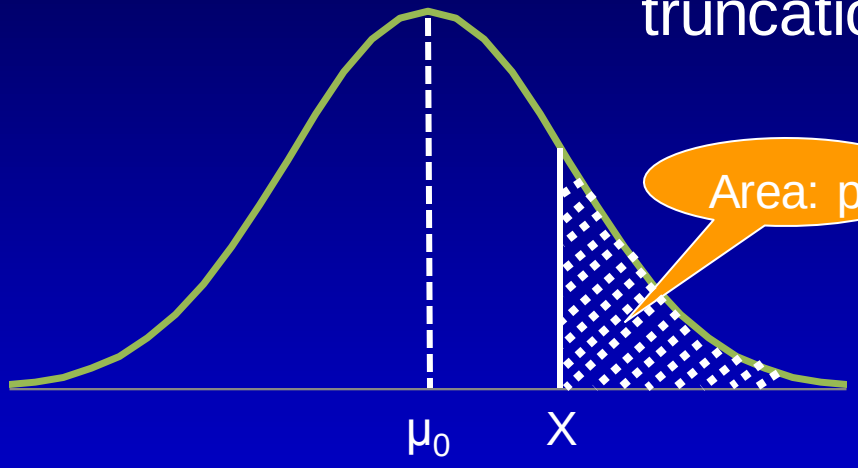
Genetic gain: response to selection

(or the change of population mean)

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

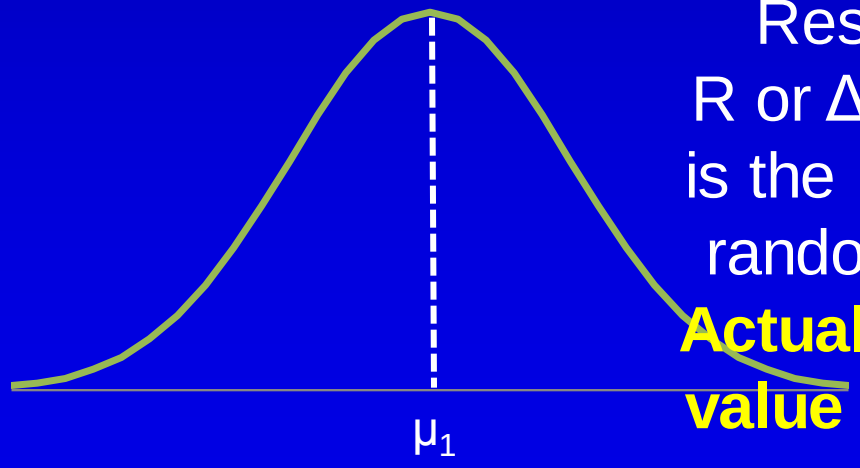
p ⇔ X

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.

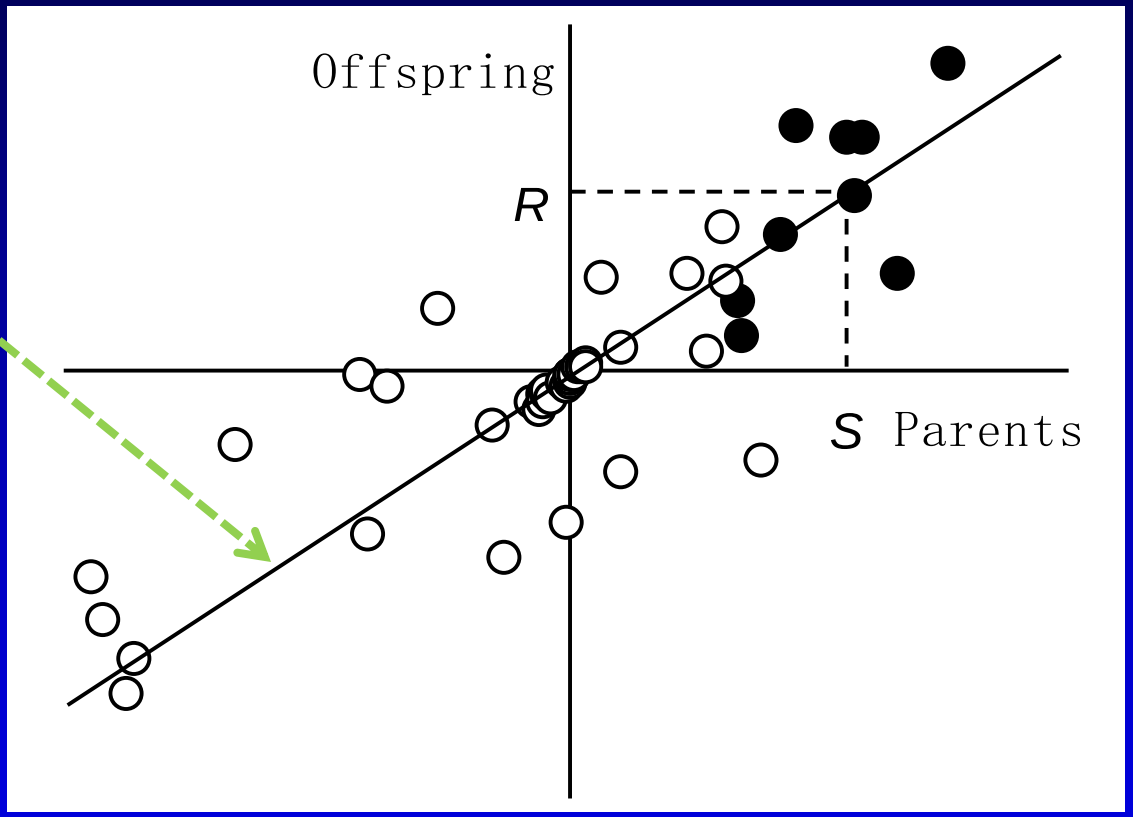
Actually, R is the breeding value of selected parents!

