# On the Genetic Origins of Phenotypes in Genome-Wide Association Studies: The SNP Allocation For Estimating Heritability (SAFE-$h^2$) Tool to Explore Additive-only Allelic Effects or Additive and Non-Additive Allelic Effects

Behrooz Darbani[1,2] and Mogens Nicolaisen[1]

[1]Department of Agroecology, Faculty of Technical Sciences, Aarhus University, DK-4200 Slagelse, Denmark
[2]Current affiliation: Molecular Diagnostic and Clinical Research Unit, University Hospital of Southern Denmark, 6200 Aabenraa, Denmark

April 10, 2025

Corresponding authors emails: behroozdarbani@gmail.com / bds@rsyd.dk; mn@agro.au.dk

**Keywords:** GWAS, SNP, heritability, additive effects, non-additive effects, false-positive hits

**Abstract**

Polygenicity requires the inclusion of all SNPs to ensure accurate estimation of SNP heritability. However, SNPs with large association $p$-values should be excluded, as they impose biologically meaningless negative contributions, *i.e.*, downward bias, to heritability estimates. This is demonstrated with simulated data and validated using real-world datasets. We introduce SAFE-$h^2$ as a solution to this dilemma. SAFE-$h^2$ performs heritability profiling and establishes a $p$-value threshold to filter out SNPs with negative contributions. By examining 74 phenotypes across eight species, SAFE-$h^2$ revealed an average negative contribution of 30 units, imposed by SNPs from the upper-bounds of $p$-value thresholds. SAFE-$h^2$ also provides a safe $p$-value interval to minimize false-positive SNP hits and facilitates capturing intra-locus ($Ref$-allele$_{\mathrm{SNPi}}$ $\times$ $Ref$-allele$_{\mathrm{SNPi}}$, $Ref$-allele$_{\mathrm{SNPi}}$ $\times$ $Alt$-allele$_{\mathrm{SNPi}}$, and $Alt$-allele$_{\mathrm{SNPi}}$ $\times$ $Alt$-allele$_{\mathrm{SNPi}}$) additive and non-additive effects within the framework of linear models. The allelic adjustment algorithm of SAFE-$h^2$ revealed considerable uncaptured phenotypic variance when relying on additive effects and improved the SNP heritability estimations by up to 56 units across 50 phenotypes in six species through the combined capture of both additive and non-additive allelic effects.

**Introduction**

Single nucleotide polymorphism (SNP) heritability is a key measure in genome-wide association studies (GWAS), reflecting the experimental power for exploitation of phenotypic variance. Estimating heritability requires both genetic variance ($\boldsymbol{\sigma_g^2}$) and phenotypic, *i.e.*, total, variance ($\boldsymbol{\sigma_p^2}$), with $\boldsymbol{\sigma_g^2} \leq \boldsymbol{\sigma_p^2}$ and heritability ($\boldsymbol{\sigma_g^2} \div \boldsymbol{\sigma_p^2}$) values ranging from 0 to 1 (0–100%). Compared to broad sense heritability ($H^2$) estimates, SNP heritability is typically estimated on additive effects and is thus considered as narrow-sense heritability[1,2]. Nevertheless, SNP heritability provides valuable insights into the genetic architecture of traits and has significant applications in plant and animal breeding, as well as medicine. SNP heritability is a critical metric not only for guiding future research and calculating response to genomic-selection but also for assessing the genetic influence on the trait-of-interest in genome-wide association studies, as reporting significant SNP associations alone is insufficient. Therefore, accurate estimation of SNP heritability is

crucial, as both underestimation and overestimation can lead to substantial genetic misinterpretations. For instance, a significant underestimation of SNP heritability for a trait, *i.e.*, estimated $h^2{}_{\text{SNP}} \approx 10\%$–$30\%$ while actual $h^2{}_{\text{SNP}} \approx 70\%$–$90\%$, can shift the future focus from genetic research toward environmental factors (*e.g.*, diet, smoking, humidity, and temperature) and vice versa.

Phenotypic variance consists of two main components: genetic and environmental factors. Environmental physiochemical interaction with organisms and their influences on traits are well addressed aspects[3–5]. When phenotypic variation is observed for genotypes (individuals) within a single, specific environment, the study is considered a single-environment study. However, if observations are made across multiple environments, the study is classified as a multi-environment study, which is most relevant to animal and plant genetic research. In contrast, this concept does not apply to humans due to the lack of genotypic replication and controlled lifestyle or living conditions; individuals have their own unique lifestyle and environment. Similar to the inclusion of new genotypes, incorporating additional environments into a multi-environment study can alter the variance and heritability levels. Consistent with this, heritability shifts have been observed across different experimental conditions such as age[6]. In single-environment experiments, genetic variance is expected to account for the majority of phenotypic variance, approaching 100% heritability, with any deviations typically attributed to sampling errors. However, SNP heritability estimates in single-environment experiments are exceedingly underestimated. Experimental challenges that can affect heritability estimates include inadequate genotypic coverage, partial genomic coverage, imprecise phenotypic scoring scales, assignment of unassociated SNPs to estimation models, exclusion of associated SNPs from estimation models, neglecting epistatic interactions, and ignoring non-additive allelic effects. For example, approximately $3^{100,000}$ individuals with different genotypes are required for precise genetic estimations in a diploid genome with 100,000 biallelic SNPs, given that mutations are independent and conditional lethal SNPs are exceedingly rare. Various models have been developed to estimate SNP heritability, for example, using the minor allele frequency (MAF) and the linkage disequilibrium based weighted SNP effects[2,7–9]. Despite genotyping by millions of variants, SNP heritability still captures 30%–50% less phenotypic variance compared to heritability estimates derived from family and twin studies[1,9–19]. This underestimation is even more evident when compared to the much higher levels of heritability estimates, typically expected in single-environment studies that utilize millions of SNPs. Approaching total genomic and genotypic coverage, along with more robust scoring systems, such as subclasses for qualitative traits, can help bridge the gap. Capturing epistatic variance[20] and controlling population structure[21] have also been introduced to reduce the missing heritability levels. However, these strategies alone may not sufficiently close the gap unless SNPs are properly allocated to estimation models that account for both additive and non-additive effects between reference (*Ref*) and alternative (*Alt*) alleles.

Polymorphic genomic positions, should either be neutral (such as unassociated SNPs) or contribute positively (such as positive and negative associated SNPs) when included in models estimating SNP heritability. However, this study shed light on the negative contributions to heritability estimates imposed by SNPs with large association $p$-values through the commonly used genome-wide relatedness matrices. Family-based heritability estimations employ the true population statistic of genetic similarities (Fig. 1), *e.g.*, monozygotic twins are genetically identical by descent and fraternal twins share, on average, 50% of their genomes[22]. In contrast, SNP heritability calculations rely on genome-wide estimates of genetic similarities, which are not phenotype-specific and are therefore susceptible to substantial error, as each phenotype is governed by a small fraction of the genome (Fig. 1). In Figure 1, we present an example where the addition of one unassociated SNP increases the genetic similarity from 50% (obtained based on an associated SNP) to 62.5%, introducing a downward bias in the phenotype-specific genetic variance. Despite the identification of this common miscalculation over a decade ago[23,24], the majority of studies have continued to rely on the relatedness matrices derived genome-wide using all SNPs. It is evident that unassociated SNPs, *i.e.*, SNPs from the upper-bound of a dataset specific $p$-value threshold, result in overestimation of genetic similarities (Fig. 1), ultimately leading to the underestimation of SNP heritability. Consistent with this, studies report remarkably low SNP heritability estimates (Average = 9.6%, Median = 6.6%, Fig. 2a). Accepting these estimates would confirm that many of the traits studied are purely environmental characteristics rather than genuine traits governed by genomes. This, of course, cannot be true and clearly indicates a downward bias in

genetic variance estimates. Relying on a predefined $p$-values threshold such as $10^{-8}$ is also prone to underestimations of heritability given that polygenicity is a common phenomenon and many relevant variants may fall above this arbitrary cutoff. These technical challenges have led to the widespread practice of using either all SNPs or only those with $p$-value $\leq 10^{-8}$, depending on which approach yields the highest heritability estimate in every study. One of the remarkable examples is the heritability estimates for human height[15,16]. Furthermore, false-positive SNP associations and the neglect of non-additive genetic effects can lead to overestimations and underestimations of heritability, respectively. This study is accordingly focused on excluding unassociated SNPs with negative contributions to heritability, minimizing false-positive SNP hits, and finally capturing both additive and non-additive effects between *Ref* and *Alt* alleles. We introduce SAFE-$h^2$ (SNP Allocation For Estimating Heritability) method and application tool (Supplementary Fig. 1) that performs a $p$-value based SNP and heritability profiling to identify a dataset-specific $p$-values threshold, preventing biologically meaningless negative contributions to SNP heritability estimates. SAFE-$h^2$ also defines a safe $p$-value interval to mitigate the heritability contributions of false-positive SNP hits. Furthermore, SAFE-$h^2$ introduces an allelic adjustment strategy based on the observed genetic effects for every SNP to jointly capture intra-locus non-additive and additive allelic effects (*Ref*-allele$_{\text{SNPi}}$ $\times$ *Ref*-allele$_{\text{SNPi}}$, *Ref*-allele$_{\text{SNPi}}$ $\times$ *Alt*-allele$_{\text{SNPi}}$, and *Alt*-allele$_{\text{SNPi}}$ $\times$ *Alt*-allele$_{\text{SNPi}}$), enhancing the precision of SNP heritability estimates. SAFE-$h^2$ can leverage heritability estimation models used in EMMAX[25], LDAK-GCTA, LDAK-Thin[7], GCTA-GREML[26], and GEMMA[27].

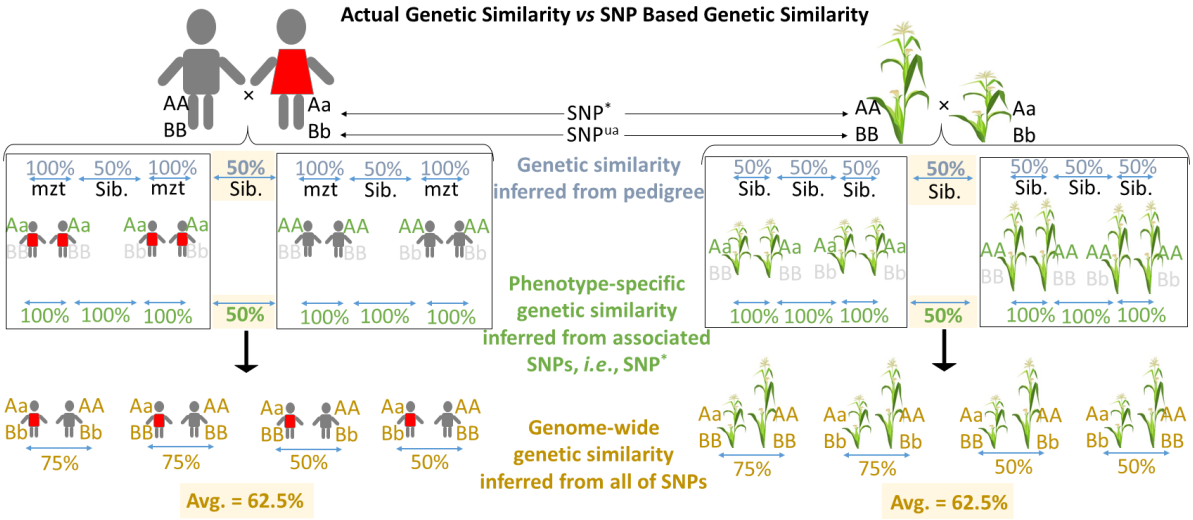



Figure 1: **Unassociated SNPs contribute to a downward bias in genetic variance estimates.** Among genotypes which are differentiated based on a given phenotype-of-interest, family-based studies apply actual genetic similarities which are directly reflected in the genetic similarities inferred from associated SNPs (SNP$^*$). However, genome-wide genetic similarities are overestimated when they are based on all SNPs, *i.e.*, associated (SNP$^*$) and unassociated (SNP$^{\text{ua}}$). The mzt and Sib refer to monozygotic twins and siblings, respectively.

## Results

### *Within an array of association p-values, SNPs with large p-values neutralize the contribution of SNPs with small p-values to heritability estimates*

The meaningless negative contributions to heritability depend on levels of association $p$-values in relation to each other. Therefore, we simulated phenotypes as random integers to generate arrays of association $p$-values, rather than representing true associations, for illustration of negative contributions. Analyses were performed on 1,000 and 10,000 simulated genotypes with 100, 200k, and 2,000k SNPs (MAF = 1.6%–10%)

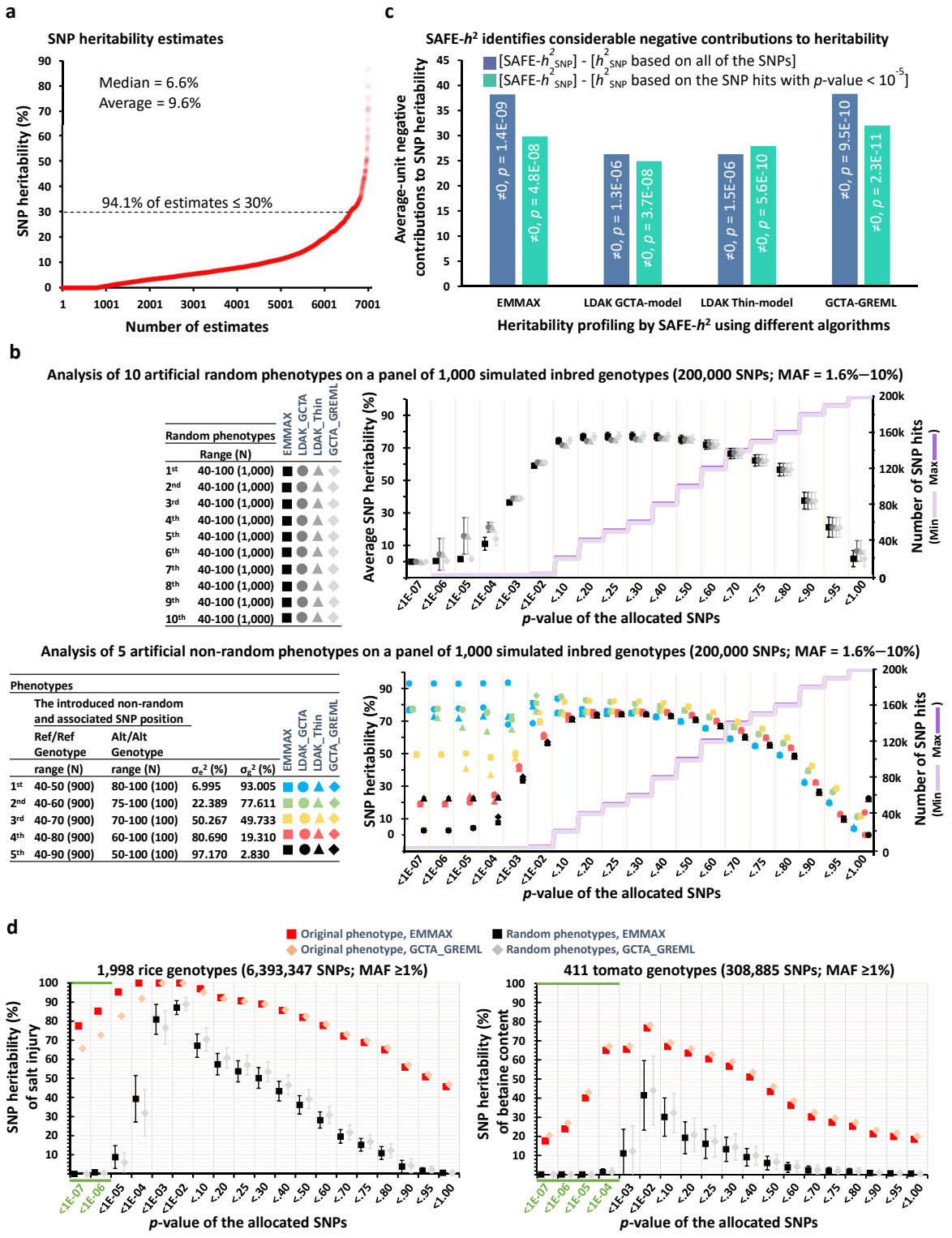



Figure 2: **SAFE-$h^2$ ensures accurate SNP heritability estimates by discarding meaningless negative contributions and restricting false-positive inputs.** a) SNP heritability estimates were obtained from 110 recently published studies (see Supplementary Table 1), encompassing 7,008 independent estimates across various traits. b) EMMAX, LDAK-GCTA, LDAK-Thin, and GCTA-GREML were applied

Figure 2 legend continues: to estimate SNP heritability for the clusters of SNPs identified by their association $p$-values. By ensuring additive-only effects, five different levels of associations were simulated for a predesignated SNP position and as a result, the SNP heritability levels were predictable at or above the predefined levels, given that other SNPs could also, by chance, account for additional phenotypic variance. N represents the number of genotypes. The error bars represent 99% confidence intervals. c) SAFE-$h^2$ improves the heritability estimations for 14 and 10 phenotypes in *Arabidopsis* and rice, respectively, by preventing the negative contributions of weakly-associated SNPs. The $p$-values are from two-tailed paired t-tests. d) SAFE-$h^2$ minimizes false-positive contributions. The safe-intervals for $p$-value are highlighted in green representing the area with minimum false-positive-SNPs contribution to heritability. The average and standard deviation of estimations are illustrated for six random phenotypes simulated within the range of original phenotype. b–d) GCTA-GREML was performed based on the "inbred" algorithm. MAF: minor allele frequency, ${\sigma_g}^2$: inter-genotype variances, ${\sigma_e}^2$: intra-genotype variations.

in linkage equilibrium as well as the human HapMap phase 1 genomic data[28] with 1,092 individuals and 177,675 SNPs (MAF $\geq$ 1%) on chromosome 22. Analyses *via* four different models of EMMAX[25], LDAK-GCTA, LDAK-Thin[7], and GCTA-GREML[26] revealed a consistent pattern of negative contribution to heritability estimates imposed by the SNPs having large association $p$-values (Fig. 2b, Supplementary Fig. 2a,c and 3). These analyses were replicated after incorporating 1, 12, 24, or 48 predesignated SNPs, each explaining 2.8% to 94.3% of the phenotypic variance through additive effects (Fig. 2b, Supplementary Fig. 2b,d and 4). It is virtually impossible to find multiple SNPs in linkage equilibrium and with large contributions to phenotypic variance and $p$ values $< 10^{-20}$ in real-world experiments with 200k SNPs and 1,000 genotypes. Nevertheless, there were 400 to 39,000 other SNPs with random associations and positive contributions to the heritability estimates among the 200k examined SNPs. Biologically implausible negative contributions were again observed for SNPs with large association $p$-values, leading to substantial underestimations of SNP heritability across all scenarios (Fig. 2b, Supplementary Fig. 2b,d and 4).

