Population and quantitative genetics introduction

Prof. Roberto Fritsche-Neto

rfneto@agcenter.lsu.edu

Baton Rouge, Jan 23rd, 2023

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

Number of Favorable Alleles
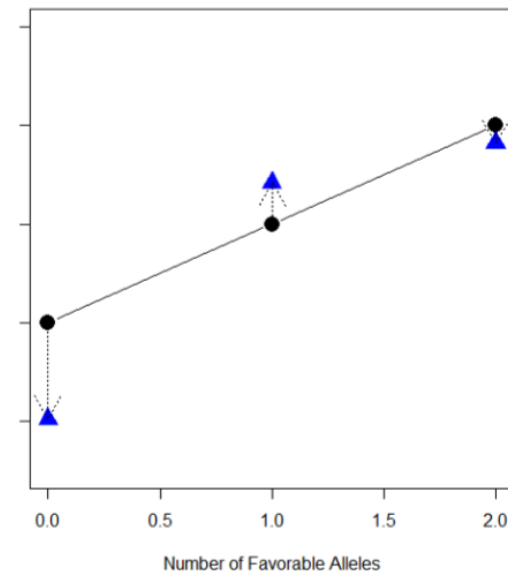
Transmitting ability

- 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
- cov(a, y) = cov(a, a+e) = cov(a, a) + cov(a, e)
- Because a and 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 x and 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 => $E(c) = c$
- 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
- $COV(Y_{ij}, g_i)$
- $= E[Y_{ij} - E(Y_{ij})] . E[g_i - E(g_i)]$
- $= E[u + g_i + e_{ij} - u] . E[g_i - 0]$
- $= E(g_i)^2 + E(e_{ij}).E(g_i) = E(g_i)^2 = Vg$

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

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

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

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

Expectation between two observations

- The same genotype evaluated in different replicates
- $COV(Y_{ij}, Y_{ij'}) = E[Y_{ij} - E(Y_{ij})] \cdot E[Y_{ij'} - E(Y_{ij'})]$
- $= E[u + g_i + r_j + e_{ij} - u] \cdot E[u + g_i + r_{j'} + e_{ij'} - u] = E[(g_i + r_j + e_{ij}) \cdot (g_i + r_{j'} + e_{ij'})]$
- $= E(g_i)^2 + E(g_i).E(r_{ij'}) + ...$
- $= Vg$
- 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
- $COV(Y_{ijk}, Y_{ijk'}) = E[Y_{ijk} - E(Y_{ijk})] \cdot E[Y_{ijk'} - E(Y_{ijk'})]$
- $= E[u + g_i + l_j + r_{k/l} + gl_{ij} + e_{ijk} - u - l_j] \cdot E[u + g_i + l_j + r_{k'/l} + gl_{ij} + e_{ijk'} - u - l_j]$
- $E[g_i + r_{k/l} + gl_{ij} + e_{ijk}] \cdot E[g_i + r_{k'/l} + gl_{ij} + e_{ijk'}]$
- $= E(g_i)^2 + E(gl_{ij})^2 + E(g_i).E(r_{k'/l}) + ...$
- $= Vg + Vge$
- Overestimated the heritability – there is a confusion between these two components
- Solution – evaluate in more than one location
- 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
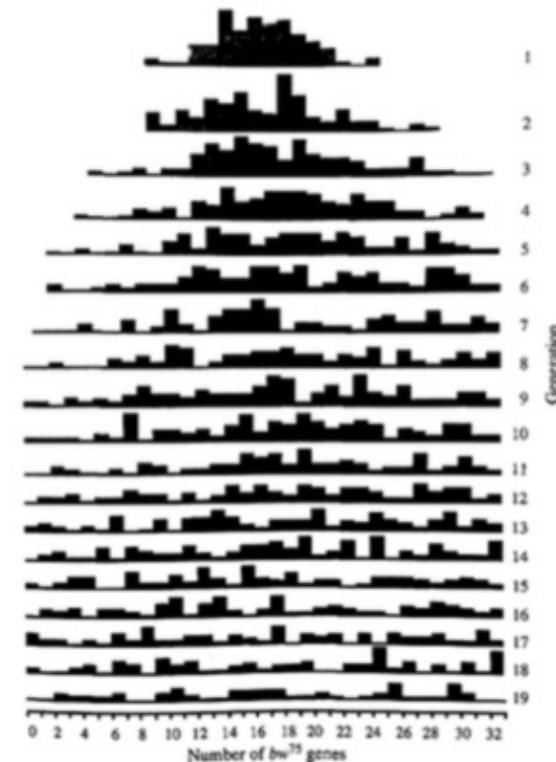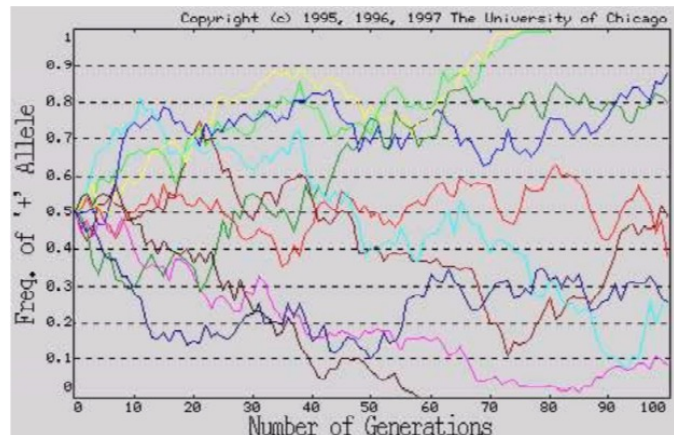
Magnitude of genetic drift

- 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 $p_0 q_0/2Ne$
- *where Ne 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
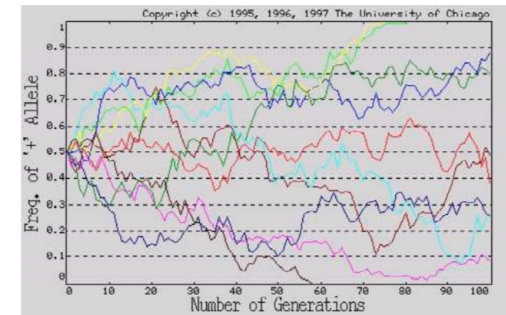
Inbreeding

- 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)

- **Δ**F 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 (Ne)

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

- However, with unequal numbers of females and males
- $Ne = 4Nm.Nf/(Nm + Nf)$

- Thus, for half-sibs we have
- $Ne = 4\infty.1/(\infty + 1) \approx 4\infty/(\infty) = 4$
- $\Delta F = 1/2Ne = 1/(2.4) = 1/8$

- And for full-sibs we have
- $Ne = 4.1.1/(1 + 1) = 4/2 = 2$
- $\Delta F = 1/2Ne = 1/(2.2) = 1/4$

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