LOUISIANA STATE UNIVERITY
College of Agriculture
School of Plant, Environmental, and Soil Sciences
AGRO 7075 Prediction-based Breeding

Mating designs

Prof. Roberto Fritsche-Neto

rfneto@agcenter.lsu.edu

Baton Rouge, Mar 27th, 2023

Importance of mating designs

- **Estimate components of variance**

- **Understand the genetic control**

- **Identify:**
- *the best parents,*
- *population structure (heterotic groups),*
- *testers, and*
- *the best combinations (hybrids)*
- Support decisions – *populations and breeding schemes*

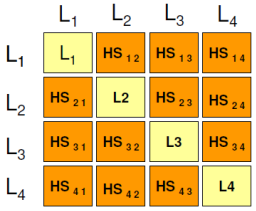
Principle and ideal scenario

Principle: generate various levels of relatedness among the relative progenies by effecting a series of crosses among individuals of a random mating population that caused different statistical components of variation from which genetic components of variations were estimated.

Ideal: all possible crosses, but generally not practical and not economically or biologically necessary

Balanced mating designs are preferred to estimate general combining ability (GCA) and specific combining ability (SCA)

- Therefore, breeders have moved from full diallel to top cross over the years due to practical issues.
- E.g., 49 lines, divided into two groups (34 and 15)



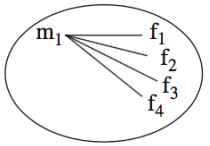
$$SC = n(n-1)/2$$

$$SC = \frac{49(48)}{2} = 1{,}176$$

	P1	P2
P3	F1(1,3)	F1(2,3)
P4	F1(1,4)	F1(2,4)

$$SC = na\ x\ n_b$$

$$SC = 34\ x\ 15 = 510$$



$$SC = t_b n_a + t_a n_b$$

$$SC = 1x34 + 1x15 = 49$$

Mating Designs: A Survey

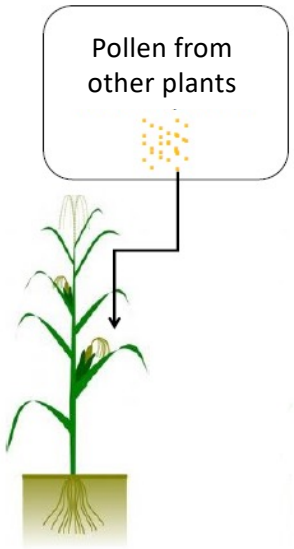
- **Two schools have contributed the most:** North Carolina State University, USA, and the University of Birmingham, England (UK)
- *The former concentrated on the analyses of random mating populations. They are extensions of the one-way, two-way, and nested anova.*
- *The latter worked on a series of mating designs that can be easily selfed or crossed*

- **Two important issues should be considered before choosing any mating design:**
- ***The kind of relatives that will be available for analysis:*** certain types of relatives are observed more readily in some species than others, and some types of covariance between relatives are more likely to approximate the desired quantities than others.
- ***The degree of precision:*** number of individuals per family, genetic distance between parents, number of parents, number of families, traits, genetic effects controlling the traits, and polyploid *vs.* diploid

Sib Analysis – progeny test

- An attractive alternative to parent-offspring regression in estimating genetic parameters
- **There are three types of sib analysis**: full-sib families, half-sib families, and; combinations of the half-sib and full-sib families
- A comparison between families and within families' variances provides an estimate of VA (additive genetic variance), in some cases, dominance (VD)
- The common environmental effects are the main drawbacks of any sib analysis.
- Let’s see the overall way to estimate GCA and SCA in a full diallel

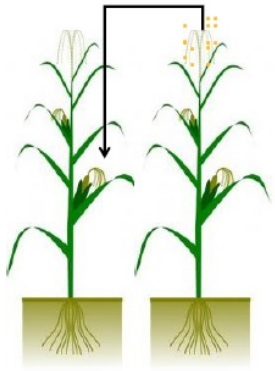
Progeny types



Half-sibs (HS)

Open pollination

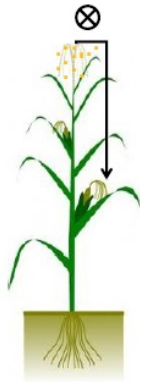
$$\sigma_g^2 = \frac{1}{4}\sigma_A^2$$



Full-sibs (FS)

Manual pollination

$$\sigma_g^2 = \frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_D^2$$



Self-sibs (Sn)

Natural or manual pollination

$$\sigma_g^2 = \frac{1}{2}\sigma_A^2$$

Sib Analysis – progeny test

Table 1. ANOVA table of polycross design with many replicated.