We observed the same pattern of negative contributions to heritability in linkage disequilibrium simulations, *i.e.*, 200k SNPs on a 200kb chromosome (MAF = 1.6%–10%, Supplementary Fig. 5). Although, HapMap SNPs had MAFs up to 50%, we further simulated genotypes with MAFs ranging from 1.6%–50% and found the same patterns of negative contributions (Supplementary Fig. 6). Notably, LDAK-GCTA and LDAK-Thin models failed to accurately estimate the predefined heritability levels of 93.4%, 93%, 3.7%, and 2.8% when the estimations were solely based on the predesignated SNPs (*i.e.*, $p$ values $< 10^{-7}$, Fig. 2b, Supplementary Fig. 2b,d and 4–6). In contrast, EMMAX and GCTA provided more stable estimates, yielding the closest approximations.

In any conceivable scientific experiment, replication serves to reaffirm the accuracy of the estimated statistics. When this confirmation cannot be achieved, it suggests an ineffective and flawed sampling/replication strategy. In other words, it indicates that the samples do not accurately represent the population from which they were drawn, as seen when comparing two datasets, both simulated independently from scratch. Therefore, it is important to note that performing replicated analysis—i.e., obtaining $p$-values from one simulated dataset and applying them to a second simulated dataset—is not a valid approach (see Supplementary Text 1). This is because adding a specific level of noise to the second dataset will result in the same estimates, plus or minus the introduced noise, which does not provide any additional information beyond what is already presented in Figure 2b and Supplementary Figures 2-6. Furthermore, $p$-values from one simulated dataset cannot be applied to another dataset that is independently simulated from scratch, as this would involve using $p$-values derived for one phenotype (e.g., weight) to estimate heritability for a different phenotype (e.g., a specific disease).

### ***SAFE-$h^2$ effectively prevents the illogical negative contributions of SNPs to heritability estimations***

To ensure accurate estimation of SNP heritability, we developed SAFE-$h^2$ to perform SNP heritability profiling across $p$-value-guided SNP clusters, identifying an optimal $p$-value threshold for excluding unassociated

SNPs with negative contributions. As a model plant, different traits have been studied through GWAS in *Arabidopsis*[3,29]. Here, heritability profiling was conducted for 14 phenotypes in *Arabidopsis* using 2,621k to 2,840k SNPs[30–32] as well as 10 phenotypes in rice using 6,131k to 6,414k SNPs[33–35]. SAFE-$h^2$ improved the accuracy of heritability estimates by excluding negative contributions. Notably, SNPs with negative contributions masked the heritability estimates by an average of 32.3 units (5.6–64.7, Fig. 2c, Supplementary Table 2). Heritability levels were also underestimated by an average of 28.7 units (13.6–63.7) when based solely on SNPs with $p$-values $\leq 10^{-5}$ (Fig. 2c, Supplementary Table 2). We used SAFE-$h^2$ to analyze 50 additional phenotypes from three plant and three animal species[36–44] and found SNPs with large association $p$-values imposing an average negative contribution of 30 units to the heritability estimates (Supplementary Table 3). SAFE-$h^2$ also outperformed BayesR estimator[45] in the simulated (linkage equilibrium and linkage disequilibrium scenarios) and the real datasets. BayesR either failed to detect or underestimated the genetic variances, while being 8–10 times slower (Supplementary Tables 4–6).

***SAFE-h2 reports a safe p-value interval to minimize positive contributions of false-positive SNP hits to heritability estimates***

False-positives are particularly challenging not only because of the resource-intensive nature of GWAS but also due to the impracticality of near-complete ($\approx 2^{(\text{no. SNPs})}$ or $\approx 3^{(\text{no. SNPs})}$ in the haploid or the diploid state, under a biallelic scenario) genotypic coverage. SAFE-$h^2$ uses simulated phenotypes to pinpoint a safe $p$-value interval with minimum SNP heritability level at the lower-tail of heritability profile. For example, we can set the $p$-value interval from 0 up to a maximum $p$-value, determined at a user selected noise level, e.g., 0% or 5%. The heritability noise levels are obtained by analyzing simulated phenotypes within the range of the phenotype-of-interest . The SNP heritability for the original phenotype is subsequently estimated at the upper-end of this interval to minimize false-positive contributions. Heritability estimates using all SNPs were consistently underestimated not only when compared to the SAFE-$h^2$ heritability estimates with excluded negative contributions but also when compared to the SAFE-$h^2$ heritability estimates with excluded negative and false-positive contributions (Fig. 2d, Supplementary Fig. 7).

***SNP heritability estimation through joint-analysis of intra-locus allelic effects of additive and non-additive***

The significance of intra-locus non-additive allelic effects ($Ref$-allele$_{\text{SNPi}}$ $\times$ $Ref$-allele$_{\text{SNPi}}$, $Ref$-allele$_{\text{SNPi}}$ $\times$ $Alt$-allele$_{\text{SNPi}}$, and $Alt$-allele$_{\text{SNPi}}$ $\times$ $Alt$-allele$_{\text{SNPi}}$) has been reviewed[46] and methods like GCTA-GREMLd has been introduced to address dominance effects in heritability estimations[47]. However, the estimated small intra-locus non-additive effects have led to the widespread assumption that these effects are negligible[46]. Similar to pervasive non-linear interaction of genes[20,48], there is no inherent genetic or biological rationale and the estimated small non-additive effects between reference and alternative alleles may be attributable to methodological limitations and artifacts, aligning with our results.

When linear models are augmented with some allelic adjustments tailored to the observed non-additive effects, they capture genetic variance at levels comparable to those of the exponential model (Fig. 3a–f). SAFE-$h^2$ employs the same allelic adjustment algorithm to extract non-additive effects within the framework of linear models. For every SNP position, SAFE-$h^2$ examines different scenarios, *i.e.*, additive (original genotypes), adjusted genotypes for dominance effects, adjusted genotypes for overdominance effects, and adjusted genotypes for heterosis-like effects, to find the best fitted linear model. Before heritability profiling, combinatory genotype datasets can then be built based on one selected adjustment or the original genotype, *i.e.*, the one with smallest $p$-value, for every SNP position. The SAFE-$h^2$ allelic adjustment improved the SNP heritability estimations up to 9.7 units for the 10 random phenotypes assigned to 1,092 HapMap genotypes (Fig. 3g), as well as for 10,000 simulated genotypes derived from the 1,092 HapMap genotypes (Fig. 3h). In contrast, the GCTA-GREMLd either showed no improvement or resulted in an average underestimation of 3.46 units compared to the additive approach (Fig. 3g,h). The SAFE-$h^2$ allelic adjustment proved robust enough that no further improvements in heritability estimates were observed even after increasing the cohort size from 1,092 to 10,000 (Fig. 3i). On the contrary, the additive and GREMLd based estimates were improved after cohort enlargement (Fig. 3i). The SAFE-$h^2$ additive and non-additive joint-analysis was also evaluated

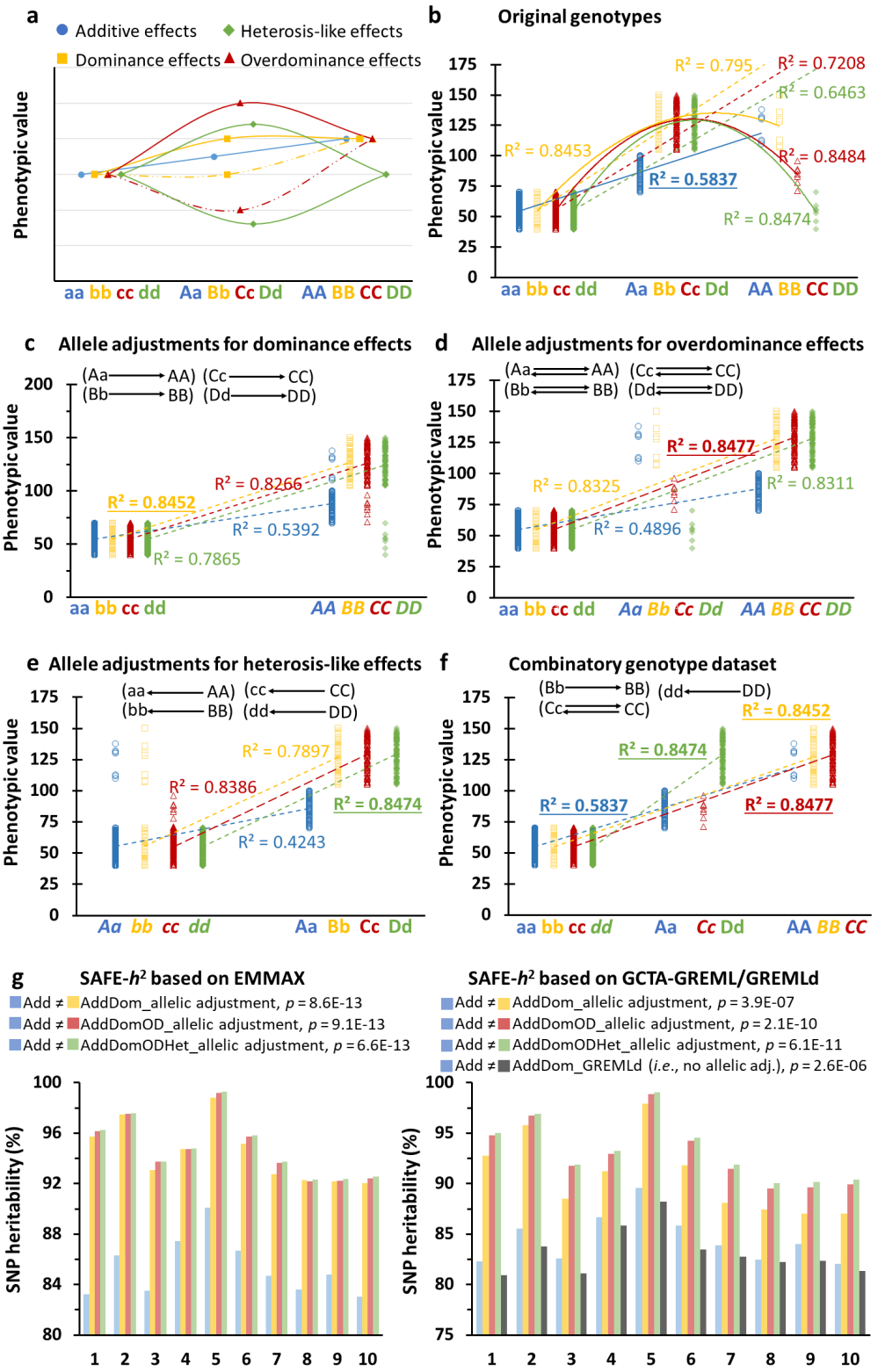



Figure 3 continued…

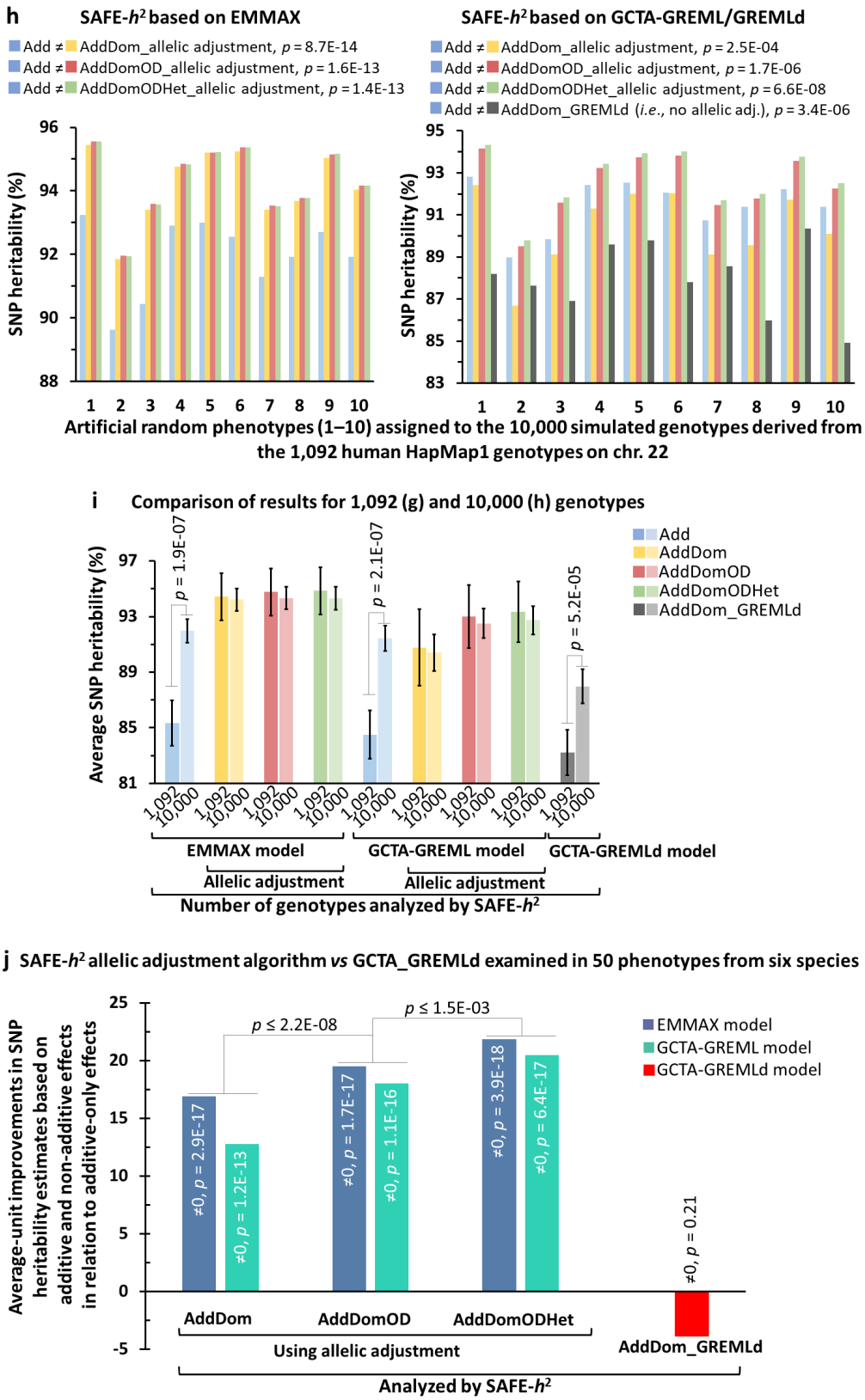



Figure 3: Allele adjustments capture intra-locus non-additive allelic effects. a–f) Demonstration of four

**Figure 3 legend continues:** biallelic loci; locus A/a with additive effect, locus B/b with dominance effect, locus C/c with overdominance effect, and locus D/d with heterosis-like effect. The best linear-fits are highlighted by bold underlined $R^2$ values. The adjusted alleles are in italic. Single arrows indicate allele adjustment by conversion and double arrows indicate allele adjustment by replacement. a) Allelic effects in a biallelic scenario. Dominance and overdominance effects can be on either alternative or reference alleles (*Alt* or *Ref*) which are differentiated by continues and dashed extrapolations. b) The artificial phenotype no. 3 (see Supplementary Fig. 2d), and its derivatives were used to represent the additive and non-additive effects, respectively. Continues lines represent the most fitted models; this was linear for the locus A/a. Low-fit linear approximates (dashed lines for B/b, C/c, and D/d loci) are also shown. c) Allelic adjustment for the dominance effects of Ref alleles (A, B, C, & D) leads to an improved linear-fit only on the B/a locus having dominance effect. d) Allelic adjustment for the overdominance effects of Ref alleles (A, B, C, & D) leads to an improved linear-fit only on the C/c locus having overdominance effect. e) Allelic adjustment for the heterosis-like effects leads to an improved linear-fit on the D/d locus with heterosis-like effect. f) A combinatory genotype dataset contains every SNP position in either of its original or one of the adjusted states, depending on which provides the best linear-fit, *i.e.*, highest $R^2$. g–i) Capturing additive and non-additive effects for the 10 artificial random phenotypes described in Supplementary Fig. 2c. j) The SAFE-$h^2$ allelic adjustment algorithm efficiently captures additive and non-addtive effects. g–j) The $p$-values are from two-tailed paired t-tests. The error bars represent 95% confidence intervals. Add: heritability based on additive effects, AddDom: heritability based on additive and dominance effects, AddDomOD: heritability based on additive, dominance and overdominance effects, AddDomODHet: heritability based on additive, dominance, overdominance, and heterosis-like effects, GREMLd: heritability estimated by capturing additive and dominance effects without allelic adjustment, MAF: minor allele frequency.

using a biallelic SNP associated at different rates and through additive, dominance, overdominance, and heterosis-like genetic effects to simulated phenotypes (Table 1). When the underlying genetic effects were additive, the estimates of SNP heritability based on joint additive and non-additive effects were identical to those derived solely from additive effects (Table 1). When the underlying genetic effects were non-additive, the allelic adjustments for non-additive effects by SAFE-$h^2$ improved the heritability estimations using all the four models, *i.e.*, EMMAX, GCTA-GREML, LDAK-Thin, and LDAK-GCTA (Table 1). GCTA-GREML and LDAK were however less effective at capturing additive-only or joint additive and non-additive effects at different association levels (Table 1). SAFE-$h^2$ obtained the best approximates of the genetic variances, *i.e.*, here equivalent to the broad-sense heritability, when applying the EMMAX and GEMMA models (Table 1, Supplementary Table 7). SAFE-$h^2$ improved estimates by capturing 6.22% extra broad-sense heritability on the allelic effects of dominance, 16.22% extra broad-sense heritability on the allelic effects of overdominance, and 24.31% extra broad-sense heritability on the heterosis-like effects on average. Analyzing 50 phenotypes observed in plant and animal species[36–44], further validated the SAFE-$h^2$ allelic adjustment by extracting non-additive effects as well as the SAFE-$h^2$ superior and stable performance versus GREMLd (Fig. 3j, Supplementary Fig. 8). It is important to emphasize that the overestimation of SNP heritability can only occur through positive contributions from false-positive SNP hits, which depends on statistical power levels. At an identical statistical power level, *i.e.*, identical genomic and genotypic coverage, the most effective heritability estimation method is the one that captures the greatest proportion of phenotypic variation.

