Introduction to Quantitative Genetics

Dr Loïc Yengo

Institute for Molecular Bioscience
The University of Queensland

August 13th, 2020

Contents

The Laws of inheritance

Pioner works of Mendel (mid-1800's) established laws of inheritance for discrete characters: size (tall vs. short), color, etc.

He introduced the concepts of dominance/recessiveness which clarified the notions of genotypes and phenotypes.

- Example: Two individuals may have the same phenotype without having the same genotype (Mendel's peas).

Example of known prediction from Mendel's Laws
Under sexual reproduction, if parents have the genotypes AB and ab, then the children have the genotypes Aa, Ab, Ba and Bb with probability 1/4.

The Laws of inheritance

Pioner works of Mendel (mid-1800's) established laws of inheritance for discrete characters: size (tall vs. short), color, etc.

He introduced the concepts of dominance/recessiveness which clarified the notions of genotypes and phenotypes.

- Example: Two individuals may have the same phenotype without having the same genotype (Mendel's peas).

Example of known prediction from Mendel's Laws

Under sexual reproduction, if parents have the genotypes AB and ab, then the children have the genotypes Aa, Ab, Ba and Bb with probability 1/4.

From Mendel to Fisher

Extension of Mendel's Laws

Quantitative Genetics theory tries to deduce the consequences of the laws of inheritance for quantitative traits[1] in a population.

Founding contributions to the field

- Fisher (1918): infinitesimal model.
- Wright (1921)
- Haldane (1932)

[1] Other extensions exist for discrete traits.

From Mendel to Fisher

Extension of Mendel's Laws

Quantitative Genetics theory tries to deduce the consequences of the laws of inheritance for quantitative traits[1] in a population.

Founding contributions to the field

- Fisher (1918): infinitesimal model.
- Wright (1921)
- Haldane (1932)

[1] Other extensions exist for discrete traits.

Infinitesimal Model[2]

- Large number of gene variants
- All having small effects
- Acting additively and independently
- According to Mendel's Laws

[2] Barton et al. *Theor. Pop. Biol.* (2017)

Infinitesimal Model[2]

- Large number of gene variants
- All having small effects
- Acting additively and independently
- According to Mendel's Laws

[2] Barton et al. *Theor. Pop. Biol.* (2017)

Infinitesimal Model[2]

- Large number of gene variants
- All having small effects
- Acting additively and independently
- According to Mendel's Laws

[2] Barton et al. *Theor. Pop. Biol.* (2017)

Infinitesimal Model[2]

- Large number of gene variants
- All having small effects
- Acting additively and independently
- According to Mendel's Laws

[2] Barton et al. *Theor. Pop. Biol.* (2017)

Some questions addressed by Q.G. theory

- Can we predict from genetic data how well an individual would perform?
- How much can we predict of the risk of diseases from genetic data?
- How to build breeding programs to enhance a particular character in the population?

Basic model of Quantitative Genetics

Definition

We classically assume that a quantitative phenotype P results from the contribution of genetic factors G and environmental factors E:

$$P = G + E \tag{1}$$

In Equation (1) G, also referred to as **Genotypic value**, is the average phenotype over all possible environments.

Example: 2 environments

- If the phenotypic value is 10 in Environment 1 and 20 in Environment 2
- $\Longrightarrow$ then the genotypic value is $G = 15$.

Basic model of Quantitative Genetics

Definition

We classically assume that a quantitative phenotype P results from the contribution of genetic factors G and environmental factors E:

$$P = G + E \tag{1}$$

In Equation (1) G, also referred to as **Genotypic value**, is the average phenotype over all possible environments.

Example: 2 environments

- If the phenotypic value is 10 in Environment 1 and 20 in Environment 2
- $\Longrightarrow$ then the genotypic value is $G = 15$.

Non transmission of the genotypic value

More definitions

Under sexual reproduction, parents only pass along SINGLE ALLELES to their offspring. The genotypic value of each parent is therefore partially transmitted to the next generation.