Source	df	MS	EMS	Variance components
Progenies	$g-1$	M_1	$\sigma_e^2 + r\sigma_{prog}^2$	$\sigma^2 prog = Cov(HS) = \frac{1+F}{4}\sigma^2 A$
Blocks	$r-1$	M_2	-	-
error	$(g-1)(r-1)$	M_3	σ_e^2	$\sigma^2{}_e = \sigma^2$

Source: Wricke and Weber, 1986

$$RS = i.\frac{\frac{1}{4}\sigma_a^2}{\sqrt{\sigma_G^2+\sigma_D^2+\sigma_E^2/\mathrm{r}}}$$

Generation Mean Analysis - Mather (1949)

- It is an extensive analysis of family means, universally known as the "basic generation analysis"
- Six generations (P1, P2, F1), F2, F2xP1, and F2xP2
- Test and estimate the additive, dominance, and epistatic effects of genes at the means level
- Provided a rationale to estimate VA, VD, and VE, thus heritability
- Vp = F2
- Ve = V(P1 + P2 + F1) / 3
- Vg = Vp - Ve

Mean analysis = m + a

- **Two consecutive generations.** ***E.g., F_1 and F_2***
- **$F_1 = m + a + d$**
- **$F_2 = m + a + 1/2d$**
- **$F_3 = m + a + 1/4d$**
- **...**
- **F_∞ - homozygosity – the lines performance will be *m + a***
- **So, progenies with higher *m + a* are "superior"**

$$m + a = 2F_2 - F_1$$

$$m + a = 2(m + a + \frac{d}{2}) - (m + a + d)$$

Cross	F1 (Kg ha)	F2 (Kg ha)	2F2 – F1	m + a
L1 x L2	2,0	2,2	2 * 2,2 - 2,0	2,4
L3 x L4	2,5	2,5	2 * 2,5 - 2,5	2,5
L5 x L6	2,2	3,0	2 * 3,0 - 2,2	3,8

North Carolina Designs

Comstock and Robinson (1952) proposed three mating designs, I, II, and III

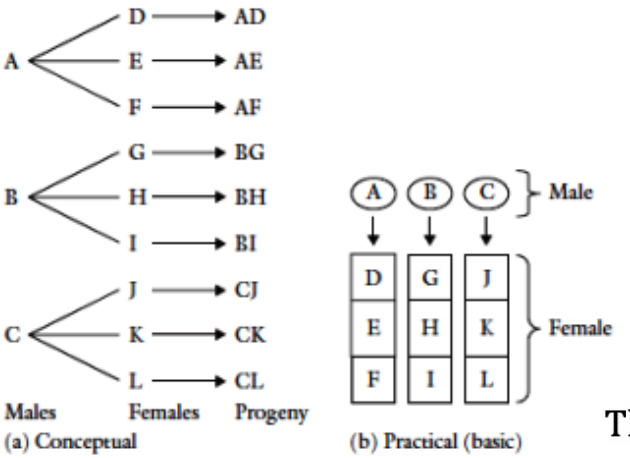


Table 2. Skeleton of general ANOVA for North Carolina design I.

Source of variation	df	MS	Expected mean squares
Males	$(m-1)$	M_1	$\sigma^2 w + r\sigma^2_{mf} + rf\sigma^2_m$
Females	$m(f-1)$	M_2	$\sigma^2 w + r\sigma^2_{mf}$
Within plots	$mf(r-1)$	M_1	$\sigma^2 w$
Total	$rmf-1$		

Source: Acquaah, 2012.

The parameter $\sigma^2 w$ refers to the average variance within the full sib families and is given as;

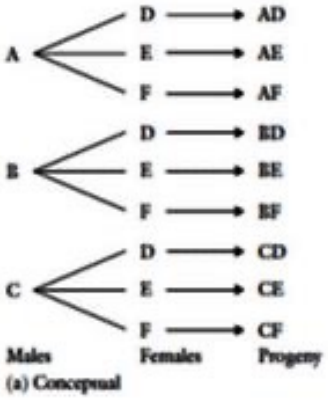
$$\sigma^2 w = M_1 = \tfrac{1}{2}V_A + \tfrac{3}{4}V_D + V_E$$

$$\sigma^2_m = (M_1 - M_2)/rf = \tfrac{1}{4}V_A$$

$$r\sigma^2_{mf} = (M_2 - M_3)/r = \tfrac{1}{4}V_A + \tfrac{1}{4}V_D$$

$$RS.males = i.\frac{\frac{1}{4}\sigma_a^2}{\sqrt{\sigma_G^2+\sigma_D^2+\sigma_E^2/\mathrm{r}}}$$

