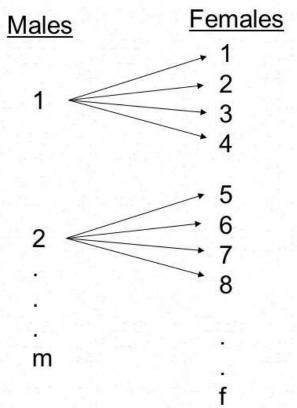
AGR 6322 Advanced Plant Breeding Fall 2018

Mating Designs – Part I



Goals for today

Introduction to Mating Designs – Part I

Previous class

- 1. Demo: Felipe will run the solutions later!
- 2. PCA vs. PCoA
- 3. When do we apply Population Genetic theory in our breeding program?

Syllabus review

Population development



Population Evaluation



Trait Integration



Product Commercialization, Marketing and Supply



Mendel → discrete traits. Followed the ratios from a cross between two distinct pure breeding lines, one could easily understand the genetics of such traits.

Several traits where the variation is more of a continuous nature (quantitative traits) → biometrical approaches.

Mating designs \rightarrow estimate different genetic components of variation.

- Predict performance of hybrids and populations
- Identify breeding methods designed to utilize different types of gene actions

Role of mating designs \rightarrow elucidating the nature of quantitative variation for their use in plant breeding.

Definition: schematic cross/es between two groups or strains of plants OBJECTIVES

- Obtain information and understand genetic control of a trait
- Obtain base population for development of plant cultivars



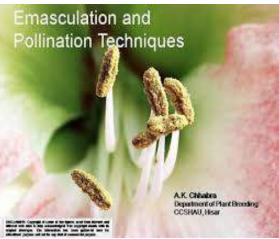


Choice of Mating Design

- Pollination method (self- or cross-pollinated)
- Type of crossing to be used (artificial or natural)
- Type of pollen dissemination (wind or insect)
- Presence of a male-sterility system
- Purpose of the project (pure breeding or genetic studies)
- Size of the population required

<u>Key questions:</u> How significant is genetic variation? How much of the variation is heritable? Breeder → identify plants with superior <u>genotypes</u> as judged by the performance of their progeny.











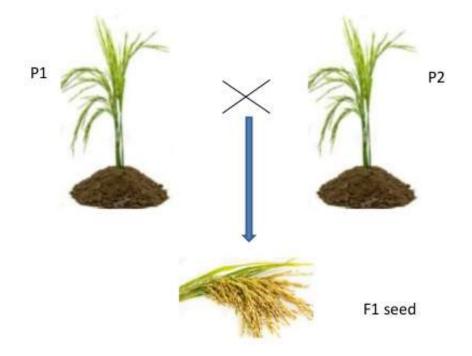
Choice of Mating Design

- Pollination method (self- or cross-pollinated)
- Type of crossing to be used (artificial or natural)
- Type of pollen dissemination (wind or insect)
- Presence of a male-sterility system
- Purpose of the project (pure breeding or genetic studies)
- Size of the population required

<u>Key questions:</u> How significant is genetic variation? How much of the variation is heritable? Breeder → identify plants with superior <u>genotypes</u> as judged by the performance of their progeny.

Mating design examples

<u>Biparental Mating</u>: plants are selected and crossed in pairs. They produce 1/2n full sib families. Progenies are tested by ANOVA into between and within families.



Mating design examples

<u>Polycross:</u> intermating a group of cultivars/lines by natural crossing in isolated blocks in the same nursery.

Most suitable design for obligate cross pollinators.

Equal opportunity to each and every clone to cross with each other, but lines must be equally represented and randomly arranged.

Polycross is used in breeding to produce cultivars, or for evaluating the GCA.



Mating design examples

<u>Polycross:</u> half sibs are generated (common parents) and progenies from individual plants (half sibs) are tested.

Covariance within progenies = Cov = 1+F 4 σ^2_A ; where: F = inbreeding coefficient of genotypes tested. Selection is done on half sib family means.

Random mating and Insufficient statistics to estimate all parameters.

No control on pollen source.

Expected genetic gains are reduced.

