Relationship and inbreeding

Lecture 4 ▪ Quantitative Genetics

Grum Gebreyesus ▪ ASAP-Bio Knowledge Hub

asap-bio.github.io/lec-qg-relationship.html

Contents

1 Intended learning outcomes

- Explain how shared environments and shared genes contribute to the similarities among relatives.
- Compute relatedness and inbreeding coefficients using pedigree information.

2 Relatedness and the base population

RELATEDNESS

Two individuals are related if

- they share one or more ancestors, or
- one is descended from the other.

2.1 Why relatedness matters in breeding

- Estimating and managing inbreeding
- QTL mapping
- Breeding value estimation
- Prediction of response to selection

2.2 Two essential features of any measure of relatedness

- It can only be defined **relative to a specific frame of reference**, the *base population*.
- It is based on the concept of **identity by descent**.

The first point is frequently overlooked and is a common source of confusion. All members of a population share ancestors at *some* point in the phylogeny, so relatedness is undefined without a stated reference point. In practice we consider recent ancestry rather than ancient common ancestry: if no further information is available for the top generation of a pedigree, those individuals are taken as the **base animals** and treated as unrelated and non-inbred. Every relationship coefficient you compute is therefore a statement about ancestry *since that base*.

3 Identity by descent

- Genes that are copies of the same gene in a past generation are **identical by descent (IBD)**.
- Genes where one is a copy of the other through inheritance are also IBD.
- Genes that merely "look" the same are **identical by state (IBS)**.
- Genes may be IBS *without* being IBD. Genes that are IBD are necessarily also IBS.

INBREEDING AT A LOCUS

- An individual is **inbred at a locus** if the two alleles at that locus are IBD.
- If two genes are IBD they are also IBS, so the individual is homozygous at the locus.

The converse does not hold: a homozygote need not be inbred, since the two identical-looking alleles may descend from different ancestral copies.

4 Why relatives resemble one another

Individuals in a population vary considerably. Those differences may arise from variation in genes, from environmental experience, or from both. For most traits it is both. To the extent that a trait is

influenced by heritable effects, phenotypic similarity should track genetic similarity. We can make that precise by decomposing the covariance between two individuals' phenotypes.

Decomposing the covariance between two phenotypes

1. Write each phenotype in the familiar form $y_i = \mu + g_i + e_i$. Then $\mathrm{Cov}(y_i, y_j) = \mathrm{Cov}(\mu + g_i + e_i,\ \mu + g_j + e_j)$.
2. Expanding, $\mathrm{Cov}(y_i, y_j) = \mathrm{Cov}(g_i, g_j) + \mathrm{Cov}(g_i, e_j) + \mathrm{Cov}(e_i, g_j) + \mathrm{Cov}(e_i, e_j)$.
3. Typically $\mathrm{Cov}(g_i, e_j) = \mathrm{Cov}(e_i, g_j) = 0$: one individual's genes are not associated with another's environment.
4. Hence $\mathrm{Cov}(y_i, y_j) = \mathrm{Cov}(g_i, g_j) + \mathrm{Cov}(e_i, e_j)$, that is, similarity is caused by **shared genes** and **shared environments**.

This result is the basis of quantitative genetic estimation from relatives, and also its principal limitation: relatives usually share environments as well as genes, so $\mathrm{Cov}(e_i, e_j) \neq 0$ inflates estimated heritability unless the design controls for it.

5 The coefficient of coancestry

Take a random allele at a locus from individual i and a random allele at the same locus from individual j.

DEFINITIONS

- θ_{ij} is the probability that these two alleles are IBD: the **coefficient of coancestry**.
- The **coefficient of relatedness** in diploids is $F_{ij} = 2\theta_{ij}$.
- F_{ij} is the expected number of IBD copies, in j, of a random allele drawn from i.