North Carolina II



	P1	P2
P3	F1(1,3)	F1(2,3)
P4	F1(1,4)	F1(2,4)

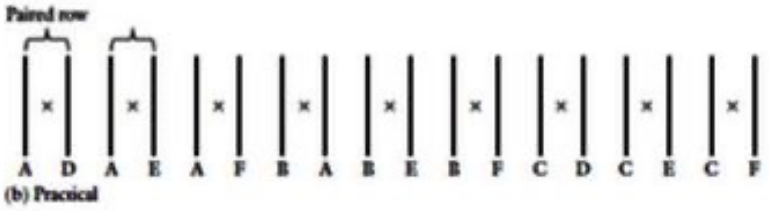


Table 5. Skeleton of general ANOVA of NC II.

Source of variation	df	MS	Expected mean squares
Replications	$r-1$		
Males	$m-1$	M_1	$\sigma^2w+r\sigma^2_{mf}+rf\sigma^2_m$
Females	$f-1$	M_2	$\sigma^2w+r\sigma^2_{mf}+rm\sigma^2_f$
Males x females	$(m-1)(f-1)$	M_3	$\sigma^2w+r\sigma^2_{mf}$
Within progenies	$mf(r-1)$	M_4	σ^2w
Error	$(r-1)(mf-1)$	M_5	σ^2
Total	$rmf-1$		

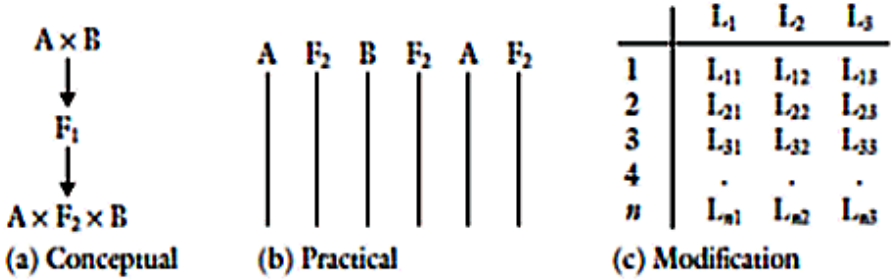
Source: Kearsey and Pooni, 1996

$$RS.males = i.\frac{\frac{1}{4}\sigma_a^2}{\sqrt{\sigma_G^2+\sigma_D^2+\sigma_E^2/r}} \qquad RS.females = i.\frac{\frac{1}{4}\sigma_a^2}{\sqrt{\sigma_G^2+\sigma_D^2+\sigma_E^2/r}}$$

Where: σ^2w is the within progenies genetic and environmental variances. In the absence of epistasis and common environmental effects, σ^2_{mf} is a function of dominance variance V_D only (Kearsey and Pooni, 1996). If there is environmental variation between FS families, this could be due to general and specific maternal effects. The general maternal effects (V_{EM}) will appear in $\sigma^2_f = \frac{1}{4}V_A + V_{EM}$, while the specific maternal effects $V_{EC} - V_{EM}$ will appear in $\sigma^2_{mf} = \frac{1}{4}V_D + (V_{EC} - V_{EM})$ and will be confounded with V_D (Kearsey and Pooni, 1996). If the number of males and females is the same, $n_1 = n_2 = n$, we can have a test of maternal effects by comparing M_1/M_2 as a variance ratio.

North Carolina III

- A random sample of F2 plants is backcrossed to the two inbred lines from which the F2 was descended - triple testcross
- Capable of testing non-allelic (epistatic) interactions



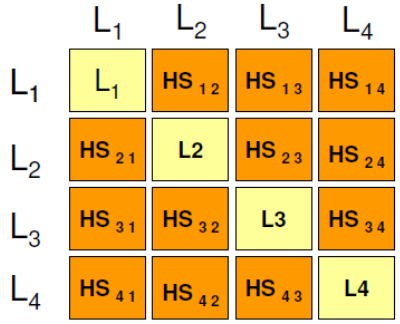
(a) Conceptual (b) Practical (c) Modification

Table 7. Skeleton of NC III ANOVA.