Mating design examples

North Carolina designs were developed in order to obtain more information about combining ability without much labor compare to full diallel designs.

- North Carolina Design I
- North Carolina Design II
- North Carolina Design III

Mating design examples

North Carolina Design I: used to estimate additive and dominance variances, and for the evaluation of full- and half-sib recurrent selection.

For self- and cross-pollinated species.

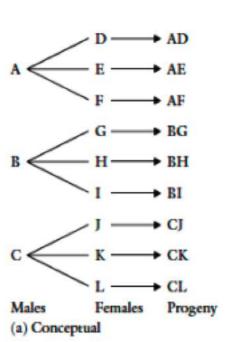
Each member of a group of parents used as males is mated to a

different group of parents.

The progenies include both full-sibs and half-sibs.

Test of significance for the additive genetic variance.

It is higly used for tree breeding.



Mating design examples

North Carolina Design II: each member of a group of parents used as males is mated to each member of another group of parents used as females. It is used to evaluate inbred lines for combining ability. The design is most adapted to plants that have multiple flowers.

The design is essentially a two-way ANOVA.

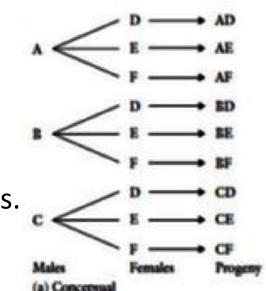
Allows measure of both GCA and SCA.

Every progeny family has half sib relationships

through both common males and common females.

Reciprocal crosses may be carried out to

analyze maternal effects

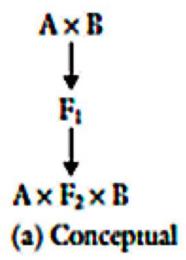


Mating design examples

North Carolina Design III: it is considered the most powerful of all the three NC designs. A random sample of F2 plants is backcrossed to the two inbred lines from which the F2 was descended.

A modification called the triple test cross has also been introduced, i.e adding a third tester not just the two inbreds.

It is capable of testing non-allelic (epistatic) interactions, also capable of estimating additive and dominance variance.



Mating design examples

<u>Full Diallel</u>: full diallel design consists of parents crossed in all possible combinations, one set of F1's and reciprocal F1's. This system gives n² genotypes.

	P1	P2	Р3	P4
P1	X	X	X	X
P2	X	X	X	X
Р3	X	X	X	X
P4	x	X	X	Х

<u>Partial Diallel</u>: This method includes parents and one set of F1's without reciprocal F1's. This design gives p(p+1)/2 genotypes.

	P1	P2	Р3	P4
P1	X	x	X	X
P2		X	X	X
Р3			X	X
P4				X

- Selection of suitable parents and good mating designs are keys to success of plant breeding schemes.
- Selection of mating design depends on various factors: stages of breeding program
- In terms of amount of information, diallel > NC II > NC III > NC I > BIPs Diallel mating design is the most important for GCA and SCA.
- The proper choice and use of a mating design will provide the most valuable information for breeding

Estimating Variance Components

Essential for a breeder to estimate the variance components (var. comp.) in their breeding populations

Many techniques have been develop to estimate var. comp. for quantitative traits

Two main considerations:

- What kind of relatives to use in the study?
- Experimental design

The degree of precision is a function of

- Number of families
- Number of individuals within families

Estimating Variance Components

Essential for a breeder to estimate the variance components (var. comp.) in their breeding populations

- Parent-Offspring regression
- Half-Sib Sire model
- Full-Sib Nested model
- Cross-Classified (factorials) Designs
 - North Carolina
 - Diallels

Estimating Variance Components

1. Regression of offspring phenotypes on those of their parents is one of the most common used method for estimating heritability.

