UNIVERSIDADE DE SÃO PAULO
ESCOLA SUPERIOR DE AGRICULTURA
“LUIZ DE QUEIROZ”
DEPARTAMENTO DE GENÉTICA
LGN5825 Genética e Melhoramento de Espécies Alógamas

Population and quantitative genetics (a brief review)

Prof. Roberto Fritsche-Neto

roberto.neto@usp.br

Piracicaba, March 9th, 2018

Objectives

- **Course objective**
- Gain the ability to design effective and sustainable breeding programs of cross-pollination species and to implement modern selection tools

- **Learning outcomes:**
- 1) Be able to predict response from selection in complex cross-pollination breeding programs
- 2) Understand the dynamics of cross-pollination populations under selection
- 3) Be able to use best linear unbiased prediction (BLUP) for both conventional and genomic selection

- **Be aware that**
- No basic concepts will be covered during the lectures
- I assume that all the students have a basic knowledge in genetics and plant breeding

Requisites

- Methods of plant breeding
- Population genetics
- Quantitative genetics
- Biometry
- Mixed models and components of variance
- Biometry of molecular markers
- R

Schedule

Workflow - LGN5825 - 2018 On Fridays, 8 -12 pm

Week	Date	Lectures	Labs
1	9-Mar	Population and quantitative review	Data quality control
2	16-Mar	Population structure and genetic effects	Population genetics and structure
3	23-Mar	Covariance between relatives	Pedigree
4	6-Apr	Response to selection	Kinship
5	13-Apr	Inbreeding, heterosis, and hybrids between populations	Mixed Model Equations
6	20-Apr	Hybrids between lines	REML/BLUP (I and A)
7	27-Apr	Test I	
8	4-May	Lines, testers and testcrosses	Diallells
9	11-May	Base populations and breeding schemes	Optimized Training Sets
10	18-May	GWAS	GWAS
11	25-May	Genomic Selection	GS (GBLUP)
12	8-Jun	Recurrent Selection	GS (Bayes + GE)
13	15-Jun	Reciprocal Recurrent Selection	GS (MOOB)
14	22-Jun	Test II	
15	29-Jun	Test on R	

classes Moodle STOA

References

- BERNARDO, R. Breeding for quantitative traits in plants. 2ed. Stemma Press, 2010. 369 p.
- BOS, I; CALIGARI, P. Selection methods in plant breeding. 2ed. Springer, 2008, 471 p.
- HALLAUER, AR; CARENA, MJ; MIRANDA FILHO, LB. Quantitative genetics in maize breeding. 2 ed. Springer, 680 p.
- MRODE, RA. Linear models for the prediction of animal breeding values. 3ed. CABI, 360 p.
- SOUZA JR., C.L. Componentes da variância genética e suas implicações no melhoramento vegetal. FEALQ, Piracicaba. 1989. 134 p.
- SOUZA JR., C.L. Melhoramento de espécies alógamas. In: Nass, LL; Valois, A.C.C.; Melo, I.S.M.; Valadares-Inglis, M.C. (eds.). Recursos genéticos & Melhoramento-Plantas. Fundação MT. 2001. Capítulo 8: 159-199.
- VENCOVSKY, R; BARRIGA, P. Genética biométrica no fitomelhoramento. SBG, 1992, 486 p.
- WALSH, B. Evolutionary quantitative genetics. University of Arizona, 371.

