Heritability:
the nature and nurture of phenotypes
(part 1)

Most physiological and morphological traits are quantitative (or continuous)

- Weight.
- Size.
- Perception of depth.
- Flowering time.
- Age at first reproduction.
- Amount of resources consumed.
- Levels of parental care.
- Blood pressure.
- Level of hormones.
- Strength.
- Bone resistance to fractures.
- Facial features.
- Reaction time.
- Canine length

Most psychological or personality traits are quantitative

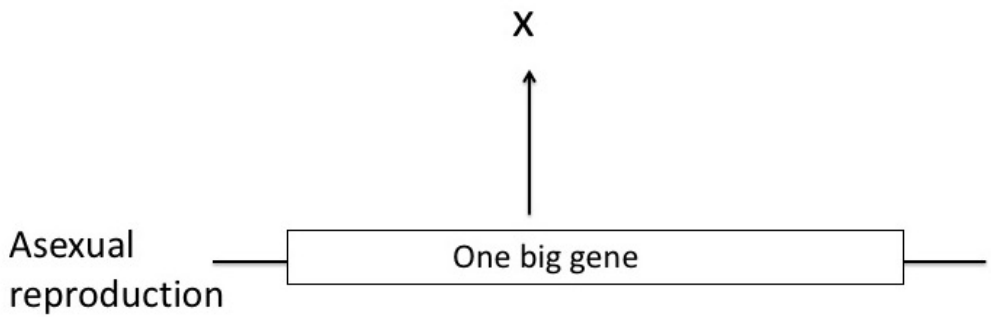
- Aggressiveness.
- Self-esteem.
- IQ.
- Harm avoidance.
- Novelty seeking.
- Perfectionism.
- Rigidity.
- Psychoticism.
- Openness.
- Impulsivity.
- Disinhibition.
- Obsessionallity.
- Sensory processing sensitivity.
- Alexithymia (inability to express emotions).

The big five personality traits are: openness to experience, conscientiousness, extraversion, agreeableness, neuroticism.

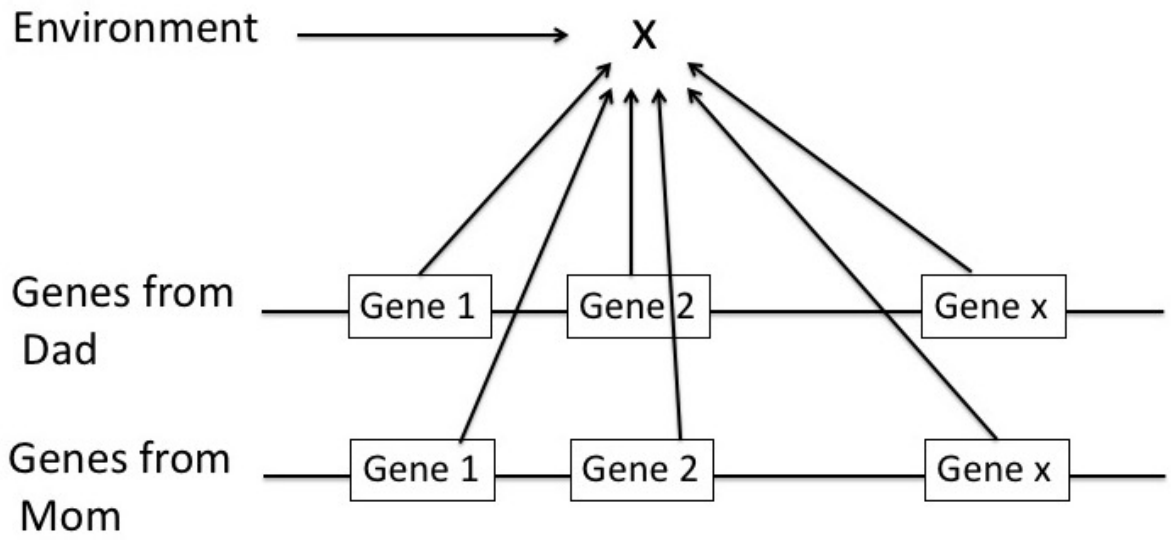
Simple genetic determinism

So far we looked at phenotypic evolution in a simplified way:

- We considered a single gene underlying the phenotype and to each different phenotypic value we assigned a different allele.
- We considered full genetic determinism.



Both genes and the environment determine phenotype



In general:

1. All phenotypes have both a genetic and an environmental component.
2. Many genes are likely to underly a given phenotype.