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）

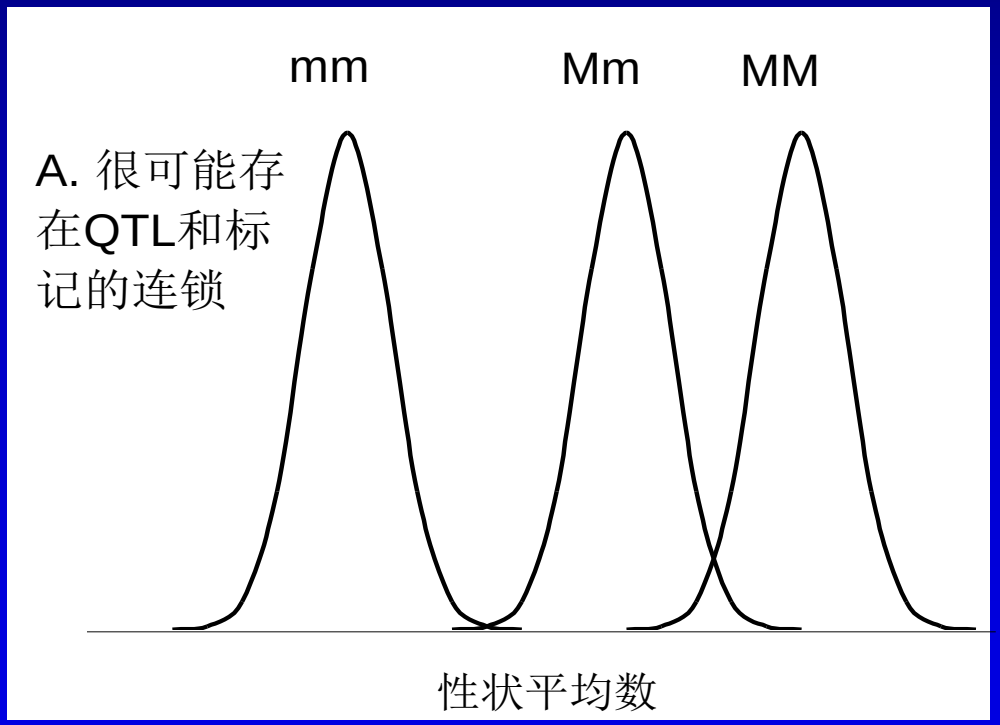
Example: 10 RILs of Rice (Linkage Map of Chr. 5)

