

The 9th Workshop on QTL Mapping and Breeding Simulation
The University of Sydney, Cobbitty NSW, 7-9 March 2012

Introduction of Quantitative Genetics

Jiankang Wang, CAAS and CIMMYT China

**E-mail: wangjk@caas.net.cn;
jkwang@cgiar.org**

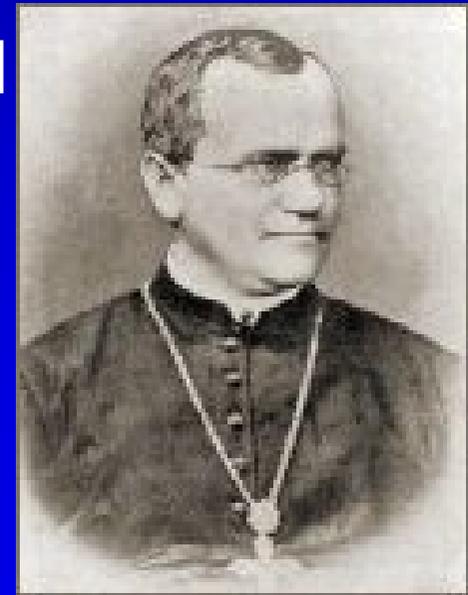
Outline of the presentation

- Brief History of Quantitative Genetics
- Contents of Quantitative Genetics
- Applied Quantitative Genetics in Plant Breeding
- Modern Quantitative Genetics

Brief History of Quantitative Genetics

Rediscovery of Mendel's hybridization experiments in garden pea

- *Experiments with Plant Hybrids* (1866)
 - 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 (20-50cm) vs 277 short (185-230cm)
- Rediscovered in 1900



Quarrel on the inheritance of quantitative traits

- **Biometrical school** (led by K. Pearson): the continuously quantitative traits are important in evolution; Mendelian laws cannot be used on the inheritance of quantitative traits
- **Mendelian school** (led by W. Bateson): qualitative traits are important in evolution; continuously quantitative traits are not heritable
- 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.

Ear length of maize (East 1911)

	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	Mean
P1	4	21	24	8														7
F1					1	12	12	14	17	9	4							12
P2									3	11	12	15	26	15	10	7	2	17
F2			4	5	22	56	80	145	129	91	63	27	17	6	1			13

➤ P1: 7cm; P2: 17cm

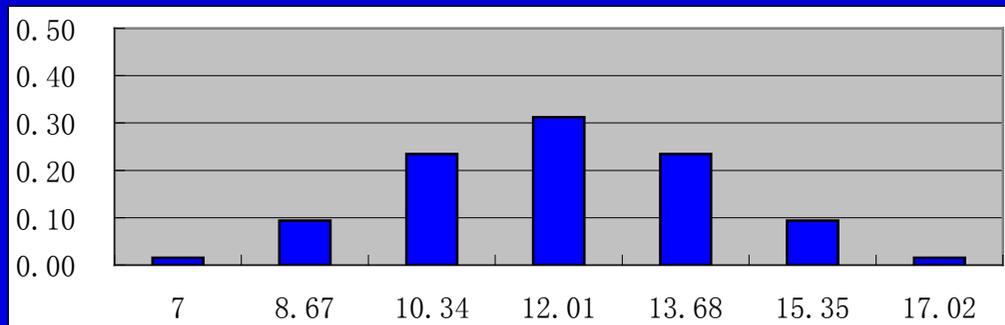
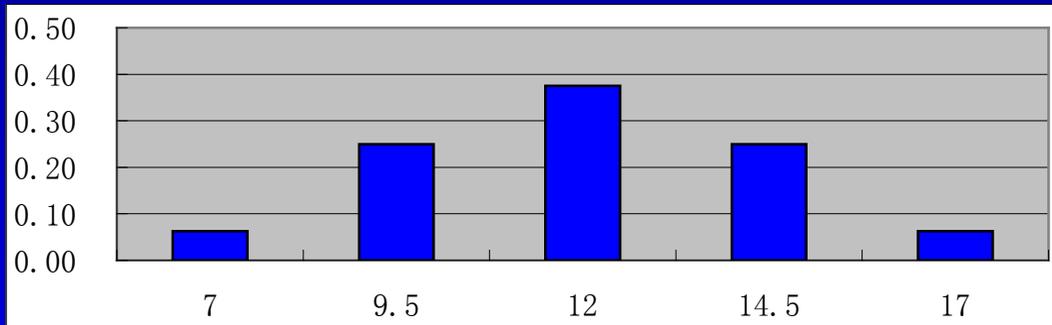
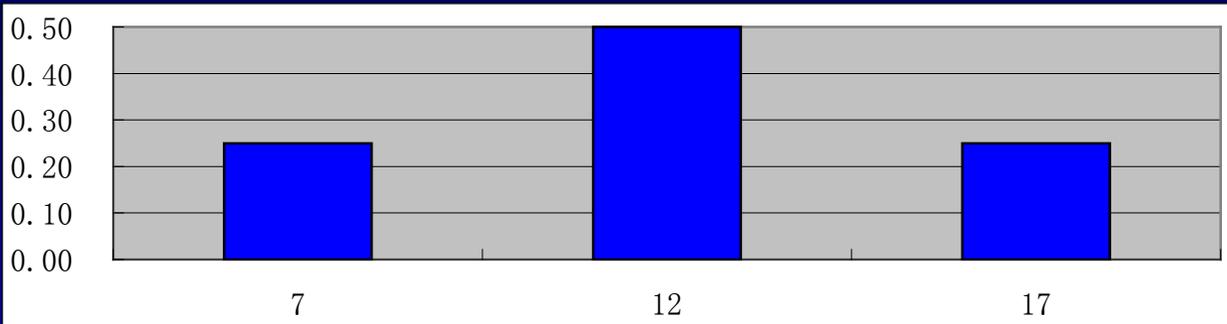
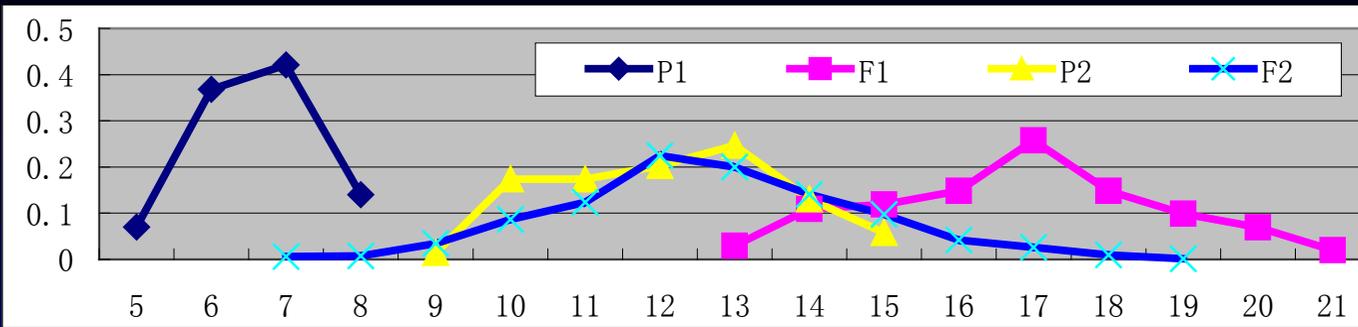
➤ One locus