Several reasons:

- For many species is the most easily identified relationship in the field
- Computations are based on LS regression
- Neither dominance or linkage influence the covariance between parents and offspring
- Unbiased by selection of parents

Estimating Variance Components

- Regression of offspring phenotypes on those of their parents is one of the most common used method for estimating heritability.
- Single offspring and single parent in each family:

$$z_{oi} = \alpha + \beta_{op} z_{pi} + e_i$$

 z_{oi} =offspring phenotype for the ith family

 z_{pi} = parent phenotype for the ith family

 β_{op} = regression coefficient, estimated by b_{op} = $Cov(z_o, z_p)/Var(z_p)$

 $e_i = residual$

Assuming no environmental causes of resemblance between parent and offspring then:

$$E(b_{op}) = \frac{\sigma(z_o, z_p)}{\sigma^2(z_p)} = \frac{(\sigma_A^2/2) + (\sigma_{AA}^2/4) + (\sigma_{AAA}^2/8) + \dots}{\sigma_z^2}$$

Estimating Variance Components

 Regression of offspring phenotypes on those of their parents is one of the most common used method for estimating heritability.

$$z_{oi} = \alpha + \beta_{op} z_{pi} + e_i$$

$$E(b_{op}) = \frac{\sigma(z_o, z_p)}{\sigma^2(z_p)} = \frac{(\sigma_A^2/2) + (\sigma_{AA}^2/4) + (\sigma_{AAA}^2/8) + \dots}{\sigma_z^2}$$

Then assuming no epistasis

Expression varies depending if two parents are considered and if offspring is unbalance

Estimating Variance Components

1. Regression of offspring phenotypes on those of their parents is one of the most common used method for estimating heritability.

Sometimes is impossible to acquired data for offspring/parents simultaneously:

- Age specific character (e.g. forestry)
- Species with non-overlapping generations
- Parents are not possible to be identified

Alternatively, analysis of contemporary relatives can be used to estimate genetic parameters. Three types: Half-sibs, Full-sibs, Combination.

Goal: partition the total phenotypic variance into within- and among-family components. ANOVA (balance data), REML, ML.

Estimating Variance Components

2. Half-Sibs (HS)

 Under random mating and negligible epistasis and common environmental effect, four times the genetic covariance between halfsibs (HS) is equal to the additive variance:

$$\sigma(HS) = (\sigma_A^2/4) + (\sigma_{AA}^2/16) + \dots$$

$$4*\sigma(HS) = \sigma_A^2$$

 Problems may arise if maternal effect or common environmental effect are significant. Thus the additive variance may be overestimated.

Estimating Variance Components

2. Half-Sibs (HS)

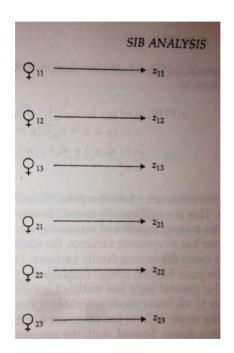
One-way ANOVA used to calculate var. comp.

$$z_{ij} = \mu + s_i + e_{ij}$$

 z_{ij} =phenotype of jth offspring ith father

 s_i =effect of the ith sire (parent)

 e_{ij} =residual (dominance, genetic variance among mothers and environmental variance



Estimating Variance Components

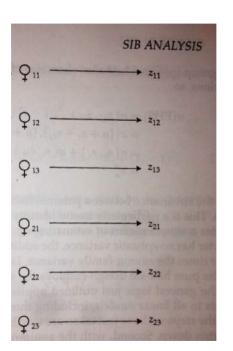
2. Half-Sibs (HS)

One-way ANOVA used to calculate var. comp.

$$z_{ij} = \mu + s_i + e_{ij}$$

- Sires assumed to be a random sample of population with E(s)=0 and variance= $\sigma_s^2 \rightarrow$ among-family variance.
- Residual assumed random with common variance = σ_e²
 →within-family variance.
- Assuming independent random factors (covariance=zero)

$$\sigma_z^2 = \sigma_s^2 + \sigma_e^2 \qquad \qquad h^2 = \frac{4 * \sigma_s^2}{\sigma_z^2}$$



Estimating Variance Components

2. Half-Sibs (HS)

One-way ANOVA used to calculate var. comp.

