Unravelling nature's blueprint: Understanding Heritability and genetic influence in plant and animal breeding

Fikret Isik
May 17, 2023

Narrow-sense Heritability

Imagine you walk into a field trial of trees. The age of the trees is the same, let's say five years from planting. You notice that trees have different heights. You may wonder what the causes of variation are. Obviously, soil fertility, sunlight, and water availability vary from one corner of the field to another, no matter how much we carefully prepare the site. Pests and pathogens affect tree growth, and their effects are not uniform across the site. We lump all these and other unknown factors as the "environment" effect causing variation in height.

The other important factor is the different genetic makeup of trees, even if they share the same parents—many genes with small (additive) effects contribute to the variation we observe for the trait. Heritability is a measure to understand the relative effects of the environment and additive effects of genes on a trait like height.

Heritability plays a major role in quantitative genetics and in plant and animal breeding. We use the h^2 symbol for heritability. Heritability is defined as a proportion of total genetic variance that is attributable to the differences between the breeding values (additive effects of genes) of individuals. Breeding values and/or additive effects represent the heritable genetic effects that parents pass on to their progenies when mated at random to their population. When we talk about heritability, we usually refer to **narrow-sense heritability**. It is simply a ratio of additive genetic variance, σ_A^2, to total phenotypic variance, σ_P^2.

You may wonder why the h^2 symbol is used for heritability. This is because Sewall Wright, one of the founders of quantitative genetics, used h as the correlation between the phenotype and breeding values. To define the percent of genetic variation for the trait, he used h^2.

Narrow-sense heritability can be used to predict the genetic gain from selecting individuals based on their phenotypes in plant and animal breeding programs. Imagine we select a subset of the trees for height in the above experiment. Their mean is $\bar{x}$. The deviation of selected individuals from the population mean μ is:

$$S = \bar{x} - \mu$$

The S is the selection differential. Suppose we crossed the selected trees to produce a new population. Narrow-sense heritability tells us about the expected improvement in the trait mean (response to selection). The new population mean (u^*) is predicted to be:

$$u^* = \mu + h^2(\bar{x} - \mu) = \mu + \mathbf{h^2 S}$$

The change ($h^2 S$) compared to the original population mean (μ) is caused by the additive effects of genes.

Heritability is a Squared Correlation

The square root of the heritability is a correlation between the breeding values and the phenotype. As indicated earlier, the square root of the heritability is a correlation between the breeding values and the phenotype.

Let $\mathbf{A}$ be the additive effects of genes (breeding values), and let P is the phenotype, which is the sum of additive genetic ($\mathbf{A}$), dominance genetic deviations (D), and the environment (E) effects:

$$P = \mathbf{A} + D + E$$

The correlation coefficient is a scaled covariance between two traits. To obtain the correlation coefficient, We divide the covariance by the square root of the multiplications of variances (or by standard deviations). So, the correlation between breeding values $\mathbf{A}$ and the phenotypes P is

$$r_{AP} = \frac{Cov(\mathbf{A}, P)}{\sqrt{\sigma_A^2 \sigma_P^2}} = \frac{Cov(\mathbf{A}, P)}{\sigma_A \sigma_P}$$

Let's plug the phenotypic variance ($P = \mathbf{A} + D + E$) into the equation above.

$$r_{AP} = \frac{Cov(\mathbf{A}, A + D + E)}{\sigma_A \sigma_P}$$

The covariance of the breeding values ($\mathbf{A}$), with each element of phenotypic variance is

$$Cov(\mathbf{A}, A + D + E) = \textcolor{red}{Cov(\mathbf{A}, A)} + Cov(\mathbf{A}, D) + Cov(\mathbf{A}, E)$$

The covariances $Cov(A, D)$ and $Cov(A, E)$ are zero because A is independent of D and E (if you have done a properly randomized experiment). This leaves only $Cov(A, A)$.

By definition, the covariance of x with itself is the variance of x. $Cov(A, A) = \sigma_A^2$

so the correlation is

$$r_{AP} = \frac{Cov(A, A) + 0 + 0}{\sigma_A \sigma_P} = \frac{\sigma_A^2}{\sigma_A \sigma_P}$$

We divide the variance of breeding values (σ_A^2) by the standard deviations of breeding values (σ_A), and the phenotypes (σ_P). With simplification, the correlation coefficient becomes a square root of heritability.

$$r_{AP} = \frac{\sigma_A^2}{\sigma_A \sigma_P} = \frac{\sigma_A \sigma_A}{\sigma_A \sigma_P} = \frac{\sigma_A}{\sigma_P} = h$$

The variance (σ_A^2) is the squared standard deviation (σ_A). If the heritability of a trait is high, the correlation between the phenotypes and the breeding values will be high, and our confidence in selecting the winners will be high.

Properties of Heritability

The heritability of a trait is **strictly population-specific**. Further, it may be specific to the experimental conditions! For example, the heritability of a particular trait can vary when measured at a different age or developmental stage.

Heritability is environment specific. Heritability estimated in one environment might be very different from an estimate for the same population grown in a different environment. Genotype-by-environment interactions can affect heritability. Heritability estimates from a single environment are likely biased (overestimated) because of shared GxE interactions. In this case, the numerator of the heritability is the sum of the additive genetic and genotype by environment interaction variances.

$$h^2 = (\sigma_A^2 + \sigma_{GxE}^2) / \sigma_P^2$$

However, we can account for GxE when calculating heritability if the experiment has been conducted in multiple environments (See chapter 8 of Isik, Holland and Maltecca 2017 for examples).