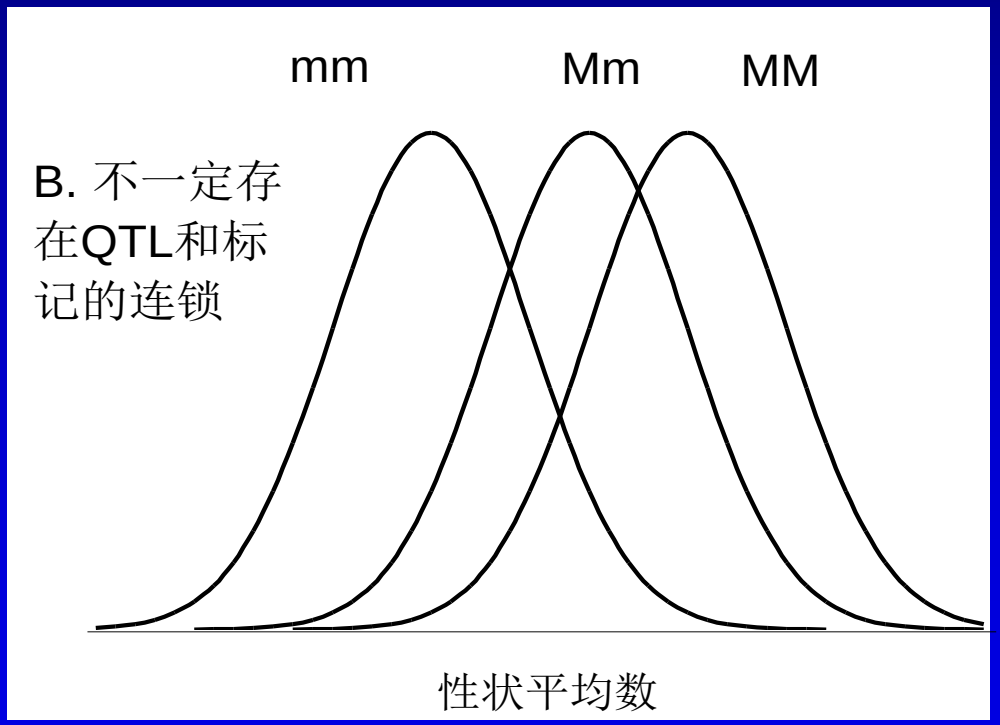
Marker	C263	R830	R3166	XNpb387	R569	R1553	C128	C1402	XNpb81	C246	R2953	C1447	Grain width (mm)
Position (cM)	0.0	3.5	8.5	19.5	32.0	66.6	74.1	78.6	81.8	91.9	92.7	96.8	
RIL1	1	1	1	1	1	1	1	1	1	1	1	1	2.33
RIL2	2	2	2	2	2	1	1	1	1	2	2	2	1.99
RIL3	1	2	2	2	2	2	2	2	2	2	2	2	2.24
RIL4	1	1	1	1	1	1	2	2	2	2	2	2	1.94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