The nature-nurture question

- How much of the phenotypic variation that is observed in a population for a given phenotype (say in IQ "intelligence quotient", disinhibition, or political opinion) is due to genes and how much is due to the environment?
- This is a fundamental question in biology and increasingly so in the social sciences.
- The way schools, organizations, and societies are managed depends to some extent at least on the response to this question.

Two broad answers to these questions have been proposed.

The standard social science model (SSSM): all nurture

The standard model of behavior in the social science goes approximately as follows.

1. We are born with a mind with no innate traits thus allowing us to make choices free from biology: we have "blank slates".
2. We are born good but become corrupted through society: we are "noble savages".

We will address the first point in this course. The second point will be addressed in the course "The Evolution of Cooperation: from Genes to Culture" of the next semester.

(Pinker 2002)

The natural sciences say differently: nature and nurture

The three **empirical** laws of behavior genetics:

1. All human behavioral traits are heritable.
2. The effect of being raised in the same family is smaller than the effect of the genes.
3. A substantial portion of complex human traits is not accounted for by the effect of genes or families.

What exactly do these statements mean? The concept of heritability is key here.

(Turkheimer, 2000)

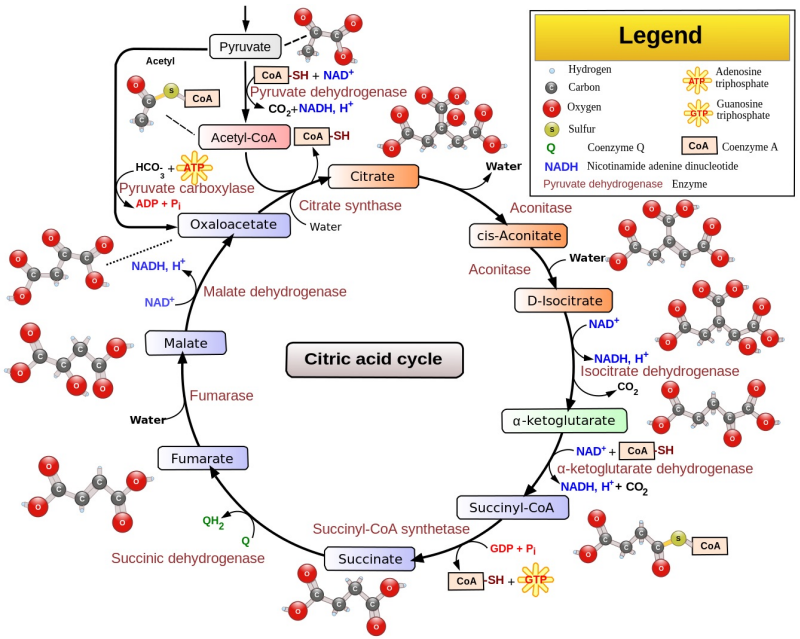
Objective of this lesson

- Understand what heritability means.
- For this we will build a tractable model to quantify environmental and genetic effects on phenotype. This model is from the field called quantitative genetics.
- This will allow us to understand statements such has IQ or aggressiveness in males has a heritability 0.5.

More generally quantitative genetics will become increasingly more relevant to medicine, economics, and societal issues at large (e.g., the field of "geneconomics").

Phenotypes are complex

- The expression of phenotypes depends in a complicated way on environmental and genetic effects.
- There is an unknown mapping from genotype and environment to the phenotype.



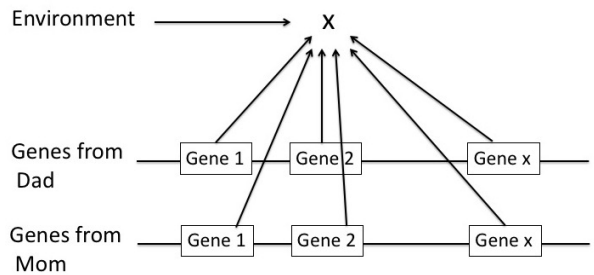
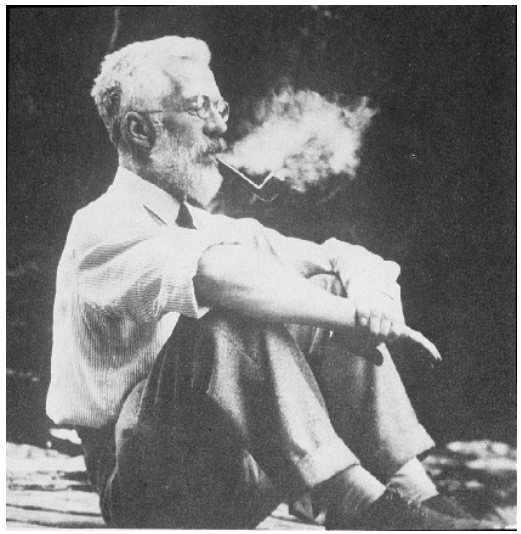


Figure: Citric acid (or Kreps) cycle underlying the production of energy in cells.