We define the **average effect of an allele A (α_A)** as the (mean) phenotypic difference between offspring inheriting allele A compared to all offspring:

$$\alpha_A = \begin{array}{c}\text{Mean phenotype} \\ \text{in offspring with A allele}\end{array} - \begin{array}{c}\text{Mean phenotype} \\ \text{in all offspring}\end{array} \quad (2)$$

Non transmission of the genotypic value

More definitions

Under sexual reproduction, parents only pass along SINGLE ALLELES to their offspring. The genotypic value of each parent is therefore partially transmitted to the next generation.

We define the **average effect of an allele A (α_A)** as the (mean) phenotypic difference between offspring inheriting allele A compared to all offspring:

$$\alpha_A = \begin{array}{c}\text{Mean phenotype} \\ \text{in offspring with A allele}\end{array} - \begin{array}{c}\text{Mean phenotype} \\ \text{in all offspring}\end{array} \quad (2)$$

Fisher's (1918) decomposition of the genotypic value

Decomposition of G

One of Fisher's key insights was that G consists of a fraction that can be passed from parent to offspring and a fraction that can cannot.

Let $A_{im}^{(\ell)}$ and $A_{if}^{(\ell)}$ be the alleles that individual i inherited from their mother and father respestively at the locus ℓ.The genotypic value $G[A_{im}^{(\ell)} A_{if}^{(\ell)}]$ of individual i at the locus ℓ can be decomposed (Fisher) as:

$$G[A_{im}^{(\ell)} A_{if}^{(\ell)}] = \mu_G^{(\ell)} + \underbrace{\left(\alpha[A_{im}^{(\ell)}] + \alpha[A_{if}^{(\ell)}]\right)}_{\text{passed along to offspring}} + \underbrace{\delta[A_{im}^{(\ell)} A_{if}^{(\ell)}]}_{\text{not passed along}} \quad (3)$$

where $\mu_G^{(\ell)}$ is the mean value of that genotype in the population. $\alpha[A_{im}^{(\ell)}]$ and $\alpha[A_{if}^{(\ell)}]$ contribute negative $\mu_G^{(\ell)}$.

Fisher's (1918) decomposition of the genotypic value

Additive Genetic Value (A)

In Equation (3), $\mu_G^{(\ell)} + \left(\alpha[A_{im}^{(\ell)}] + \alpha[A_{if}^{(\ell)}]\right)$ is the expected value of an individual with the $A_{im}^{(\ell)}A_{if}^{(\ell)}$ genotype and $\delta[A_{im}^{(\ell)}A_{if}^{(\ell)}]$ is the deviation from that expectation also called dominance deviation.

We define the **Additive Genetic Value (A)** or **Breeding Value** of an individual i as the sum over all m loci influencing the trait of the $\left(\alpha[A_{im}^{(\ell)}] + \alpha[A_{if}^{(\ell)}]\right)$:

$$A_i = \sum_{\ell=1}^{m} \left(\alpha[A_{im}^{(\ell)}] + \alpha[A_{if}^{(\ell)}]\right) \tag{4}$$

A_i is also twice the mean value of all (potential/possible) offspring of individual i with different co-parent.

Genetic variance

Writing the genotypic value as

$$G[A_{im}^{(\ell)} A_{if}^{(\ell)}] = \mu_G^{(\ell)} + \left(\alpha[A_{im}^{(\ell)}] + \alpha[A_{if}^{(\ell)}]\right) + \delta[A_{im}^{(\ell)} A_{if}^{(\ell)}]$$

We can calculate the variance of genetic values or simply the **genetic variance** as

$$\begin{aligned} \sigma_G^2 &= var\left(\sum_{\ell=1}^{m} G[A_m^{(\ell)} A_f^{(\ell)}]\right) \\ &= \underbrace{\textstyle\sum_{\ell=1}^{m} var\left(\alpha[A_m^{(\ell)}] + \alpha[A_f^{(\ell)}]\right)}_{\text{Additive Variance}} + \underbrace{\textstyle\sum_{\ell=1}^{m} var\left(\delta[A_m^{(\ell)} A_f^{(\ell)}]\right)}_{\text{Dominance Variance}} \\ &= \sigma_A^2 + \sigma_D^2 \end{aligned} \tag{5}$$

Note that $\text{cov}[\left(\alpha[A_m^{(\ell)}] + \alpha[A_f^{(\ell)}\right)], \delta[A_m^{(\ell)} A_f^{(\ell)}]] = 0$ by design.