▪ $a=(17-7)/2=5$; F2: $1/4$ aa (7) + $2/4$ Aa (12) + $1/4$ AA (17)

➤ Two locus

▪ $a=(17-7)/4=2.5$

▪ F2: $1/16$ (7) + $4/16$ (9.5) + $6/16$ (12) + $4/16$ (14.5) + $1/16$ (17)



Merge of the Biometrical and Mendelian Schools

- Pure Line Theory (W. L. Johannsen 1903)
 - Phenotype = Genotype + Random environmental error
- Hybridization experiments on quantitative traits
 - Grain color in wheat (Nilsson-Ehle 1909)
 - Ear length in maize (East 1911)
 - Flower length in tobacco (East 1913)
- Multi-factorial hypothesis (or polygene system) on the inheritance of quantitative traits

Multi-factorial hypothesis (or polygene system)

- 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

Estimation of the effective number

- Assumptions (k is the number of genes)
 - Additive model ($d=0$)
 - Equal effect
 - No linkage
 - All positive alleles in one parent P1, negative alleles in another parent P2

➤ Genetics variance of F2

$$V_{F2} = \frac{1}{2} ka^2$$

Castle-Wright formula (1921)

$$\bar{P}_1 = ka$$

$$\bar{P}_2 = -ka$$

$$k = \frac{(\bar{P}_1 - \bar{P}_2)^2}{8[V_{F2} - \frac{1}{2}(V_{P1} + V_{P2})]}$$

