

Population and quantitative genetics (a brief review)

Prof. Roberto Fritsche-Neto

roberto.neto@usp.br

Piracicaba, August 14th, 2019

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 - 2019

On Wednesday, 8 -12 pm

Week	Date	Lectures	Labs
1	14-Aug	Population and quantitative review	Data quality control
2	21-Aug	Population structure and genetic effects	Population genetics and structure
3	28-Aug	Covariance between relatives	Pedigree
4	11-Sep	Response to selection	Kinship
5	28-Sep	Inbreeding, heterosis, and hybrids between populations	Mixed Model Equations
6	09-Oct	Hybrids between lines	REML/BLUP (I, A and K)
7	09-Oct	Test I	
8	16-Oct	Lines, testers and testcrosses	Diallell
9	23-Oct	Base populations and breeding schemes	Phenotype correction and Optimized Training Sets
10	30-Oct	GWAS	GWAS
11	06-Nov	Genomic Selection	GS and GS multi-trait (GBLUP)
12	13-Nov	Recurrent Selection	GS (Bayes alphabet)
13	27-Nov	Reciprocal Recurrent Selection	GS (Bayes GE models)
14	27-Nov	Test II	

Classes and tests

[Moodle STOA](#)

Homeworks

[Every week, based on R labs](#)

R labs (R Markdown file)

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 $\geq 95\%$

- Mechanisms

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

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

- **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

- $P = G + E$

$$y_i = u + g_i + e_i$$

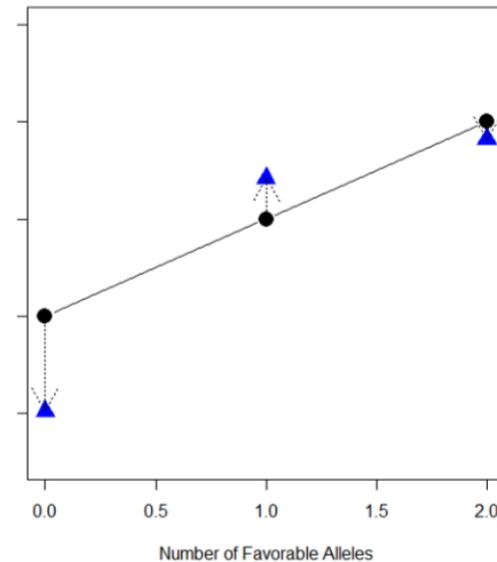
- The total **genetic value** g is the genetic value of an individual per se, and this is of key importance when selecting the best individuals to release as varieties
- An individual's genetic value can be further broken down into
 - *Additive* (g_a),
 - *Dominance* (g_d), and
 - *Epistatic* (g_i)
- **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 a_i is the additive genetic value of individual 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 p_1 and 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 $y = a + e$ where a is the additive genetic component of the phenotype (y), and e is the non-additive genetic component
- **Then**
- $\text{cov}(a, y) = \text{cov}(a, a+e) = \text{cov}(a, a) + \text{cov}(a, e)$
- Because a and e are uncorrelated $\text{cov}(a, y) = \sigma_a^2$

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 σ_x and σ_y are the standard deviations of x and y
- From standard regression theory, the regression coefficient for the regression of y on 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 $\Rightarrow E(c) = c$
- Expectation of a random variable multiplied by a constant $\Rightarrow E(cX) = cE(X)$
- **Expectation of two random random variables**
- $E(X + Y) = E(X) + E(Y)$
- $E(X \cdot Y) = E(X) \cdot E(Y) \Rightarrow$ *If they are independent*
- **Variance and covariance**
- $V(X) = E[X - E(X)]^2$
- $COV(X, Y) = E[X - E(X)] \cdot E[Y - E(Y)]$
- $V(X + Y) = V(X) + V(Y) + 2COV(X, Y)$
- $V(X - Y) = V(X) + V(Y) - 2COV(X, Y)$

Mathematical expectation applied to $P=G+E$

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

- $E(Y_{ij}) = u \Rightarrow$ the mean of experiment, considered as fixed
- $E(g_i) = 0 \Rightarrow$ deviations from the mean
- $E(e_{ij}) = 0 \Rightarrow$ deviations from the mean
- $E(g_i + e_{ij}) = E(g_i) + E(e_{ij}) = 0$
- **Phenotypic 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**
- $COV(Y_{ij}, g_i)$
- $= E[Y_{ij} - E(Y_{ij})] \cdot E[g_i - E(g_i)]$
- $= E[u + g_i + e_{ij} - u] \cdot E[g_i - 0]$
- $= E(g_i)^2 + E(e_{ij}) \cdot E(g_i) = E(g_i)^2 = Vg$