Key concepts so far

- $\alpha(A)$: average effect. Property of a single allele in a particular population (depends on allele frequency).
- A: additive genetic value.
  - A is the sum over all loci of average effects.
  - Fraction of G that parents pass along to their offspring
  - Property of an individual in a particular population
- $var(A)$: additive genetic variance.
  - Variance of genetic additive values
  - Property of a population

A and $var(A)$ can be estimated without molecular/genetic data!

(Part I)
Red pill or blue pill?

The concept of heritability

Heritability

Definition

Under the basic model

$$P = G + E$$

we define the broad sense heritability H^2 as the ratio $var(G)/var(P)$.

Heritability is central concept of QG.

Under Fisher decomposition $G = A + D$, we define the narrow sense heritability h^2 as the ratio $var(A)/var(P)$, i.e. fraction of phenotypic variance due to additive genetic values (breeding values).

Some observations...

- The amount of phenotypic resemblance among relatives for a trait is an indication of the genetic variation for the trait.
- If trait variation has a significant genetic basis, the closer the relatives, the more similar their appearance.
- The covariance between phenotyic values of relatives measures the strength of this similarity, with larger covariance meaning more similarity.

Covariance between relatives (1)

A quick exercise

We consider a locus with two alleles B and b.
We assume that the genotypic values are

$$G(BB) = 0, G(Bb) = a \text{ and } G(bb) = 2a.$$

We denote G_1 and G_2 as the genotypic value of Parent 1 and 2 respectively. The genotypic value of the offspring of (1) and (2) is $G_O = a(g_1 + g_2)$, with g_i being the number of b alleles (0 or 1) transmitted by Parent (i).

Question: what is $cov(G_i, G_O)$?

Covariance between relatives (1)

Solution

g_i is the number of b alleles (0 or 1) transmitted by Parent (i)
$\Longrightarrow$ g_i follows a Bernoulli distribution with probability $G_i/2a$.

$$
\begin{aligned}
cov(G_1, G_O) &= cov[G_1, a(g_1 + g_2)] \\
&= a \times cov[G_1, g_1] + \underbrace{a \times cov[G_1, g_2]}_{=0\text{, independence}} \\
&= a \times (E[G_1 g_1] - E[G_1]E[g_1]) \\
&= a \times (E[G_1 E[g_1|G_1]] - E[G_1]E[E[g_1|G_1]]) \\
&= \frac{a}{2a} \times \left(E[G_1^2] - E[G_1]^2\right) \\
&= \frac{1}{2} var(G_1)
\end{aligned}
$$

We have used three properties: (1) linearity of covariances, (2) definition of coavriances as function of the expectation and (3) $E[X] = E[E[X|Y]]$.

Covariance between relatives (1)

Solution

g_i is the number of b alleles (0 or 1) transmitted by Parent (i)
$\Longrightarrow$ g_i follows a Bernoulli distribution with probability $G_i/2a$.

$$
\begin{aligned}
cov(G_1, G_O) &= cov[G_1, a(g_1 + g_2)] \\
&= a \times cov[G_1, g_1] + \underbrace{a \times cov[G_1, g_2]}_{=0\text{, independence}} \\
&= a \times (E[G_1 g_1] - E[G_1]E[g_1]) \\
&= a \times (E[G_1 E[g_1|G_1]] - E[G_1]E[E[g_1|G_1]]) \\
&= \frac{a}{2a} \times \left(E[G_1^2] - E[G_1]^2\right) \\
&= \frac{1}{2} var(G_1)
\end{aligned}
$$

We have used three properties: (1) linearity of covariances, (2) definition of coavriances as function of the expectation and (3) $E[X] = E[E[X|Y]]$.