## Discussion

SNP heritability estimations are typically obtained by including all SNPs or by selecting SNPs with $p$-value less than $10^{-8}$, often depending on which approach yields the highest estimate. A remarkable example is the heritability estimates on human height[15,16]. Functional, regional, linkage disequilibrium, and MAF stratified SNPs have also been employed in SNP heritability calculations[8–10,12,13,15]. Nevertheless, these approaches provide estimates of genetic variance at sub-genome levels rather than phenotype-specific estimates (*i.e.*, associated SNP-specific estimates). It is evident that all of these SNP heritability estimation strategies are vulnerable to the inclusion of unassociated SNPs with upper-tail association $p$-values, potentially introducing illogical negative contributions to heritability estimates. Although the genetic relatedness matrices ideally

**Table 1: SNP heritability estimations on additive or additive + non-additive effects of a biallelic locus.**

| Phenotypes | | | | | | SAFE-$h^2$ estimates of SNP-heritability | | | | | | | | | | | | | | | | GCTA GREMLd -model |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Ref/Ref Genotype | Ref/Alt Genotype | Alt/Alt Genotype | | | using EMMAX model | | | | using GCTA GREML-model | | | | using LDAK GCTA-model | | | | using LDAK Thin-model | | | | |
| Data_Genetic effect | range (N) | range (N) | range (N) | $\sigma_\varepsilon^2$ (%) | $\sigma_g^2$ (%) | A[a] | AD[b] | ADO[c] | ADOH[d] | A[a] | AD[b] | ADO[c] | ADOH[d] | A[a] | AD[b] | ADO[c] | ADOH[d] | A[a] | AD[b] | ADO[c] | ADOH[d] | |
| 1st_Additice | 40-50 (982) | 80-100 (103) | 120-150 (7) | 5.7 | 94.3 | 94.29 | 94.29 | 94.29 | 94.29 | 93.91 | 93.91 | 93.91 | 93.91 | 77.22 | 77.22 | 77.22 | 77.22 | 77.22 | 77.22 | 77.22 | 77.22 | 93.92 |
| 1st_Dominance | 40-50 (982) | 120-150 (103) | 120-150 (7) | 2.16 | 97.84 | 92.8 | ▲ 97.83 | ▲ 97.83 | ▲ 97.83 | 92.77 | ▲ 95.74 | ▲ 95.74 | ▲ 95.74 | 77.16 | ▲ 77.53 | ▲ 77.53 | ▲ 77.53 | 77.16 | ▲ 77.53 | ▲ 77.53 | ▲ 77.53 | 97.89[e] |
| 1st_Overdominance | 40-50 (982) | 120-150 (103) | 80-100 (7) | 2.35 | 97.65 | 84.77 | ▲ 95.9 | ▲ 97.65 | ▲ 97.65 | 84.29 | ▲ 92.12 | ▲ 95.48 | ▲ 95.48 | 85.21 | ▼ 77.35 | ▼ 77.51 | ▼ 77.51 | 85.21 | ▼ 77.35 | ▼ 77.51 | ▼ 77.51 | 97.77[e] |
| 1st_Heterosis-like | 40-50 (982) | 120-150 (103) | 40-50 (7) | 6.11 | 93.89 | 70.51 | ▲ 86.47 | ▲ 91.72 | ▲ 93.86 | 69.34 | ▲ 76.16 | ▲ 84.93 | 94.14[e] | 70.77 | ▲ 76.61 | ▲ 77.02 | ▲ 77.19 | 70.77 | ▲ 76.61 | ▲ 77.02 | ▲ 77.19 | 94.24[e] |
| 2nd_Additive | 40-60 (982) | 75-100 (103) | 112-150 (7) | 21.04 | 78.96 | 78.86 | 78.86 | 78.86 | 78.86 | 77.7 | 77.7 | 77.7 | 77.7 | 78.84 | 78.84 | 78.84 | 78.84 | 78.84 | 78.84 | 78.84 | 78.84 | 77.53 |
| 2nd_Dominance | 40-60 (982) | 112-150 (103) | 112-150 (7) | 12.75 | 87.25 | 82.1 | ▲ 87.23 | ▲ 87.23 | ▲ 87.23 | 81.2 | ▼ 77.34 | ▼ 77.34 | ▼ 77.34 | 82.22 | ▼ 76.67 | ▼ 76.67 | ▼ 76.67 | 82.22 | ▼ 76.67 | ▼ 76.67 | ▼ 76.67 | 87.57[e] |
| 2nd_Overdominance | 40-60 (982) | 112-150 (103) | 75-100 (7) | 14.7 | 85.3 | 71.58 | ▲ 82.7 | ▲ 85.13 | ▲ 85.13 | 70.33 | ▲ 70.49 | ▲ 74.44 | ▲ 74.44 | 71.74 | ▲ 82.69 | ▲ 76.5 | ▲ 76.5 | 71.74 | ▲ 82.69 | ▲ 76.5 | ▲ 76.5 | 86.02[e] |
| 2nd_Heterosis-like | 40-60 (982) | 112-150 (103) | 40-60 (7) | 8.1 | 91.9 | 70.56 | ▲ 85.58 | ▲ 90.28 | ▲ 91.89 | 69.36 | ▲ 74.78 | ▲ 82.54 | 92.26[e] | 70.79 | ▲ 76.54 | ▲ 76.91 | ▲ 77.04 | 70.79 | ▲ 76.54 | ▲ 76.91 | ▲ 77.04 | 92.35[e] |
| 3rd_Additive | 40-70 (982) | 70-100 (103) | 105-150 (7) | 41.5 | 58.5 | 58.34 | 58.34 | 58.34 | 58.34 | 56.67 | 56.67 | 56.67 | 56.67 | 58.32 | 58.32 | 58.32 | 58.32 | 58.32 | 58.32 | 58.32 | 58.32 | 56.31 |
| 3rd_Dominance | 40-70 (982) | 105-150 (103) | 105-150 (7) | 15.47 | 84.53 | 79.48 | ▲ 84.51 | ▲ 84.51 | ▲ 84.51 | 78.45 | ▼ 73.15 | ▼ 73.15 | ▼ 73.15 | 79.57 | ▲ 84.49 | ▲ 84.49 | ▲ 84.49 | 79.57 | ▲ 84.49 | ▲ 84.49 | ▲ 84.49 | 84.91[e] |
| 3rd_Overdominance | 40-70 (982) | 105-150 (103) | 70-100 (7) | 15.16 | 84.84 | 72.05 | ▲ 82.65 | ▲ 84.76 | ▲ 84.76 | 70.82 | ▼ 70.41 | ▲ 73.88 | ▲ 73.88 | 72.21 | ▲ 82.63 | ▲ 84.74 | ▲ 84.74 | 72.21 | ▲ 82.63 | ▲ 84.74 | ▲ 84.74 | 85.57[e] |
| 3rd_Heterosis-like | 40-70 (982) | 105-150 (103) | 40-70 (7) | 16.44 | 83.56 | 62.59 | ▲ 76.85 | ▲ 81.56 | ▲ 83.52 | 61.12 | ▲ 62.39 | ▲ 69.24 | 84.19[e] | 62.73 | ▲ 76.83 | ▲ 81.54 | ▲ 83.5 | 62.73 | ▲ 76.83 | ▲ 81.54 | ▲ 83.5 | 84.38[e] |
| 4th_Additive | 40-80 (982) | 60-100 (103) | 90-150 (7) | 72.1 | 27.9 | 27.34 | 27.34 | 27.34 | 27.34 | 26.01 | 26.01 | 26.01 | 26.01 | 27.33 | 27.33 | 27.33 | 27.33 | 27.33 | 27.33 | 27.33 | 27.33 | 25.07 |
| 4th_Dominance | 40-80 (982) | 90-150 (103) | 90-150 (7) | 33.36 | 66.64 | 62.6 | ▲ 66.6 | ▲ 66.6 | ▲ 66.6 | 61.01 | ▼ 49.91 | ▼ 49.91 | ▼ 49.91 | 62.61 | ▲ 66.59 | ▲ 66.59 | ▲ 66.59 | 62.61 | ▲ 66.59 | ▲ 66.59 | ▲ 66.59 | 67.29[e] |
| 4th_Overdominance | 40-80 (982) | 90-150 (103) | 60-100 (7) | 32.11 | 67.89 | 56.78 | ▲ 65.73 | ▲ 67.72 | ▲ 67.72 | 55.15 | ▼ 48.93 | ▼ 51.63 | ▼ 51.63 | 56.82 | ▲ 65.71 | ▲ 67.7 | ▲ 67.7 | 56.82 | ▲ 65.71 | ▲ 67.7 | ▲ 67.7 | 69.16[e] |
| 4th_Heterosis | 40-80 (982) | 90-150 (103) | 40-80 (7) | 35.4 | 64.6 | 51.1 | ▲ 60.96 | ▲ 63.84 | ▲ 64.5 | 49.34 | ▼ 43.82 | ▼ 47.32 | 65.64[e] | 51.03 | ▲ 60.94 | ▲ 63.82 | ▲ 64.49 | 51.03 | ▲ 60.94 | ▲ 63.82 | ▲ 64.49 | 65.98[e] |
| 5th_Additive | 40-90 (982) | 50-100 (103) | 75-150 (7) | 91.42 | 8.58 | 7.7 | 7.7 | 7.7 | 7.7 | 7.23 | 7.23 | 7.23 | 7.23 | 22.95[e] | 27.95[e] | 22.95[e] | 22.95[e] | 22.95[e] | 27.95[e] | 22.95[e] | 22.95[e] | na |
| 5th_Dominance | 40-90 (982) | 75-150 (103) | 75-150 (7) | 61.67 | 38.33 | 36.61 | ▲ 38.24 | ▲ 38.24 | ▲ 38.24 | 35.04 | ▼ 23.62 | ▼ 23.62 | ▼ 23.62 | 36.59 | ▲ 38.22 | ▲ 38.22 | ▲ 38.22 | 36.59 | ▲ 38.22 | ▲ 38.22 | ▲ 38.22 | 38.65[e] |
| 5th_Overdominance | 40-90 (982) | 75-150 (103) | 50-100 (7) | 55.12 | 44.88 | 34.61 | ▲ 41.87 | ▲ 44.1 | ▲ 44.82 | 33.09 | ▼ 26.46 | ▼ 28.64 | 46.06[e] | 34.61 | ▲ 41.84 | ▲ 44.08 | ▲ 44.8 | 34.61 | ▲ 41.84 | ▲ 44.08 | ▲ 44.8 | 46.34[e] |
| 5th_Heterosis-like | 40-90 (982) | 75-150 (103) | 40-90 (7) | 56 | 44 | 34.25 | ▲ 41.23 | ▲ 43.33 | ▲ 43.93 | 32.73 | ▼ 25.95 | ▼ 28 | 45.16[e] | 34.24 | ▲ 41.21 | ▲ 43.3 | ▲ 43.91 | 34.24 | ▲ 41.21 | ▲ 43.3 | ▲ 43.91 | 45.46[e] |
| 6th_Additive | 40-95 (982) | 45-100 (103) | 50-150 (7) | 97.56 | 2.44 | 2.3 | 2.3 | 2.3 | 2.3 | 2.15 | 2.15 | 2.15 | 2.15 | 22.59[e] | 22.59[e] | 22.59[e] | 22.59[e] | 22.59[e] | 22.59[e] | 22.59[e] | 22.59[e] | na |
| 6th_Dominance | 40-95 (982) | 50-150 (103) | 50-150 (7) | 76.6 | 23.4 | 20.67 | ▲ 23.03 | ▲ 23.31 | ▲ 23.31 | 19.57 | ▼ 13.02 | ▼ 13.39 | ▼ 13.39 | 20.65 | ▲ 23.05 | ▲ 23.29 | ▲ 23.29 | 20.65 | ▲ 23.05 | ▲ 23.29 | ▲ 23.29 | 24.16[e] |
| 6th_Overdominance | 40-95 (982) | 50-150 (103) | 45-100 (7) | 76.64 | 23.36 | 19.5 | ▲ 22.56 | ▲ 23.24 | ▲ 23.24 | 18.45 | ▼ 12.71 | ▼ 13.35 | ▼ 13.35 | 19.44 | ▲ 22.55 | ▲ 23.22 | ▲ 23.22 | 19.49 | ▲ 22.55 | ▲ 23.22 | ▲ 23.22 | 24.29[e] |
| 6th_Heterosis-like | 40-95 (982) | 50-150 (103) | 40-95 (7) | 76.57 | 23.43 | 16.19 | ▲ 20.62 | ▲ 22.57 | ▲ 23.21 | 15.28 | ▼ 11.49 | 60.84[e] | 24.11[e] | 16.18 | ▲ 20.61 | ▲ 22.56 | ▲ 23.19 | 16.18 | ▲ 20.61 | ▲ 22.56 | ▲ 23.19 | 24.31[e] |

Green triangles indicate improved estimation when compared to the analysis on the original genotypes (a). Red triangles indicate underestimations when compared to the analysis on the original genotypes (a). $\sigma_\varepsilon^2$ & $\sigma_g^2$ represent the the enviromental and genetic variances, respectively. $\sigma_g^2$ is equivalent to the broad-sense heritability. [a]On original genotypes to capture additive effects; [b]On adjusted genotypes for dominance effects to capture additive and dominance effects; [c]On adjusted genotypes for dominance and overdominance effects to capture additive, dominance, and overdominace effects; [d]On adjusted genotypes for dominance,overdominance, and heterosis-like effects to capture additive, dominance, overdominace, and heterosis-like effects; [e]Overestimated.

require phenotype-associated SNPs and not irrelevant variants, the trade-off between the detrimental effects of excluding associated SNPs and including irrelevant variants has made the SNP selection challenging, often resulting in inaccurate genomic predictions, also addressed previously by Lippert et al. (2013)[24]. The SAFE-$h^2$ method bypasses this trade-off by identifying a $p$-value threshold below which no SNP can be found with deflationary effects on genetic variance. Finally, SAFE-$h^2$ generates a list of $p$-value-guided SNP clusters and outputs a profiling graph of heritability estimates (Supplementary Fig. 9a–c). The EMMAX and GEMMA algorithms were found as the most stable models for accurate estimation of SNP heritability using SAFE-$h^2$ (Fig. 2, Table 1, Supplementary Fig. 2,4–6, Supplementary Table 7). For EMMAX-based estimations, SAFE-$h^2$ utilizes the Balding-Nichols (BN) kinship matrix, which has been shown to be highly correlated with FaST-LMM and GEMMA kinship matrices[49]. Compared to the analyses based on $p$-values from EMMAX, we observed even larger negative contributions to heritability estimations when the analyses were based on $p$-values calculated by GEMMA (Supplementary Table 8), which employs an exact mixed model[27].

A second key challenge in GWAS studies is the low statistical power due to insufficient genotypic coverage, leading to false-positive results. SAFE-$h^2$ enables the identification of a second $p$-value threshold at which contributions of false-positive SNP hits to heritability estimations can be restricted efficiently (Fig. 2d, Supplementary Fig. 7 and 9e). This approach can also be applied in GWAS to define significant SNPs. The hidden heritability due to false-negative SNP hits has also been discussed previously[50]. SAFE-$h^2$ ensures the inclusion of weakly associated SNPs, which might otherwise be excluded (*i.e.*, leading to false-negatives) due to a predefined $p$-value cut-off, *e.g.*, $10^{-8}$. Replication of experiments, *e.g.,* using effect-sizes of associated SNPs obtained in one cohort to calculate polygenic risk scores for individuals/genotypes in a second cohort, is a common approach in GWAS to mitigate overfitting[51] caused by inadequate genotypic coverage. However, replication to reduce overfitting is not applicable in heritability estimations. It is important to note that overfitting occurs at genotype-level, while heritability is a population- or cohort-level statistic that is specific to datasets and experiments, varying with new genotypic and environmental inputs, as outlined in the introduction. An interesting example in this case is the observed heritability shifts across different experimental conditions such as age[6].

The recently developed tools such as Next-Gen GWAS[20] and GWAIS[52] offer major improvements in the processing of epistatic (SNP × SNP) effects. To complement, SAFE-$h^2$ is capable of extracting intra-locus

(*i.e.*, *Ref*-allele$_{\text{SNPi}}$ × *Ref*-allele$_{\text{SNPi}}$, *Ref*-allele$_{\text{SNPi}}$ × *Alt*-allele$_{\text{SNPi}}$, and *Alt*-allele$_{\text{SNPi}}$ × *Alt*-allele$_{\text{SNPi}}$) non-additive allelic effects. An example of SAFE-$h^2$ output graphs for joint analysis of additive and non-additive allelic effects is shown in Supplementary Figure 9d. SAFE-$h^2$ achieved the best performance using EMMAX and GEMMA models to jointly capture additive and non-additive allelic effects of the predefined single biallelic-locus, without overestimations (Table 1, Supplementary Table 7). In contrast, GCTA-GREMLd, while still providing useful estimates, resulted in small overestimation errors across all measurements (Table 1). GCTA-GREMLd was also prone to SNP heritability underestimations, as its estimates for joint additive and non-additive effects were smaller than those for additive-only effect (Fig. 3g,h,j, Supplementary Fig. 8). Furthermore, SAFE-$h^2$ extracts not only the dominance but also overdominance and heterosis-like allelic effects.

In summary, SAFE-$h^2$ provides phenotype-specific estimates of SNP heritability rather than genome or sub-genome specific estimates, and minimize the likelihood of false-positive contributions. SAFE-$h^2$ further refines the estimations through joint analysis of intra-locus non-additive and additive allelic effects. By adjusting alleles, SAFE-$h^2$ generates genotypic datasets that can facilitate linear association tests to address intra-locus additive and non-additive effects jointly.