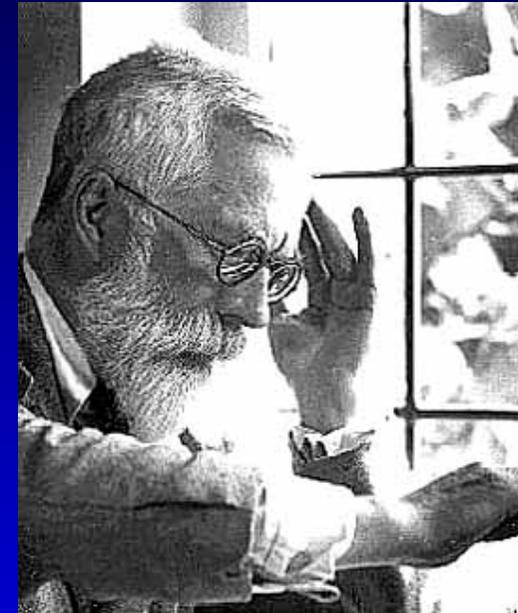
When dominance is present

$$V_A = \frac{1}{2} \Sigma a^2 = \frac{1}{2} k a^2$$

$$k = \frac{(\bar{P}_1 - \bar{P}_2)^2}{8V_A}$$

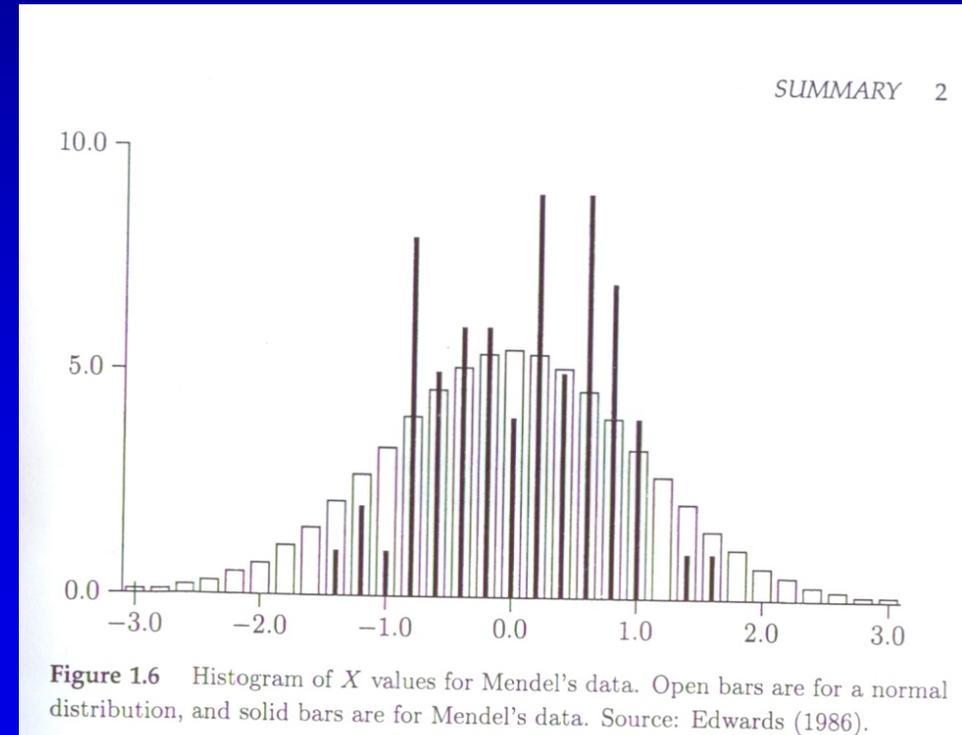
Founding of the conventional quantitative genetics (1920s-1940s)

- R. A. Fisher (1918) The correlation between relatives on the supposition of Mendelian inheritance
- J. B. S. Haldane (1924-1927) A mathematical theory of natural and artificial selection I-V
- S. Wright (1921) Systems of Mating
- J. L. Lush (1940) Animal Breeding Plan (heritability)
- G. Malecot: Inbreeding and Inbreeding Coefficient
- G.F. Sprague and L.A. Tatum (1942): General Combining Ability (GCA) and Specific Combining Ability (SCA)



Mendel and Fisher

- Fisher(1936). Has Mendel's work been rediscovered? *Annals 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



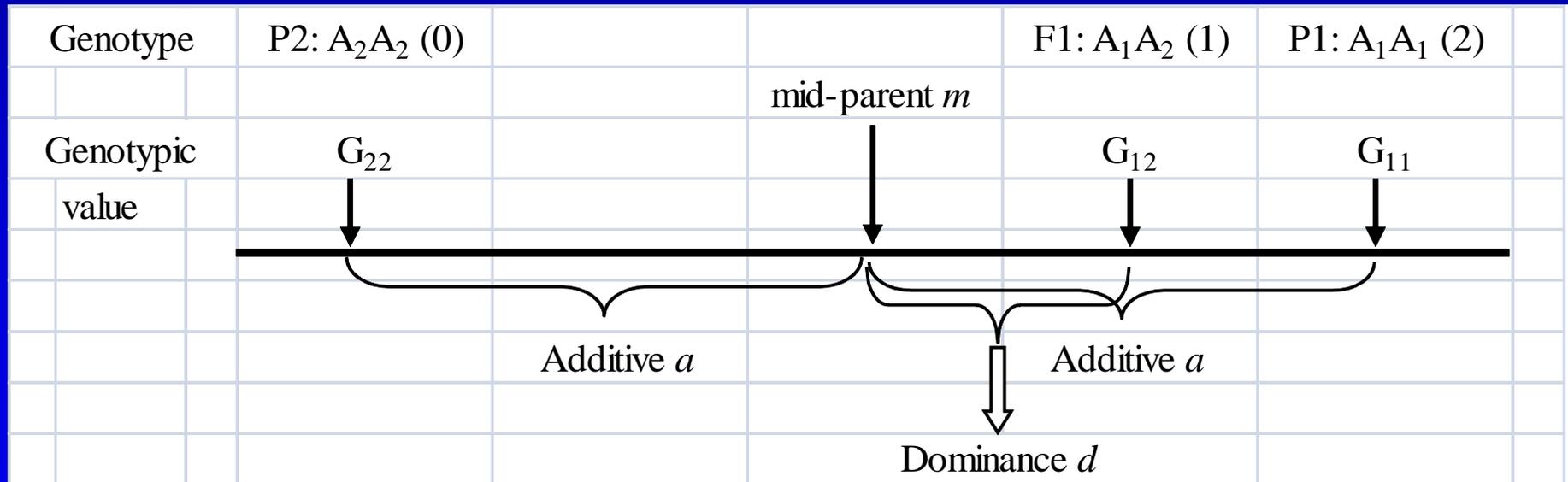
Contents of Quantitative Genetics

Population Genetics

- Population structure: gene frequency (allele A and a) and genotypic frequency (genotypes AA, Aa and aa)
- Population structure under various mating systems (selfing, backcrossing, random mating etc.)
- Hardy-Weinberg Equilibrium (HWE)
- Linkage and linkage disequilibrium (LD)
- Mutation and migration
- Change of population structure under selection
- Limited population size and random drift
- Related by ancestry and inbreeding