A zero heritability does not mean genes do not control the trait. In fact, the trait might be controlled 100% by genetics (fixed). It means there is no genetic variation between the evaluated individuals for the given trait. For example, if you walk into a spruce progeny test and assess the stem straightness, you may not be able to see any variation between trees because they all would have straight stems. In fact, stem straightness of spruces is very likely fixed (totally controlled by genetics) during the evolutionary process, and there is no variation to select for. Spruce species are naturally found in snowy environments of the northern hemisphere. The snow may have worked against the trees with crooked stems or large crowns because they would keep more snow on the crowns and break before they reach maturation to reproduce. **A perfect heritability** suggests that the phenotype is a perfect predictor of breeding values, which can happen for traits with simple Mendelian control.

Can we change the heritability? The short answer is yes. Since heritability is a proportion of the variance attributed to the genetic effects, we can increase heritability by using better experimental designs and better management of the trials (this is by reducing the error variance). We can have a genetically more diverse population (this is by increasing the genetic variance), which in turn can increase heritability. Heritability can decrease if the population is genetically more uniform or if we do a poor job when establishing and maintaining the field experiment.

Heritability is not a good estimator of response to selection beyond a few generations. This is because selection changes the allele frequencies (genetic variance) from one generation to the other and therefore changes the heritability. A heritability estimate for a population might be valid for a few generations. The heritability may increase for a population because of the increase of rare favorable alleles. Selection may decrease the heritability for another population because of fixing alleles or genetic drift.

Other Types of Heritability

In plant breeding, **broad-sense heritability** might be of interest. Traditionally H^2 is used for broad-sense heritability. It is the ratio of additive plus non-additive genetic variance to the phenotypic variance. Non-additive genetic variance includes both dominance and epistatic genetic variances.

$$H^2 = \frac{\sigma_A^2 + \sigma_D^2}{\sigma_A^2 + \sigma_D^2 + \sigma_E^2}$$

Broad-sense heritability is used to predict the response to selection when we have clones. The variance associated with the genotype effect is partitioned into additive and non-additive genetic effects.

You may have also heard about the **repeatability** of a trait. It is a ratio of the phenotypic variance explained by differences in repeated measures of the same individuals. For example, measuring the growth of trees over a certain number of years or milk yield over the lifetime of cows. The estimation for animals requires accounting for the permanent environmental effects. So the phenotypes measured repeatedly on an individual includes genetics, permanent environment, and residual effects. In plant breeding, measuring multiple genetically identical copies of cloned individuals is similar to measuring the same individuals over time.

The selection unit might be parents (sires in cattle breeding) or full-sib families. In this case, were are interested in exploiting the genetic variation among the parents. The appropriate heritability to calculate the response to the selection of parents is the **heritability of means**. The formula for the heritability of means can be quite different for different mating and field designs. This is especially true for calculating the phenotypic variance of means (the denominator of the formula). This is illustrated below.

Suppose we have simple mating and experimental design. The between groups (e.g., families) and within groups (i.e., residual) are the only two variance components. Suppose each family has n number of individuals (offspring) in the experiment. The expected mean of a family is:

$$\bar{y}_. = \mu + f_i + w_.$$

Hence, the variance of family means is

$$Var(\bar{y}_.) = Var(\mu + f_i + w_.) = \sigma_f^2 + \frac{1}{n}\sigma_w^2$$

The heritability of family means is

$$h_{\bar{f}}^2 = \frac{\sigma_f^2}{\sigma_f^2 + \frac{\sigma_w^2}{n}}$$

where σ_f^2 is the variance of family means, σ_w^2 is the residual variance averaged over the number of offspring per parent in the data. The heritability of means is almost always higher than the narrow-sense heritability. This is because we average the residual variance using the average number of offspring per parent. Here, we are dealing with the variance of means, not individual observations, and the means are more reliable than individual data points (the environment has a smaller effect).

Conclusions

Heritability is a crucial concept in genetics that measures the extent to which genetic factors contribute to the variation in a trait within a population. It represents the proportion of phenotypic variation that can be attributed to genetic differences. However, it's important to consider the specific definition, calculation method, and population under study when interpreting heritability estimates. Heritability is a population-specific measure and cannot be directly applied to individuals. It does not identify specific genes or their individual contributions, and it does not imply that a trait is solely determined by genetics or unchangeable. Environmental factors and gene-environment interactions also play significant roles. Therefore, understanding heritability requires careful consideration of its definition, calculation method, and the population context in which it is being applied.

About the author

Dr. Fikret Isik is a quantitative geneticist at North Carolina State University, Department of Forestry and Environmental Resources. He serves as the co-director of the Cooperative Tree Improvement Program and the president of the North America Forest Genetics Society.

Currently, Dr. Isik is a Fulbright visiting professor at the University of Pretoria, South Africa, where he teaches courses on quantitative genetics and plant molecular breeding.

Dr. Isik has more than 25 years of experience in genetic data analysis for breeding and forest tree breeding strategies. He is the co-author of 86 manuscripts, three book chapters, and the textbook "Genetic Data Analysis for Plant and Animal Breeding". Dr. Isik's research focuses on developing genotyping platforms for forest trees, genomic selection, molecular tree breeding strategies, and plant-biotic interactions, specifically the genetic basis of fungal disease resistance in forest trees. Email: fisik@ncsu.edu

Acknowledgment

I am thankful to Nanette Christie (University of Pretoria), James Holland (North Carolina State University), and Carey Langley (VSNi) for reading the draft of this blog and providing comments to improve it.