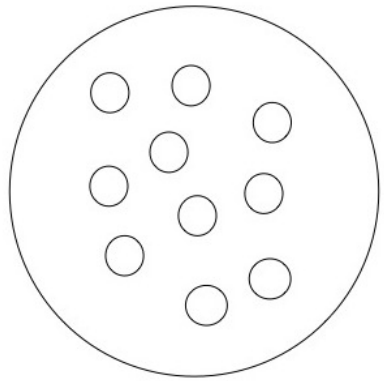
Phenotypes are too complex to be described mechanistically and need to be accounted for statistically

- The effect of genes is so complex that there is presently no hope to built a detailed mechanistic model of genetic and phenotypic expression.
- We will look at genetic effects in a statistical way.



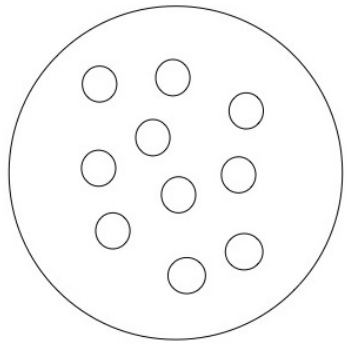
R. A. Fisher, one of the founder of modern statistics and quantitative evolutionary biology.

Consider a population for a quantitative trait



- We index individuals in the population by i and denote by x_i the value of the phenotype of individual $i \in \{1, 2, 3, ..., n\}$ of the population.
- The population is described by the vector $\{x_1, x_2, ..., x_n\}$, e.g., for size in a population of $n = 5$ this could be $\{1.64, 1.70, 1.66, 1.8, 1.75\}$.

Statistical descriptors of the population



The average and variance in the population are respectively

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad V_x = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

For the example with $\{1.6, 1.7, 1.65, 1.85, 1.75\}$ this would be

$$\bar{x} = 1.71 \qquad V_x = 0.007$$

The additive model of phenotypic determinism

We decompose the phenotype into two separate effects

$$x_i = g_i + e_i$$

- g_i is the genetic value of the trait of individual i and represents the net contribution of all genetic factors.
- e_i is the environmental value of individual i and represents the net contribution of all environmental factors.
- In all models so far we (implicitly) assumed that $e_i = 0$.

Two major sources of environmental effects

1. Special (or idiosyncratic) environmental effects refers to effects that is specific to each individual independently from each other (e.g., quality of air in one's sleeping room).
2. General (or systematic) environmental effects refers to environmental effects that are shared by several or all individuals in the population simultaneously (e.g., school). These include effects due to the family or the socioeconomic strata in which an individual resides.

We will not distinguish between these effects explicitly and just consider that each individual is exposed to a net environmental effect e_i including all relevant environmental factors.

Environment is assumed to be independent of genes

We assume no covariance between gene and environment

$$\mathrm{Cov}[g_i, e_i] = \frac{1}{n}\sum_{i=1}^{n}(g_i - \bar{g})(e_i - \bar{e}) = 0$$

- For several traits there is no covariance.
- One can think of g_i as the linear regression of phenotype on genetic effects (due to expressing the genotype) so that e_i is the "error" (or residual) in the prediction of phenotype given the genes.
- e_i is then uncorrelated to g_i and in a planned experiment we can control $\mathrm{Cov}[g_i, e_i]$ to be equal to zero. This maximizes the contribution of the genes to phenotype.

Hence, if we look at things statistically, we can write $x_i = g_i + e_i$ eventhough there are interactions between genes and environment.

Mean in the population

Taking the average of the phenotype

$$x_i = g_i + e_i$$

over all individuals in the population gives the mean phenotype $\bar{x}$ in the population as

$$\bar{x} = \bar{g} + \bar{e}$$

- $\bar{g}$ is the mean genetic value of the trait in the population.
- $\bar{e}$ is the mean environmental value of the trait.