Additive and dominant model

- Single locus model (alleles A_1 and A_2)
 - p^2 : A_1A_1 , $2pq$: A_1A_2 , q^2 : A_2A_2
 - Genotype A_iA_j has value G_{ij}
 - Mid-parental: $m = (G_{11} + G_{22}) / 2$
 - Additive effect: $a = G_{11} - m$
 - Dominant effect: $d = G_{12} - m$



Genotypic values in the additive and dominant inheritance model

➤ A1A1

$$G_{11} = m + a$$

➤ A1A2

$$G_{12} = m + d$$

➤ A2A2

$$G_{22} = m - a$$

Average allele effect in a population

Allele	Genotype and genotypic values of its progeny			Mean of progeny	Population mean	Allele effect
	A1A1 <i>a</i>	A1A2 <i>d</i>	A2A2 <i>-a</i>			
A1	<i>p</i>	<i>q</i>		$pa+qd$	$(p-q)a+2pqd$	$q[a+(q-p)d]$
A2		<i>p</i>	<i>q</i>	$-qa+pd$	$(p-q)a+2pqd$	$-p[a+(q-p)d]$

Breeding value and dominance deviation

- Single-locus model (alleles A_1 and A_2)
 - Population mean: $\mu = m + (p - q)a + 2pqd$
 - Allele effect: $\alpha_1 = q[a + (q - p)d]$; $\alpha_2 = -p[a + (q - p)d]$
 - $G_{ij} = \mu + \alpha_i + \alpha_j + \delta_{ij}$
 - Breeding value: $A_{ij} = \alpha_i + \alpha_j$
 - Dominance deviation: $D_{ij} = \delta_{ij}$
- Easy to extend to multiple loci and multiple alleles

Epistatic deviation

- Two-loci model (A_i, A_j and B_k, B_l)

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

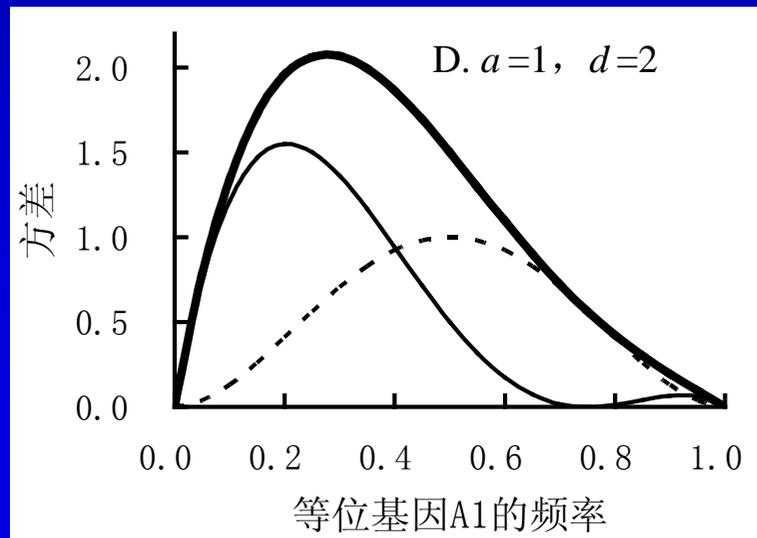
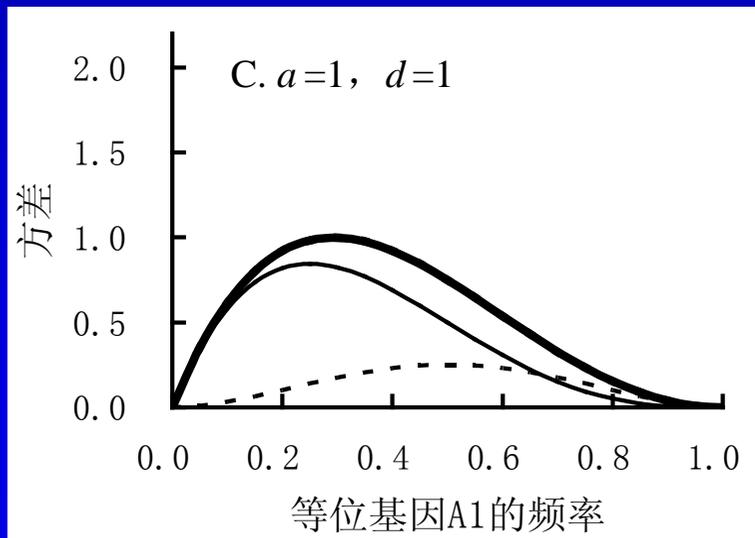
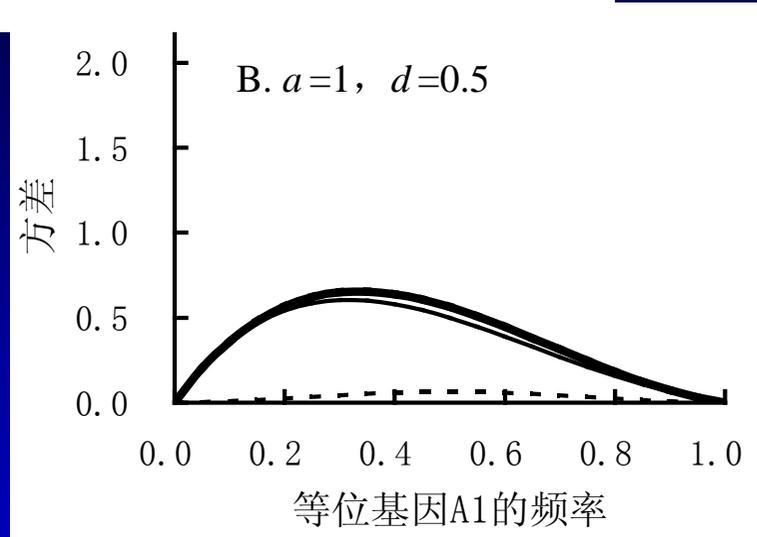
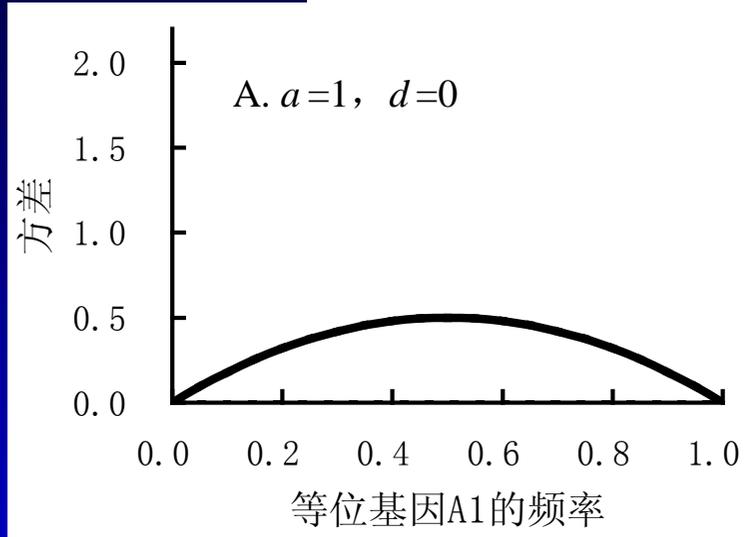
- Breeding value: $A = \sum \alpha$
- Dominance deviation: $D = \sum \delta$
- Epistatic deviation: $I = \sum I$
- $P = G + \varepsilon = \mu + A + D + I + \varepsilon$
- $V_P = V_G + V_\varepsilon = V_A + V_D + V_I + V_\varepsilon$