Factor	d.f	SS	MS	E(MS)
Among-families (sire)	S-1	$SS_s = \sum_i n_i (\overline{z}_i - \overline{z})^2$	$MS_s = SS_s / S-1$	$\sigma_e^2 + n\sigma_s^2$
Within-families	N-S	$SS_e = \sum_{i} \sum_{j} (z_{ij} - \overline{z}_i)^2$	$MS_e = SS_e / N-S$	σ_e^2
Total	N-1	$\sum_{i}\sum_{j}(z_{ij}-\overline{z})^{2}$		

Thus:

$$Var(s) = \frac{MS_s - MS_e}{n}$$

$$Var(e) = MS_e$$

$$Var(z) = Var(s) + Var(e)$$

$$h^2 = 4 * Var(s) / Var(z)$$

Estimating Variance Components

- Half-Sibs (HS)
 - Avoid designs with families smaller than 4 individual per family
 - Broad recommendation is to use between 5-20 families
 - ANOVA will be bias for unbalance data. Use ML or REML on those cases (Mixed models)
 - ANOVA can yield negative estimations of heritability.
 - ANOVA procedure will not give breeding values, but BLUP will

Estimating Variance Components

Full-Sibs (FS)

- Same statistical methods than HS applied to FS analysis, but we need to adjust the interpretations of the within- and among-family variance components.
- Under no dominance and no common environmental effect, twice the genetic covariance between full-sibs provide an estimate of the additive variance:

$$\sigma(FS) = (\sigma_A^2 / 2) + (\sigma_D^2 / 4) + (\sigma_{AA}^2 / 4) + (\sigma_{AD}^2 / 8) + (\sigma_{DD}^2 / 16)..$$
$$2 * \sigma(FS) = \sigma_A^2$$

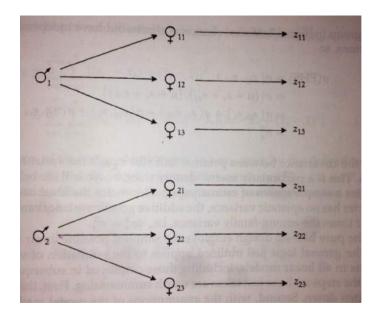
 Problems if dominance or common environmental effect are significant. Thus the additive variance may be overestimated.

Estimating Variance Components

Full-Sibs (FS)

- Offspring under same female are FS
- Offspring under same male are HS
- Linear model:

$$z_{ijk} = \mu + s_i + d_{ij} + e_{ijk}$$



 z_{ij} =phenotype of kth offspring from the family of the ith sire and jth dam s_i =effect of the ith sire (parent)

 d_{ij} =effect of the jth dam mated to the ith sire (parent)

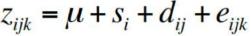
 e_{iik} =residual deviation

Estimating Variance Components

Full-Sibs (FS)

- Offspring under same female are FS
- Offspring under same male are HS
- Linear model:

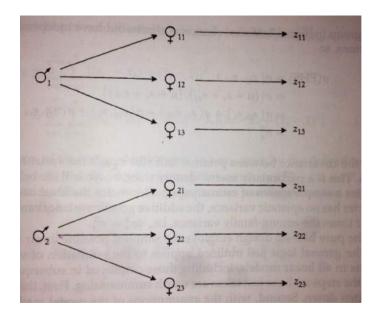
$$z_{ijk} = \mu + s_i + d_{ij} + e_{ijk}$$



Random effects s, d and e, with expectation equal to zero, and:

$$\sigma_z^2 = \sigma_s^2 + \sigma_d^2 + \sigma_e^2$$

- $\sigma_{\rm s}^2$ is the variance among sires
- σ_d^2 is the variance among dams within sire
- σ_s^2 is the variance within full-sib families



Estimating Variance Components

Full-Sibs (FS)