6 Covariance among relatives

Ignoring epistasis, split each genotypic value into its additive and dominance parts, $g_i = a_i + \delta_i$:

$$\begin{aligned}\mathrm{Cov}(g_i, g_j) &= \mathrm{Cov}(a_i + \delta_i,\ a_j + \delta_j) \\ &= \mathrm{Cov}(a_i, a_j) + \underbrace{\mathrm{Cov}(a_i, \delta_j)}_{=0} + \underbrace{\mathrm{Cov}(\delta_i, a_j)}_{=0} + \mathrm{Cov}(\delta_i, \delta_j) \\ &= \mathrm{Cov}(a_i, a_j) + \mathrm{Cov}(\delta_i, \delta_j)\end{aligned}$$

The dominance term $\mathrm{Cov}(\delta_i, \delta_j)$ is normally only appreciable if i and j are **full sibs**, since only full sibs can share *both* alleles at a locus IBD. Generally, therefore, we focus on $\mathrm{Cov}(a_i, a_j)$.

Additive genetic similarity

1. Let α_{ik} be the additive effect of allele k in individual i, so $a_i = \alpha_{i1} + \alpha_{i2}$.
2. $\mathrm{Cov}(a_i, a_j) = \mathrm{Cov}(\alpha_{i1} + \alpha_{i2},\ \alpha_{j1} + \alpha_{j2})$, which expands into four allele-by-allele covariance terms.
3. Each such pair contributes $V(\alpha)$ if the two alleles are IBD (probability θ_{ij}) and 0 otherwise, giving $4\left[(1 - \theta_{ij}) \cdot 0 + \theta_{ij} V(\alpha)\right]$.
4. Since $V(\alpha) = \frac{1}{2} V_a$ (the gametic variance is half the additive variance), this is $4 \times \frac{1}{2}\theta_{ij} V_a = 2\theta_{ij} V_a$.

CENTRAL RESULT

$$\mathrm{Cov}(a_i, a_j) = 2\theta_{ij} V_a = F_{ij} V_a$$

For a few common outbred relative pairs this gives:

Relationship	θ_{ij}	$\mathrm{Cov}(a_i, a_j)$
Parent and offspring	$1/4$	$\frac{1}{2}V_a$
Half sibs	$1/8$	$\frac{1}{4}V_a$
Full sibs	$1/4$	$\frac{1}{2}V_a$ (plus $\frac{1}{4}V_d$)

7 Parent average and Mendelian sampling

Parents have breeding values a_s and a_d. An offspring's expected deviation from the population mean is the **parent average breeding value**:

$$a_{\bar{p}} = \tfrac{1}{2}(a_s + a_d)$$

7.1 Variance of the parent average

$$\begin{aligned}\mathrm{Var}(a_{\bar{p}}) &= \mathrm{Var}\Big(\tfrac{1}{2}(a_s + a_d)\Big) = \tfrac{1}{4}\mathrm{Var}(a_s) + \tfrac{1}{2}\mathrm{Cov}(a_s, a_d) + \tfrac{1}{4}\mathrm{Var}(a_d) \\ &= \tfrac{1}{4}V_a + \tfrac{1}{4}V_a = \tfrac{1}{2}V_a \qquad \text{(assuming } \mathrm{Cov}(a_s, a_d) = 0\text{, i.e. no assortative mating)}\end{aligned}$$

7.2 Mendelian sampling

An offspring is not exactly the average of its parents. Which particular allele each parent transmits is a coin toss, and that randomness has a name and a variance. With no selection and random mating, V_a is constant across generations, so:

$$\begin{aligned} a_o &= \tfrac{1}{2}(a_s + a_d) + a_M = a_{\bar{p}} + a_M \\ V(a_o) &= \tfrac{1}{4}(V_s + V_d) + V_M \\ V_M &= V_a - \tfrac{1}{2}V_a = \tfrac{1}{2}V_a \end{aligned}$$

DISTRIBUTION OF OFFSPRING

Given parents with breeding values a_s and a_d, the two sources of variance among their offspring are $V_M = \frac{1}{2}V_a$ and V_e. Hence

$$a_o \sim N\Big(a_{\bar{p}},\ \tfrac{1}{2}V_a\Big), \qquad y_o \sim N\Big(a_{\bar{p}},\ \tfrac{1}{2}V_a + V_e\Big)$$

This is why even an elite mating produces a range of offspring, and why selection among full sibs is possible at all.

8 The inbreeding coefficient

If the two parents are related, there is a chance they transmit IBD alleles to their offspring. That probability is the **inbreeding coefficient** F_i of the offspring. Note the logic: relatedness is a property of the *parents*, inbreeding is a property of the *offspring*.

