

## Quantitative genetics

Quantitative genetics is the study of the inheritance of quantitative/continuous phenotypic traits, like human height and body size, grain colour in winter wheat or beak depth in birds. These phenotypes are more commonly found in nature than Mendelian traits. Quantitative traits are influenced by both (many) loci (i.e. they are **oligogenic** or **polygenic**) and the environment. The alleles at all loci contributing to the **phenotypic variation** act predominantly **additively**, leading to a **normal distribution** of **phenotypic values** in a population. Under additivity, the effect of each allele is not affected by what other allele is present at the same locus, nor by what alleles are present at other loci. Additive effects do not have to be the same for each allele.

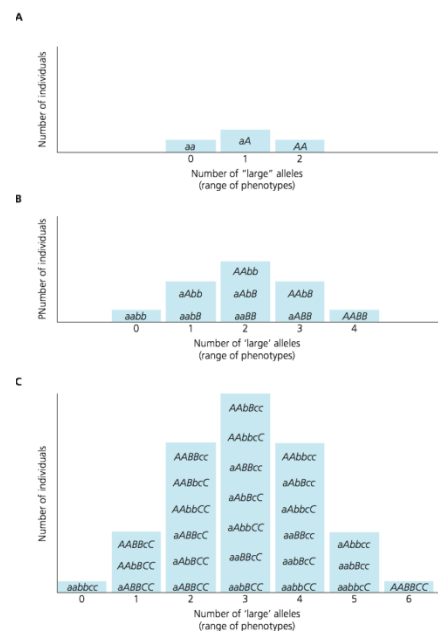


Figure 1: Distribution of phenotypic values in a population, where alleles at (A) one, (B) two or (C) three loci act additively. The phenotype of each individual is obtained by summing the effects of individual alleles. The more loci contribute to the phenotype, the more the distribution is approximately normal.

The two components of the **phenotypic value P** of an individual are the **genotypic value G** and the **environmental deviation E**, such that

$$P = G + E$$

The mean environmental deviation in a population is taken to be zero, such that the expected phenotypic value is the genotypic value. In practice, **we can normally measure only P**. G and E are artificial constructs that have to be estimated using statistical models (can you think of a way how G could be measured directly?). Interestingly, in order to do so, we do not need to know the loci contributing to the phenotypic value. Actually, most of the theory of quantitative genetics was developed before it was even known that DNA is the genetic material of inheritance. So how does this work then?

In quantitative genetics, we make use of the **resemblance between relatives** to estimate how much of the variation in a phenotypic trait has a genetic basis. For example, we can use the resemblance

between parents and their offspring (using linear regression), between half-sibs and full-sibs, monozygotic and dizygotic twins, or between all members in a complex **pedigree** (using mixed-effects models that are called “**animal models**”). In clonal or highly selfing organisms the resemblance is due to  $G$ , because individuals pass their diploid genotypes on to their offspring intact. In sexually outcrossing populations, each parent contributes half of its genome to the offspring and genotypes are created anew. Then the resemblance is primarily due to the additive effects of alleles (actually, it is the *average effect* because whenever allele frequencies are unequal, non-additive gene action contributes to additive genetic variance and therefore the resemblance between parents and offspring, but this goes too far here) but not all alleles act entirely additively and we need to further decompose  $G$  into its **additive genetic component A**, the **dominance deviation D** and the **interaction (epistatic) deviation I**, such that

$$G = A + D + I \text{ and } P = A + D + I + E$$

$A$  is also called the **breeding value** of an individual (and is twice the mean deviation of its offspring from the population mean).  $D$  is due to interactions of alleles within a locus and  $I$  is due to interactions of alleles between loci (**epistasis**). We call  $D$  and  $I$  deviations because they depend on the genetic background and cannot be transmitted from parents to offspring. Thus, they lead to a deviation of the offspring phenotypic value from their parents.

We next partition the total **phenotypic variance  $V_P$**  (the variance in  $P$ ,  $P$  *measured* in multiple individuals in a population) into its **genotypic ( $V_G$ )** and **environmental ( $V_E$ ) variance component**. Assuming that there is no correlation or interaction between the genotypes and environment, it turns out that the variance of a sum of independent variables is equal to the sum of their individual variances, such that

$$V_P = V_G + V_E \text{ and } V_P = V_A + V_D + V_I + V_E$$

where  $V_A$  is the **additive genetic variance component** (= variance in breeding values),  $V_D$  is the dominance genetic variance component and  $V_I$  is the interaction genetic variance component (the latter two are jointly denoted as **non-additive genetic variance**). In sexually reproducing species, it is  $V_A$  that causes the resemblance between parents and offspring, meaning that it is the easiest of the genetic variance components to be measured using the resemblance between relatives. It is also the most important genetic variance component for changes in mean phenotypic values across generations (due to **artificial** or **natural selection**). We say that it is “visible to selection”.