- Offspring under same female are FS
- Offspring under same male are HS
- Linear model:

$$z_{ijk} = \mu + s_i + d_{ij} + e_{ijk}$$

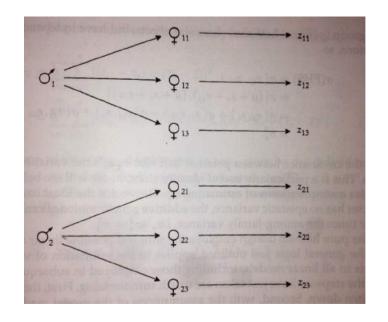
· Assuming null epistasis effect

$$\sigma_s^2 \approx (\sigma_A^2/4)$$

$$\sigma_d^2 \approx (\sigma_A^2/4) + (\sigma_D^2/4) + \sigma_{Ec}^2$$

$$\sigma_e^2 \approx (\sigma_A^2/2) + (3\sigma_D^2/4) + \sigma_{Es}^2$$

$$2*\sigma_s^2 = \sigma_A^2$$
$$\sigma_d^2 - \sigma_s^2 = \sigma_D^2$$



Estimating Variance Components

3. Full-Sibs (FS)

Factor	d.f	SS	MS	E(MS)
Sires	S-1	$SS_s = \sum_{i} \sum_{j} n_{ij} (\overline{z}_i - \overline{z})^2$	$MS_s = SS_s/S-1$	$\sigma_e^2 + k_2 \sigma_d^2 + k_3 \sigma$
Dams(sires)	S(D-1)	$SS_d = \sum_i \sum_j n_{ij} (\overline{z}_{ij} - \overline{z}_i)^2$	$MS_d = SS_d/S(D-1)$	$\sigma_e^2 + k_1 \sigma_d^2$
Sibs(dams)	N – SD	$SS_w = \sum_i \sum_j \sum_k (z_{ijk} - \overline{z}_{ij})^2$	MS _e = SS _e / N-SD	σ_{*}^{2}
Total	N-1	$\sum_{i} \sum_{j} \sum_{k} (z_{ijk} - \overline{z})^{2}$		e

$$Var(s) = \frac{MS_s - MS_e - (k_2 - k_1)(MS_d - MS_e)}{k_3}$$

$$Var(d) = \frac{MS_d - MS_e}{k_1}$$

$$Var(e) = MS_e$$

Estimating Variance Components

North Carolina II

This design involves all possible crosses between two sets of individuals.

Here the group of females (N_d) are different than the group of males (N_s)

Plants need to produce multiple flowers.

Number of crosses limited by the number of flowers a female can produce or by the number of emasculations that can be achieved

Estimating Variance Components

North Carolina II

Model for Design II is:
$$z_{ijk} = \mu + s_i + d_j + I_{ij} + e_{ijk}$$

 s_{i} , d_{i} =additive effects (breeding value) of the ith sire and jth dam

 I_{ij} =non-additive (interaction) effect due to the combination of genes from parents ith and jth

 e_{ijk} =residual deviation

Estimating Variance Components

North Carolina II

Model for Design II is:
$$z_{ijk} = \mu + s_i + d_j + I_{ij} + e_{ijk}$$

 s_{i} , d_{i} =additive effects (breeding value) of the ith sire and jth dam

 I_{ij} =non-additive (interaction) effect due to the combination of genes from parents ith and jth

 e_{ijk} =residual deviation

Estimating Variance Components

North Carolina II

Model for Design II is:
$$z_{ijk} = \mu + s_i + d_j + I_{ij} + e_{ijk}$$

Assumptions:

Parents sampled randomly from a random-mating population

All effects are independent with zero expectation and variances such that: $\sigma_z^2 = \sigma_d^2 + \sigma_s^2 + \sigma_I^2 + \sigma_e^2$

Effects are:
$$s_i = \mu_i - \mu$$

$$d_j = \mu_j - \mu$$

$$I_{ij} = \mu_{ij} - \mu - s_i - d_j$$

$$e_{ijk} = z_{ijk} - \mu - s_i - d_j - I_{ij}$$

Estimating Variance Components

North Carolina II

- Variance component for sire and dam are equivalent. So their estimates should be similar (accounting for sampling). However, the covariance between maternal half-sibs will be inflated by any genetic or environmental maternal effect.
- Thus, the difference between Var(d) and Var(s) is an estimate of the maternal effects

 The interaction variance, Var(I) is equivalent to Cov(FS) – Cov(MHS) – Cov(PHS)