Allogamous (cross-pollination)

- Species
- Cross-pollination ≥ 95%
- Mechanisms

monoecy, dioecy, protogyny, protandry, self-incompatibility, morphological

- Evolution – *some advantages of being heterozygous*
- Utilize the heterosis and avoid the inbreeding depresssion

- Populations
- Group of individuals that constitute a set of genes and are maintained using cross-fertilization at the same place and time
- Parents do not transfer the entire genotype to offspring, which is randomly formed each generation
- Although the phenotype is evaluated, the alleles are selected

Variation in breeding populations

- A phenotypic observation on a single individual is determined by the environment, genetic effects, and residual effects
- P = G + E

$$y_i = u + g_i + e_i$$

- The total genetic value **g** is the genetic value of an individual per se, and this is of key importance when selecting the best individuals to release as varieties
- An individual's genetic value can be further broken down into
- *Additive* (g_a),
- *Dominance* (g_d), *and*
- *Epistatic* (g_i)

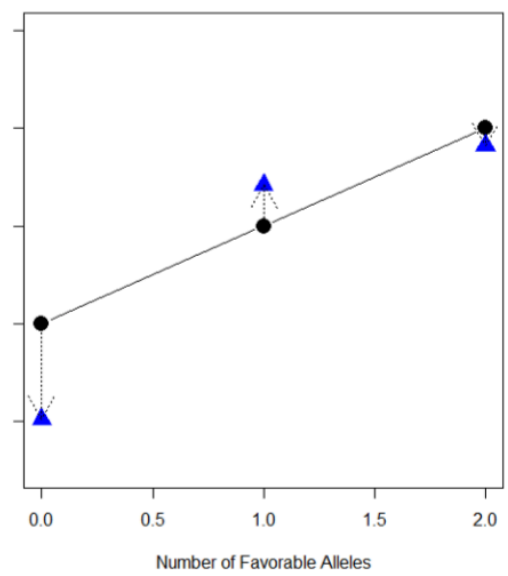
- Breeding value (BV)
- Only additive allelic effects can be transmitted from parent to offspring
- BV is the sum total of the additive allele effects
- It is also the value of an individual as a parent

Additive effect of an allele

- The additive effect at a locus is the linear effect of allele dosage on the phenotypic value
- Note: Loci that are dominant or that interact epistatically with other loci still have an additive effect
- In most cases, dominance and epistasis are assumed insignificant, and are included in the **e** error term
- The general model then becomes

$$y_i = u + a_i + e_i$$

- where $\boldsymbol{a_i}$ is the additive genetic value of individual $\boldsymbol{i}$



Transmitting ability

- The average effect of a random sample of half of an individual's alleles
- Equals one half of an individual's total breeding value (***a/2***)

- Breeding values of parents and progeny
- Each parent contributes half of its alleles to the progeny
- Average breeding value of progeny is the average breeding value of the two parents

$$E(a) = \frac{1}{2}a_{p1} + \frac{1}{2}a_{p2}$$

- where $\boldsymbol{p_1}$ and $\boldsymbol{p_2}$ are parents one and two respectively

- Progeny breeding values vary due to random sampling
- **BV** of individual ***i*** deviates from the parental mean due to random sampling of alleles
- This random term is referred to as the '**Mendelian sampling**"

$$E(a_i) = \frac{1}{2}a_{p1} + \frac{1}{2}a_{p2} + m_i$$

Heritability

- The degree of correspondence between the phenotypic values and the breeding values
- Indicates how well the trait will respond to selection
- Ratio of additive genetic variance to phenotypic variance

$$h_a^2 = \frac{\sigma_a^2}{\sigma_y^2}$$

- Is also the regression of the breeding value on the phenotypic value

$$h_a^2 = b_{ay} = \frac{\sigma_{ay}}{\sigma_y^2} = r_{ay}\frac{\sigma_y}{\sigma_a}$$

- This is because $\boldsymbol{y} = \boldsymbol{a} + \boldsymbol{e}$ where $\boldsymbol{a}$ is the additive genetic component of the phenotype ($\boldsymbol{y}$), and $\boldsymbol{e}$ is the non-additive genetic component
- Then
- **cov(a,y) = cov(a,a+e) = cov(a,a) + cov(a,e)**
- Because $\boldsymbol{a}$ and $\boldsymbol{e}$ are uncorrelated **cov(a, y) =** σ_a^2