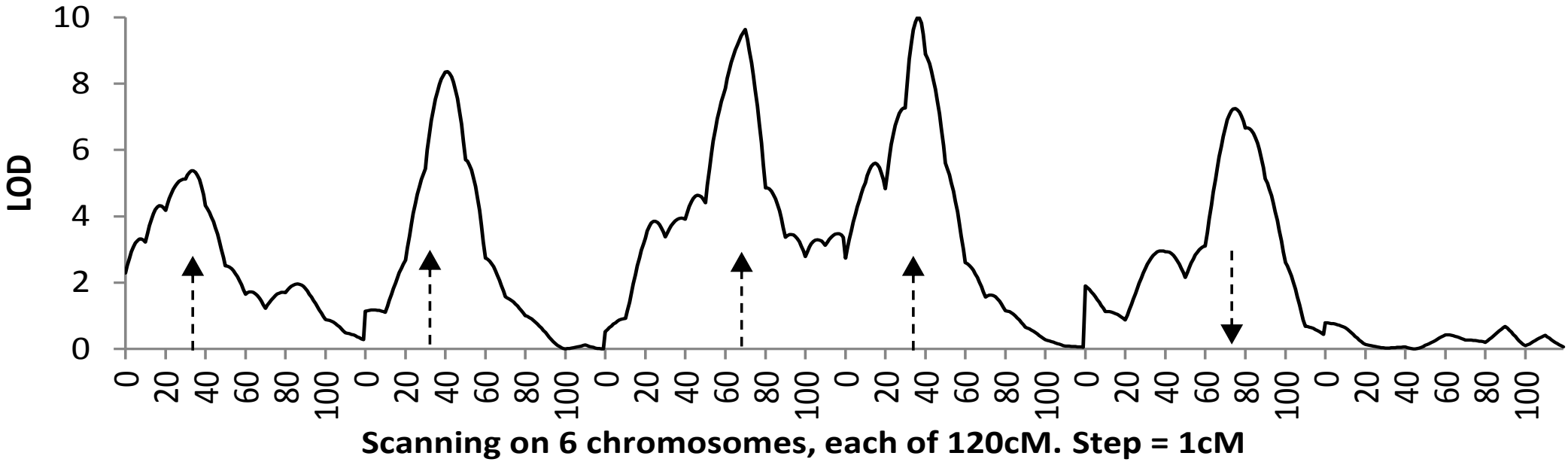
一个标记位点上3种基因型的性状平均数





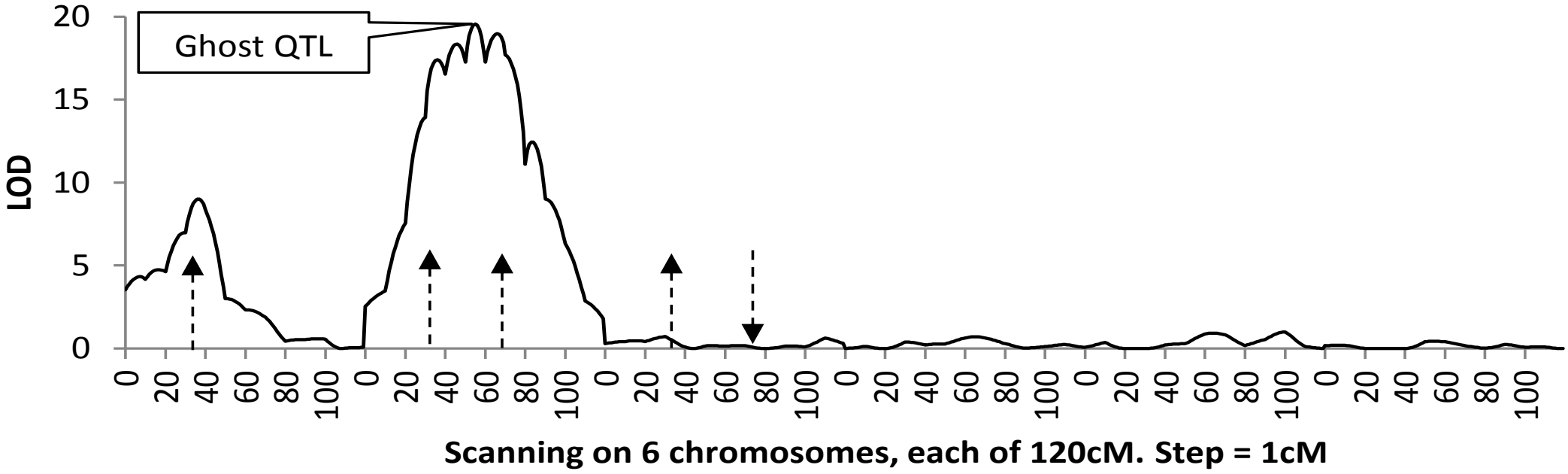
Simple Interval Mapping

- Multiple peaks when QTLs are unlinked



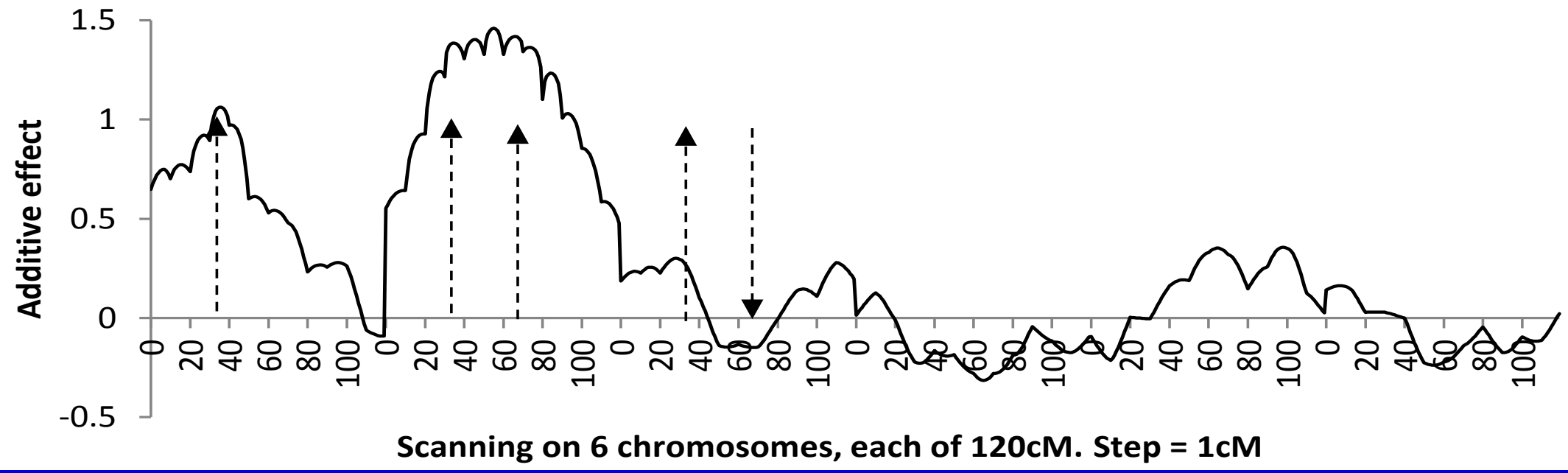
Problems with Simple Interval Mapping

- Ghost QTL when two QTLs are linked



Problems with Simple Interval Mapping