## Materials and Methods

### *Datasets*

The simulated homozygote genotypes were arranged across 29 chromosomes, with 300kb distance between SNPs to meet linkage equilibrium. For each genotype, reference and alternative alleles were randomly assigned at different chromosomal positions. VCF files for 1,000 and 10,000 genotypes were assembled with 200k SNPs with MAF ranging from 1% to 10%. Within these datasets, one of the 200k SNP positions was designated as non-random to be associated with the simulated non-random phenotypes. For having multiple predefined associations, 11, 23, or 47 extra predefined SNPs were integrated into the same genomic datasets. Genotypes with 100 and 2,000k SNPs were also simulated. Further proof-of-concept analyses were accomplished using human chromosome 22 from the HapMap phase 1 genomic data[28]. This included 1,092 genotypes and 177,675 SNPs after dataset filtering (MAF $\geq$ 1%). As for the simulated genotypes, one simulated non-random SNP position was introduced among the HapMap variants to be associated with the simulated non-random phenotypes. The human chr.22 SNP-set was also extended by simulation to 10,000 individuals. For this, the original SNP dataset was shuffled three times among the chromosomal positions to get a new genotype set. This was repeated to have 9 new datasets. The original and new datasets were then horizontally assembled to build a dataset of 10,000 genotypes represented by 177,675 SNPs. The final simulated dataset had 177,445 variants after filtering for MAF of $\geq$ 1%. The artificial phenotypes were produced as random or non-random integers assigned randomly among the genotypes or among the genotypic groups of the simulated SNPs with varying levels of associations. Comprehensive analyses were also conducted using *Arabidopsis* and rice public datasets. For *Arabidopsis*, 14 different phenotypes were obtained from the AraPheno database[30]. These phenotypes were from a panel of 328 to 1,058 genotypes. The imputed genomic variants for these panels[31,32] contained 2,621,362 to 2,840,072 post-filtering SNPs. The public rice data was obtained from the rice SNP-Seek database[33]. The analyses were performed using 10 phenotypes that were previously recorded in 1,422 to 2,102 rice genotypes. The rice genomic data was also publicly available in rice SNP-Seek database[33–35]. The genotypes contained from 6,131,325 to 6,414,800 post-filtering SNPs. Additional datasets from both animal and plant species were analyzed to examine the negative contributions to heritability as well as the inclusion of intra-locus allelic effects of non-additive in the SNP heritability estimates. These included clutch size in 314 wild avians[36] genotyped by 40,766 SNPs (MAF $\geq$ 0.01), four fruit and flowering related phenotypes in 96 tomato genotypes[37] genotyped by 16,782 SNPs (MAF $\geq$ 0.01), the contents of 15 secondary metabolites in 411 tomato genotypes[43] genotyped by 308,885 SNPs (MAF $\geq$ 0.01), the contents of 7 secondary metabolites in 136 apple genotypes[38] genotyped by 98,584 SNPs (MAF $\geq$ 0.01), apple fruit-weight, apple sucrose, fructose, and tartrate contents in 368 to 461 genotypes[39] represented by 3,055,912 SNPs (MAF $\geq$ 0.01), six growth related phenotypes in 752–755 goat genotypes[40] represented by 40,639 SNPs (MAF $\geq$ 0.01), the abundance of 12 microbiome sub-communities in 827 genotypes of foxtail

millet represented by 161,562 SNPs[44], and a merged dataset for susceptibility to diabetes mellitus in 1,022 genotypes represented by 826 SNPs (MAF $\geq 0.01$) that derived from two independent studies in cats[41,42].

***Data analysis***

Plink[53] and EMMAX[25] were used for the initial processing of genomic data and association analyses. MAF was $\geq 1\%$ in all the analyses. A complete record on the latitudes of geographical origins was available for *Arabidopsis* genotypes when analyzing the stem length and leaf mineral contents. This geographical information was applied as quantitative covariates in the analyses. SAFE-$h^2$ was developed to perform a dynamic SNP allocation for estimation of narrow-sense heritability. A comprehensive manual is provided together with the software. In brief, SAFE-$h^2$ begins by scanning the input files and executing the user-defined combination of heritability estimation models on $p$-value guided SNP clusters. SAFE-$h^2$ identifies a dataset- and experiment-specific $p$-value threshold at which SNPs with larger $p$-values are distributed randomly across genotypes in relation to the phenotypic variation, imposing negative contributions to heritability estimates. In essence, this $p$-value threshold is the best border that we can assume between associated and unassociated SNPs under standard experimental conditions, *i.e.*, the genomic and genotypic coverage, for accurate estimation of SNP heritability. Therefore, we define these SNPs with $p$-values larger than the threshold as "unassociated". These unassociated SNPs lead to overestimated genetic similarities within genetic relationship matrices and thereby, underestimations in genetic variance and SNP heritability levels. SAFE-$h^2$ requires a $p$-value file for SNPs, the Plink b-files ("bim", "bed", and "fam"), and optional covariate files specific for every estimation model. SAFE-$h^2$ leverages five different models of EMMAX[25], LDAK-GCTA, LDAK-Thin[7], GCTA-GREML[26], and GEMMA[27] to estimate SNP heritability. GCTA-GREML inbred algorithm can also be applied to inbred genotypes or populations with considerable levels of inbreeding. To operate, SAFE-$h^2$ requires the binary files of the mentioned tools (see the SAFE-$h^2$ manual with full description). In return, SAFE-$h^2$ generates the output data files and related graphs for clustered SNP hits and heritability profiles. It also features the $p$-value threshold and the estimated SNP heritability at that threshold on the graphs. Given that different levels of false-positive rates are typically expected in GWAS, SAFE-$h^2$ employs an innovative approach to exclude these false-positive SNP hits. SAFE-$h^2$ determines a "safe" $p$-value interval, with its upper-end to be used as an alternative $p$-value threshold for heritability estimations, ensuring minimal contributions from false-positive SNP hits. To perform the analysis, the user should simulate six independent random phenotypes within the range of the original phenotype-of-interest for all genotypes in the dataset. SAFE-$h^2$ requires several input files: a "bim" file, a "bed" file, a "fam" file containing the original phenotypic records, six "fam" files for simulated random phenotypes, SNP $p$-value files for the real (one file) and random phenotypes (six files), and optional covariate files specific for estimation models. Using these input files, SAFE-$h^2$ creates the SNP heritability profiles for all the random phenotypes as well as the real phenotype-of-interest using an identical genomic dataset. Users can select the estimation models of interest. Analyzing the random phenotypes subsequently determines a safe $p$-value interval with minimum SNP heritability level at the lower-tail of heritability profile. The upper-end of this interval can function as an alternative $p$-value threshold to omit the false-positive SNP hits from the estimation models and keep their contribution to heritability at minimum. Therefore, SAFE-$h^2$ keeps false-negative and false-positive rates at minimum by heritability profiling for the phenotype-of-interest and the simulated phenotypes, respectively. Finally, SAFE-$h^2$ generates an aligned graph of heritability profiles for the random and original phenotypes. The SNP heritability of the original phenotype can accordingly be defined at upper-end of the determined $p$-value interval to keep the false-positive contributions at its minimum.

SAFE-$h^2$ also offers allelic adjustments for linear approximation on intra-locus allelic effects including dominance, overdominance, and heterosis-like effects. This facilitates capturing of non-additive allelic effects. SAFE-$h^2$-preADOH builds two VCF files for the dominance adjustments by converting the heterozygote nucleotide positions into one of the homozygotes. It also creates two VCF files for the overdominance adjustments, where heterozygote nucleotide positions are converted into either of homozygotes and, the same homozygotes are converted into their heterozygote counterparts. For the heterosis-like effects, SAFE-$h^2$-preADOH builds an additional VCF file by converting *Ref* $\times$ *Ref* homozygotes into the corresponding homozygote of *Alt* $\times$ *Alt*, and vice versa, at each nucleotide position. Here, the user can perform an indepen-

dent association test or alternatively can command SAFE-$h^2$-preADOH to execute the Plink's glm function on the five adjusted and one original genotype datasets. To construct the combinatory genotype datasets, SAFE-$h^2$-preADOH applies the association $p$-values to select the best linear-fit (with smallest association $p$-value) among the original and the five adjusted versions for every SNP position. At the end, SAFE-$h^2$-preADOH provides three genotypic datasets that can be used to estimate the additive-dominance effects, additive-dominance-overdominance effects, and additive-dominance-overdominance-heterosis like effects. Here, every SNP-ID gets either of the following specific tags; [Add] for having additive effects on the phenotype of interest, [Dom] or [dom] for having dominance effects on the phenotype of interest, [OD] or [od] for having overdominance effects on the phenotype of interest, or [Het] for having heterosis-like effects on the phenotype of interest. Finally, SAFE-$h^2$-ADOH performs the heritability profiling as explained for SAFE-$h^2$ and on the three adjusted genotype datasets as well as the original genotype dataset. It is important to emphasize that SAFE-$h^2$ considers very SNP through either of additive, dominance, overdominance, or heterosis-like effects for inclusion in the heritability model. Therefore, the file for additive-dominance effects is used to examine the SNPs tagged with [Add] only for additive effect and the SNPs tagged with [Dom] or [dom] only for dominance effect. The file for additive-dominance-overdominance effects is used to examine the SNPs tagged with [Add] only for additive effect, the SNPs tagged with [Dom] or [dom] only for dominance effect, and the SNPs tagged with [OD] or [od] only for overdominance effect. Finally, the file for additive-dominance-overdominance-heterosis effects is used to examine the SNPs tagged with [Add] only for additive effect, the SNPs tagged with [Dom] or [dom] only for dominance effect, the SNPs tagged with [OD] or [od] only for overdominance effect, and the SNPs tagged with [Het] only for heterosis-like effect.

**Author contributions**

B.D conceived the study. B.D performed the analyses and compiled the SAFE-$h^2$ script. B.D and M.N discussed the experimental designs and results. B.D, and M.N wrote the manuscript.

**Competing interests**

The authors declare no competing interests.

**Acknowledgments**

We thank Prof. Detlef Weigel (professor in plant genetics and evolution, Department of Molecular Biology, Max Planck Institute for Biology Tübingen, Germany) for his invaluable consult and comments.

**Code availability**

SAFE-$h^2$, containing the compiled SAFE-$h^2$ scripts with related installation and usage manual, is attached to this publication as suppl. File 1. It also contains example datasets. SAFE-$h^2$ is available for download at https://github.com/SAFE-h2/SAFE-h2-v2024-1. The example dataset can also be obtained from figshare (DOI: 10.6084/m9.figshare.24025728, Link: https://figshare.com/s/7193d332894f11deb1a5).

**Supplementary figures, tables, and text**

**Supplementary Fig. 1:** SAFE-h2 pipeline.

**Supplementary Fig. 2:** The SNP heritability profiles for simulated quantitative-phenotypes using simulated SNPs or variants of chr. 22 from the human HapMap phase 1 genotypes.

**Supplementary Fig. 3:** The SNP heritability profiles for simulated random (quantitative) phenotypes and SNPs.

**Supplementary Fig. 4:** The SNP heritability profile for simulated quantitative-phenotypes and SNPs, with predesignated SNPs associated at varying rates.

**Supplementary Fig. 5:** The SNP heritability profiles for simulated quantitative-phenotypes and SNPs were analyzed comparing linkage disequilibrium to linkage equilibrium.

**Supplementary Fig. 6:** The SNP heritability profiles for simulated quantitative-phenotypes and SNPs with minor allele frequencies ranging from 1.6%–50%.

**Supplementary Fig. 7:** SAFE-$h^2$ estimates SNP heritability by minimizing the contributions from false-positive SNP hits.

**Supplementary Fig. 8:** The SAFE-$h^2$ estimates of SNP heritability using additive-only model and the combined additive and non-additive models, compared to GREMLd estimates.

**Supplementary Fig. 9:** SAFE-$h^2$ generates bar graphs for both the $p$-value-guided SNP clusters and the SNP heritability profiles.

**Supplementary Table 1:** The PubMed PMID accessions of the recently published studies used to extract SNP heritability estimates for different traits and species.

**Supplementary Table 2:** SAFE-$h^2$ improves the heritability estimations for 24 phenotypes in *Arabidopsis* and rice by preventing the negative contributions of weakly-associated genomic regions.

**Supplementary Table 3:** Heritability profiling by SAFE-$h^2$ reveals negative contributions of SNPs with large association $p$-values by examining 50 phenotypes in six species.

**Supplementary Table 4:** BayesR SNP heritability estimates for simulated quantitative-phenotypes using the 200K simulated SNPs in LE and LD.

**Supplementary Table 5:** BayesR SNP heritability estimates for simulated quantitative-phenotypes using the chr. 22 SNPs from human HapMap phase 1 genotypes.

**Supplementary Table 6:** The SNP heritability estimates for real phenotypes using BayesR and SAFE-$h^2$.

**Supplementary Table 7:** SNP heritability estimations on additive or additive + non-additive effects of a biallelic locus.

**Supplementary Table 8:** Heritability estimates based on $p$-values, calculated using EMMAX and GEMMA.

**Supplementary Text 1:** : Independently simulated datasets cannot be used as replicates.

## References


1. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).

2. Tang, M., Wang, T. & Zhang, X. A review of SNP heritability estimation methods. *Briefings in Bioinformatics* **23**, bbac067 (2022).

3. Bergelson, J. & Roux, F. Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. *Nat Rev Genet* **11**, 867–879 (2010).

4. Pujol, B. *et al.* The Missing Response to Selection in the Wild. *Trends in Ecology & Evolution* **33**, 337–346 (2018).

5. Adaptation to Climate Across the Arabidopsis thaliana Genome | Science. https://www.science.org/doi/abs/10.1126/science.1209244.

6. Cohen, N. M. *et al.* Personalized lab test models to quantify disease potentials in healthy individuals. *Nat Med* **27**, 1582–1591 (2021).

7. Speed, D., Hemani, G., Johnson, M. R. & Balding, D. J. Improved Heritability Estimation from Genome-wide SNPs. *Am J Hum Genet* **91**, 1011–1021 (2012).

8. Evans, L. M. *et al.* Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat Genet* **50**, 737–745 (2018).

9. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565–569 (2010).

10. Lee, S. H. *et al.* Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet* **44**, 247–250 (2012).

11. Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet* **51**, 793–803 (2019).

12. Speed, D., Cai, N., Johnson, M. R., Nejentsev, S. & Balding, D. J. Reevaluation of SNP heritability in complex human traits. *Nat Genet* **49**, 986–992 (2017).

13. Speed, D., Kaphle, A. & Balding, D. J. SNP-based heritability and selection analyses: Improved models and new results. *BioEssays* **44**, 2100170 (2022).

14. Baselmans, B. M. L., Yengo, L., van Rheenen, W. & Wray, N. R. Risk in Relatives, Heritability, SNP-Based Heritability, and Genetic Correlations in Psychiatric Disorders: A Review. *Biological Psychiatry* **89**, 11–19 (2021).

15. Yengo, L. *et al.* A saturated map of common genetic variants associated with human height. *Nature* **610**, 704–712 (2022).

16. Wainschtein, P. *et al.* Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nat Genet* **54**, 263–273 (2022).

17. Elks, C. E. *et al.* Variability in the Heritability of Body Mass Index: A Systematic Review and Meta-Regression. *Front Endocrinol (Lausanne)* **3**, 29 (2012).

18. Carmichael, C. M. & McGue, M. A cross-sectional examination of height, weight, and body mass index in adult twins. *J Gerontol A Biol Sci Med Sci* **50**, B237-244 (1995).

19. Young, A. I. *et al.* Relatedness disequilibrium regression estimates heritability without environmental bias. *Nat Genet* **50**, 1304–1310 (2018).

20. Carré, C. *et al.* Next-Gen GWAS: full 2D epistatic interaction maps retrieve part of missing heritability and improve phenotypic prediction. *Genome Biology* **25**, 76 (2024).

21. Brachi, B., Morris, G. P. & Borevitz, J. O. Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biology* **12**, 232 (2011).

22. Barry, C.-J. S. *et al.* How to estimate heritability: a guide for genetic epidemiologists. *International Journal of Epidemiology* **52**, 624–632 (2023).

23. Golan, D. & Rosset, S. Accurate estimation of heritability in genome wide studies using random effects models. *Bioinformatics* **27**, i317–i323 (2011).

24. Lippert, C. *et al.* The benefits of selecting phenotype-specific variants for applications of mixed models in genomics. *Sci Rep* **3**, 1815 (2013).

25. Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* **42**, 348–354 (2010).

26. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics* **88**, 76–82 (2011).

27. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**, 821–824 (2012).

28. McVean, G. A. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).

29. Atwell, S. *et al.* Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature* **465**, 627–631 (2010).

30. Seren, Ü. *et al.* AraPheno: a public database for Arabidopsis thaliana phenotypes. *Nucleic Acids Research* **45**, D1054–D1059 (2017).

31. Alonso-Blanco, C. *et al.* 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell* **166**, 481–491 (2016).

32. Arouisse, B., Korte, A., van Eeuwijk, F. & Kruijer, W. Imputation of 3 million SNPs in the Arabidopsis regional mapping population. *The Plant Journal* **102**, 872–882 (2020).

33. Mansueto, L. *et al.* Rice SNP-seek database update: new SNPs, indels, and queries. *Nucleic Acids Research* **45**, D1075–D1081 (2017).

34. McCouch, S. R. *et al.* Open access resources for genome-wide association mapping in rice. *Nat Commun* **7**, 10532 (2016).

35. The 3, 000 rice genomes project. The 3,000 rice genomes project. *GigaScience* **3**, 2047-217X-3–7 (2014).

36. Husby, A. *et al.* Genome-wide association mapping in a wild avian population identifies a link between genetic and phenotypic variation in a life-history trait. *Proc Biol Sci* **282**, 20150156 (2015).

37. Yamamoto, E. *et al.* A simulation-based breeding design that uses whole-genome prediction in tomato. *Sci Rep* **6**, 19454 (2016).

38. McClure, K. A. *et al.* Genome-wide association studies in apple reveal loci of large effect controlling apple polyphenols. *Hortic Res* **6**, 1–12 (2019).

39. Liao, L. *et al.* Unraveling a genetic roadmap for improved taste in the domesticated apple. *Molecular Plant* **14**, 1454–1471 (2021).

40. Moaeen-ud-Din, M., Danish Muner, R. & Khan, M. S. Genome wide association study identifies novel candidate genes for growth and body conformation traits in goats. *Sci Rep* **12**, 9891 (2022).

41. Forcada, Y., Boursnell, M., Catchpole, B. & Church, D. B. A genome-wide association study identifies novel candidate genes for susceptibility to diabetes mellitus in non-obese cats. *PLOS ONE* **16**, e0259939 (2021).

42. Hernandez, I. *et al.* Complex Feline Disease Mapping Using a Dense Genotyping Array. *Frontiers in Veterinary Science* **9**, (2022).

43. Zhu, G. *et al.* Rewiring of the Fruit Metabolome in Tomato Breeding. *Cell* **172**, 249-261.e12 (2018).

44. Wang, Y. *et al.* GWAS, MWAS and mGWAS provide insights into precision agriculture based on genotype-dependent microbial effects in foxtail millet. *Nat Commun* **13**, 5913 (2022).