Variance in the population

Since $\mathrm{Cov}[g_i, e_i] = 0$, taking the variance over all individuals in the population of

$$x_i = g_i + e_i$$

gives the phentoypic variance in the population as

$$V_{\mathrm{x}} = V_{\mathrm{g}} + V_{\mathrm{e}}$$

1. V_{g} is the genetic variance.
2. V_{e} is the environmental variance.

Hence, we have decomposed the variation in the phenotype into two sources of variation: (1) variation of genetic values and (2) variation of environmental values.

Broad sense heritability

The broad sense heritability of the trait is defined as

$$H^2 = \frac{V_{\mathrm{g}}}{V_{\mathrm{x}}}$$

This is the proportion of variance in the population that is explained by the genetic variance.

- If heritability is low, much of the difference between individuals in the population is explained by them having different environments.
- If heritability is high, much of the difference between individuals is explained by them having different alleles.

Broad sense heritability

The broad sense heritability determines the proportion of the phenotypic variance in the population that is determined by genes

$$1 = \underbrace{\underbrace{\frac{V_g}{V_g + V_e}}_{H^2}}_{\text{"Nature"}} + \underbrace{\underbrace{\frac{V_e}{V_g + V_e}}_{1-H^2}}_{\text{"Nurture"}}$$

- The higher the heritability the more variance explained by different individuals having different alleles.
- $1 - H^2$ is the proportion of phenotypic variance ascribed to environmental effects. This proportion can be modified by manipulation of the environment.

What does broad sense heritability measure?

- In a uniform environment $V_e = 0$. Then $H^2 = 1$.
- In a genetically monomorphic population $V_g = 0$. Then $H^2 = 0$.
- Hence, heritability is not an index of the degree of genetic determinism of any given phenotype of a given individual.

Heritability is a population concept and explains the proportion of phenotypic variance in a population due to different individuals having different alleles.

What does broad sense heritability measure?

Some examples of H^2 values:

- Fingerprint ridge count $H^2 = 0.98$.
- Height $H^2 = 0.66$.
- IQ $H^2 \in [0.5, 0.8]$.
- Social maturity score $H^2 = 0.16$.

$1 - H^2$ can be thought of as the proportion of phenotypic variance that can be modified by manipulation of the environment

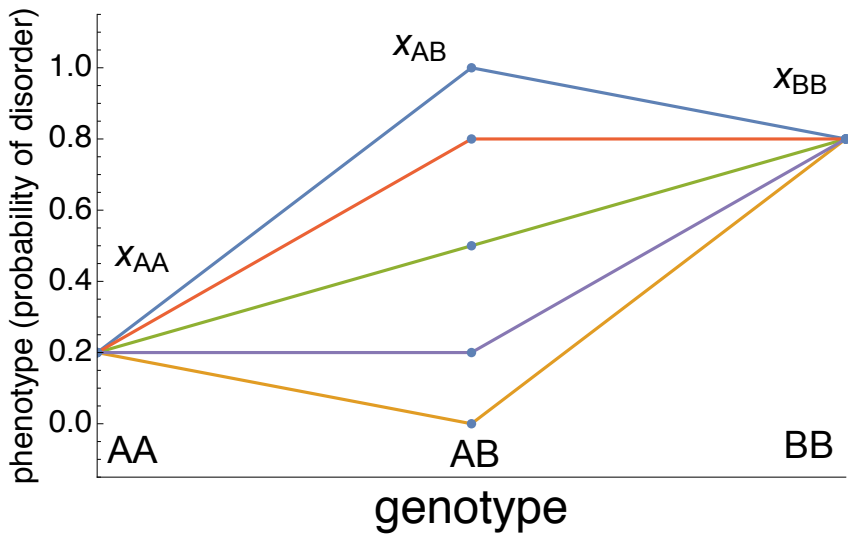
Inheritance of genes from parent to offspring

- If genes matter to determine phenotype we expect offspring to resemble their parents.
- If large parents tend to have large offspring, then the genetic value for "large size" tends to be inherited.
- How can we quantify phenotypic inheritance from parent to offspring?
- This quantification is not given by broad sense heritability where genes can interact and gene interactions cannot be transmitted from parent to offspring.

To answer the question of phenotypic inheritance we introduce a second concept of heritability: the narrow sense heritability and will derive it in the case of a biallelic trait, but the concepts are completely general.