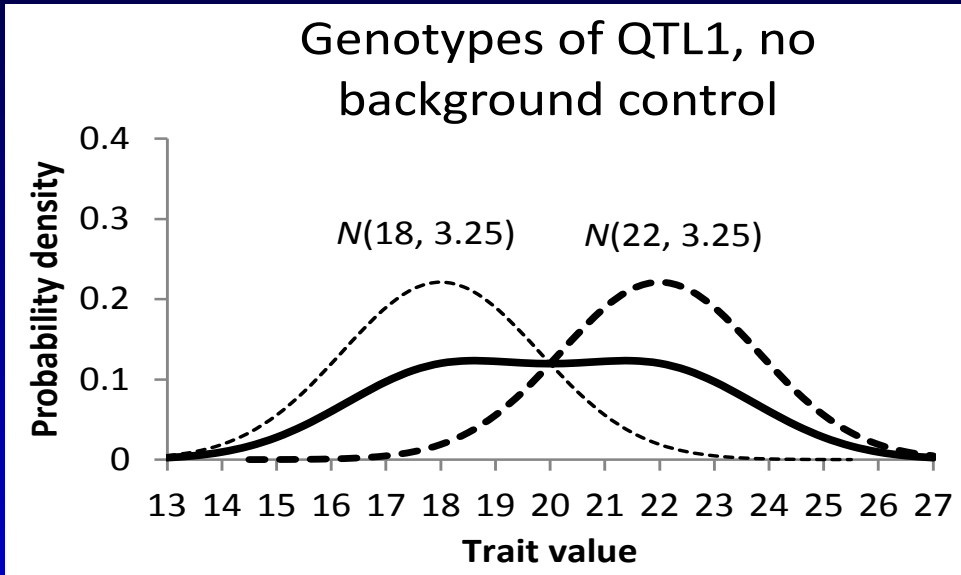
- Biased estimation of QTL effects

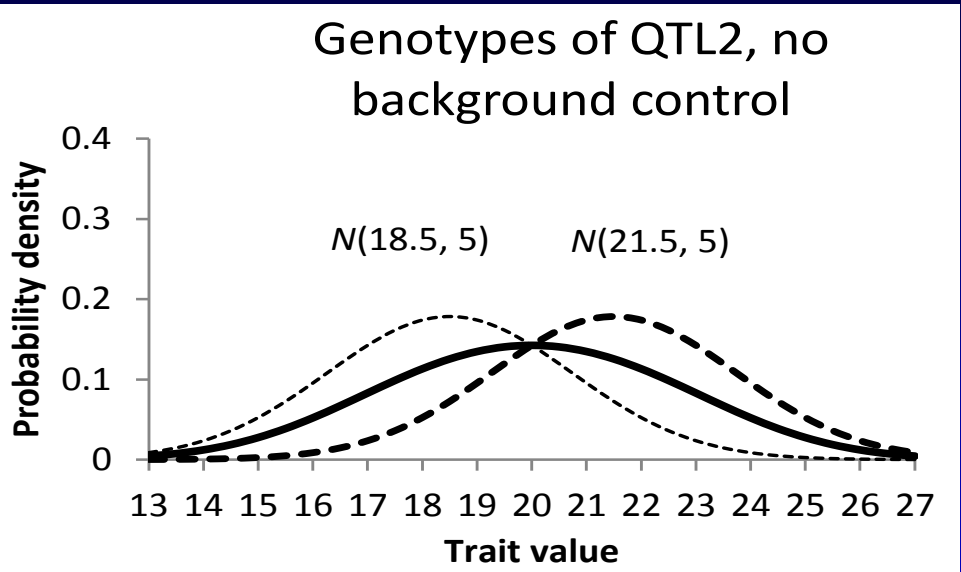


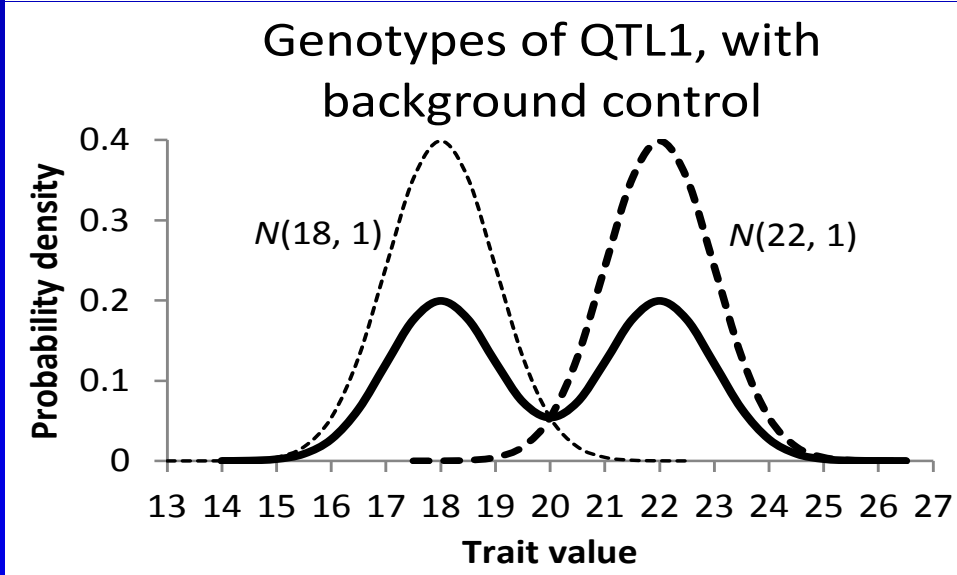
Two putative unlinked QTL

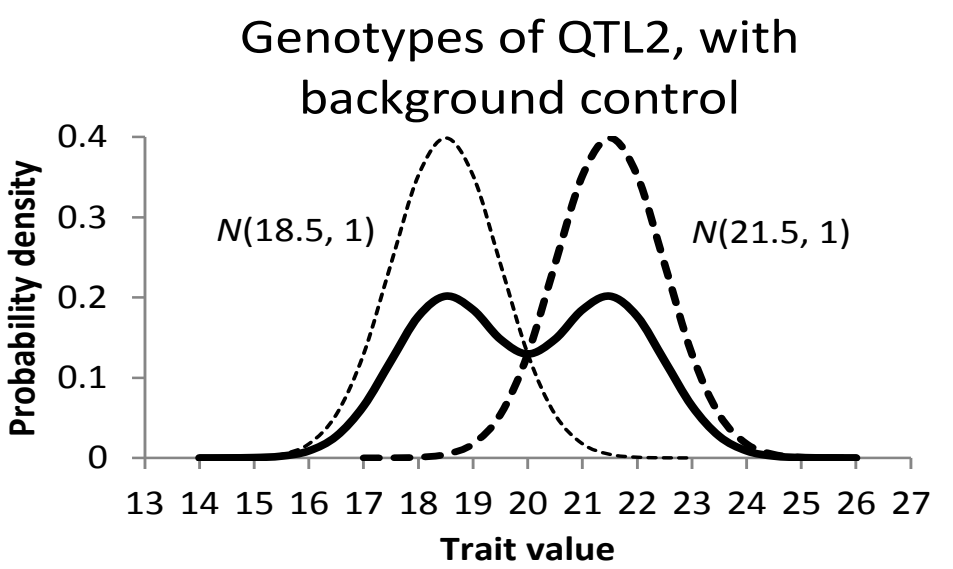
Overall mean	Additive effect		Additive variance		Genetic variance	Error variance	H^2