45. Moser, G. *et al.* Simultaneous Discovery, Estimation and Prediction Analysis of Complex Traits Using a Bayesian Mixture Model. *PLOS Genetics* **11**, e1004969 (2015).

46. Varona, L., Legarra, A., Toro, M. A. & Vitezica, Z. G. Non-additive Effects in Genomic Selection. *Frontiers in Genetics* **9**, (2018).

47. Zhu, Z. *et al.* Dominance Genetic Variation Contributes Little to the Missing Heritability for Human Complex Traits. *The American Journal of Human Genetics* **96**, 377–385 (2015).

48. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences* **109**, 1193–1198 (2012).

49. Eu-ahsunthornwattana, J. *et al.* Comparison of Methods to Account for Relatedness in Genome-Wide Association Studies with Family-Based Data. *PLoS Genet* **10**, e1004445 (2014).

50. Williams, S. M. & Haines, J. L. Correcting away the hidden heritability. *Ann Hum Genet* **75**, 348–350 (2011).

51. Subramanian, J. & Simon, R. Overfitting in prediction models - is it a problem only in high dimensions? *Contemp Clin Trials* **36**, 636–641 (2013).

52. Wienbrandt, L., Prieß, C. & Ellinghaus, D. GWAIS-Web: A Fast and Secure Web Service for Epistasis Detection in Genome-wide Association Interaction Studies. bioArxiv (2023). doi: https://doi.org/10.1101/2023.05.16.540964.

53. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, s13742-015-0047–8 (2015).

# Supplementary figures, tables, and text for:

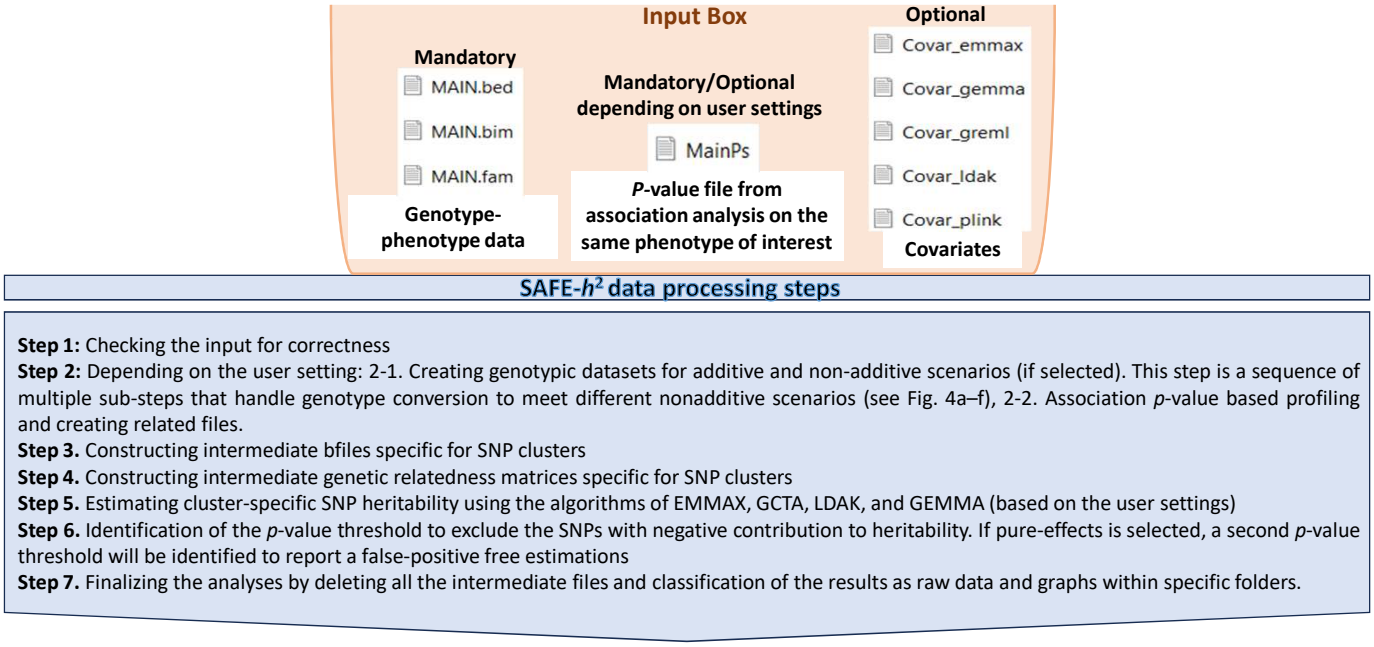
# On the Genetic Origins of Phenotypes in Genome-Wide Association Studies: The SNP Allocation For Estimating Heritability (SAFE-$h^2$) Tool to Explore Additive-only Allelic Effects or Additive and Non-Additive Allelic Effects

Behrooz Darbani*, Mogens Nicolaisen

***Department of Agroecology, Faculty of Technical Sciences, Aarhus University, DK-4200 Slagelse, Denmark***

Corresponding authors emails: behroozdarbani@gmail.com / bds@rsyd.dk; mn@agro.au.dk

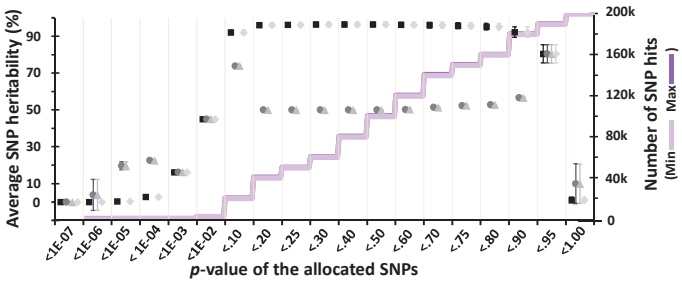
*Current affiliation: ***Molecular Diagnostic and Clinical Research Unit, University Hospital of Southern Denmark, 6200 Aabenraa, Denmark***




**Supplementary Fig. 1: SAFE-$h^2$ pipeline.**

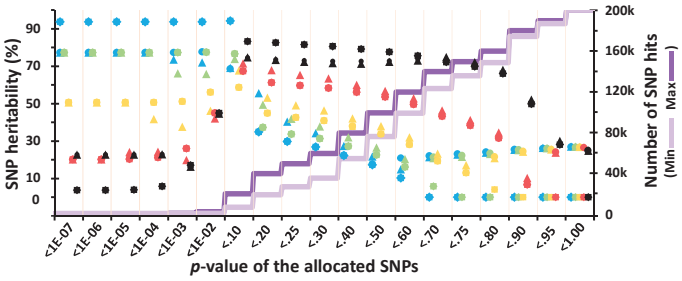
**a) Analysis of 10 artificial random phenotypes on a panel of 10,000 simulated inbred genotypes (200,000 SNPs; MAF = 1.6%–10%)**

| Random phenotypes | Range (N) | EMMAX | LDAK_GCTA | LDAK_Thin | GCTA_GREML |
|---|---|---|---|---|---|
| 1st | 40-100 (10,000) | ■ | ● | ▲ | ◇ |
| 2nd | 40-100 (10,000) | ■ | ● | ▲ | ◇ |
| 3rd | 40-100 (10,000) | ■ | ● | ▲ | ◇ |
| 4th | 40-100 (10,000) | ■ | ● | ▲ | ◇ |
| 5th | 40-100 (10,000) | ■ | ● | ▲ | ◇ |
| 6th | 40-100 (10,000) | ■ | ● | ▲ | ◇ |
| 7th | 40-100 (10,000) | ■ | ● | ▲ | ◇ |
| 8th | 40-100 (10,000) | ■ | ● | ▲ | ◇ |
| 9th | 40-100 (10,000) | ■ | ● | ▲ | ◇ |
| 10th | 40-100 (10,000) | ■ | ● | ▲ | ◇ |



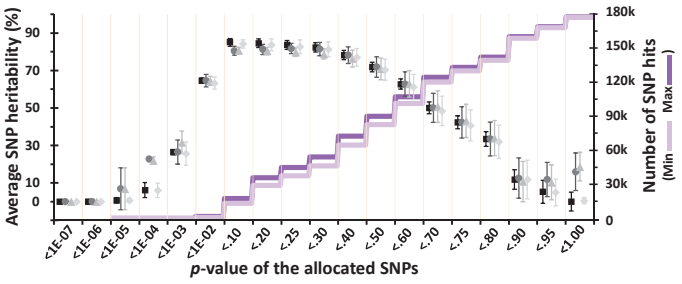

**b) Analysis of 5 artificial non-random phenotypes on a panel of 10,000 simulated inbred genotypes (200,000 SNPs; MAF = 1.6%–10%)**

| Phenotypes | The introduced non-random and associated SNP position | | | | |
|---|---|---|---|---|---|
| | Ref/Ref Genotype | Alt/Alt Genotype | | | |
| | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | EMMAX / LDAK_GCTA / LDAK_Thin / GCTA_GREML |
| 1st | 40-50 (9,000) | 80-100 (1,000) | 6.605 | 93.395 | ■ ● ▲ ◆ |
| 2nd | 40-60 (9,000) | 75-100 (1,000) | 22.738 | 77.262 | ■ ● ▲ ◆ |
| 3rd | 40-70 (9,000) | 70-100 (1,000) | 48.937 | 51.063 | ■ ● ▲ ◆ |
| 4th | 40-80 (9,000) | 60-100 (1,000) | 79.675 | 20.325 | ■ ● ▲ ◆ |
| 5th | 40-90 (9,000) | 50-100 (1,000) | 96.309 | 3.691 | ■ ● ▲ ◆ |



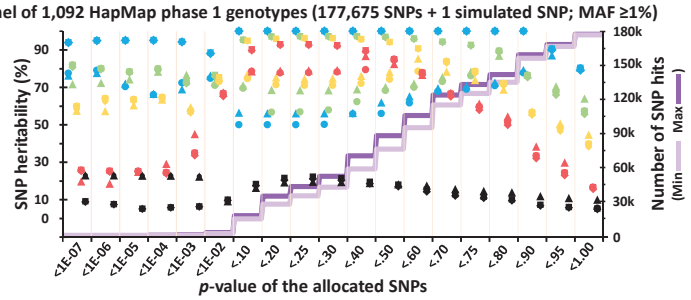

**c) Analysis of 10 artificial random phenotypes on a panel of 1,092 HapMap phase 1 genotypes (177,675 SNPs + 1 simulated SNP; MAF ≥1%)**

| Random phenotypes | Range (N) | EMMAX | LDAK_GCTA | LDAK_Thin | GCTA_GREML |
|---|---|---|---|---|---|
| 1st | 40-150 (1,092) | ■ | ● | ▲ | ◇ |
| 2nd | 40-150 (1,092) | ■ | ● | ▲ | ◇ |
| 3rd | 40-150 (1,092) | ■ | ● | ▲ | ◇ |
| 4th | 40-150 (1,092) | ■ | ● | ▲ | ◇ |
| 5th | 40-150 (1,092) | ■ | ● | ▲ | ◇ |
| 6th | 40-150 (1,092) | ■ | ● | ▲ | ◇ |
| 7th | 40-150 (1,092) | ■ | ● | ▲ | ◇ |
| 8th | 40-150 (1,092) | ■ | ● | ▲ | ◇ |
| 9th | 40-150 (1,092) | ■ | ● | ▲ | ◇ |
| 10th | 40-150 (1,092) | ■ | ● | ▲ | ◇ |



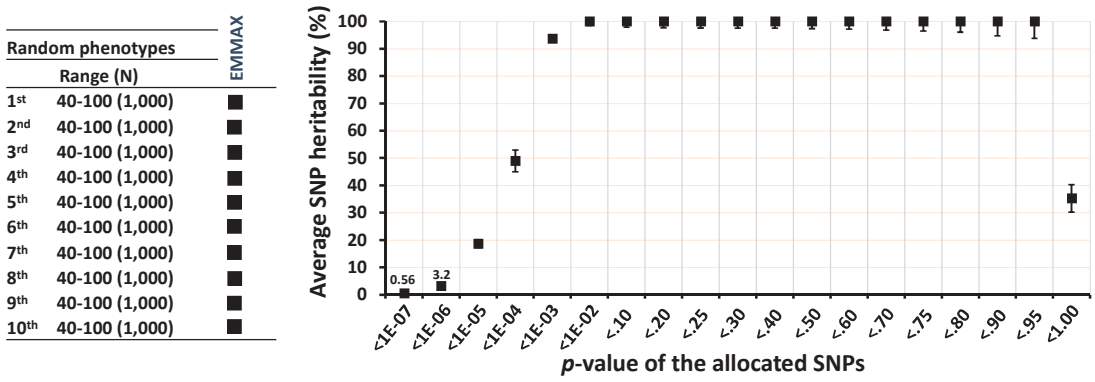

**d) Analysis of 5 artificial non-random phenotypes on a panel of 1,092 HapMap phase 1 genotypes (177,675 SNPs + 1 simulated SNP; MAF ≥1%)**

| Phenotypes | The introduced non-random and associated SNP position | | | | |
|---|---|---|---|---|---|
| | Ref/Ref Genotype | Ref/Alt Genotype | Alt/Alt Genotype | | |
| | range (N) | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | EMMAX / LDAK_GCTA / LDAK_Thin / GCTA_GREML |
| 1st | 40-50 (982) | 80-100 (103) | 120-150 (7) | 5.704 | 94.296 | ■ ● ▲ ◆ |
| 2nd | 40-60 (982) | 75-100 (103) | 112-150 (7) | 21.037 | 78.963 | ■ ● ▲ ◆ |
| 3rd | 40-70 (982) | 70-100 (103) | 105-150 (7) | 41.501 | 58.499 | ■ ● ▲ ◆ |
| 4th | 40-80 (982) | 60-100 (103) | 90-150 (7) | 72.099 | 27.901 | ■ ● ▲ ◆ |
| 5th | 40-90 (982) | 50-100 (103) | 75-150 (7) | 91.416 | 8.584 | ■ ● ▲ ◆ |

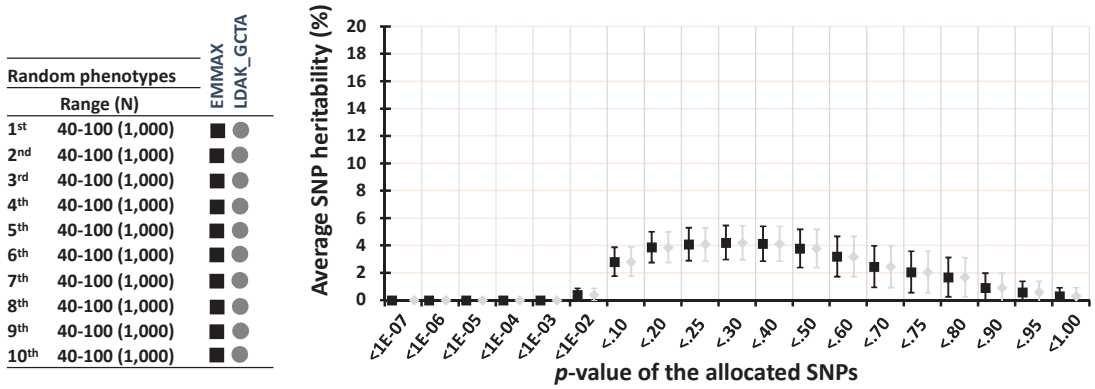



**Supplementary Fig. 2: The SNP heritability profiles for simulated quantitative-phenotypes using simulated SNPs or variants of chr. 22 from the human HapMap phase 1 genotypes.** a–d) An initial GWAS was performed using EMMAX to calculate association levels for the SNPs. The four different approaches of EMMAX, LDAK-GCTA, LDAK-Thin, and GCTA-GREML were then applied to estimate SNP heritability for the clusters of SNP hits identified using the association *p*-values. N represents the sample sizes, *i.e.*, number of genotypes. a,c) The average SNP heritability estimates were obtained for 10 independent random phenotypes within identical ranges. The error bars represent 99% confidence intervals. b,d) The SNP heritability estimates were obtained for five simulated phenotypes which were associated with a predesignated SNP position introduced within both the simulated and HapMap genomes. By ensuring additive-only effects, associations were simulated at five different levels. As a result, the SNP heritability levels were predictable at or above the predefined levels, given that other SNPs could also, by chance, account for additional phenotypic variance. a–d) For inbred genotypes, GCTA-GREML were performed based on the "inbred" algorithm. MAF: minor allele frequency, $\sigma_g^2$: inter-genotype variances, $\sigma_e^2$: intra-genotype variations.

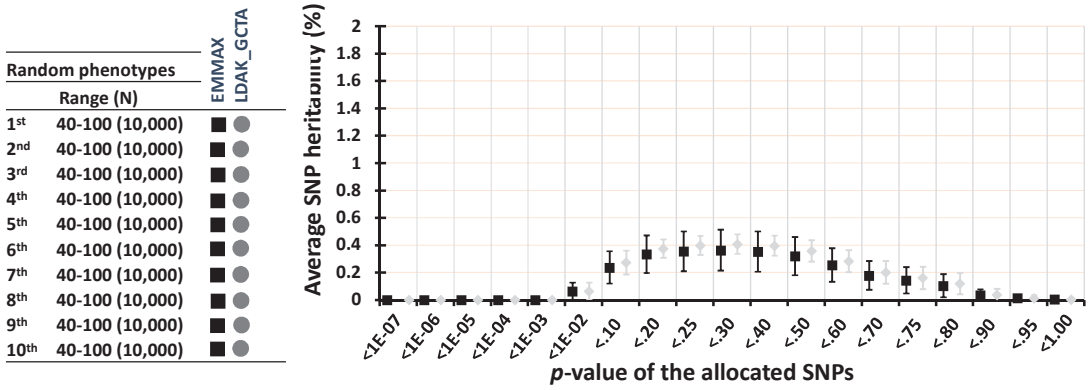
a) Analysis of 10 artificial random phenotypes on a panel of 1,000 simulated inbred genotypes (2000k SNPs; MAF = 1.6%—10%)




b) Analysis of 10 artificial random phenotypes on a panel of 1,000 simulated inbred genotypes (100 SNPs; MAF = 1.6%—10%)




c) Analysis of 10 artificial random phenotypes on a panel of 10,000 simulated inbred genotypes (100 SNPs; MAF = 1.6%—10%)

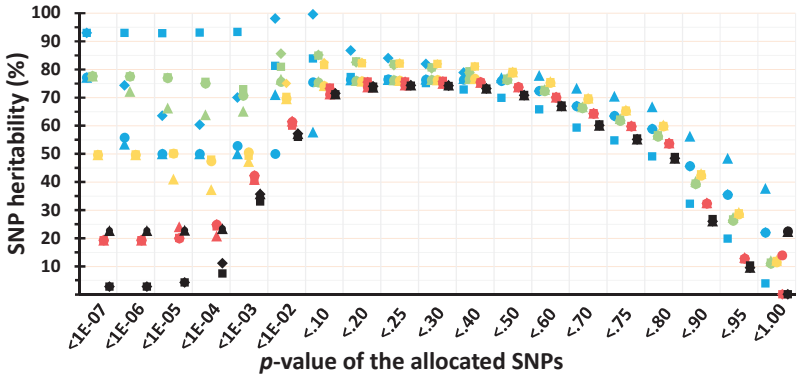



**Supplementary Fig. 3: The SNP heritability profile for simulated random phenotypes (quantitative) and SNPs.** An initial GWAS was performed using EMMAX to determine the association levels for the SNPs. The four different approaches of EMMAX, LDAK-GCTA, LDAK-Thin, and GCTA-GREML (by "inbred" algorithm) were then applied to estimate the SNP heritability for the clusters of SNP hits captured using the association *p*-values. N represents the sample sizes, *i.e.*, number of genotypes. MAF stands for the minor allele frequency. The average SNP heritability estimates were obtained for 10 independent random phenotypes within identical ranges. The error bars represent 99% confidence intervals.