Modes of gene action for a biallelic trait in graphics

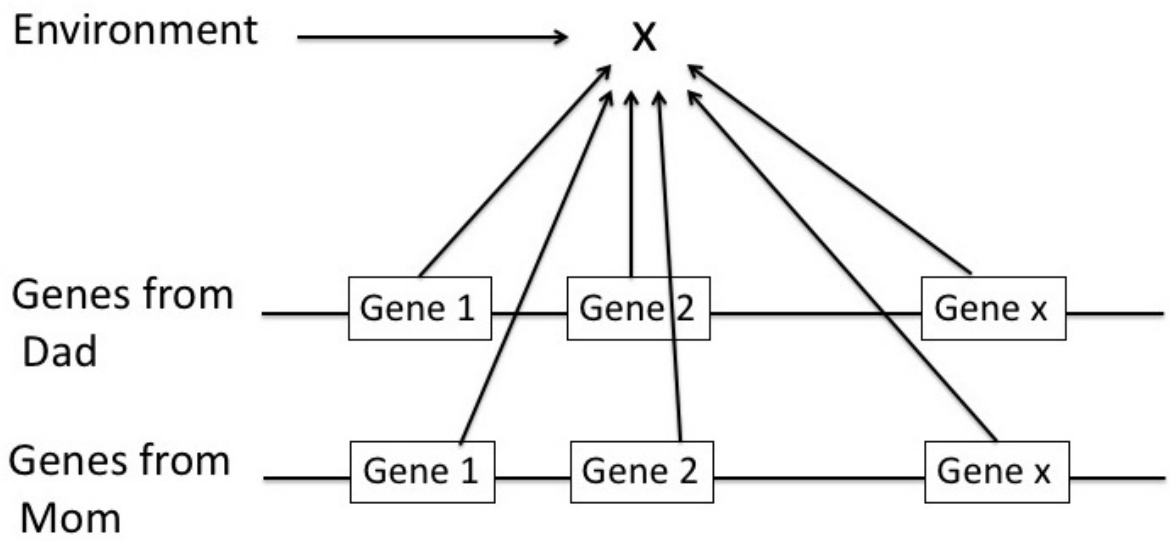
The following relationships between genotype and phenotype can be conceived.



Overdominance (blue), dominance (red), co-dominance (or additive effects) (green), recessivity (purple), and underdominace (yellow).

Except for additive gene action, all other cases involve interactions between allele to determine phenotype.

Decomposition into additive and interaction effects



- Interactions between homologous genes (like Gene 1 form Dad and Gene 1 from Mom) are called dominance effects.
- Interactions between non homologous genes (like Gene 1 form Dad and Gene 2 from Dad or Gene 1 form Dad and Gene 2 from Mom) are called epistatic effects or epistasis.

Decomposition into additive and interaction effects

We saw that we have

$$x_i = g_i + e_i$$

and we now decompose the genetic value of individual i into two terms

$$g_i = a_i + d_i \qquad \text{whereby} \qquad \bar{g} = \bar{a} + \bar{d}$$

- a_i is the additive value due to the separate contribution of each of the genes inherited from the mother and each of the genes inherited from the father. These effects are heritable.
- d_i is the interaction effect due to the interaction between patrigenes, matrigenes, and all cross interactions. These effects are not heritable.

Decomposition into additive and interaction effects

We construct the decomposition

$$g_i = a_i + d_i$$

such that:

$$V_{\mathrm{g}} = \underbrace{V_{\mathrm{a}}}_{\text{maximized}} + \underbrace{V_{\mathrm{d}}}_{\text{minimized}}$$

- V_{a} is the additive genetic variance and we try to maximize the contribution of the alleles (who are inherited) to this variance.
- V_{d} is the variance due to interaction effects.
- d_i is then uncorrelated to a_i ($\mathrm{Cov}[a_i, d_i] = 0$).

How do we do this?