Correlation and regression coefficient

- The normalized version of the covariance, the correlation coefficient, ranges from -1 to 1, and its magnitude indicates the strength of a linear relationship between two variables

$$r_{xy} = \frac{COV(x, y)}{\sigma_x \sigma_y}$$

- where $\boldsymbol{x}$ and $\boldsymbol{y}$ are the standard deviations of $\boldsymbol{x}$ and $\boldsymbol{y}$

- From standard regression theory, the regression coefficient for the regression of $\boldsymbol{y}$ on $\boldsymbol{x}$ is

$$b_{xy} = \frac{\sigma_{xy}}{\sigma_x^2} = r_{xy} \frac{\sigma_y}{\sigma_x}$$

- Covariance, correlation, and regression coefficients are important for understanding and estimating accuracy of selection

Mathematical expectation

- **It is also known as the expected value – (the mean)**

- How can we estimate the mean and variance?

- Expectation of a constant => $E(c) = c$
- Expectation of a random variable multiplied by a constant => $E(cX) = cE(X)$

- Expectation of two random random variables
- $E(X + Y) = E(X) + E(Y)$
- $E(X.Y) = E(X) . E(Y)$ => *If they are independent*

- Variance and covariance
- $V(X) = E[X - E(X)]^2$
- $COV(X, Y) = E[X - E(X)] . E[Y - E(Y)]$
- $V(X + Y) = V(X) + V(Y) + 2COV(X, Y)$
- $V(X - Y) = V(X) + V(Y) - 2COV(X, Y)$

Mathematical expectation applied to P=G+E

$$y_{ij} = u + g_i + e_{ij}$$

- $E(Y_{ij}) = u$ => the mean of experiment, considered as fixed
- $E(g_i) = 0$ => deviations from the mean
- $E(e_{ij}) = 0$ => deviations from the mean
- $E(g_i + e_{ij}) = E(g_i) + E(e_{ij}) = 0$

- Phenotipic variance
- $V(Y) = E[Y_{ij} - E(Y_{ij})]^2 = E[u + g_i + e_{ij} - u]^2 = E[g_i + e_{ij}]^2 = E(g_i)^2 + E(e_{ij})^2 + 2COV(g_i, e_{ij})$
- $E(g_i)^2 = E[g_i - E(g_i)]^2 = Vg$
- $E(e_{ij})^2 = E[e_{ij} - E(e_{ij})]^2 = Ve$
- $V(Y) = Vp = Vg + Ve$

- Heritability
- $COV(Y_{ij}, g_i)$
- $= E[Y_{ij} - E(Y_{ij})] . E[g_i - E(g_i)]$
- $= E[u + g_i + e_{ij} - u] . E[g_i - 0]$
- $= E(g_i)^2 + E(e_{ij}).E(g_i) = E(g_i)^2 = Vg$

$$r_{Y_{ij},gi} = \frac{COV(Y_{ij}, gi)}{\sigma_{Y_{ij}}\sigma_{gi}} \qquad r_{Y_{ij},gi} = \frac{\sigma_{gi}}{\sigma_{Y_{ij}}}$$

$$r_{Y_{ij},gi} = \frac{\sigma_{gi}^2}{\sigma_{Y_{ij}}\sigma_{gi}} \qquad r_{Y_{ij},gi} = \sqrt{h_g^2}$$

Expectation between two observations

- The same genotype evaluated in different replicates
- $COV(Y_{ij}, Y_{ij'}) = E[Y_{ij} - E(Y_{ij})] . E[Y_{ij'} - E(Y_{ij'})]$
- $= E[u + g_i + r_j + e_{ij} - u] . E[u + g_i + r_{j'} + e_{ij'} - u] = E[(g_i + r_j + e_{ij}) . (g_i + r_{j'} + e_{ij'})]$
- $= E(g_i)^2 + dp$
- $= Vg$
- The variance among genotypes is equal to the covariance within
- Independent of the experimental design
- The covariance between related individuals means genetic covariance

