

## 2.2 Pedigree-based vs. SNP-based heritability

Traditional heritability estimation relies primarily on pedigree information, constructing additive genetic covariance matrices among individuals based on kinship coefficients or identity-by-descent (IBD), and decomposing phenotypic variance within a linear mixed model framework (Vinkhuyzen et al., 2013; Bérénos et al., 2014). Such approaches have long played an important role in animal and plant breeding as well as in studies of natural populations. However, their estimation accuracy depends critically on the completeness and correctness of pedigree records. When shared environmental effects among related individuals are not adequately modeled, pedigree-based heritability estimates may be systematically upward biased.

With the development of high-throughput genotyping technologies and statistical genetic methods, genotype-based heritability estimation has emerged as an important complement to pedigree-based approaches. Methods represented by GCTA/GREML construct a genome-wide relationship matrix (GRM) from SNP data and estimate the additive genetic variance captured by markers within a restricted maximum likelihood (REML) framework (Speed et al., 2012; Evans et al., 2017; Yang et al., 2017).

It is important to emphasize that SNP-based heritability is not directly equivalent to the “true” heritability of a trait. Rather, it reflects the proportion of additive genetic variance that can be captured by a given set of markers under specific statistical model assumptions. Such estimates are typically obtained from samples with close relatives removed, in order to reduce confounding effects arising from shared environment and pedigree structure (Srivastava et al., 2023; Zimmermann and Distl, 2023). Therefore, differences between pedigree-based and SNP-based heritability do not necessarily imply “missing” genetic information, but more likely arise from differences in estimands, marker coverage, and modeling assumptions.

## 2.3 Sources of discrepancy and “missing heritability”

In numerous studies of complex traits, heritability estimates based on pedigree data are often higher than those derived from SNP-based approaches, giving rise to the so-called problem of “missing heritability” (Vinkhuyzen et al., 2013; Yang et al., 2017). From a statistical genetics perspective, this discrepancy should not be interpreted simply as a true loss of genetic information, but rather as a systematic difference arising from distinct estimands, data coverage, and modeling assumptions.

First, limited marker coverage is a major source of lower SNP-based heritability. Conventional genotyping arrays primarily capture common variants, while providing limited representation of rare variants, low-frequency variants, and structural variants. As a result, part of the genetic variance remains untagged, leading to downward-biased SNP heritability estimates (Wainshtein et al., 2019; Jang et al., 2022; Wainshtein et al., 2022). Recent analyses based on whole-genome sequencing data have demonstrated that rare variants can explain a portion of the previously “missing” heritability, further supporting this explanation.

Second, incomplete linkage disequilibrium (LD) constrains the ability of markers to capture the effects of causal variants. Even with high-density SNP data, LD between markers and true causal loci is often insufficient to fully reflect effect sizes, resulting in systematic underestimation of additive genetic variance (Speed et al., 2012; 2016; Evans et al., 2017). This issue is particularly pronounced in populations with complex LD structures or highly heterogeneous allele frequency distributions.

Third, confounding due to shared environmental effects may inflate pedigree-based heritability estimates. In pedigree studies, related individuals often share both genetic background and environmental conditions. If the model fails to adequately disentangle these contributions, environmental correlations may be incorrectly attributed to genetic variance, thereby inflating heritability estimates (Vinkhuyzen et al., 2013; Bérénos et al., 2014). In contrast, SNP-based methods are typically applied to samples with close relatives removed, reducing such confounding.

In addition, non-additive genetic effects and gene-environment interactions can further widen the gap between pedigree-based and SNP-based heritability estimates. Narrow-sense heritability and most SNP-based frameworks