Covariance between relatives (1)

We can prove similarly that

- Parent-offspring covariance $= \sigma_A^2/2$
- full-sib covariance $= \sigma_A^2/2 + \sigma_D^2/4$

General theorem

The genetic covariance between two individuals i and i' is

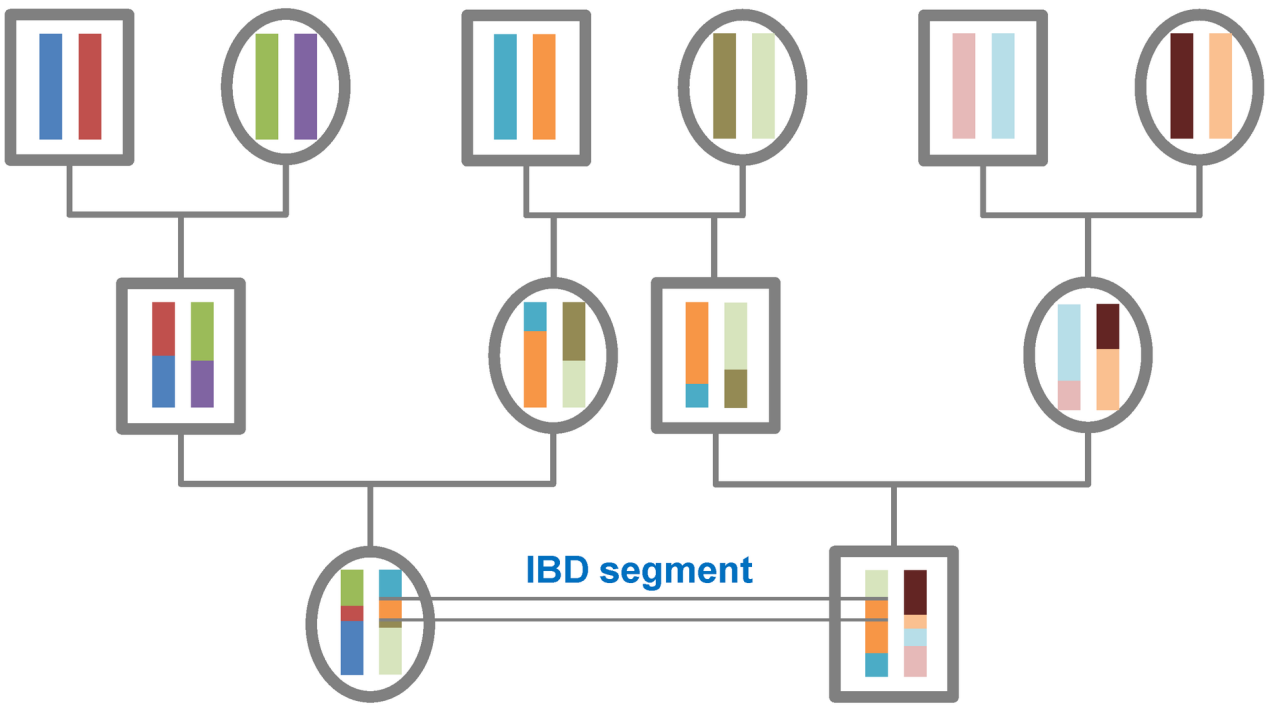
$$cov[G_i, G_{i'}] = 2\theta_{ii'}\sigma_A^2 + \Delta_{ii'}\sigma_D^2 \tag{6}$$

where $\theta_{ii'}$ is the coefficient of coancestry and $\Delta_{ii'}$ is the coefficient of fraternity. We define those two coefficients further in the lecture.

Identity by Descent

Definition

Two alleles at one locus are said Identical By Descent or IBD[3] if they both can be traced back to a common ancestor. If there is no common ancestor then those alleles are simply referred to as Identical By State or IBS.



[3] Picture from Wikipedia.

Coefficient of coancestry

Definition

The coefficient of coancestry $\theta_{ii'}$ between two individuals i and i' is the probability that two alleles picked at random in i and i' are IBD.