The genetic value decomposition in a two allele system

Suppose that only two alleles segregate in a diploid panmictic population, the usual A and B allele. We write the genetic value of each genotype as

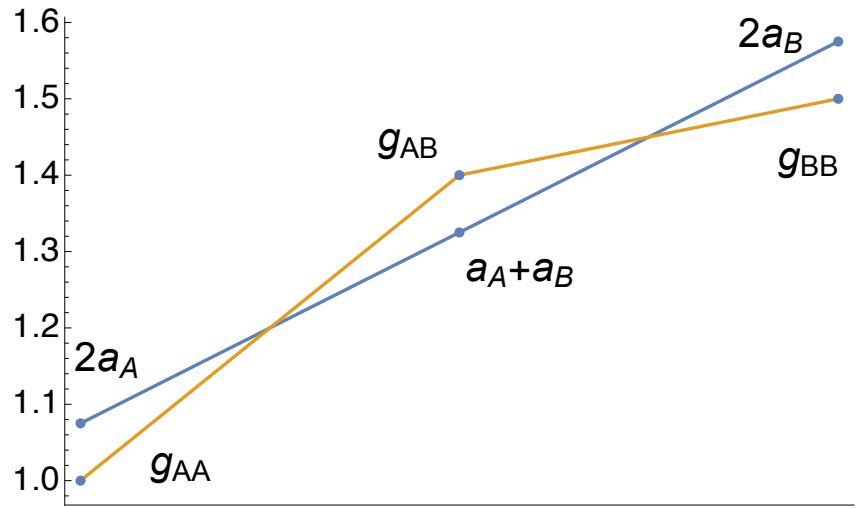
$$g_{\mathrm{AA}} = 2a_{\mathrm{A}} + d_{\mathrm{AA}}, \quad g_{\mathrm{AB}} = a_{\mathrm{A}} + a_{\mathrm{B}} + d_{\mathrm{AB}} \text{ and } g_{\mathrm{BB}} = 2a_{\mathrm{B}} + d_{\mathrm{BB}}$$

- a_{A} is the additive value from carrying one copy of allele A. This is heritable since alleles are transmitted.
- a_{B} is the additive value from carrying one copy of allele B. This is heritable since alleles are transmitted.
- d_j are the dominance deviations resulting from carrying genotype $j \in \{\mathrm{AA}, \mathrm{AB}, \mathrm{BB}\}$. Not heritable.

The a's as least square regression coefficients.

The genetic value decomposition

The idea behind the additive values a_{A} and a_{B} is to be able to describe the genetic values by fitting a straight line between number of A alleles within an individual and the genetic values.



$d_{\mathrm{BB}} = g_{\mathrm{BB}} - 2a_{\mathrm{B}}$, $d_{\mathrm{AB}} = g_{\mathrm{AB}} - (a_{\mathrm{A}} + a_{\mathrm{B}})$ and $d_{\mathrm{AA}} = g_{\mathrm{AA}} - 2a_{\mathrm{A}}$

The additive effects (a's) are chosen to be the linear regression of genetic effects on genetic values and the (d's) are the residuals in the regression.

The genetic value decomposition

Given A frequency p, the genotype frequencies, genetic values, additive genetic values, and dominance values are, respectively,

AA	AB	BB
p^2	$2p(1-p)$	$(1-p)^2$
g_{AA}	g_{AB}	g_{BB}
$2a_{\mathrm{A}}$	$a_{\mathrm{A}} + a_{\mathrm{B}}$	$2a_{\mathrm{B}}$
$d_{\mathrm{AA}} = g_{\mathrm{AA}} - 2a_{\mathrm{A}}$	$d_{\mathrm{AB}} = g_{\mathrm{AB}} - (a_{\mathrm{A}} + a_{\mathrm{B}})$	$d_{\mathrm{BB}} = g_{\mathrm{BB}} - 2a_{\mathrm{B}}$

From this we can evaluate the variances:

$$V_{\mathrm{a}} = (2a_{\mathrm{A}} - \bar{a})^2 p^2 + (a_{\mathrm{A}} + a_{\mathrm{B}} - \bar{a})^2 2p(1-p) + (2a_{\mathrm{B}} - \bar{a})^2 (1-p)^2$$

and likewise for V_{d}.

The mean and variance in the population

We choose the additive values such that they maximize their contribution to the phenotype. This requires that the additive values

- Describe the average genetic value: $\bar{g} = \bar{a}$ so that $\bar{d} = 0$.
- Explain as much as possible of the additive genetic variance in the population:

$$V_{\mathrm{g}} = \underbrace{V_{\mathrm{a}}}_{\text{maximized}} + \underbrace{V_{\mathrm{d}}}_{\text{minimized}}$$

Hence, we minimize the dominance variance by least square fitting (i.e. doing a regression analysis) the additive values a_{A} and a_{B}, which are then determined.[1]

[1] Technically we compute the a_{A} and a_{B} by solving $\frac{\partial V_{\mathrm{d}}}{\partial a_{\mathrm{A}}} = 0$ and $\frac{\partial V_{\mathrm{d}}}{\partial a_{\mathrm{B}}} = 0$ for them; see lecture notes for details.