Components of genetic value and components of genetic variance

Effect	Variance
Phenotype value (P)	Phenotype variance (V_P)
Genotype value (G)	Genotype variance (V_G)
Breeding value (A)	Additive variance (V_A)
Dominant deviation (D)	Dominant variance (V_D)
Epistatic deviation (I)	Epistatic variance (V_I)
Genotype by environment interaction (GE)	Genotype by environment variance (V_{GE})
Random error (ϵ)	Random error variance (V_ϵ)

Components of genetic variance

—— 加性方差 - - - - 显性方差 ——— 总遗传方差



Applied Quantitative Genetics in Plant Breeding

Definition of plant breeding

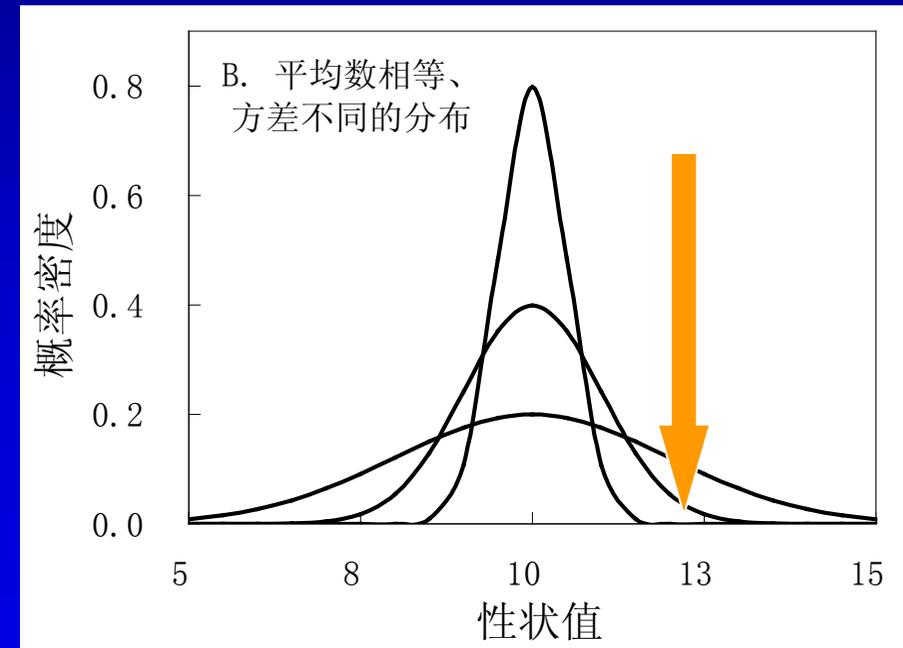
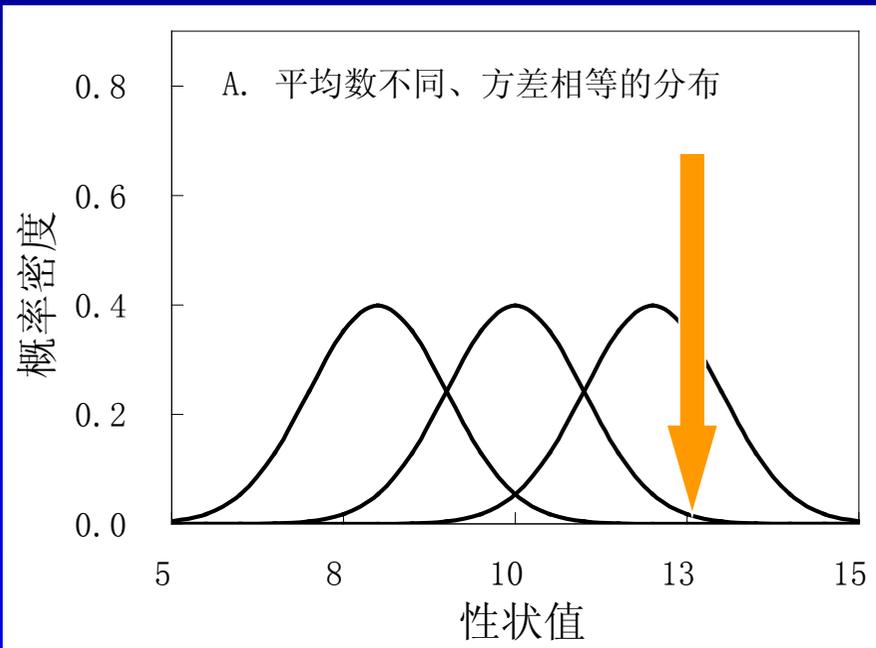
- Plant breeding is the science, art, and business of improving plants for human benefits (Rex Bernardo, 2002. Breeding for Quantitative Traits in Plants. Stemma Press, Minnesota)