Example: parent - offspring

We denote P and O as the Parent and the Offspring respectively. A_P and A_O are random allele from P and O respectively.

$$\begin{aligned}
\theta(P,O) &= P(A_O \text{ IBD } A_P) \\
&= P(A_O \text{ IBD } A_P | A_P \text{ transmitted}) P(A_P \text{ transmitted}) \\
&\quad + \underbrace{P(A_O \text{ IBD } A_P | A_P \text{ untransmitted})}_{=0} P(A_P \text{ untransmitted}) \\
&= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}.
\end{aligned}$$

Coefficient of coancestry

Other examples

- $\theta(\mathit{Full} - \mathit{sibs}) = 1/4$
- $\theta(\mathit{Half} - \mathit{sibs}) = 1/8$
- $\theta(\mathit{First} - \mathit{cousins}) = 1/16$

Calculations for full sibs

...

Coefficient of faternity

Definition

The coefficient of fraternity $\Delta_{ii'}$ between two individuals i and i' is the probability that both alleles in i and i' are IBD.

If i is the offspring of m_i and f_i and i' the offspring of $m_{i'}$ and $f_{i'}$ then $\Delta_{ii'}$ can be expressed as function of the coancestry coefficients between m_i, f_i, $m_{i'}$ and $f_{i'}$ as

$$\Delta_{ii'} = \theta\left(m_i, m_{i'}\right)\theta\left(f_i, f_{i'}\right) + \theta\left(m_i, f_{i'}\right)\theta\left(f_i, m_{i'}\right) \tag{7}$$

Example: full sibs with unrelated parents

Full sibs share the same mother m and father f. Therefore

$$\begin{aligned}\Delta(\text{Full sibs}) &= \theta\left(m, m\right)\theta\left(f, f\right) + \underbrace{\theta\left(m, f\right)\theta\left(f, m\right)}_{=0} \\ &= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}.\end{aligned}$$

Key concepts so far

- Heritability (broad sense: based on additive and dominance / narrow sense: based on additive values): $H^2 = var(G)/var(P)$ and $h^2 = var(A)/var(P)$.
- Genetic covariances arise because relatives share alleles Identical By Descent (IBD).
- Genetic covariances between relatives is a function of $var(A)$, $var(D)$ and of the coefficients of coancestry and fraternity:

$$cov(X, Y) = 2\theta_{XY} var(A) + \Delta_{XY} var(D)$$

(Part II)
Red pill or blue pill?

Estimation of genetic variances

Application

A forthcoming lecture will cover in depth how to estimate genetic variances. We aim here to put in application some of the concepts introduced in this lecture to estimate genetic variances (heritability) in simple experimental designs.