Source of variation	df	MS	Expected mean squares
Testers, p	1	M_4	$\sigma^2 + r\sigma^2_{mp} + rmK^2_p$
Males (F_2), m	$m-1$	M_3	$\sigma^2 + 2r\sigma^2_m$
Testers x parents	$m-1$	M_2	$\sigma^2 + r\sigma^2_{mp}$
Within FS families/error	$(r-1)(2m-1)$	M_1	σ^2
Total	$2mr-1$		

Source: Hallauer *et al.*, 2010

Griffing's diallel



- **A set of genotypes are crossed in all possible combinations**
- **Lead to the testing and estimation of the general (GCA) and specific (SCA) combining abilities**
- **Four diallel mating methods depend on whether parents and reciprocals are retained or excluded from a particular design:**
- Method I or full diallel: parents, one set of F1s, and reciprocal F1s. n^2 genotypes
- Method II: parents and one set of F1s without reciprocals F1s. $p(p+1)/2$ genotypes
- Method III: one set of F1s and the reciprocals are included. $p(p-1)$ genotypes
- Method IV: only one set of F1s is included. It is the most common of the diallel crossing systems. $p(p-1)/2$ genotypes

Diallel

The table 8 presents the estimates of variances for variance components both fixed and random model.

Table 8. Skeleton of ANOVA for method I diallel design.

Source	df	SS	MS	Expected mean squares	
				Model I	Model II
GCA	$p-1$	S_g	M_g	$\sigma^2+2p(\frac{1}{p-1})\Sigma g^2{}_i$	$\sigma^2+\frac{2(p-1)}{p}\sigma^2{}_g+2p\sigma^2{}_g$
SCA	$p(p-1)/2$	Ss	M_s	$\sigma^2+\frac{2}{p(p-1)}\Sigma\Sigma S_{ij}{}^2$	$\sigma^2+\frac{2(p^2-p+1)}{p^2}\sigma_s{}^2$
Reciprocal eff.	$p(p-1)/2$	S_r	M_r	$\sigma^2+2(\frac{2}{p(p-1)})\sum_{i<j}\Sigma r_{ij}{}^2$	$\sigma^2+2\sigma^2{}_r$
Error	m	S_e	M_e	σ^2	

Source: Griffing (1956b)

Table 9. The analysis of variance for method II.

Source	df	SS	MS	Expected mean squares	
				Model I	Model II
GCA	$p-1$	S_g	M_g	$\sigma^2+(2+p)(\frac{1}{p-1})\Sigma g_i{}^2$	$\sigma^2+\sigma_s{}^2+(p+2)\sigma_g{}^2$
SCA	$p(p-1)/2$	Ss	M_s	$\sigma^2+\frac{2}{p(p-1)}\sum_i\sum_j s_{ij}{}^2$	$\sigma^2+\sigma_s{}^2$
Error	m	S_e	M_e'	σ^2	

Source: Griffing (1956b)

Table 10. Skeleton of ANOVA of Diallel method III.

Source	df	SS	MS	Expected mean squares	
				Model I	Model II
GCA	$p-1$	S_g	M_g	$\sigma^2+2p(p-2)(\frac{1}{p-1})\Sigma g^2{}_i$	$\sigma^2+2\sigma_s{}^2+2(p-2)\sigma_g{}^2$
SCA	$p(p-3)/2$	Ss	M_s	$\sigma^2+\frac{2}{p(p-3)}\sum_{i<j}\sum s_{ij}{}^2$	$\sigma^2+2\sigma_s{}^2$
Reciprocal eff.	$p(p-1)/2$	S_r	M_r	$\sigma^2+2(\frac{2}{p(p-1)})\sum_{i<j}\Sigma r_{ij}{}^2$	$\sigma^2+2\sigma_r{}^2$
Error	m	S_e	M_e'	σ^2	σ^2

Source: Griffing (1956b)

Table 11. Skeleton of ANOVA for Diallel method IV.