We define the **broad-sense heritability** as

$$H^2 = V_G / V_P$$

which is the fraction of the phenotypic variance in a population that can be attributed to variance in genotypes. It is more of theoretical interest because in sexually reproducing species we cannot measure  $V_G$  (because it includes the epistatic and the dominance variance components). In clonal or selfing species, however, it is  $H^2$  that affects the evolutionary rates.  $H^2$  sets an upper bound for the (**narrow-sense**) **heritability** which is defined as

$$h^2 = V_A / V_P$$

In outbreeding species, it is  $h^2$  that affects evolutionary rates (see below). Sewall Wright used  $h$  (for heredity) to denote the correlation between genotype and phenotype. The square of that correlation (that is,  $h^2$ ) is, per definition, the proportion of variation in the phenotype that is attributable to the path from genotype to phenotype. The heritability is **not** the fraction of an individual's phenotype that is caused by its heredity versus environment but it is the fraction of the phenotypic variance that is due to additive genetic variation *in a population*. It provides information on the relative importance of heredity in determining phenotypic variance (**nature versus nurture**). Actually, there can be many loci contributing to a trait, but if they are all fixed in a population, the heritability would be zero. As an example, humans have two arms, which is undoubtedly genetically determined but there is no variation in the number of arms. The definition of heritability is based on *variation* and it is a population-specific parameter (that can vary within a species, across life stages and environments).

Heritability estimates for morphological traits typically are in the range of 0.5–0.8, for behavioural traits around 0.3, and for fitness-related traits closer to zero. The relatively low heritability of fitness-related traits is because natural selection should deplete additive genetic variation in fitness and / or because fitness-related traits are functions of morphological traits such that there is additional environmental variation in these traits. Additive genetic variation in fitness (and other traits under selection) can be maintained through **negative genetic correlations (negative pleiotropy)** between fitness traits or through **genotype-by-environment interactions**, where genotype G1 has higher fitness in environment E1 than genotype G2 but lower fitness than G2 in environment E2. Interestingly, although the heritability is a population-specific estimate that depends on the allele frequencies in a population, estimates from different populations of the same species (also between captive and field environments) tend to be highly correlated.

We can estimate the heritability of a trait by performing a **linear regression** of the phenotypic values of the offspring over the mean of the two parental phenotypic values (the “mid-parent” value). The slope of the regression line then is the heritability. When we use the phenotypic value of just one parent instead of the mid-parent value, we have to multiply the slope by two in order to obtain the heritability. The name regression actually comes from the observation that in parent-offspring regressions the slope is always less than one. This means that the offspring phenotypes “regress” back to the parental phenotypic mean.

In order to get an unbiased estimate of  $h^2$  it is important that parent and offspring environments are not correlated. This could result from humans growing up in the same family (e.g. there is a discussion on the heritability of the IQ) or birds raised in the same nest where larger birds provision the offspring with more food. To reduce this correlation eggs are often **cross-fostered** into nests of unrelated foster parents.