Narrow sense heritability

We can now define the ratio

$$h^2 = \frac{V_\mathrm{a}}{V_\mathrm{x}} = \frac{V_\mathrm{a}}{\underbrace{V_\mathrm{a} + V_\mathrm{d} + V_\mathrm{e}}_{V_\mathrm{x}}}$$

This is the narrow sense heritability.

- The narrow sense heritability is the proportion of phenotypic variance in the population that is explained by different individuals having different additive genetic effects on phenotype.
- This is the proportion of phenotypic variance that is explained by different allelic effects (which can be inherited), as opposed to different genotypic effects (which cannot be inherited).

Narrow sense heritability

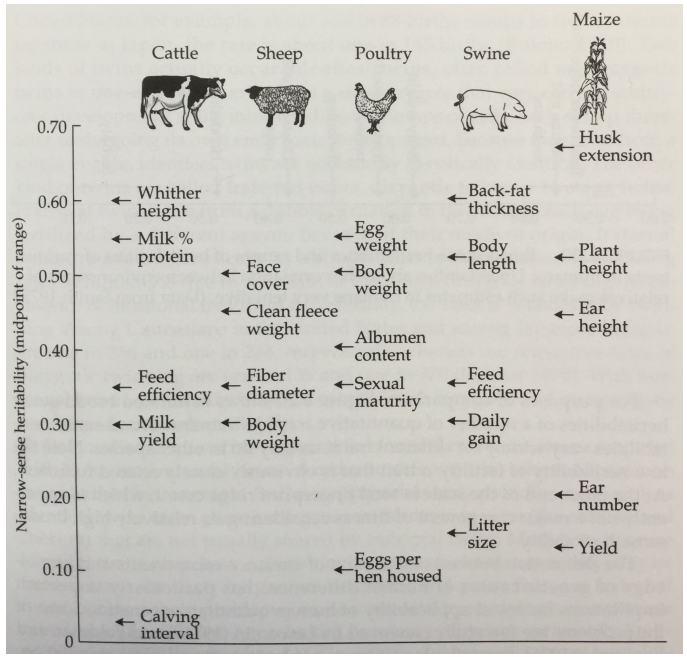
Narrow sense heritability is lower than broad sense heritability

$$h^2 = \frac{V_{\mathrm{a}}}{V_{\mathrm{x}}} \quad \leq \quad H^2 = \frac{V_{\mathrm{g}}}{V_{\mathrm{x}}} = \frac{V_{\mathrm{a}} + V_{\mathrm{d}}}{V_{\mathrm{x}}}$$

- The broad sense heritability takes into account all genetic effects.
- The narrow sense heritability gives the proportion of phenotypic variance that is heritable; namely, heritable from parent to offspring since it is ascribed to single alleles, and only alleles and not genotype are inherited.

Narrow sense heritability in plants and animals

In most domesticated plants and animals the narrow sense heritability $h^2 = V_a/V_x$ usually takes a value between 0.1 for 0.9.



h^2 for behavioral traits is generally lower than h^2 for morphological traits.

Narrow sense heritability in humans

- The narrow sense heritability is one of the most investigated quantity in human genetics.
- Meta-analysis of the heritability of human traits based on fifty years of twin studies[2], reporting variance components for 17×804 traits, from 2 748 publications, including $14 \times 558 \times 903$ partly dependent twin pairs shows that