- The same genotype evaluated in different replicates at the same local
- $COV(Y_{ijk}, Y_{ijk'}) = E[Y_{ijk} - E(Y_{ijk})] . E[Y_{ijk'} - E(Y_{ijk'})]$
- $= E[u + g_i + l_j + r_{k/l} + gl_{ij} + e_{ijk} - u - l_j] . E[u + g_i + l_j + r_{k'/l} + gl_{ij} + e_{ijk'} - u - l_j]$
- $E[g_i + r_{k/l} + gl_{ij} + e_{ijk}] . E[g_i + r_{k'/l} + gl_{ij} + e_{ijk'}]$
- $= E(g_i)^2 + E(gl_{ij})^2 + dp$
- $= Vg + Vge$
- Overestimated the heritability – there is a confusion between these two components
- Solution – evaluate in more than one place
- The number of places depends on the expected heritability and ratio of components

$$E[h_g^2 = \frac{\sigma_g^2}{\sigma_g^2+\sigma_{gl}^2/l+\sigma_e^2/rl}]$$

$$L=\left(\frac{\sigma_{gl}^2+\sigma_e^2/r}{\sigma_g^2}\right)\left(\frac{h_g^2}{1-h_g^2}\right)$$

Hardy-Weinberg law

- States that the gene and genotype frequencies are constant across generations if:
- *population infinitely large*
- *mating is random*
- *no selection, mutation or migration*

- If allele frequencies in the parents are ***p*** and ***q***, for allele ***1*** and ***2*** respectively, then the genotype frequencies in the progeny should be:
- p^2 for **homozygous** allele ***1***
- $2pq$ for **heterozygous**
- q^2 for **homozygous** allele ***2***

- Processes that change the allele frequencies in a predicable manner
- Migration, mutation, and selection

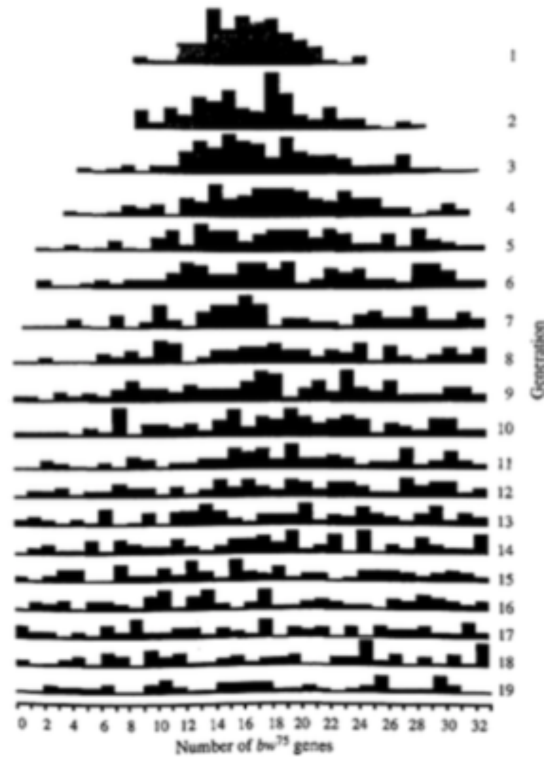
- A process that changes allele frequencies in an unpredictable manner
- Random sampling of gametes in small populations - *drift*

Random drift

- Drift is predictable in amount but not in direction
- Allele frequencies may be seen to change erratically from one generation to another

- Leads to
- Genetic differentiation between the populations
- Reduced genetic variation within each population
- Increase in homozygote genotypes at the expense of heterozygotes genotypes

- Example of changes in allele frequency in an idealized small population



Magnitude of genetic drift

- The change in allele frequency is random in that its direction unpredictable
- However, its variance can be predicted
- Thus we can only know the magnitude of change in allele frequency, but not the direction

- Across all lines in a population the allele frequencies will be distributed around $\boldsymbol{q_0}$ with a variance
- $p_0 q_0/2Ne$
- where $\boldsymbol{Ne}$ is the population size