**a) Analysis of 5 artificial non-random phenotypes on a panel of 1,000 simulated inbred genotypes (200,011 SNPs; MAF = 1.6%—10%)**

| Phenotypes | The introduced non-random and associated SNP positions | | | | |
|---|---|---|---|---|---|
| | Ref*i*/Ref*i* Genotype | Alt*i*/Alt*i* Genotype | | | EMMAX / LDAK_GCTA / LDAK_Thin / GCTA_GREML |
| | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | |
| 1st | 40-50 (900) | 80-100 (100) | 6.995 | 93.005 | |
| 2nd | 40-60 (900) | 75-100 (100) | 22.389 | 77.611 | |
| 3rd | 40-70 (900) | 70-100 (100) | 50.267 | 49.733 | |
| 4th | 40-80 (900) | 60-100 (100) | 80.690 | 19.310 | |
| 5th | 40-90 (900) | 50-100 (100) | 97.170 | 2.830 | |

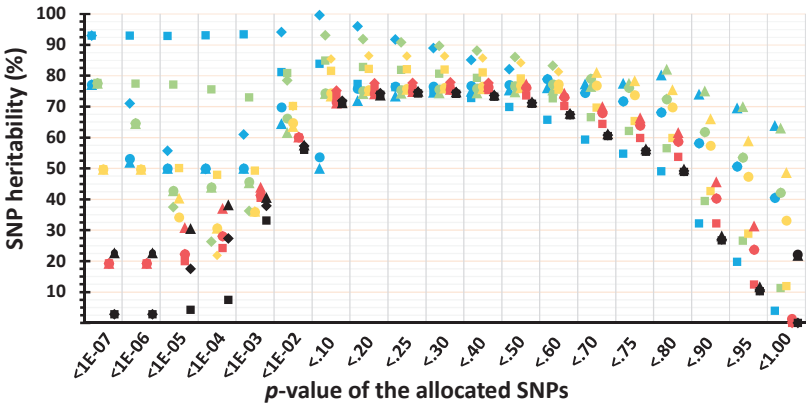
*i*: 1st-12th associated SNP.




**b) Analysis of 5 artificial non-random phenotypes on a panel of 1,000 simulated inbred genotypes (200,023 SNPs; MAF = 1.6%—10%)**

| Phenotypes | The introduced non-random and associated SNP positions | | | | |
|---|---|---|---|---|---|
| | Ref*i*/Ref*i* Genotype | Alt*i*/Alt*i* Genotype | | | EMMAX / LDAK_GCTA / LDAK_Thin / GCTA_GREML |
| | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | |
| 1st | 40-50 (900) | 80-100 (100) | 6.995 | 93.005 | |
| 2nd | 40-60 (900) | 75-100 (100) | 22.389 | 77.611 | |
| 3rd | 40-70 (900) | 70-100 (100) | 50.267 | 49.733 | |
| 4th | 40-80 (900) | 60-100 (100) | 80.690 | 19.310 | |
| 5th | 40-90 (900) | 50-100 (100) | 97.170 | 2.830 | |

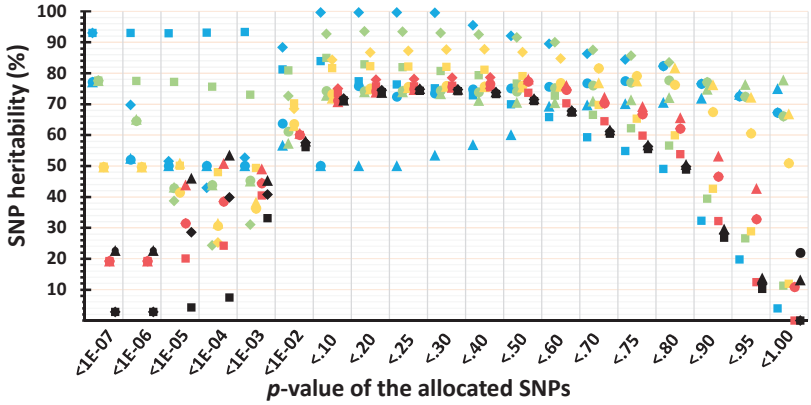
*i*: 1st-24th associated SNP.




**c) Analysis of 5 artificial non-random phenotypes on a panel of 1,000 simulated inbred genotypes (200,047 SNPs; MAF = 1.6%—10%)**

| Phenotypes | The introduced non-random and associated SNP positions | | | | |
|---|---|---|---|---|---|
| | Ref*i*/Ref*i* Genotype | Alt*i*/Alt*i* Genotype | | | EMMAX / LDAK_GCTA / LDAK_Thin / GCTA_GREML |
| | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | |
| 1st | 40-50 (900) | 80-100 (100) | 6.995 | 93.005 | |
| 2nd | 40-60 (900) | 75-100 (100) | 22.389 | 77.611 | |
| 3rd | 40-70 (900) | 70-100 (100) | 50.267 | 49.733 | |
| 4th | 40-80 (900) | 60-100 (100) | 80.690 | 19.310 | |
| 5th | 40-90 (900) | 50-100 (100) | 97.170 | 2.830 | |

*i*: 1st-48th associated SNP.

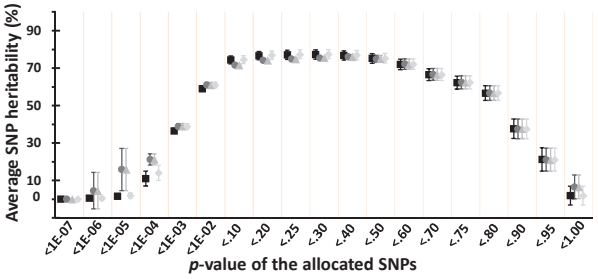



**Supplementary Fig. 4: The SNP heritability profile for simulated quantitative-phenotypes and SNPs, with predesignated SNPs associated at varying rates.** The SNP heritability estimations were predictable at or above the predefined levels, given that other SNPs could also, by chance, account for additional phenotypic variance. An initial GWAS was performed using EMMAX to determine the association levels for the SNPs. The four different approaches of EMMAX, LDAK-GCTA, LDAK-Thin, and GCTA-GREML (by "inbred" algorithm) were then applied to estimate the SNP heritability for the clusters of SNP hits captured using the association *p*-values. N represents the sample sizes, *i.e.*, number of genotypes. MAF stands for the minor allele frequency. Ensuring additive-only effects, the non-random phenotypes were defined among the genotypic groups of the introduced associated SNP positions, *i.e.*, Ref*i*/Ref*i* and Alt*i*/Alt*i* for the simulated inbred genotypes. The intra-genotype assignment of the phenotypes, while being within their specific ranges, was random. The inter-genotype variances of the phenotypes are shown as $\sigma_g^2$. $\sigma_e^2$ represents the intra-genotype variations.

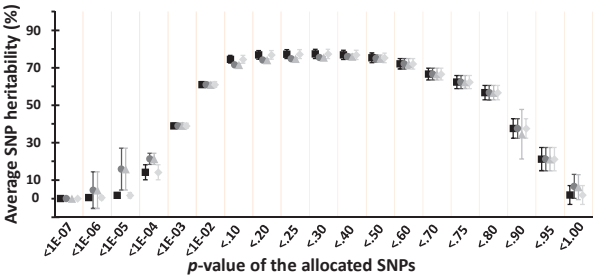
**Analysis of 10 artificial random phenotypes on a panel of 1,000 simulated inbred genotypes (200,000 SNPs; MAF = 1.6%–10%)**

| Random phenotypes | Range (N) | EMMAX | LDAK_GCTA | LDAK_Thin | GCTA_GREML |
|---|---|---|---|---|---|
| 1st | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 2nd | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 3rd | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 4th | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 5th | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 6th | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 7th | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 8th | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 9th | 40-100 (1,000) | ■ | ● | ▲ | ◆ |
| 10th | 40-100 (1,000) | ■ | ● | ▲ | ◆ |

Linkage Equilibrium; 300kb distance between adjacent SNPs



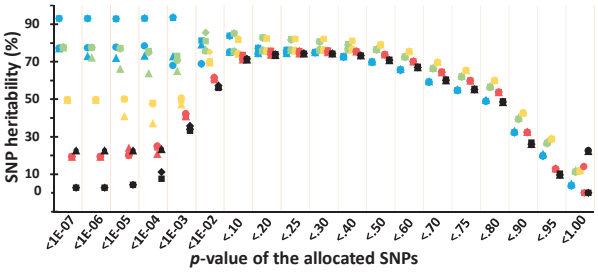

Linkage Disequilibrium; no distance between adjacent SNPs



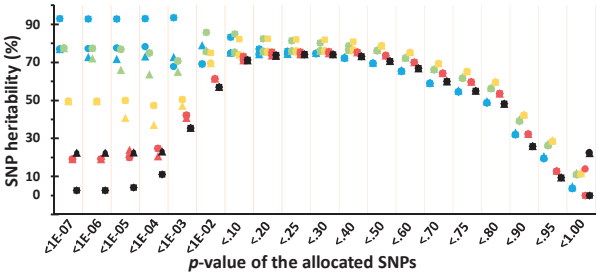

**Analysis of 5 artificial non-random phenotypes on a panel of 1,000 simulated inbred genotypes (200,000 SNPs; MAF = 1.6%–10%)**

| Phenotypes | The introduced non-random and associated SNP position | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Ref/Ref Genotype | Alt/Alt Genotype | | | EMMAX | LDAK_GCTA | LDAK_Thin | GCTA_GREML |
| | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | | | | |
| 1st | 40-50 (900) | 80-100 (100) | 6.995 | 93.005 | ■ | ● | ▲ | ◆ |
| 2nd | 40-60 (900) | 75-100 (100) | 22.389 | 77.611 | ■ | ● | ▲ | ◆ |
| 3rd | 40-70 (900) | 70-100 (100) | 50.267 | 49.733 | ■ | ● | ▲ | ◆ |
| 4th | 40-80 (900) | 60-100 (100) | 80.690 | 19.310 | ■ | ● | ▲ | ◆ |
| 5th | 40-90 (900) | 50-100 (100) | 97.170 | 2.830 | ■ | ● | ▲ | ◆ |

Linkage Equilibrium; 300kb distance between adjacent SNPs



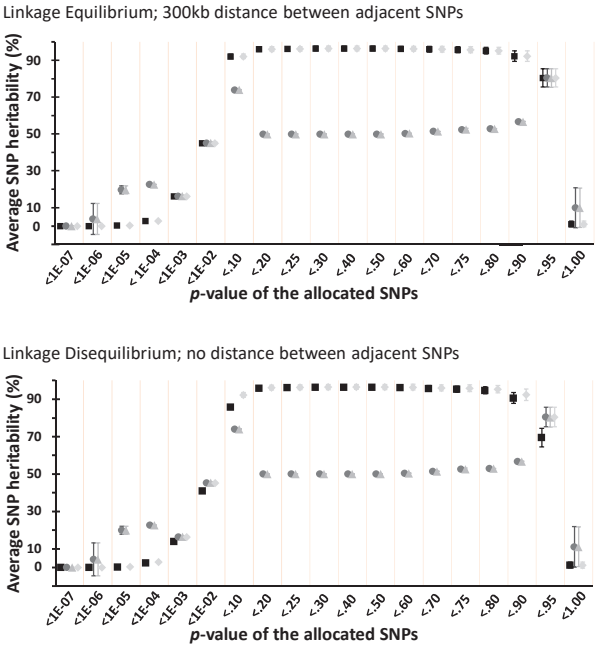

Linkage Disequilibrium; no distance between adjacent SNPs




**Supplementary Fig. 5 continues…**

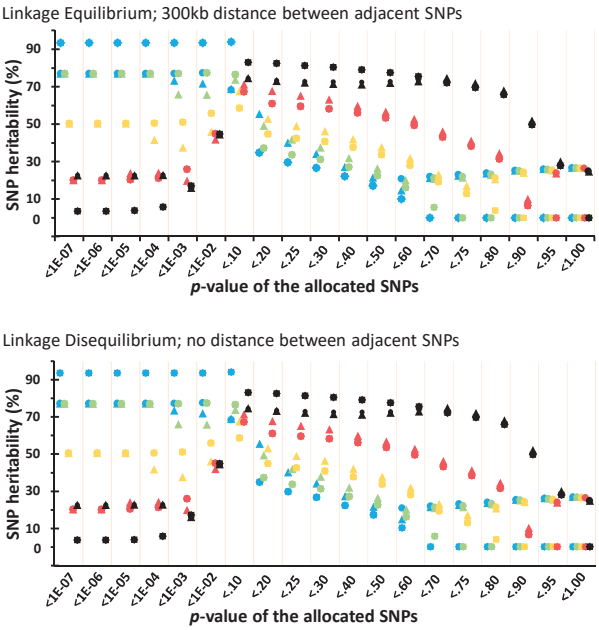
**Analysis of 10 artificial random phenotypes on a panel of 10,000 simulated inbred genotypes (200,000 SNPs; MAF = 1.6%–10%)**

| Random phenotypes | Range (N) | EMMAX | LDAK_GCTA | LDAK_Thin | GCTA_GREML |
|---|---|---|---|---|---|
| 1st | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 2nd | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 3rd | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 4th | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 5th | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 6th | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 7th | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 8th | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 9th | 40-100 (10,000) | ■ | ● | ▲ | ◆ |
| 10th | 40-100 (10,000) | ■ | ● | ▲ | ◆ |




**Analysis of 5 artificial non-random phenotypes on a panel of 10,000 simulated inbred genotypes (200,000 SNPs; MAF = 1.6%–10%)**

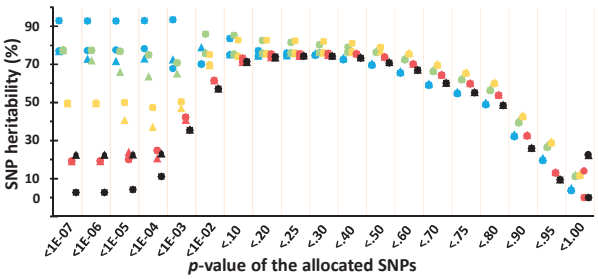
| Phenotypes | The introduced non-random and associated SNP position | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Ref/Ref Genotype range (N) | Alt/Alt Genotype range (N) | $\sigma_g^2$ (%) | $\sigma_e^2$ (%) | EMMAX | LDAK_GCTA | LDAK_Thin | GCTA_GREML |
| 1st | 40-50 (9,000) | 80-100 (1,000) | 6.605 | 93.395 | ■ | ● | ▲ | ◆ |
| 2nd | 40-60 (9,000) | 75-100 (1,000) | 22.738 | 77.262 | ■ | ● | ▲ | ◆ |
| 3rd | 40-70 (9,000) | 70-100 (1,000) | 48.937 | 51.063 | ■ | ● | ▲ | ◆ |
| 4th | 40-80 (9,000) | 60-100 (1,000) | 79.675 | 20.325 | ■ | ● | ▲ | ◆ |
| 5th | 40-90 (9,000) | 50-100 (1,000) | 96.309 | 3.691 | ■ | ● | ▲ | ◆ |