The origin of linear regression

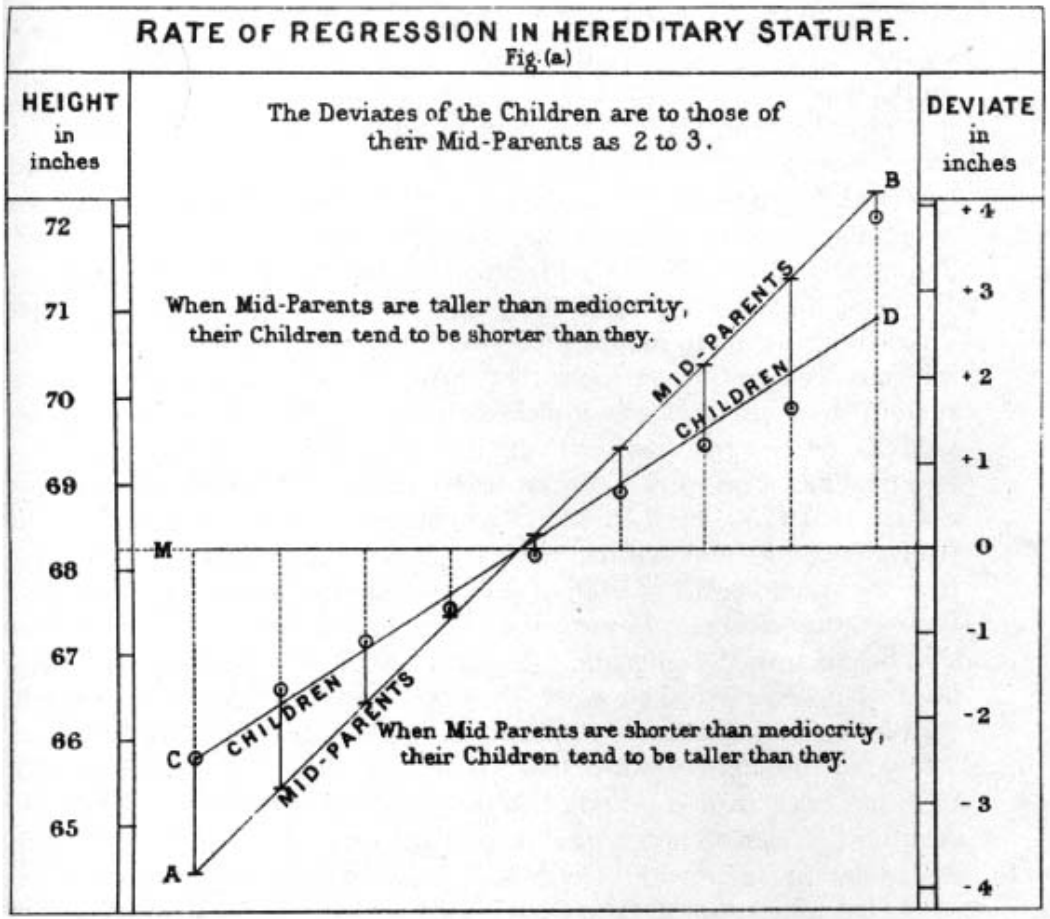


Figure 8.8. Galton's graphical illustration of regression; the circles give the average heights for groups of children whose midparental heights can be read from the line AB. The difference between the line CD (drawn by eye to approximate the circles) and AB represents regression toward mediocrity. (From Galton, 1886a.)

Parent-offspring regression

We have previously derived the genetic covariance (covariance of breeding/additive value) between parent and offspring.

Since breeding values are not observable, can we estimate heritability using phenotypic values? We use below the subscripts p and o respectively for Parent and Offspring:

$$\begin{aligned} cov(P_p, P_o) &= cov(G_p + E_p, G_o + E_o) \\ &= cov(G_p, G_o) + cov(E_p, E_o) \\ &= 1/2 var(A) + cov(E_p, E_o) \end{aligned}$$