Source	df	SS	MS	Expected mean squares	
				Model I	Model II
GCA	$p-1$	S_g	M_g	$\sigma^2+(p-2)(\frac{1}{p-1})\sum_i g_i{}^2$	$\sigma^2+2\sigma^2{}_s+(p-2)\sigma_g{}^2$
SCA	$p(p-3)/2$	Ss	M_s	$\sigma^2+\frac{2}{p(p-3)}\sum_{i<j}\sum s_{ij}{}^2$	$\sigma^2+\sigma_s{}^2$
Error	m	S_e	M_e'	σ^2	σ^2

Source: Griffing (1956b)

Line x tester - topcross

- Involves hybridization between lines (f) and wide based testers in one-to-one fashion generating f x m = fm hybrids
- It provides both full-sibs and half-sibs simultaneously as opposed to topcross which provides only half-sibs
- Estimates SCA of each cross, but not provides GCA of lines only for testers

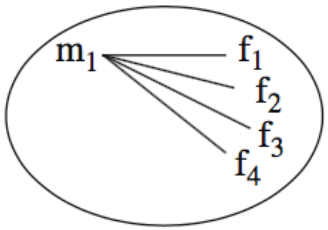


Table 3. Skeleton of ANOVA for Line X Tester Design

Source	df	MS	Expected mean squares	
			Model I	Model II
Replication	$r-1$			
Lines	$m-1$	M_1	$\sigma^2 + rf\frac{1}{m-1} + \sum_i g_i^2$	$\sigma^2 + v_{sca} + rf_{gca(m)}$
Testers	$f-1$	M_2	$\sigma^2 + rm(\frac{1}{f-1})\sum_j g_j^2$	$\sigma^2 + rv_{sca} + rm_{gca(f)}$
Line x tester	$(m-1)(f-1)$	M_3	$\sigma^2 + r\left[\frac{1}{(m-1)(f-1)}\right]\sum_i\sum_j s_{ij}$	$\sigma^2 + rv_{sca}$
Error	$(r-1)(mf-1)$	M_4	σ^2	σ^2

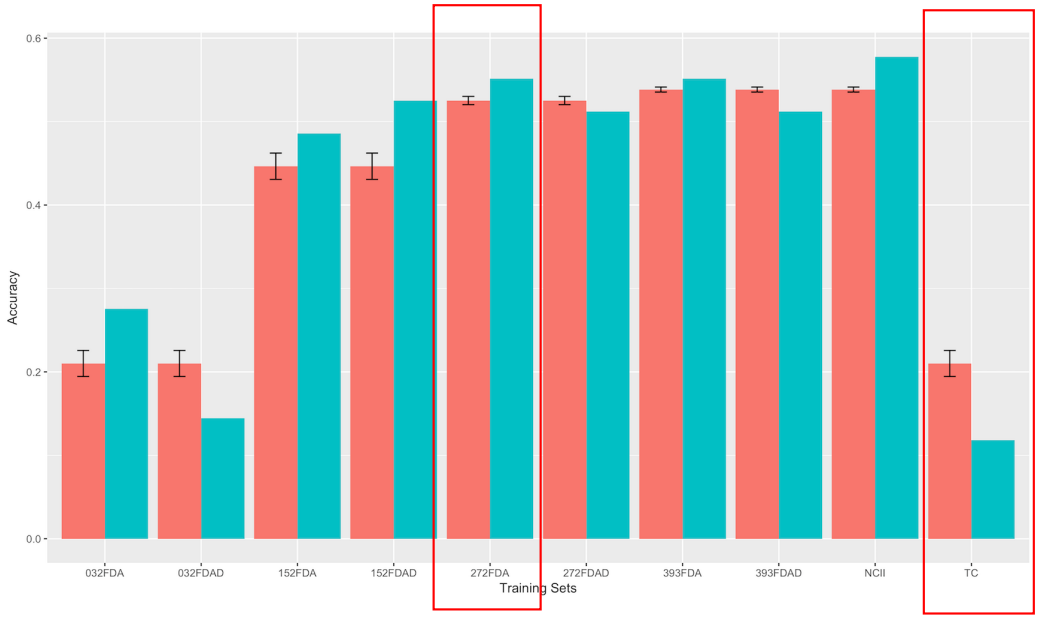
Source: Sharma (2006)

Theoretical and Applied Genetics (2018) 131:1153–1162
https://doi.org/10.1007/s00122-018-3068-8

ORIGINAL ARTICLE

Accuracy of genomic selection to predict maize single-crosses obtained through different mating designs

Roberto Fristche-Neto[1] · Deniz Akdemir[2] · Jean-Luc Jannink[3]



- Testcross is the worst mating design to be used as TS
- At least six crosses per parent to separate GCA and SCA