Hybridization breeding after the rediscovery of Mendel's experiments

- Two resistance genes, and 3 yield genes
 - P1: **R1R1** r2r2 y1y1 **Y2Y2 Y3Y3**: Medium R, medium Y
 - P2: r1r1 **R2R2 Y1Y1** y2y2 y3y3: Medium R, low Y
- Best genotype: **R1R1 R2R2 Y1Y1 Y2Y2 Y3Y3**
- Possible genotypes in fixed progeny: $2^5=32$
 - Low R, high Y: r1r1 r2r2 y1y1 y2y2 y3y3
 - High R, low Y: **R1R1 R2R2** y1y1 y2y2 y3y3
 - And so on
- The objective of breeding is to identify the ideal genotypes out of the 32 genotypes

The ideal breeding populations

- High population mean
- Large genetic variance



Many decisions to make as a breeder

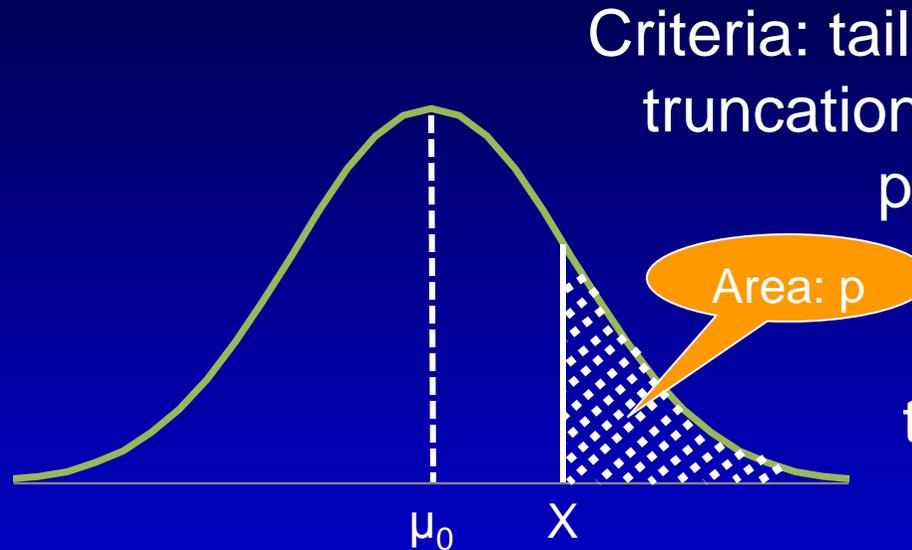
- Choose of parents (or parental selection)
- Crossing schemes
 - Single cross
 - Backcross
 - Topcross (or three-way cross)
 - Double cross
- Selection schemes
- New technologies: DH, MAS (MAS pyramiding, MABC, MARS, and GS), transformation etc.

How quantitative genetics helps plant breeding?