If environmental effects are not correlated between parent and offspring, then

$$corr(P_p, P_o) = \frac{h^2}{2}.$$

Monozygotic *vs* Dizygotic twins

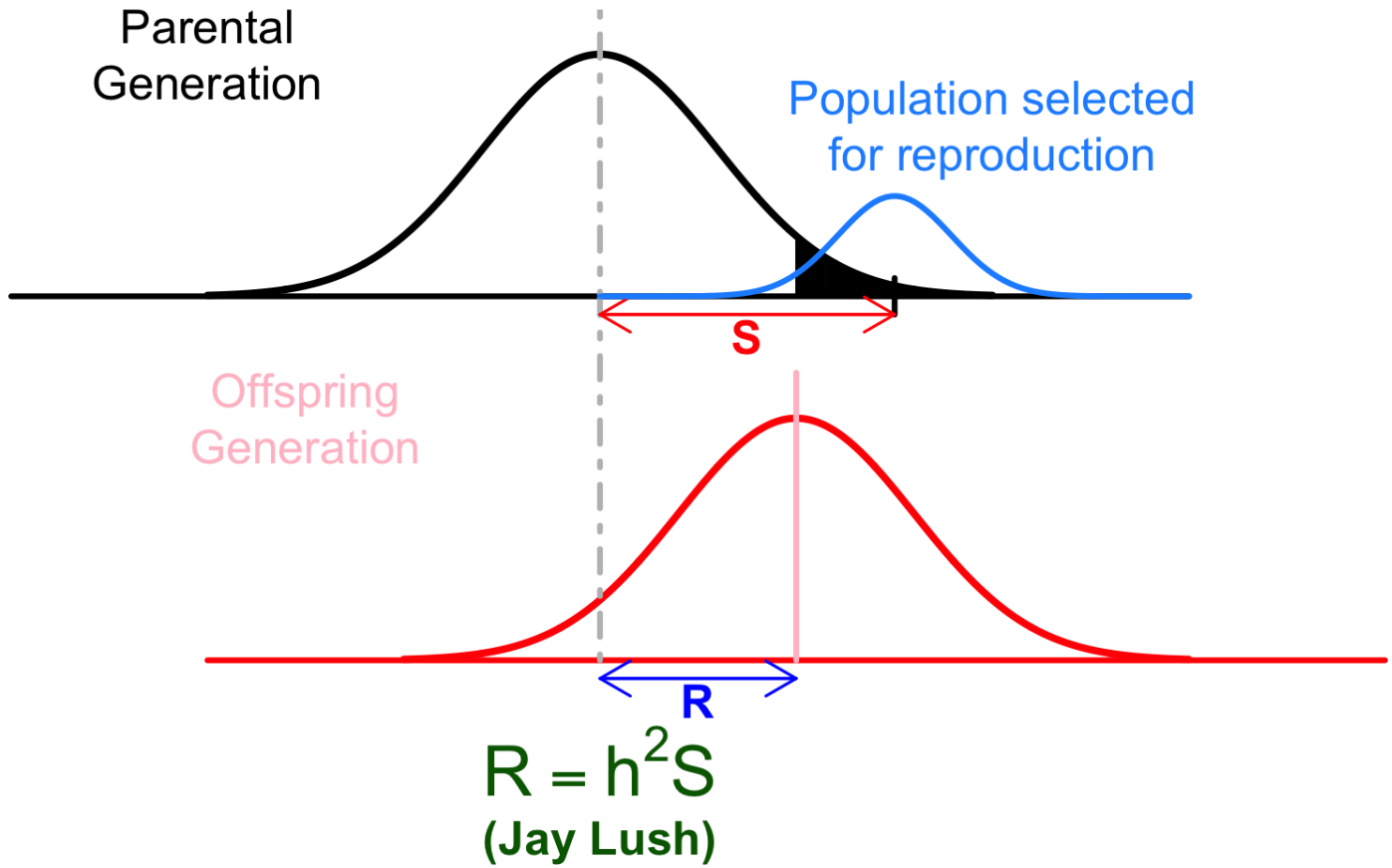
We could refine the estimates of heritability using the following observations and assumptions

- The phenotypic covariance between MZ twins is $cov_p(MZ) = var(A) + var(D) + var(\text{Shared Environment})$
- "DZ twins are full sibs that share a common environment to approximately the same extent as MZ twins".
- Dominance variance is small.

Therefore an **approximate** estimate of h^2 can be obtained as $\widehat{h^2} = 2(cor_p[MZ] - cor_p[DZ]) = [var(A) + \underbrace{var(D)/2}_{\text{if small}}]/var(P)$.

Response to selection (alternative definition)

Heritability is important to predict response to selection



Summary (1/2)

Quantitative Genetics theory answers questions such as

- How would an individual perform depending on its genotype (e.g. breeding program)
- How/why do quantitative phenotypes correlate between relatives?

We discussed key concepts such as the fact that

- "Genotypic values are not passed on to the offspring, only the average (allelic) value."
- "The breeding value (or additive value) is the key quantity to make predictions."

Summary (2/2)

We introduced the notion of heritability h^2 as the ratio between the genetic variance (variance of breeding values) over the phenotypic variance.

Heritability can also be approached through the phenotypic correlation between relatives (e.g. as correlation between mid-parent and offspring) or through the breeder's Equation.

Finally we showed few examples to estimate heritability using phenotypic information from related individuals.

References

Textbooks

Falconer, S.D. and Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics*, 4th Edition.

Lynch, M. and Walsh, B. (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer.

Bulmer, M. (1980) *The Mathematical Theory of Quantitative Genetics*. Clarendon Press.

Article(s)

Cesarini, D. and Visscher, P.M. (2017) *Genetics and educational attainment*, Science of Learning.

Barton, N.H., Etheridge A.M. and Veber A. (2017) *The infinitesimal model: Definition, derivation, and implications*, Theor. Pop. Biol. **118**:50-73.