- $p_0 q_0/2Ne$ is also the variance of $\boldsymbol{q_1}$, the allele frequency in different lines after one generation under drift
- It expresses the magnitude

- In the next generation the sampling process is repeated

- The effect of this continued sampling of successive generations is that the allele frequencies in lines fluctuates irregularly, and lines become more and more differentiated

Drift over generations

- Allele frequencies
- Eventually alleles in a small population will reach a frequency of one or zero.
- Alleles that reach a frequency of one are said to be fixed and those that reach zero are said to be lost
- Genotype Frequency
- As lines drift apart in allele frequency they also drift apart in genotype frequencies
- There is an increase in homozygous and a decrease in heterozygous genotypes
- Within a single line, the relationship between allele and genotype frequencies follows Hardy-Weinberg
- The genotype frequencies across all lines, when considered together as one population are no longer in HWE

Inbreeding

- Inbreeding is the mating together of individuals that are related to each other by ancestry
- It depends on the population size - *number of possible ancestors*
- In a population of bisexual organisms, each individuals has two parents, four grandparents, etc.
- Thus, t generations back an individual has 2^t ancestors

- Identity by descent (IBD)
- Two mating individuals that share a ancestor may carry replicates of alleles from the common ancestor
- These replicates can then be passed on to the offspring from both parents
- Leading to homozygous in the progeny, with both alleles being identical by descent (IBD)

- The coefficient of inbreeding (F)
- The probability that two alleles at any locus in an individual are IBD
- Degree of relationship between an individual's parents
- At random mating F is the probability that two gametes taken at random from the population are IBD
- Each individual will have its own F, but the average F is of main interest as a measure of random drift

Rate of inbreeding

- F can be estimated based on the population size
- In the first generation of mating from the base population, there are N individuals and $2N$ different gametes
- Then, the probability that any given gamete unites with an identical gamete is $1/2N$
- In the second generation there are two classes of gametes that can be sampled
- The first is a gamete identical to the gamete of interest and its probability is $1/2N$
- The second is a gamete that is not identical based on the current replication with probability $(1 - 1/2N)F_1$
- Thus, the new inbreeding is $\mathbf{F_2 = 1/2N + (1 - 1/2N)F_1}$

- The coefficient of inbreeding in generation t is $\mathbf{F_t = 1/2N + (1 - 1/2N)F_t}$

- The F is made up of two parts, one attributable to new inbreeding and another to previous inbreeding
- The new inbreeding is $\Delta F = 1/2N$
- Then, we can rewrite as $F_t = \Delta F + (1- \Delta F)F_{t-1}$ and rearrange it as $\Delta F = (F_t - F_{t-1})/(1-F_{t-1})$

- ΔF is the rate of inbreeding
- ΔF provides a means of comparing the inbreeding effects of different breeding systems

Effective population size (Ne)

- Drift can also be evaluated in terms of the variance of gene frequencies or the rate of inbreeding
- As **Δ**F can be estimated by looking at the IBD, then ***Ne*** can be estimated by
- $\mathrm{Ne} = 1/2\mathbf{\Delta}\mathrm{F}$
- When the breeding structure is known, ***Ne*** can be derived (approximately) from the actual number ***N***

- However, with unequal numbers of females and males
- $\mathrm{Ne} = 4\mathrm{Nm.Nf}/(\mathrm{Nm} + \mathrm{Nf})$
- Thus, for half-sibs we have
- $\mathrm{Ne} = 4\infty.1/(\infty + 1) \approx 4\infty/(\infty) = 4$
- $\mathbf{\Delta}\mathrm{F} = 1/2\mathrm{Ne} = 1/(2.4) = 1/8$
- And for full-sibs we have
- $\mathrm{Ne} = 4.1.1/(1 + 1) = 4/2 = 2$
- $\mathbf{\Delta}\mathrm{F} = 1/2\mathrm{Ne} = 1/(2.2) = 1/4$

- When there are unequal numbers in successive generations ***Ne*** is the harmonic mean of the ***N*** in each generation