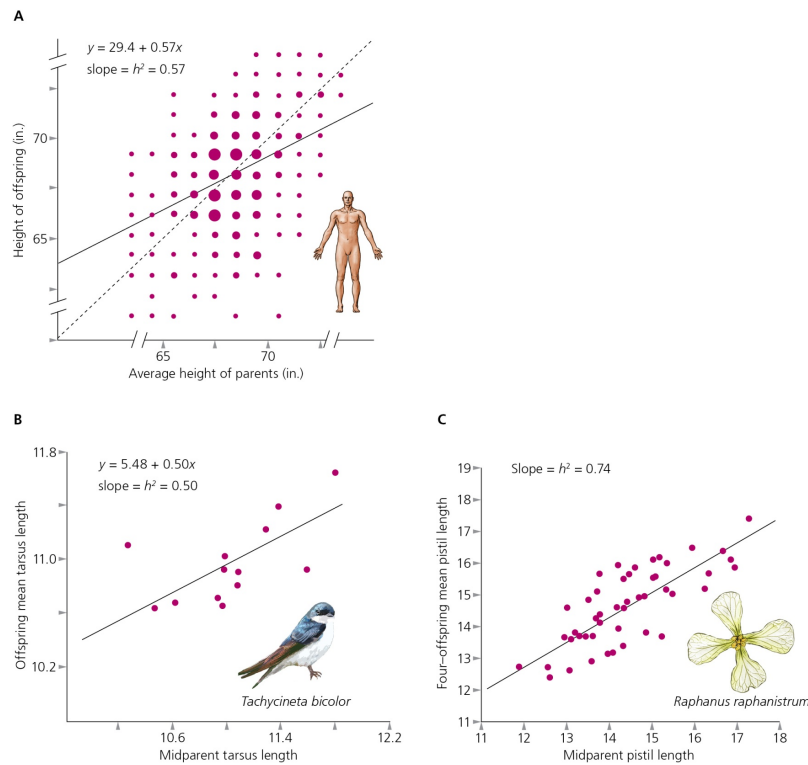


Figure 2: Estimating the heritability of morphological traits using parent-offspring regression. (A) Human height (data taken from Galton 1889), (B) tarsus length in tree swallows and (C) pistil length in wild radish.

Up to now we have been concerned with the description of the genetic properties of phenotypic traits in a population (mostly under random mating). Why should we be interested in decomposing the phenotypic variance into additive and non-additive variation? And why is the heritability one of the most often reported estimates in quantitative genetics? Once we know the sources of variation in a phenotypic trait we can study how it will evolve. If some members of a population breed and others do not and you compare the mean phenotypic value for the breeders with the mean value of the whole population, then the difference between the two is the **selection differential S**.

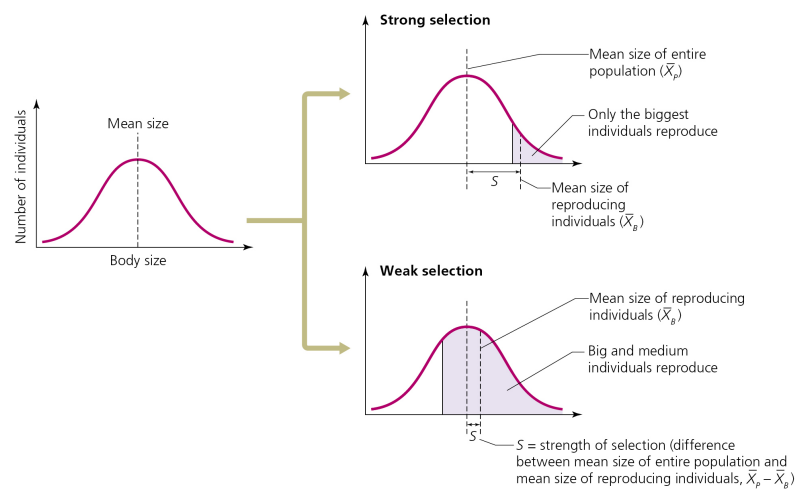


Figure 3: An illustration of different selection differentials acting on a phenotypic trait.

If **selection** favours certain values of a phenotypic trait, then we might expect the population to evolve in response. As we have said earlier, it is the additive genetic variance that causes resemblance

between parents and offspring and the heritability is a measure of what fraction of the phenotypic variation is visible to selection. Consequently, the heritability plays a key role in predicting how phenotypes will change over time as a response to natural or artificial selection. It can be shown that the response to selection  $R$  can be described by the **Breeder's equation**

$$R = h^2 \times S$$

The **response to selection  $R$**  is the difference between the mean phenotypic trait value in the offspring of those parents that were allowed to reproduce and the mean value of the whole population prior to selection, which is the change in trait value from one generation to the next. All else being equal, the response to selection is proportional to the heritability of the selected trait.