F can be computed from pedigree data in two ways: by direct computation of IBD probabilities using **Wright's path method**, or by the **tabular method**, which leads to the numerator relationship matrix below.

WRIGHT'S PATH METHOD

Count the number of meiosis links in each loop through a common ancestor:

$$F = \sum_i \left(\tfrac{1}{2}\right)^{n_i - 1}$$

and if the ancestor in loop i is itself inbred with coefficient F_{A_i},

$$F = \sum_i \left(\tfrac{1}{2}\right)^{n_i - 1} (1 + F_{A_i})$$

n_i is the number of meioses (links) in loop i. Each meiosis halves the probability of transmitting the same allele, which is where the factor of one half comes from.

9 What inbreeding does to variance

When an individual is inbred, its two alleles A_1 and A_2 at a locus are no longer independent, and neither are their effects α_1 and α_2.

$$\begin{aligned} \mathrm{Var}(a) &= \mathrm{Var}(\alpha_1 + \alpha_2) = \mathrm{Var}(\alpha_1) + 2\,\mathrm{Cov}(\alpha_1, \alpha_2) + \mathrm{Var}(\alpha_2) \\ &= 2\,\mathrm{Var}(\alpha) + 2\,[(1 - F) \cdot 0 + F\,\mathrm{Var}(\alpha)] \\ &= (1 + F)\,V_a \end{aligned}$$

So inbreeding *increases* the additive genetic variance among individuals by a factor $(1 + F)$. This is not extra genetic variation being created; it is the same variation redistributed, more of it now expressed between individuals because fewer alleles are hidden in heterozygotes.

9.1 Mendelian sampling variance under inbreeding

Inbred parents generate less variation among their offspring, because at loci where a parent is already homozygous by descent, segregation produces nothing new:

- A non-inbred parent generates $\frac{1}{4}V_a$.
- An inbred parent generates $\frac{1}{4}(1 - F)V_a$.

$$V_M = \tfrac{1}{4}(1 - F_s)V_a + \tfrac{1}{4}(1 - F_d)V_a = \tfrac{1}{2}\left(1 - \frac{F_s + F_d}{2}\right) V_a$$

10 Inbreeding at population level

Worked example: average inbreeding

Start from a population of non-inbred animals. In the next generation, 10% of matings are between half sibs and 5% between full sibs. Half-sib mating gives offspring $F = 0.125$; full-sib mating gives $F = 0.25$. Then

$$\bar{F} = 0.85(0) + 0.10(0.125) + 0.05(0.25) = 0.025$$

10.1 Genotype frequencies with inbreeding

Let a fraction $\bar{F}$ of individuals be inbred at a locus. Among the $1 - \bar{F}$ that are not, genotypes are in Hardy-Weinberg proportions p^2, $2pq$, q^2. Among the $\bar{F}$ that are, the two alleles are IBD, so the individual is homozygous: a fraction p are AA and q are aa, with *no* heterozygotes.

10.2 Estimating inbreeding from a heterozygote deficiency

Deriving $\bar{F}$ from observed heterozygosity

1. Under Hardy-Weinberg we expect $H_{\text{exp}} = 2pq$ heterozygotes.
2. With average inbreeding $\bar{F}$, only the non-inbred fraction contributes heterozygotes: $H_{\text{obs}} = (1 - \bar{F})2pq$.
3. Therefore $\dfrac{H_{\text{exp}} - H_{\text{obs}}}{H_{\text{exp}}} = \dfrac{2pq - (1 - \bar{F})2pq}{2pq} = \bar{F}$.

Worked example: computing $\bar{F}$ from genotype frequencies

A population shows 50% A_1A_1, 20% A_1A_2, 30% A_2A_2.
$p = 0.5 + 0.2/2 = 0.6$, so $q = 0.4$. Expected heterozygosity $H_{\text{exp}} = 2(0.6)(0.4) = 0.48$, observed $H_{\text{obs}} = 0.20$.

$$\bar{F} = \frac{0.48 - 0.20}{0.48} \approx 0.58$$

Worked example: why breeders worry about small $\bar{F}$

Take $\bar{F} = 0.025$ from the earlier example, and a recessive disease allele at $q = 0.001$.
Under Hardy-Weinberg the disease frequency would be $q^2 = 10^{-6}$. With inbreeding it becomes $q^2 + \bar{F}pq \approx 2.6 \times 10^{-5}$.
An average inbreeding of just 2.5% raises the incidence roughly **26-fold**. This is why rare recessive defects surface in closed nucleus herds long before inbreeding depression in production traits becomes obvious.