	QTL1	QTL2	QTL1	QTL2			
20	2	1.5	4	2.25	6.25	1	0.86

Importance of the background control









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。

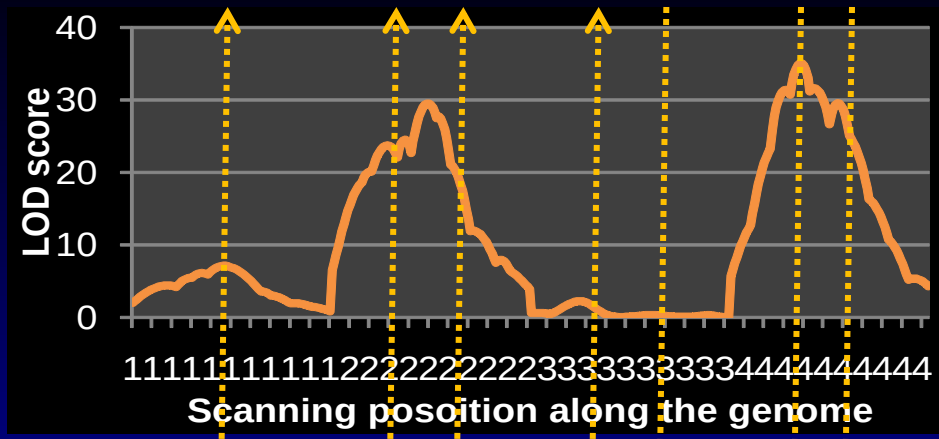
完备区间作图方法（ICIM）

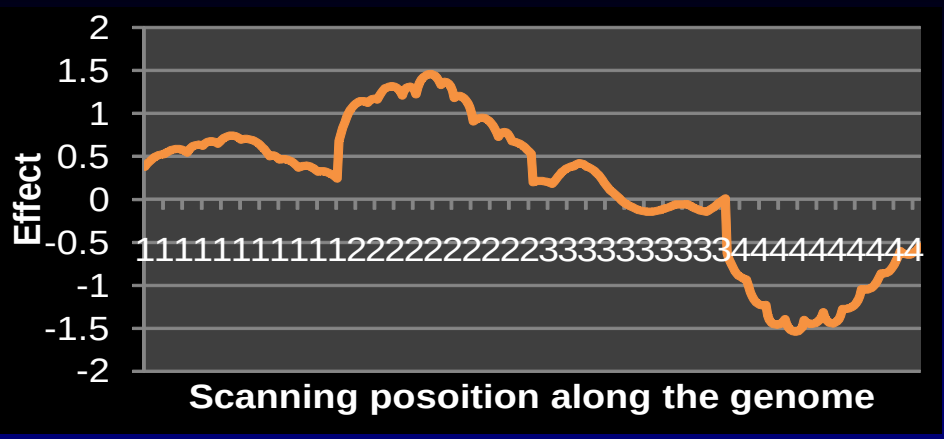
- One-dimensional scanning (interval mapping)