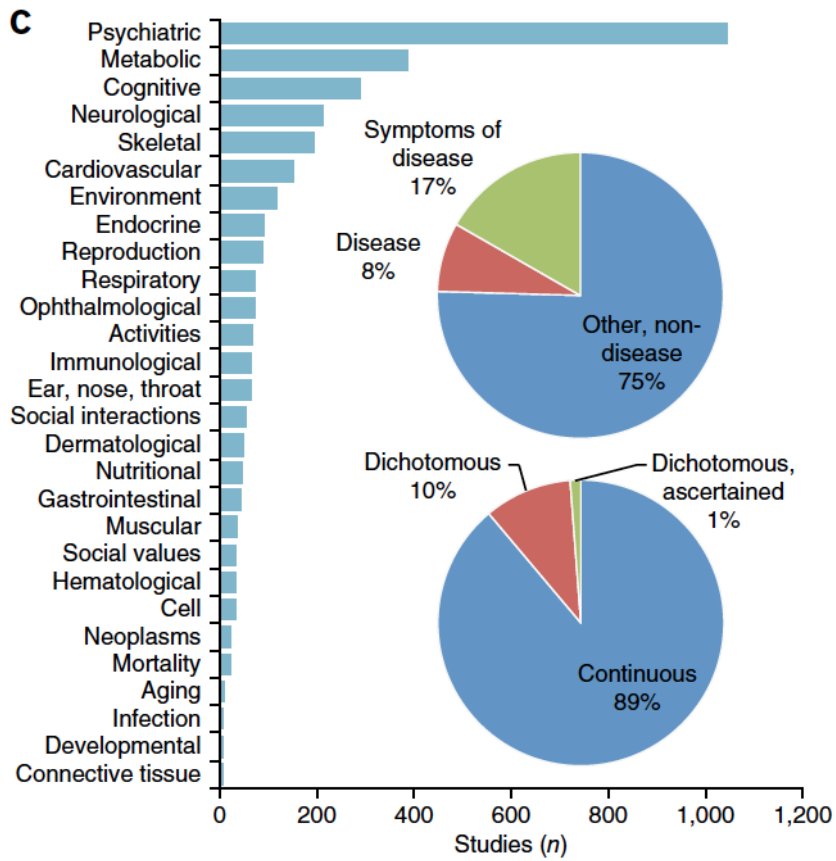
Average narrow sense heritability across traits in humans is 0.5.

(Polderman et al. 2015)

[2] By comparing twins living in same and different environments one can estimate heritability.

Narrow sense heritability in humans

Many different type of traits have been investigated for heritability.



Not a single investigated trait had a narrow sense heritability of zero: all humans traits are heritable. This refutes the concept of the blank slate and tells us that genes matter for essentially any phenotype, be it physiological, morphological, or behavioral.

Can the genes underlying heritability be identified?

- It is of fundamental medical interest to identify the genes and alleles responsible for diseases and neurological disorders underlying the observed heritability of these traits.
- Identifying this is at the heart of medical and behavioral genetics and has become one of the most active area of current research in science.
- This goes under the heading of genomics, which aims at the collective characterization of the effect of the genome[3] on the functioning of the cellular and bodily machinery of organisms.

[3] The genome of an individual is its entire genetic DNA material: genes, non-coding DNA regions, mitochondrial DNA (inherited only from Mom).

Whole genome sequencing

It has now become routine to sequence the whole genome of organisms. The Human Genome Project did this for humans (completed in 2004 and spawned the "Big Data" revolution), and thousands of species have now been fully sequenced.

By comparing the sequences of different individuals, the aim is to identify regions of the genome affecting traits, often the genome is scanned for single nucleotide (or base) polymorphism (SNP).[4]

[4] A SNP is variation in a single nucleotide (a single DNA site) that occurs at a specific position in the genome. This can be thought as a genome position where two alleles segregate, say A and B. At the single base (or nucleotide) level (A, T, G or C), there is always a maximum of two alleles, since A pairs with T and G with C.

The missing heritability problem: small effect size of genes

- The Human Genome Project led to the forecast that the genetic contribution of many diaereses would be easily identified.

- It turned out to be very difficult to actually identify single alleles (be it SNP's) that account for the observed heritability. Hence "a gene for phenotype X" is very difficult to identify.

This leads to the dilemma of the "missing heritability". The solution is that probably most genes have very small effects on the phenotype (infinitesimal model of quantitative genetics), which is thus underlain by myriads of small genetic contributions. This squares well with animal and plant breeding experiments, and the Darwin's view of evolution we investigated in previous lessons (each mutation result in a small phenotypic effect).

Summary on heritability

1. Broad sense heritability H^2 gives the proportion of phenotypic variance in a population that is explained by genes. $1 - H^2$ can be thought of as the proportion of phenotypic variance that can be modified by manipulation of the environment.

2. Narrow sense heritability h^2 gives the proportion of phenotypic variance in a population that is explained by additive genetic effects. It can be thought of as the proportion of phenotypic variance that is heritable and can be transmitted across generations.

Both concepts of heritability (h^2 and H^2) refer to the population and explain population variances. They are not indices of the importance of any individual's alleles in determining its phenotype, which is an open question and a very active area of modern science.