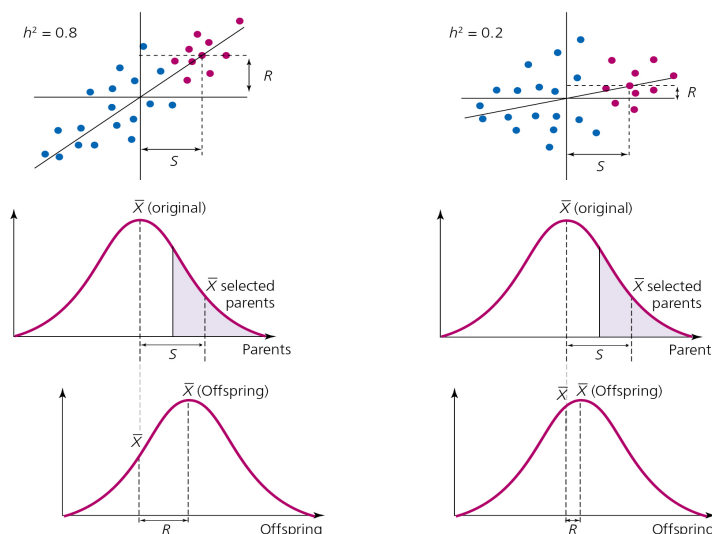


Figure 4: The response to selection  $R$  is proportional to the heritability of the phenotypic trait under selection and can be predicted from the parent-offspring regression.

**Example:** Consider a population of birds, in which the mean beak depth is 15 mm. After a severe drought, only those birds with the largest beaks survive and reproduce. The mean beak depth in these birds is now 20 mm. In the following generation the mean beak depth is 18 mm. What is the heritability of beak depth in this population?

Selection differential  $S = 20 - 15 = 5$  mm

Response to selection  $R = 18 - 15 = 3$  mm

Using the breeder's equation we can estimate the heritability as

$$R = h^2 \times S$$

$$\Leftrightarrow h^2 = R / S$$

$$h^2 = 3 / 5 = 0.6$$

Today much of the studies in quantitative genetics focus on **mapping** (locating) and estimating the effects of the underlying genetic variants contributing to variation in a phenotypic trait. This is either done through “quantitative trait locus mapping” (**QTL mapping** or **linkage analysis**) or through genome-wide association studies (**GWAS** or **association mapping**). Both techniques use data on molecular markers (today usually **single nucleotide polymorphisms**, **SNPs**) that are either genotyped within families / pedigrees (QTL mapping) or in a random sample of (unrelated) individuals from a population (GWAS). In QTL mapping studies, we focus on the co-inheritance of a marker with the phenotypic trait under study. **Meiotic recombination** usually breaks up the **genetic linkage** between molecular markers and the alleles having an effect on the phenotype (i.e. the “**causal alleles**”). However, if a marker co-segregates with a causal allele, it explains some variation in the phenotypic trait and we call this region a **quantitative trait locus (QTL)**. GWAS make use of the historical recombination patterns in a population, which lead to **linkage disequilibrium** between molecular markers and causal alleles. Researchers test for an association between the phenotypic trait values and the marker alleles, for example using a linear regression model. In a linkage analysis, researchers typically use an interval mapping approach, whereas in a GWAS each marker is fitted singly in a statistical model. Because hundreds or even thousands of statistical tests are performed, *P*-values of these tests need to be adjusted (for example, Bonferroni corrected). QTL mapping studies tend to be more powerful but less precise than GWAS.

Nevertheless, the number of individuals included in a QTL study needs to be high for sufficient power to robustly detect loci. It is important to remember that we are sampling the population, and that estimates of effect size may be less accurate for *smaller* sample sizes (this is true for statistical tests more generally). Because putative QTL with ‘larger’ effects are more likely to be detected as significant, and because we only estimate effect sizes for QTL that have been detected, there is a bias towards over-estimates of effect size in QTL studies (especially when the sample sizes are relatively small). It is also important to remember that QTL studies detect *regions* in LD with causative loci – they are *not* genes.

---

**Example:** Human body height has a heritability of around 0.8. Using height measurements on 183,727 individuals that were genotyped for more than one million SNPs, Lango Allen *et al.* (2010) identified 180 loci that were significantly associated with variation in human height. Despite this massive sample size, all 180 loci combined explained only approximately 10% of the phenotypic variance. The unexplained fraction of the heritability is called the “**missing heritability**”. Possible explanations for the missing heritability are loci with very small effects that were non-significant in current GWAS, many causal loci at low allele frequencies (and possibly larger effects) and gene-gene-interactions (epistasis).

---

### Literature

- Falconer D, Mackay T (1996) *Introduction to quantitative genetics*, 4th edn. Longman, Harlow, UK.  
Lango Allen, H *et al.* (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467, 832–838.  
Lynch M, Walsh B (1998) *Genetics and analysis of quantitative traits*. Sinauer, Sunderland, MA.  
Roff D (1997) *Evolutionary quantitative genetics*. Chapman & Hall, London, UK.