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

$$r_{Y_{ij}, g_i} = \frac{\sigma_{g_i}}{\sigma_{Y_{ij}}}$$

$$r_{Y_{ij}, g_i} = \frac{\sigma_{g_i}^2}{\sigma_{Y_{ij}} \sigma_{g_i}}$$

$$r_{Y_{ij}, g_i} = \sqrt{h_g^2}$$

Expectation between two observations

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

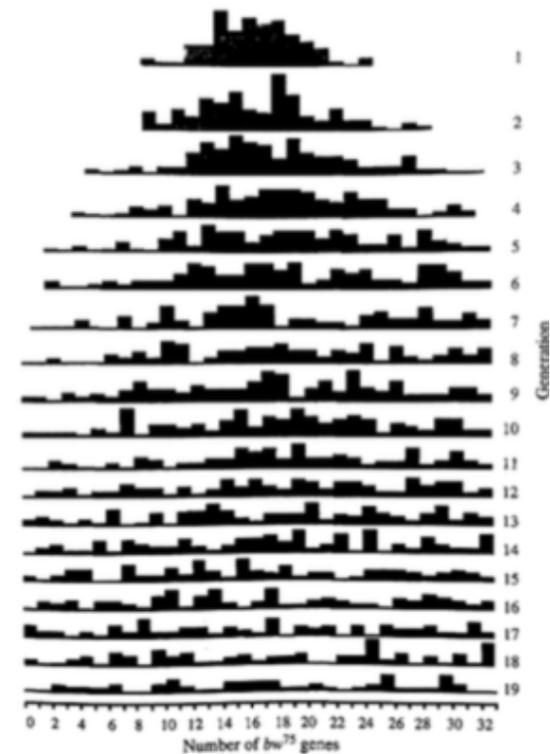
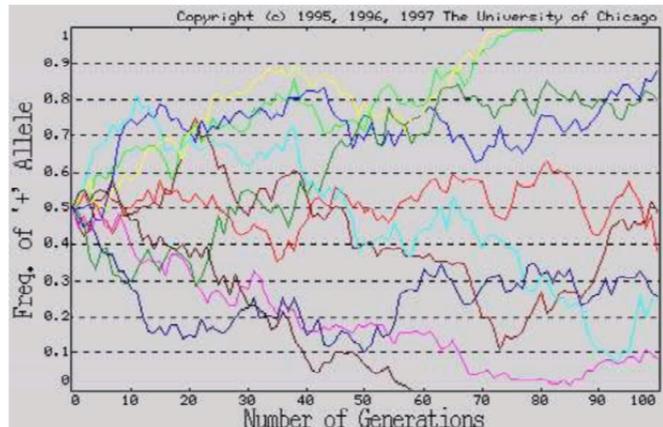
- The same genotype evaluated in different replicates at the same local
- $\text{COV}(Y_{ijk}, Y_{ijk'}) = E[Y_{ijk} - E(Y_{ijk})] \cdot E[Y_{ijk'} - E(Y_{ijk'})]$
- $= E[u + g_i + l_j + r_{k/l} + gl_{ij} + e_{ijk} - u - l_j] \cdot E[u + g_i + l_j + r_{k'/l} + gl_{ij} + e_{ijk'} - u - l_j]$
- $E[g_i + r_{k/l} + gl_{ij} + e_{ijk}] \cdot E[g_i + r_{k'/l} + gl_{ij} + e_{ijk'}]$
- $= E(g_i)^2 + E(gl_{ij})^2 + E(g_i) \cdot E(r_{k'/l}) + \dots$
- $= V_g + V_{ge}$
- 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

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*

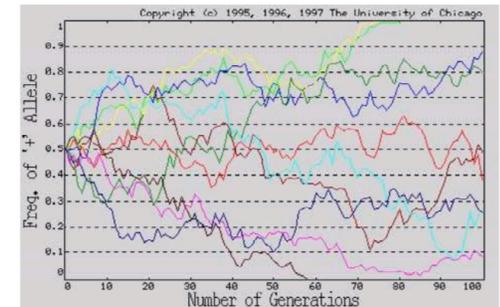
Random drift

- 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
 - Fix or loss alleles
 - Increase in homozygote genotypes at the expense of heterozygotes genotypes
 - Creates a uniform distribution



Magnitude of genetic drift

- The change in allele frequency is random in that its direction unpredictable
- However, its variance can be predicted (**magnitude but not the direction**)
- After one generation, the magnitude of drift between all lines is due to $p_0 q_0 / 2N_e$
- *where N_e is the population size*
- *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
- 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*
- **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 (ΔF)

- ΔF provides a means of comparing the inbreeding effects of different breeding systems
- 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 $F_2 = 1/2N + (1 - 1/2N)F_1$
- The coefficient of inbreeding in generation t is $F_t = 1/2N + (1 - 1/2N)F_{t-1}$
- 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 $\Delta F = (F_t - F_{t-1}) / (1 - F_{t-1})$

Effective population size (N_e)

- As ΔF can be estimated by looking at the IBD, then N_e can be estimated by
- $N_e = 1/2\Delta F$
- When the breeding structure is known, N_e can be derived (**approximately**) from the actual number N

- **However, with unequal numbers of females and males**
- $N_e = 4N_m N_f / (N_m + N_f)$

- **Thus, for half-sibs we have**
- $N_e = 4\infty \cdot 1 / (\infty + 1) \approx 4\infty / (\infty) = 4$
- $\Delta F = 1/2N_e = 1/(2 \cdot 4) = 1/8$

- **And for full-sibs we have**
- $N_e = 4 \cdot 1 \cdot 1 / (1 + 1) = 4/2 = 2$
- $\Delta F = 1/2N_e = 1/(2 \cdot 2) = 1/4$

- Under unequal numbers in successive generations N_e is the harmonic mean of the N in each generation