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

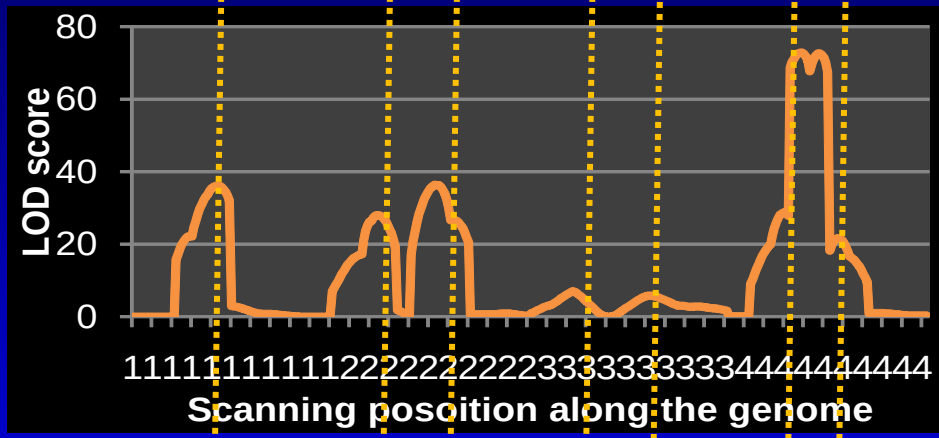
- Two-dimensional scanning (interval mapping)

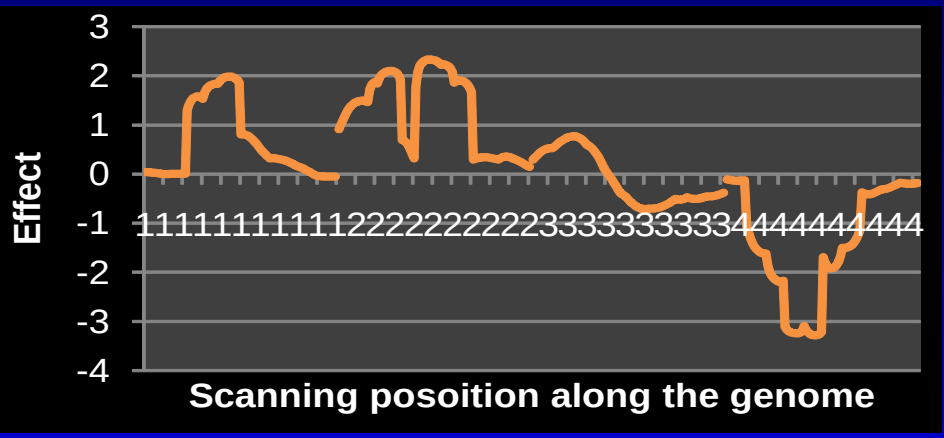
$$\Delta y_i = y_i - \sum_{r \neq j, j+1, k, k+1} \hat{b}_r x_{ir} - \sum_{\substack{r \neq j, j+1 \\ s \neq k, k+1}} \hat{b}_{rs} x_{ir} x_{is}$$