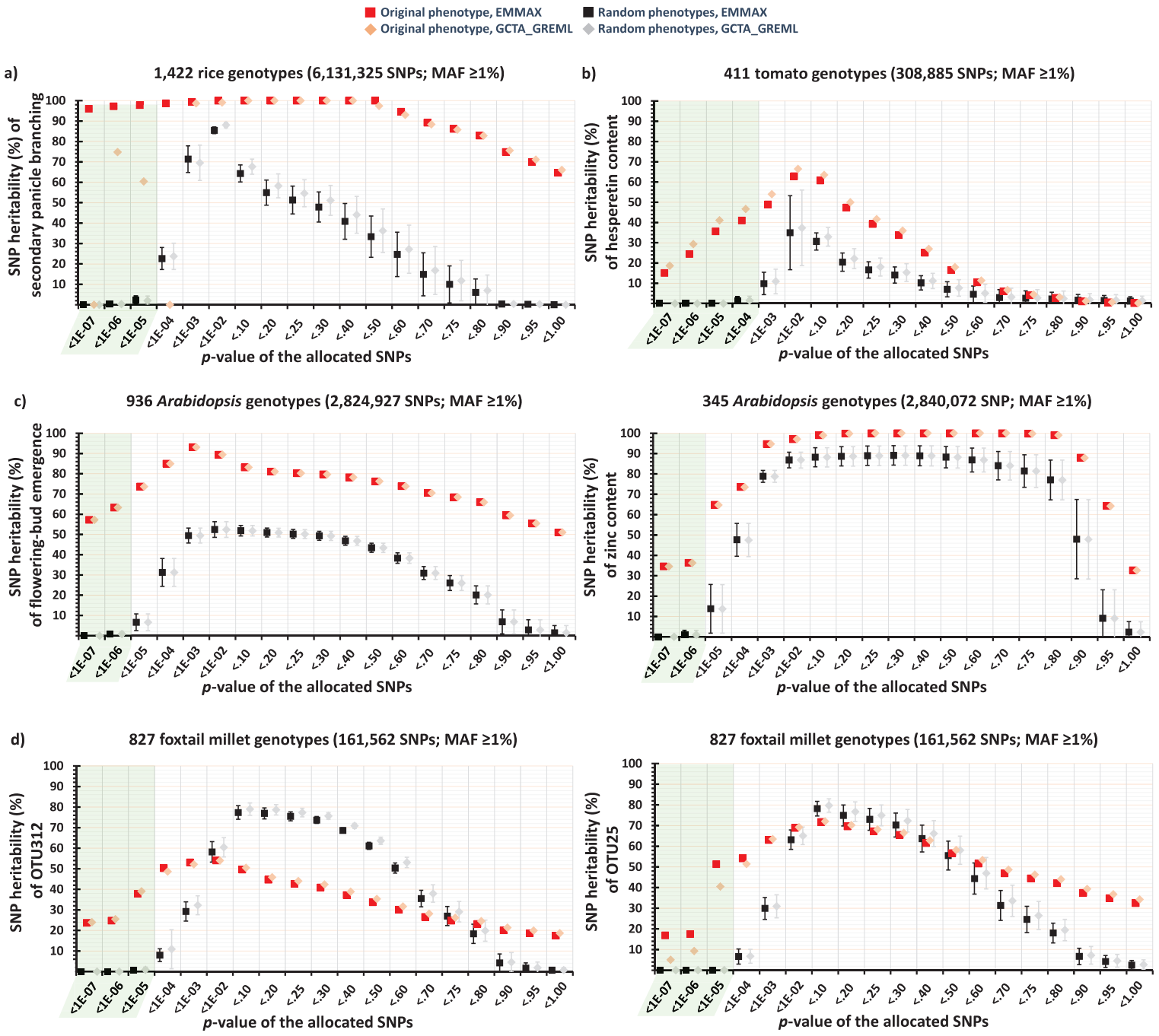

**Supplementary Fig. 5: The SNP heritability profiles for simulated quantitative-phenotypes and SNPs were analyzed comparing linkage disequilibrium to linkage equilibrium.** The analysis compares linkage disequilibrium (i.e., 200k SNPs on a 200kb chromosome) and linkage equilibrium (i.e., 300kb distance between adjacent SNPs). Initial GWAS was performed using EMMAX to realize association levels for the SNPs. The four different approaches of EMMAX, LDAKGCTA, LDAK-Thin, and GCTA-GREML were then applied to estimate SNP heritability for the clusters of SNP hits captured using the association p-values. N represents the sample sizes, i.e., number of genotypes. MAF stands for the minor allele frequency. The average SNP heritability estimates were obtained for 10 independent random phenotypes within identical ranges. The error bars represent 99% confidence intervals. The SNP heritability estimates were obtained for five artificial non-random phenotypes which were associated with a simulated non-random SNP position introduced within the simulated genotypes. By ensuring additive-only effects, the non-random phenotypes were defined at different quantitative ranges among the genotypic groups of the introduced associated SNP position, i.e., Ref/Ref and Alt/Alt for the simulated inbred genotypes. The intragenotype assignment of the phenotypes, while being within their specific ranges, was still random. The inter-genotype variances of the phenotypes are shown as $\sigma_g^2$. $\sigma_e^2$ represents the intra-genotype variations. The analyses of inbred genotypes using GCTA-GREML were performed based on the "inbred" algorithm of GCTA.

**Analysis of 5 artificial non-random phenotypes on a panel of 1,000 simulated inbred genotypes (200,000 SNPs; MAF = 1.6%–50%)**

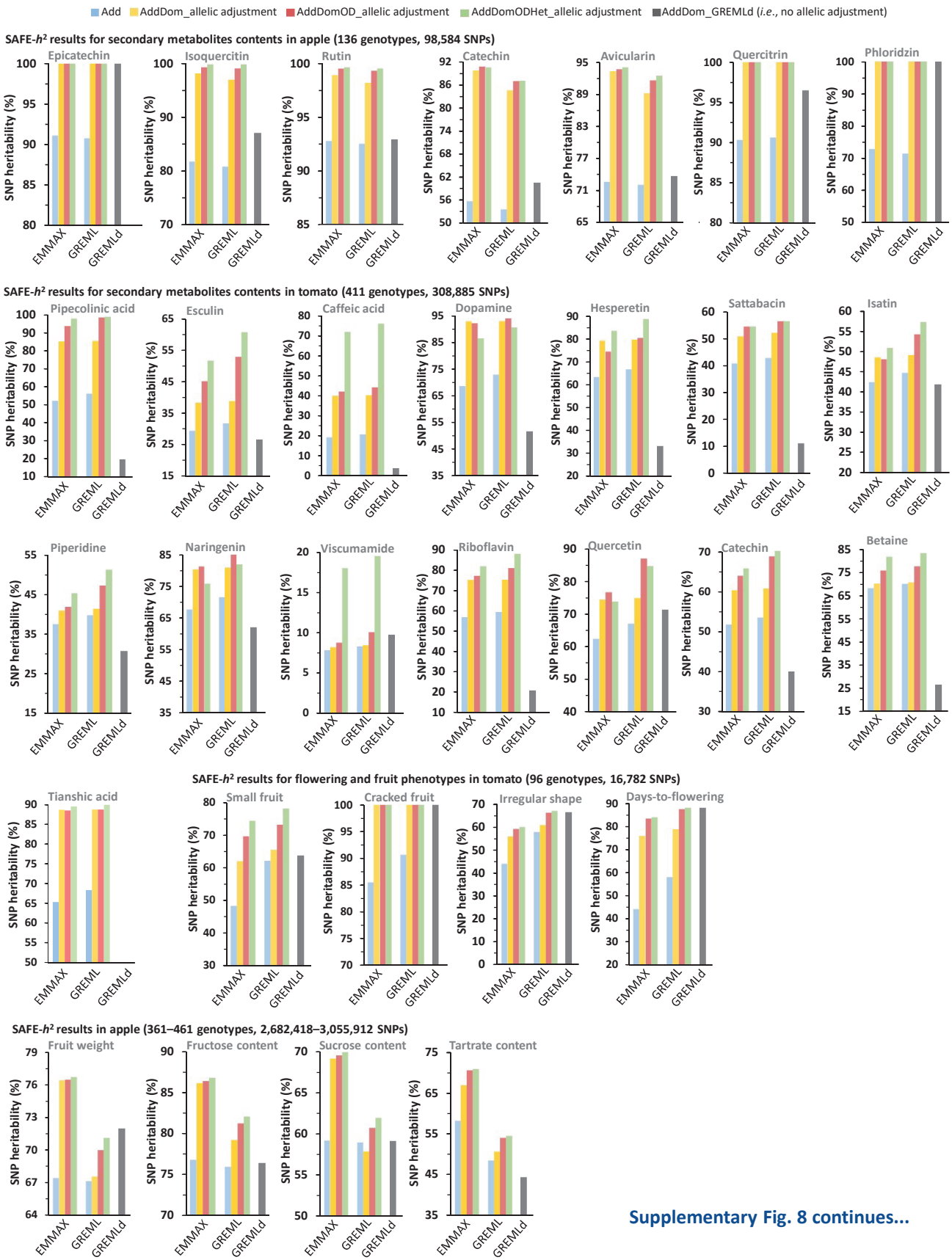
| Phenotypes | The introduced non-random and associated SNP position | | | | |
|---|---|---|---|---|---|
| | Ref/Ref Genotype | Alt/Alt Genotype | | | |
| | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | EMMAX LDAK_GCTA LDAK_Thin GCTA_GREML |
| 1st | 40-50 (900) | 80-100 (100) | 6.995 | 93.005 | ■ ● ▲ ◆ |
| 2nd | 40-60 (900) | 75-100 (100) | 22.389 | 77.611 | ■ ● ▲ ◆ |
| 3rd | 40-70 (900) | 70-100 (100) | 50.267 | 49.733 | ■ ● ▲ ◆ |
| 4th | 40-80 (900) | 60-100 (100) | 80.690 | 19.310 | ■ ● ▲ ◆ |
| 5th | 40-90 (900) | 50-100 (100) | 97.170 | 2.830 | ■ ● ▲ ◆ |



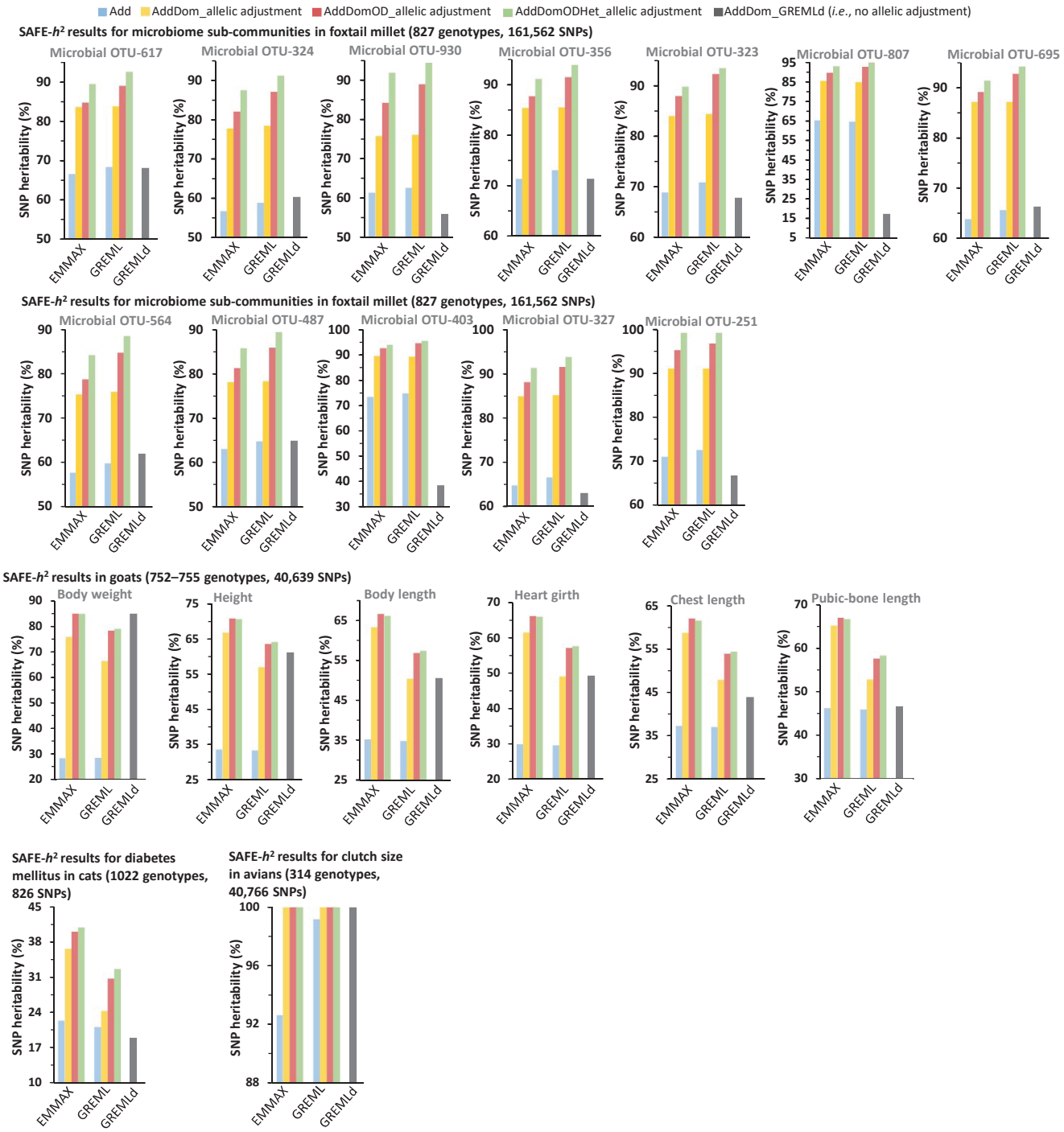

**Supplementary Fig. 6: The SNP heritability profiles for simulated quantitative-phenotypes and SNPs with minor allele frequencies ranging from 1.6%–50%.** Initial GWAS was performed using EMMAX to realize association levels for the SNPs. The four different approaches of EMMAX, LDAKGCTA, LDAK-Thin, and GCTA-GREML were then applied to estimate SNP heritability for the clusters of SNP hits captured using the association *p*-values. N represents the sample sizes, i.e., number of genotypes. MAF stands for the minor allele frequency. The SNP heritability estimates were obtained for five artificial non-random phenotypes which were associated with a simulated non-random SNP position introduced within the simulated genotypes. By ensuring additive-only effects, the non-random phenotypes were defined at different quantitative ranges among the genotypic groups of the introduced associated SNP position, i.e., Ref/Ref and Alt/Alt for the simulated inbred genotypes. The intragenotype assignment of the phenotypes, while being within their specific ranges, was still random. The inter-genotype variances of the phenotypes are shown as $\sigma_g^2$. $\sigma_e^2$ represents the intra-genotype variations. The analyses of inbred genotypes using GCTA-GREML were performed based on the "inbred" algorithm of GCTA.



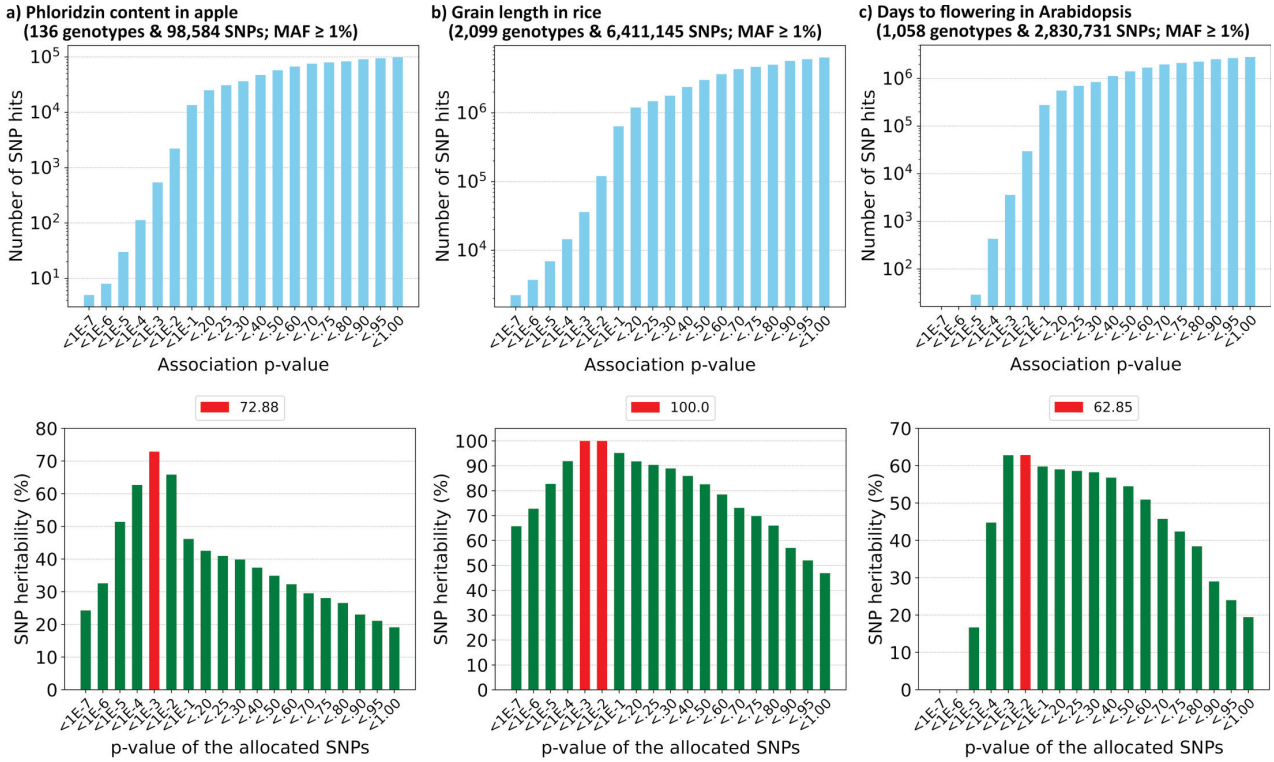

**Supplementary Fig. 7: SAFE-$h^2$ estimates SNP heritability by minimizing the contributions from false-positive SNP hits.** a–d) The safe-intervals for *p*-value are highlighted in light green representing the area with minimum contributions to heritability by false-positive SNP hits. The two different approaches of EMMAX and GCTA-GREML were applied to estimate the SNP heritability for the clusters of SNP hits captured using the association *p*-values. MAF stands for the minor allele frequency. The average estimations are illustrated for six independent random phenotypes (simulated within the range of original phenotype) and the error bars represent standard deviations among the six estimations. GCTA-GREML was performed based on the "inbred" algorithm. OUT: Operational Taxonomic Unit.