- Decomposition of genetic variance
 - Help determine the breeding objectives
- Estimation of heritability
 - Help determine when to select what
- Estimation of genetic gain
 - Compare selection methods
- Selection of multiple correlated traits
- Hybrid vigor theory and prediction

Genetic gain: response to selection (or the change of population mean)

Distribution
of parents

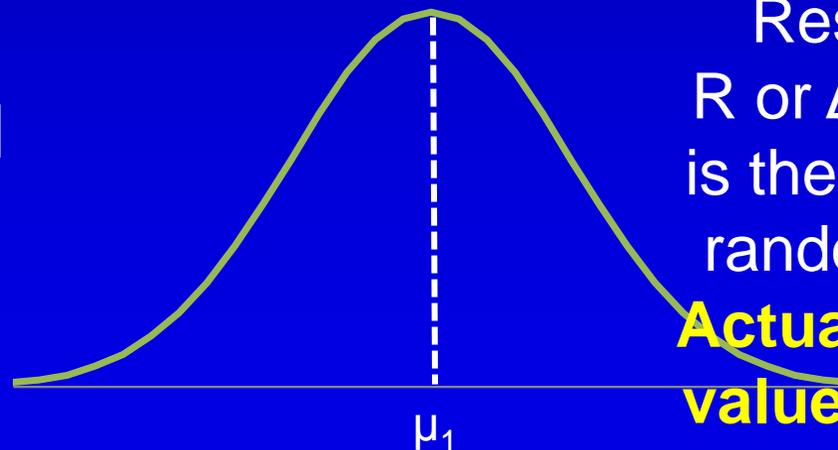


Criteria: tail selection (p) or
truncation selection (X)

$p \leftrightarrow X$

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!**

Heritability (h^2) in the narrow sense

➤ Definition: $h^2 = \frac{V_A}{V_P}$

➤ Equal to the regression coefficient of breeding value on phenotype

$$b_{(\alpha_i + \alpha_j), P_{ij}} = \frac{\text{Cov}[(\alpha_i + \alpha_j), (\alpha_i + \alpha_j + \delta_{ij} + \varepsilon_{ij(k)})]}{V(\alpha_i + \alpha_j + \delta_{ij} + \varepsilon_{ij(k)})} = \frac{V_A}{V_P} = h^2$$

➤ Square of heritability (h) equal to the correlation coefficient between

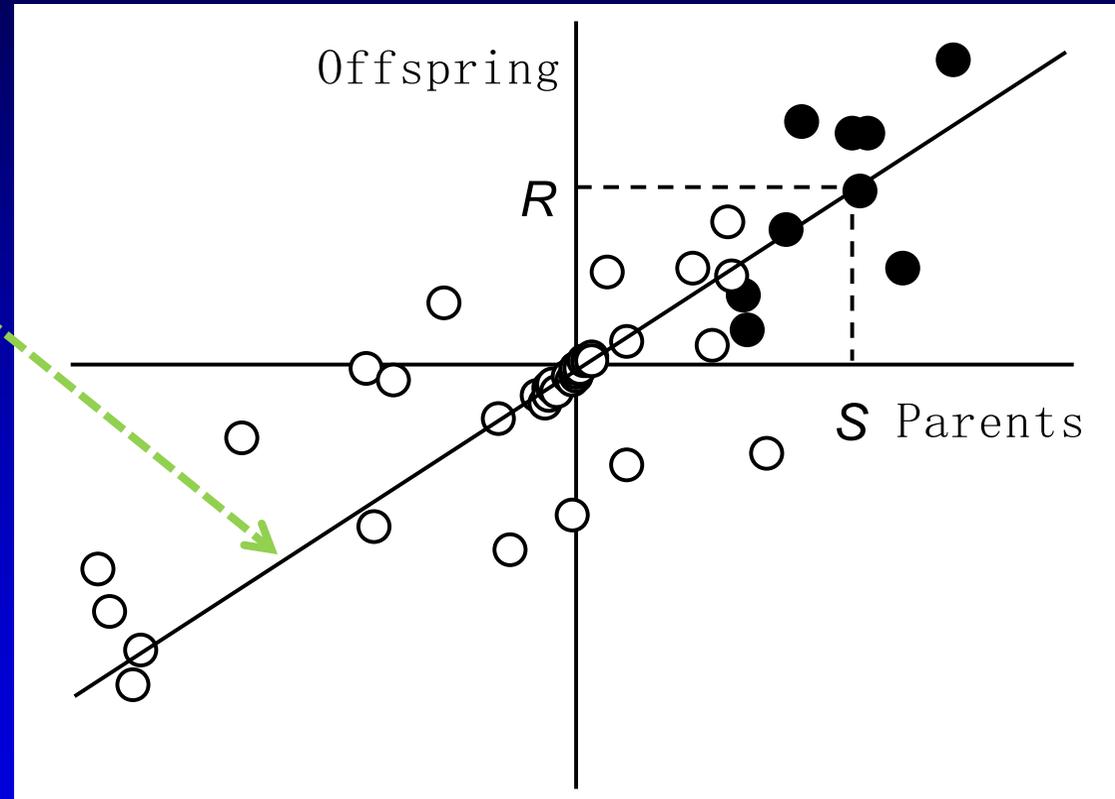
$$r_{(\alpha_i + \alpha_j), P_{ij}} = \frac{V_A}{\sqrt{V_A \times V_P}} = h$$

Estimation of R based on heritability

➤ $y = b x = h^2 x$

➤ Response to selection or Genetic gain:

R or $\Delta G = \mu_1 - \mu_0 = h^2 S$



Standardization of S

- Intensity of selection: standardized selection differential

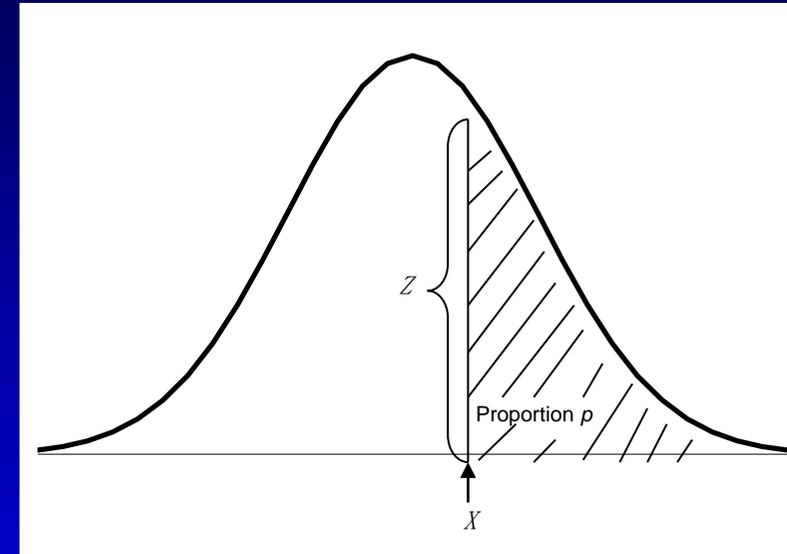
$$k_p = \frac{S}{\sqrt{V_P}}$$

$$S = k_p \sqrt{V_P}$$

- p (proportion) ↔ S (differential) ↔ k_p (intensity)

Relationship between selection intensity and selected proportion

$$k_p = \frac{Z}{p} = \frac{f(X)}{p}$$



$$\int_X^{+\infty} f(x)dx = p$$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2}$$

Density of the standardized normal distribution N (0, 1)

Selection intensity (k_p) on selected proportion (p)

Proportion (p)	0.5	0.4	0.3	0.2	0.1	0.05	0.01	0.001	0.0001
X	0.0000	0.2533	0.5244	0.8416	1.2816	1.6449	2.3263	3.0902	3.7190
Z = f(X)	0.3989	0.3863	0.3477	0.2800	0.1755	0.1031	0.0267	0.0034	0.0004
Intensity ($k_p = Z/p$)	0.7979	0.9659	1.1590	1.3998	1.7550	2.0627	2.6652	3.3671	3.9585

Estimation of genetic gain

$$R = h^2 S = k_p h^2 \sqrt{V_P}$$

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

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

Ways to increasing the response to selection

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

➤ Increase selection intensity

- $p=10\%$, $k_p=1.755$; $p=1\%$, $k_p=2.665$. So, $R(p=1\%) = 1.52 R(p=10\%)$.
- Limitation in increasing selection intensity
 - Strong selection needs much larger a population.
 - Certain amount of individuals is needed for retaining genetic variation for future genetic gain, and avoiding genetic drift
 - When 30 individuals are needed to form the next generation of breeding population, 300 individuals in the parental population are needed for $p=10\%$; 3000 individuals are needed for $p=1\%$.

Ways to increasing the response to selection

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

- Increase the coefficient of additive variance (V_A) in recurrent selection
 - By pollen control in selection
 - By recombination of S1 instead of half-sib families in half-sib family selection
 - By selection S2 instead of S1 families

$$Cov_{OP} = Cov_{OP\bar{P}} = \frac{1}{2} V_A$$

$$Cov_{HS} = \frac{1}{4} V_A$$

$$Cov_{FS} = \frac{1}{2} V_A + \frac{1}{4} V_D$$

Ways to increasing the response to selection

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

- Increase additive variance (V_A) itself
 - By introgressing other germplasm into the breeding population
- Increase heritability
 - By reducing non-genetic effects

Correlated response to selection

- Indirect selection: selection applied to traits other than the one which is desired to improve
- Assume trait 1 and trait 2 are correlated, and we want to improve trait 1. The selection on trait 2 will cause the genetic gain on trait 1
- The indirect response is estimated by

$$R_1^* = r_{G12} \sqrt{\frac{V_{G2}}{V_{G1}}} R_2 = r_{G12} k \sqrt{V_{P1} h_1^2 h_2^2}$$

Relative gain from indirect and direct selections (same selection intensity applied)

$$\frac{R_1^*}{R_1} = r_{G12} \sqrt{\frac{h_2^2}{h_1^2}}$$

- In theory, indirect selection can be better than direct selection, when
 - the trait in interest (i.e. trait 1) has a low heritability
 - the secondary trait (i.e. trait 2) has a substantially higher heritability
 - the genetic correlation is high

Secondary traits can be useful...

- If the desired trait is difficult to measure with precision
- If the desired trait is costly to measure
- When the desired trait is yield in farmers' field, the selection on yield in breeders' field is an indirect selection
- When the desired trait is yield under drought, the selection under well-water environment is an indirect selection
- To make the indirect selection effective, the genetic correlation must be significant!

$$R_1^* = r_{G12} \sqrt{\frac{V_{G2}}{V_{G1}}} R_2 = r_{G12} k \sqrt{V_{P1} h_1^2 h_2^2}$$

Modern Quantitative Genetics

➤ $P = G + E + e$

➤ $G = A + D + I$

➤ $G =$ Major gene + minor genes

➤ **$G =$ Sum of QTL effects**

- MAS: MAS pyramiding, MABC, MARS, GS

➤ Decomposition of genotypic effect

➤ Decomposition of genotypic variance