11 Inbreeding depression

The general observation is that inbred individuals perform worse, with an approximately **linear** decrease in performance as F increases.

11.1 Why

- An excess of deleterious homozygotes.
- A deficiency of advantageous heterozygotes.

Formally, inbreeding depression occurs if the dominance deviations δ_{Aa}, summed across loci, are *typically positive*, that is, if heterozygotes tend to be above the midpoint of the two homozygotes.

Supplementary: the size of the depression

1. In Falconer and Mackay's notation, with no inbreeding the mean is $\mu + (p-q)a + 2pq\,d$.
2. With average inbreeding $\bar{F}$, the heterozygote frequency falls to $(1-\bar{F})2pq$ and the mean becomes $\mu + (p-q)a + 2(1-\bar{F})pq\,d$.
3. The difference is $-2\bar{F}pq\,d$ per locus, so summed over loci the total inbreeding depression is $D = \bar{F}\sum_i 2p_iq_id_i$, linear in $\bar{F}$, as observed.

12 The numerator relationship matrix

For breeding value estimation we need a matrix representation of all pairwise relationships. The natural one is the **numerator relationship matrix $\mathbf{A}$**:

$$\mathbf{A} = \begin{pmatrix} 1+F_{11} & \cdots & F_{1n} \\ & \ddots & \\ \vdots & 1+F_{ii} & \vdots \\ & & \ddots \\ F_{n1} & \cdots & 1+F_{nn} \end{pmatrix}$$

The off-diagonal element F_{ij} $(i \neq j)$ is the coefficient of relatedness $F_{ij} = 2\theta_{ij}$; the diagonal element is $1 + F_{ii}$, where F_{ii} is the individual's inbreeding coefficient. The diagonal exceeding one is exactly the $(1+F)V_a$ result derived above.

EXTENSION TO BLUP

$\mathbf{A}$ is what makes BLUP possible: the additive genetic covariance structure of an entire pedigree is $\mathbf{A}V_a$. In genomic prediction the same role is played by a genomic relationship matrix $\mathbf{G}$, estimated from markers rather than deduced from the pedigree.

13 Assortative mating

What happens when we mate the best with the best, phenotypically? This is what breeding programmes usually do, so the consequences matter. Assortative mating means $\mathrm{Cov}(y_s, y_d) = r > 0$.

Derivation

1. Association between *phenotypes* induces association between *breeding values*: $\mathrm{Cov}(a_s, a_d) = r\,h^2V_a$.
2. The parent average is therefore more variable than before: $\mathrm{Var}(a_{\bar{p}}) = \frac{1}{4}\left(V_a + 2\,\mathrm{Cov}(a_s, a_d) + V_a\right) = \frac{1}{2}V_a\left(1 + rh^2\right)$.
3. Adding Mendelian sampling, $\mathrm{Var}(a_o) = \mathrm{Var}(a_{\bar{p}}) + V_M = \frac{1}{2}V_a(1+rh^2) + \frac{1}{2}V_a = V_a\left(1 + \frac{1}{2}rh^2\right)$.

In the first generation assortative mating **increases** the additive variance. There is, however, a counteracting effect. Mating superior with superior and inferior with inferior induces inbreeding at the genes affecting the trait, which reduces V_M. Over time the effect reverses:

Generation	Additive genetic variance	Heritability
0	V_a	h^2
1	$V_a\left(1+\frac{1}{2}rh^2\right)$	$h^2\dfrac{1+\frac{1}{2}rh^2}{1+\frac{1}{2}r\left(h^2\right)^2}$
∞	$V_a\left(1-rh^2\right)$	$h^2\dfrac{1-rh^2}{1+r\left(h^2\right)^2}$

CONSEQUENCES

- In the **short term**, assortative mating increases V_a, heritability and selection response.
- In the **longer term** the effect is the opposite, because of the inbreeding induced at the trait genes.