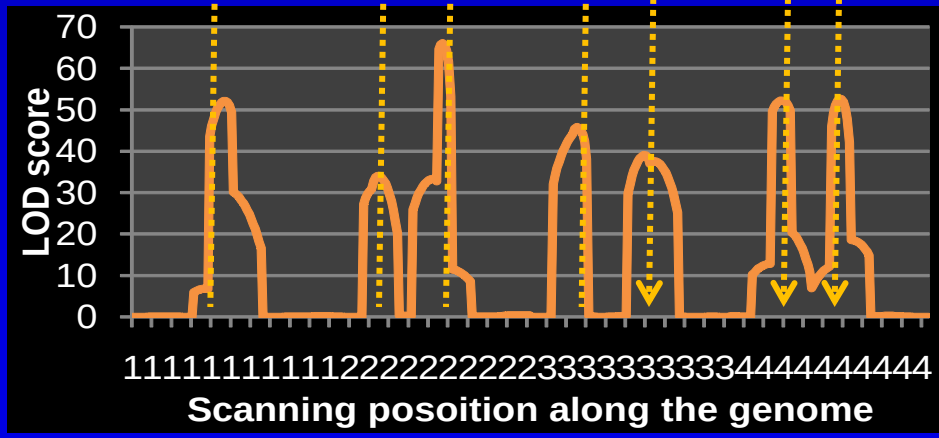


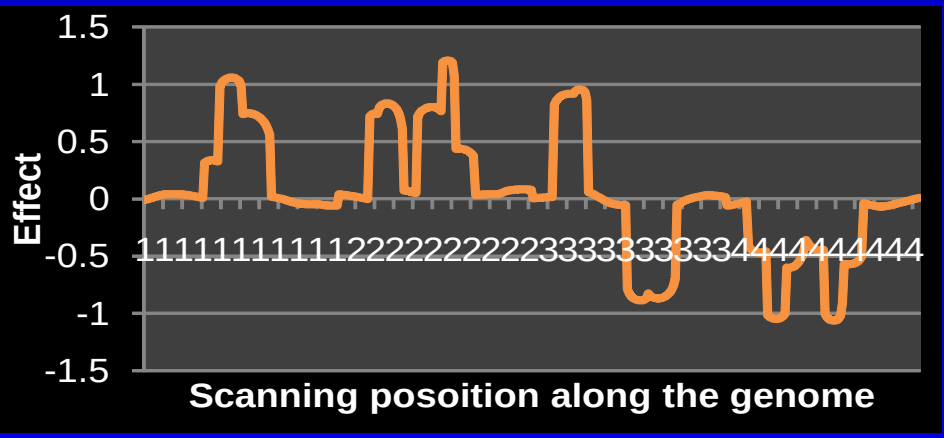
IM
1989





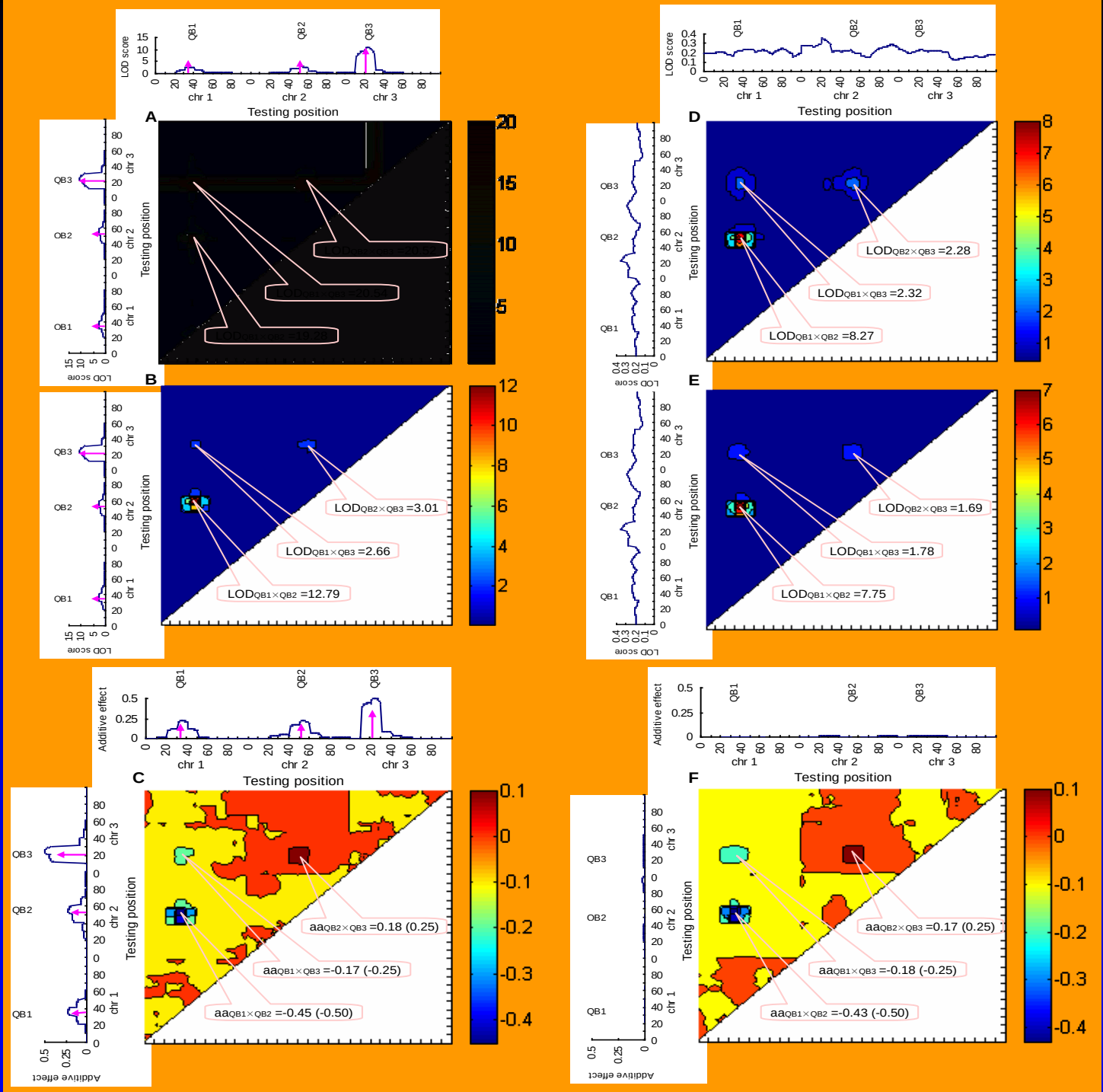
CIM
1994





ICIM
2006

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



QTL IciMapping: 遗传图谱构建和基因定位集成软件