Estimating Variance Components

North Carolina II

Using the known covariance among relatives, then:

$$\sigma_{s}^{2} = \sigma(PHS) \approx (\sigma_{A}^{2}/4) + (\sigma_{AA}^{2}/16)$$

$$\sigma_{d}^{2} = \sigma(MHS) \approx (\sigma_{A}^{2}/4) + (\sigma_{AA}^{2}/16) + \sigma_{Gm}^{2} + \sigma_{Ec}^{2}$$

$$\sigma_{I}^{2} = \sigma(FS) - \sigma(PHS) - \sigma(MHS) \approx (\sigma_{D}^{2}/4) + (\sigma_{AA}^{2}/8) + (\sigma_{AD}^{2}/8) + (\sigma_{DD}^{2}/16)$$

- G_m genetic maternal effects
- E_c common (maternal) environmental effect

Estimating Variance Components

North Carolina II

Factor	df	Sums of Squares	E(MS)
Sires	N_s-1	$nN_d\sum_i(\bar{z}_i - \bar{z})^2$	$\sigma_e^2 + n\sigma_I^2 + nN_d\sigma_e^2$
Dams	N_d-1	$nN_s \sum_j (\bar{z}_{\cdot j} - \bar{z})^2$	$\sigma_e^2 + n\sigma_I^2 + nN_s\sigma_I^2$
Interaction	$(N_d - 1)(N_s - 1)$	$n\sum_{i,j}(\bar{z}_{ij}-\bar{z}_{i\cdot}-\bar{z}_{-j}+\bar{z})^2$	$\sigma_e^2 + n\sigma_I^2$
Error	$N_sN_d(n-1)$	$\sum_{i,j,k}(z_{ijk}-\bar{z}_{ij})^2$	σ_c^2
Var(e)	$= MS_e$ $= \frac{MS_I - MS_e}{2}$	$Var[Var(e)] = \frac{2(MS_e)^2}{df_e + 2}$ $Var[Var(I)] = \frac{2}{n^2} \left[\frac{(MS_I)^2}{df_I + 2} \right]$	$\frac{2}{1+\frac{(MS_e)^2}{MS_e}}$
	$= \frac{MS_d - MS_I}{nN_s}$	$Var[Var(d)] = \frac{2}{(nN_s)^2} \left[\frac{(N_s)^2}{df} \right]$	
Var(s)	$=\frac{MS_s-MS_I}{nN_d}$	$Var[Var(s)] = \frac{2}{(nN_d)^2} \left[\frac{(N_d)^2}{df_d} \right]$	$\frac{(S_s)^2}{1+2} + \frac{(MS_I)^2}{df_I + 2}$

Estimating Variance Components

Diallel

In North Caroline Designs two sets of different parents (males and females) are used.

This design involves all possible crosses between the same set of parents n both axes, with the same individuals (or inbred lines) as male and females.

Considering N parents there are N² potential crosses.

Number of crosses will be defined by:

if all potential crosses, or reciprocals or selfed (within lines or diagonals) are considered.

Estimating Variance Components

Diallel

Two approaches to analyse the data depending if parents are considered fixed or random

Number of crosses, same as in NCD, limited by the number of flowers a female can produce or by the number of emasculations that can be achieved.

Estimating Variance Components

Diallel

Half-Diallel or Pooled reciprocals with no self crosses

	BSSS								LSC					MO	
		B14A	B37	B73	B84	Lo950	Lo951	C103	C123	Lo881	Mo17	Va59	H55	Pa91	
B S S	B14A		×	×	×	×	×	8	8	0	0	8	0	8	
	B37			×	×	×	×	8	8	⊗	8	8	8	0	
	B73				×	×	×	8	8	8	8	8	8	0	
	B84					×	×	8	0	8	8	8	8	8	
	Lo950						×	8	8	8	8	8	8	8	
	Lo951							8	8	8	8	8	8	8	
L S C	C103								×	×	×	×	8	8	
	C123									. ×	×	×	8	8	
	Lo881										×	×	8	8	
	Mo17											×	8	8	
	Va59												8	8	
M O	H55													8	
	Pa91														

N(N-1)/2 crosses

Assuming no maternal or paternal effects, reciprocal crosses are expected to yield equivalent progeny phenotype distribution.