14 Summary

- Resemblance among relatives is due to shared genes, and to shared environments.
- Mating among relatives causes inbreeding.
- Inbreeding leads to inbreeding depression.
- The numerator relationship matrix summarises degrees of allele sharing.

15 Exercises

Exercise 4.1

Regression on parent average. Show that the regression of an offspring's phenotype on the parent average phenotype equals the heritability, that is $\mathrm{Reg}(y_o, \overline{P}) = h^2$, where $\overline{P} = \frac{1}{2}(y_s + y_d)$. You may use the result from the lecture that the covariance between one parent and its offspring is $\frac{1}{2}V_a$.

We need $\mathrm{Reg}(y_o, \overline{P}) = \dfrac{\mathrm{Cov}(y_o, \overline{P})}{\mathrm{Var}(\overline{P})}$, so compute both.

Denominator. Assuming the parents are unrelated and each has phenotypic variance V_p,

$$\mathrm{Var}(\overline{P}) = \mathrm{Var}\left(\frac{y_s + y_d}{2}\right) = \tfrac{1}{4}\left(\mathrm{Var}(y_s) + \mathrm{Var}(y_d)\right) = \tfrac{1}{4}\cdot 2V_p = \tfrac{1}{2}V_p$$

Numerator. Using $\mathrm{Cov}(y_o, y_s) = \mathrm{Cov}(y_o, y_d) = \frac{1}{2}V_a$,

$$\mathrm{Cov}(y_o, \overline{P}) = \tfrac{1}{2}\left[\mathrm{Cov}(y_o, y_s) + \mathrm{Cov}(y_o, y_d)\right] = \tfrac{1}{2}\left(\tfrac{1}{2}V_a + \tfrac{1}{2}V_a\right) = \tfrac{1}{2}V_a$$

Ratio.

$$\mathrm{Reg}(y_o, \overline{P}) = \frac{\frac{1}{2}V_a}{\frac{1}{2}V_p} = \frac{V_a}{V_p} = h^2$$

Note the contrast with regression on a *single* parent, which gives $\frac{1}{2}h^2$. Averaging the two parents halves the variance of the predictor while leaving the covariance unchanged, which doubles the regression coefficient.

Exercise 4.2

Path method. Two half sibs share a single common ancestor and are mated together. Compute the inbreeding coefficient of their offspring, first assuming the common ancestor is not inbred, then assuming it has $F_A = 0.20$.

Trace the loop from one parent up to the common ancestor and back down to the other parent. For half sibs mated together the loop passes through 3 meioses, so $n = 3$.

$$F = \left(\tfrac{1}{2}\right)^{n-1} = \left(\tfrac{1}{2}\right)^{2} = 0.125$$

With an inbred common ancestor,

$$F = \left(\tfrac{1}{2}\right)^{n-1}(1 + F_A) = 0.125 \times 1.20 = 0.15$$

The ancestor's own inbreeding raises the chance that the two alleles it passes down are IBD, so it inflates the descendant's F proportionally.

Exercise 4.3

Heterozygote deficiency. A population is scored at one locus: 36% A_1A_1, 48% A_1A_2, 16% A_2A_2. Is there evidence of inbreeding? Now repeat for a second locus scored as 45% A_1A_1, 30% A_1A_2, 25% A_2A_2.

Locus 1. $p = 0.36 + 0.48/2 = 0.60$, $q = 0.40$. $H_{\text{exp}} = 2(0.6)(0.4) = 0.48$, and $H_{\text{obs}} = 0.48$. $\bar{F} = (0.48 - 0.48)/0.48 = 0$. No evidence of inbreeding; the locus is in Hardy-Weinberg proportions.
Locus 2. $p = 0.45 + 0.30/2 = 0.60$, $q = 0.40$. $H_{\text{exp}} = 0.48$, $H_{\text{obs}} = 0.30$.

$$\bar{F} = \frac{0.48 - 0.30}{0.48} = 0.375$$

A substantial apparent inbreeding. Note the caution, though: both loci have identical allele frequencies, so the difference is entirely in the genotype distribution. A heterozygote deficiency at a *single* locus can also be produced by population substructure (the Wahlund effect), null alleles, or genotyping error, so $\bar{F}$ estimated this way should be averaged over many loci before it is believed.