So either pooling or performing only half of the diallel, excluding selfed.

Vuylsteke et al 2000 - Maize

Estimating Variance Components

Diallel

Model for Design II is:
$$z_{ijk} = \mu + g_i + g_j + s_{ij} + e_{ijk}$$

 $g_{i,j}g_{j}$ = general combining ability (GCAs) of parents i and j s_{ij} =specific combining ability (SCA) of the cross $i \times j$

GCA – describes the average performance of a parent when cross with other genotypes (outcrossings) or in hybrid combination with other genotypes (inbreds). Equivalent to BV of an individual

SCA – describes the degree to which specific parental combinations lead to deviations in progeny phenotypes from the expected based on the average performance of the parents

Estimating Variance Components

Diallel

Model for half- Diallel is:
$$z_{ijk} = \mu + g_i + g_j + s_{ij} + e_{ijk}$$

Assumptions:

Under the random interpretation; parents sampled randomly from a random-mating population

All effects are independent with zero expectation and variances such that: $\sigma_z^2 = 2\sigma_{GCA}^2 + \sigma_{SCA}^2 + \sigma_e^2$

The GCA variance is equivalent to the Cov(HS):

$$\sigma(HS) = \sigma(z_{ijk}, z_{ij'k'}) = \sigma^2(g_i) = \sigma_{GCA}^2$$

The SCA variance is equivalent to the Cov(FS)-2Var(GCA):

$$\sigma(FS) = \sigma(z_{ijk}, z_{ijk'}) = 2\sigma^2(g_i) + \sigma^2(s_{ij}) = 2\sigma_{GCA}^2 + \sigma_{SCA}^2$$

Estimating Variance Components

Thus:

$$Var(GCA) = \frac{MS_{GCA} - MS_{SCA}}{n(N-2)}$$

$$Var(SCA) = \frac{MS_{SCA} - MS_{e}}{n}$$

$$Var(e) = MS_{e}$$

Estimating Variance Components

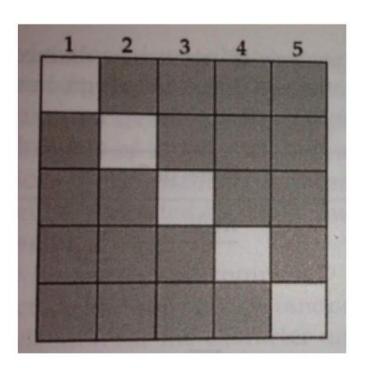
Diallel

Reciprocals with no self crosses

NxN crosses performed without selfed.

Different ways to analyse it!

Use them when parent use as female or male matters



Estimating Variance Components

Diallel

Partial or Incomplete Diallels

The number of crosses varies

Circular design

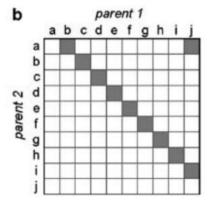
Single-round robin

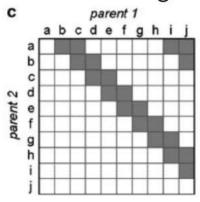
Double-round robin

Table 1 - Scheme of	in interpopulational	l circulant diallel table,	with $p = 34$ and $s = 5$.

Flint inbred lines	Dent inbred lines									
	1	2	3	4	5	6	7	***	33	34
1	Y1,1	Y1,2	Y1,3	Y1,4	Y1,5					
2		У2,2	У2,3	Y2,4	У2,5	У2,6				
3			У3,3	Уз,4	У3,5	У3,6	У3,7			
4				Y4,4	y 4,5	У4,6	Y4,7	***		
5					Y5,5	У5,6	Y 5,7	***		
6						У6,6	У6,7	***		
7							Y7,7	***		
33	У33,1	У33,2	У33,3						У33,33	y 33,34
34	У34,1	У34,2	Y34,3	У34,4						y 34,34

Use when restriction of flowers or logistic does not allow all crosses





tos et al 2005 - Maize Verhoeven et al 2006