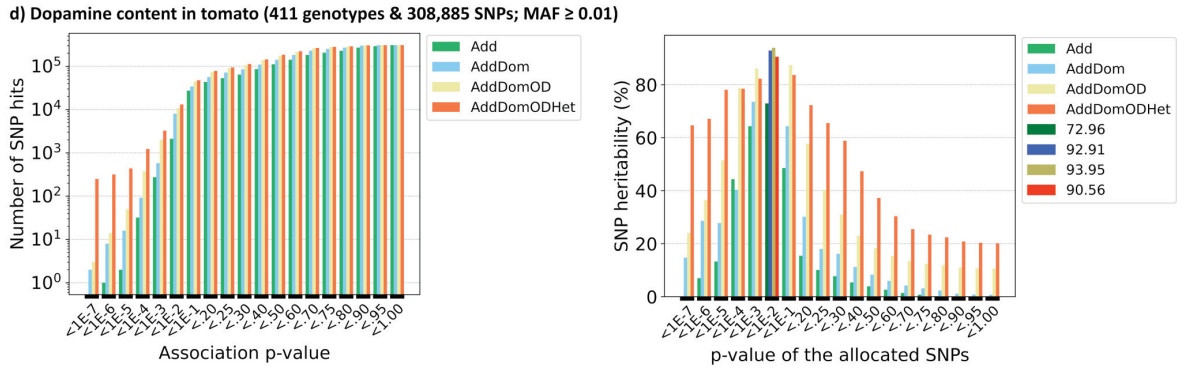


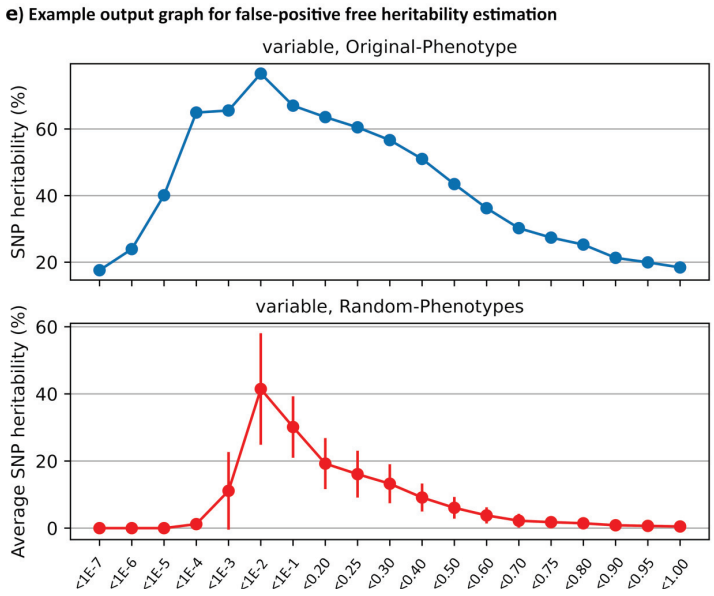

**Supplementary Fig. 8 continues...**




**Supplementary Fig. 8: The SAFE-$h^2$ estimates of SNP heritability using additive-only model and the combined additive and non-additive models, compared to GREMLd estimates.** The analyses of tomato and foxtail millet genotypes using GCTA-GREML/GREMLd models were performed based on the "inbred" algorithm of GCTA. The average inbreeding coefficients (Fhat) of tomato and foxtail millet genotypes were > 0.8. Add: heritability based on additive effects, AddDom: heritability based on additive and dominance effects, AddDomOD: heritability based on additive, dominance and overdominance effects, AddDomODHet: heritability based on additive, dominance, overdominance, and heterosis-like effects, GREMLd: heritability estimated by GREMLd model on additive and dominance effects (*i.e.*, without allelic adjustment), MAF: minor allele frequency, OUT: Operational Taxonomic Unit. MAF was ≥ 1% in all the analyses.




**Supplementary Fig. 9: SAFE-$h^2$ generates bar graphs for both the *p*-value-guided SNP clusters and the SNP heritability profiles.** The SNP heritability estimates are highlighted by red color in (a), (b), (c), and by darker colors in (d). e) Bars represent the standard deviations among six random phenotypes. Add: heritability based on additive effects. AddDom: heritability based on additive and dominance effects. AddDomOD: heritability based on additive, dominance and overdominance effects. AddDomODHet: heritability based on additive, dominance, overdominance, and heterosis-like effects. MAF: minor allele frequency.

**Supplementary Table 1. The PubMed PMID accessions of the recently published studies used to extract SNP heritability estimates for different traits and species.**

| | | | |
|---|---|---|---|
| PMID: 35697868 | PMID: 35115689 | PMID: 36071172 | PMID: 36702997 |
| PMID: 35393594 | PMID: 35115687 | PMID: 35927488 | PMID: 36658437 |
| PMID: 35256806 | PMID: 34741163 | PMID: 35835914 | PMID: 36411363 |
| PMID: 39707004 | PMID: 39578436 | PMID: 39572784 | PMID: 39551815 |
| PMID: 37808854 | PMID: 37961248 | PMID: 39567701 | PMID: 39543288 |
| PMID: 37905151 | PMID: 39567492 | PMID: 39567769 | PMID: 39537646 |
| PMID: 38811692 | PMID: 39281743 | PMID: 39377282 | PMID: 38839884 |
| PMID: 38279346 | PMID: 38127959 | PMID: 37993527 | PMID: 38741014 |
| PMID: 38937676 | PMID: 39348781 | PMID: 39054457 | PMID: 38671320 |
| PMID: 38604911 | PMID: 37877466 | PMID: 38410450 | PMID: 38637617 |
| PMID: 38336809 | PMID: 38849341 | PMID: 39216264 | PMID: 38632349 |
| PMID: 38712091 | PMID: 38559127 | PMID: 39314956 | PMID: 38036780 |
| PMID: 39333847 | PMID: 38746362 | PMID: 38110866 | PMID: 37798380 |
| PMID: 38410442 | PMID: 38259487 | PMID: 39457566 | PMID: 37709864 |
| PMID: 38352337 | PMID: 38987242 | PMID: 39530122 | PMID: 37653029 |
| PMID: 38548983 | PMID: 38872309 | PMID: 38271974 | PMID: 37592024 |
| PMID: 31891238 | PMID: 33516007 | PMID: 39202351 | PMID: 37550530 |
| PMID: 33531474 | PMID: 38466113 | PMID: 37985822 | PMID: 37337107 |
| PMID: 28761083 | PMID: 38880944 | PMID: 39192095 | PMID: 37248441 |
| PMID: 38070845 | PMID: 39256365 | PMID: 38689001 | PMID: 37156999 |
| PMID: 35835912 | PMID: 35210625 | PMID: 36266505 | PMID: 37081215 |
| PMID: 38341521 | PMID: 38766033 | PMID: 38915500 | PMID: 39572783 |
| PMID: 38040712 | PMID: 38075686 | PMID: 38826224 | PMID: 39371125 |
| PMID: 38721342 | PMID: 39072021 | PMID: 38168200 | PMID: 39358599 |
| PMID: 38643935 | PMID: 37790574 | PMID: 37678792 | PMID: 39349817 |
| PMID: 38710707 | PMID: 38136945 | PMID: 39368282 | PMID: 39039282 |
| PMID: 38496634 | PMID: 37326842 | PMID: 35663382 | PMID: 38862855 |
| PMID: 39532956 | PMID: 39511343 | | |

**Supplementary Table 2. SAFE-$h^2$ improves the heritability estimations for 24 traits in *Arabidopsis* and rice by preventing the negative contributions of weakly-associated genomic regions.**

| | | Number of Genotypes/Variants | SAFE-$h^2_{SNP}$ ([SAFE-$h^2_{SNP}$] - [$h^2_{SNP}$ based on all the SNPs]; [SAFE-$h^2_{SNP}$] - [$h^2_{SNP}$ based on the SNP hits with $p$-value < $10^{-5}$])[a] | | | | |
|---|---|---|---|---|---|---|---|
| | | | EMMAX | LDAK GCTA-model | LDAK Thin-model | GCTA-GREML[b] | Average |
| *Arabidopsis thaliana* phenotypes | CLN | 904 / 2,807,143 | **83.4%** (44.5%; 30.8%) | **75.4%** (36.4%; 22.8%) | **73.5%** (32.2%; 24.5%) | **83.4%** (44.5%; 30.8%) | **78.9%** (39.4%; 27.2%) |
| | FBE | 936 / 2,824,927 | **93.1%** (42.1%; 19.4%) | **77.0%** (26.0%; 3.3%) | **80.1%** (17.0%; 12.1%) | **93.1%** (42.1%; 19.4%) | **85.9%** (31.8%; 13.6%) |
| | DTF | 1,058 / 2,830,731 | **62.9%** (43.4%; 46.2%) | **62.8%** (43.4%; 46.1%) | **61.5%** (40.5%; 38.1%) | **62.9%** (43.4%; 46.2%) | **62.5%** (42.7%; 44.1%) |
| | FN | 409 / 2,628,158 | **62.2%** (62.2%; 53.0%) | **62.1%** (39.8%; 39.0%) | **65.9%** (44.2%; 42.8%) | **62.1%** (62.1%; 52.9%) | **63.1%** (52.1%; 46.9%) |
| | GR | 420 / 2,631,629 | **73.4%** (66.4%; 40.4%) | **73.4%** (66.4%; 40.4%) | **66.0%** (59.6%; 32.4%) | **73.4%** (66.4%; 40.4%) | **71.5%** (64.7%; 38.4%) |
| | RLN | 850 / 2,820,131 | **89.4%** (23.2%; 13.1%) | **80.9%** (14.7%; 4.6%) | **81.1%** (11.7%; 24.5%) | **89.4%** (23.2%; 13.1%) | **85.2%** (18.2%; 13.8%) |
| | RL | 328 / 2,621,362 | **71.5%** ( 71.5%; 49.6%) | **71.5%** (49.6%; 49.7%) | **72.6%** (59.8%; 44.6%) | **71.4%** (71.4%; 46.9%) | **71.8%** (63.1%; 48.4%) |
| | SL | 680 / 2,789,485 | **54.6%** (54.6%; 30.9%) | **54.6%** (54.6%; 31.0%) | **56.2%** (53.7%; 40.4%) | **54.6%** (54.6%; 30.9%) | **55.0%** (54.4%; 33.3%) |
| | As | 344 / 2,839,190 | **100%** (30.8%; 41.3%) | **78.8%** (9.8%; 20.2%) | **93.5%** (30.1%; 34.6%) | **99.5%** (30.3%; 40.9%) | **93.0%** (25.3%; 34.3%) |
| | Cd | 345 / 2,840,072 | **99.3%** (44.0%; 24.9%) | **79.4%** (24.1%; 5.0%) | **83.2%** (18.3%; 10.8%) | **99.3%** (44.0%; 24.9%) | **90.3%** (32.7%; 16.4%) |
| | Li | 345 / 2,840,072 | **97.6%** (56.2%; 69.8%) | **87.6%** (46.3%; 59.8%) | **84.8%** (49.9%; 55.2%) | **97.6%** (56.2%; 69.8%) | **91.9%** (52.2%; 63.7%) |
| | Mn | 345 / 2,840,072 | **100%** (45.5%; 65.3%) | **79.4%** (24.9%; 44.8%) | **84.5%** (31.4%; 50.6%) | **100%** (45.5%; 65.4%) | **91.0%** (36.8%; 56.5%) |
| | Zn | 345 / 2,840,072 | **100%** (67.4%; 35.2%) | **88.0%** (55.4%; 23.2%) | **81.5%** (56.4%; 35.4%) | **100%** (67.4%; 35.2%) | **92.4%** (61.7%; 32.3%) |
| | P | 345 / 2,840,072 | **96.8%** (44.1%; 46.5%) | **85.8%** (33.4%; 35.6%) | **80.0%** (32.3%; 44.0%) | **96.8%** (44.2%; 46.6%) | **89.8%** (38.5%; 43.2%) |
| Rice (*O. sativa*) phenotypes | DTF | 2,102 / 6,414,800 | **99.9%** (10.0%; 16.1%) | **90.0%** (0.0%; 18.1%) | **92.2%** (1.5%; 19.2%) | **99.9%** (11.7%; 17.8%) | **95.6%** (5.6%; 17.8%) |
| | GL | 2,099 / 6,411,145 | **99.9%** (14.0%; 16.0%) | **88.3%** (2.4%; 16.3%) | **91.3%** (3.1%; 17.3%) | **99.9%** (15.2%; 17.2%) | **94.9%** (8.6%;16.7%) |
| | GWE | 2,098 / 6,410,047 | **99.9%** (13.5%; 17.4%) | **88.8%** (2.4%; 16.6%) | **91.4%** (3.3%; 17.3) | **99.9%** (14.6%; 19.1%) | **95.0%** (8.4%; 17.6%) |
| | PL | 1,762 / 6,269,207 | **99.9%** (17.3%; 14.7%) | **86.9%** (4.2%; 14.4%) | **88.9%** (3.3%; 14.3%) | **99.9%** (17.8%; 14.7%) | **93.9%** (10.7%; 14.5%) |
| | GWI | 2,099 / 6,411,145 | **99.9%** (14.8%; 15.7%) | **87.6%** (2.4%; 15.4%) | **90.6%** (3.1%; 16.5%) | **99.9%** (15.9%; 16.8%) | **94.6%** (9.1%; 16.1%) |
| | SD (EC18)[c] | 1,998 / 6393347 | **99.9%** (54.3%; 4.6%) | **85.6%** (39.9%; 16.6%) | **88.3%** (29.7%; 20.1%) | **99.9%** (53.1%; 17.3%) | **93.5%** (44.2%; 14.6%) |
| | SD (EC12)[c] | 1,998 / 6,393,347 | **99.9%** (23.5%; 19.4%) | **88.4%** (11.9%; 15.11%) | **90.5%** (9.8%; 13.5%) | **99.9%** (23.8%; 19.9%) | **94.7%** (17.3%; 17.0%) |
| | SH | 1,743 / 6,254,739 | **99.9%** (16.4%; 13.7%) | **87.6%** (4.0%; 14.5%) | **89.5%** (3.1%; 15.0%) | **99.9%** (17.0%; 14.7%) | **94.3%** (10.1%; 14.4%) |
| | SPB | 1,422 / 6,131,325 | **99.9%** (35.3%; 2.2%) | **94.6%** (29.9%; 27.2%) | **98.3%** (26.9%; 30.0%) | **99.9%** (33.9%; 39.6%) | **98.2%** (31.5%;24.8%) |
| | SF | 2,100 / 6,411,911 | **99.9%** (21.2%; 27.7%) | **90.8%** (12.0%; 18.5%) | **92.5%** (10.4%; 16.2%) | **99.9%** (21.8%; 28.2%) | **95.8%** (16.3%; 22.6%) |
| | **Average** | | ..... (38.2%; 29.8%) | ..... (26.4%; 24.9%) | ..... (26.3%; 27.9%) | ..... (38.3%; 32.0%) | ..... **(32.3%; 28.7%)** |
| | **Minimum** | | ..... (10.0%; 2.2%) | ..... (0.0%; 3.3%) | ..... (1.5%; 10.8%) | ..... (11.7%; 13.1%) | ..... **(5.6%; 13.6%)** |
| | **Maximum** | | ..... (71.5%; 69.8%) | ..... (66.4%; 59.8%) | ..... (59.8%; 55.2%) | ..... (71.4%; 69.8%) | ..... **(64.7%; 63.7%)** |

CLN: cauline leaf number; FBE: days until the emergence of flowering buds; DTF: days until first open flower; FN: number of fruits per plant; GR: average growth rate; RLN: rossette leaf number; RL: root length; SL: stem length; As: arsenic content in leaf; Cd: cadmium content in leaf; Li: lithium content in leaf; Mn: manganese content in leaf; Zn: zinc content in leaf; P: phosphorus content in leaf; GL: grain length; GWE: grain weigth; PL: panicle length; GWI: grain width; SD (EC18): salt injury at EC of 18; SD (EC12): salt injury at EC of 12; SH: seedling height; SPB: secondary panicle branching; SF: spikelet fertility. [a]A minimum MAF of 0.01 was applied in all analyses. [b]Analyzing inbred genotypes of rice and Arabidopsis was performed under the "inbred" algorithm of GCTA. [c]The electrical conductivity.

**Supplementary Table 3. Heritability profiling by SAFE-$h^2$ reveals negative contributions of SNPs with large association *p*-values by examining 50 phenotypes in six species.**

| | Phenotypes | Number of Genotypes/Variants | [SAFE-$h^2_{SNP}$] - [$h^2_{SNP}$ based on all SNPs] | |
|---|---|---|---|---|
| | | | EMMAX | GCTA-GREML[a] |
| **Tomato** | Cracked fruit | 96 / 16,782 | 3.95% | 2.37% |
| | Day-to-flowering | 96 / 16,782 | 0.00% | 0.00% |
| | Small fruit | 96 / 16,782 | 17.96% | 18.66% |
| | Iregular shape | 96 / 16,782 | 2.71% | 2.52% |
| | Caffeic acid | 411 / 308,885 | 12.16% | 12.80% |
| | Catechin | 411 / 308,885 | 22.25% | 21.92% |
| | Tianshic acid | 411 / 308,885 | 65.31% | 68.33% |
| | Dopamine | 411 / 308,885 | 68.64% | 72.95% |
| | Esculin | 411 / 308,885 | 19.53% | 21.04% |
| | Hesperetin | 411 / 308,885 | 61.11% | 66.61% |
| | Naringenin | 411 / 308,885 | 51.78% | 54.29% |
| | Piperidine | 411 / 308,885 | 8.53% | 8.64% |
| | Viscumamide | 411 / 308,885 | 2.83% | 2.82% |
| | Betaine | 411 / 308,885 | 49.64% | 50.07% |
| | Isatin | 411 / 308,885 | 1.42% | 1.28% |
| | Pipecolinic acid | 411 / 308,885 | 52.12% | 56.12% |
| | Riboflavin | 411 / 308,885 | 48.99% | 50.73% |
| | Sattabacin | 411 / 308,885 | 26.69% | 27.50% |
| | Quercetin | 411 / 308,885 | 62.37% | 67.00% |
| **Apple** | Quercetin | 136 / 98,584 | 30.78% | 32.19% |
| | Catechin | 136 / 98,584 | 13.03% | 11.99% |
| | Avicularin | 136 / 98,584 | 72.53% | 71.99% |
| | Phloridzin | 136 / 98,584 | 53.78% | 52.99% |
| | Rutin | 136 / 98,584 | 90.21% | 90.07% |
| | Isoquercetin | 136 / 98,584 | 46.78% | 46.85% |
| | Epicatechin | 136 / 98,584 | 60.93% | 61.52% |
| | Fruit-weight | 368 / 3,055,912 | 2.10% | 2.21% |
| | Sucrose | 361 / 2,927,501 | 20.20% | 20.60% |
| | Fructose | 448 / 2,682,418 | 15.52% | 15.95% |
| | Tartrate | 461 / 2,767,199 | 19.39% | 10.73% |
| **Foxtail millet** | OUT-327 | 827 / 161,562 | 29.10% | 28.84% |
| | OUT-403 | 827 / 161,562 | 44.80% | 44.31% |
| | OUT-487 | 827 / 161,562 | 15.17% | 14.61% |
| | OUT-564 | 827 / 161,562 | 19.23% | 19.09% |
| | OUT-323 | 827 / 161,562 | 49.12% | 49.73% |
| | OUT-356 | 827 / 161,562 | 41.99% | 41.79% |
| | OUT-930 | 827 / 161,562 | 35.71% | 35.13% |
| | OUT-324 | 827 / 161,562 | 29.64% | 29.98% |
| | OUT-617 | 827 / 161,562 | 26.30% | 25.78% |
| | OUT-807 | 827 / 161,562 | 53.93% | 52.17% |
| | OUT-695 | 827 / 161,562 | 39.36% | 39.63% |
| | OUT-251 | 827 / 161,562 | 49.04% | 49.31% |
| **Cats** | Dibetes mellitus | 1,022 / 826 | 12.07% | 11.39% |
| **Goats** | Height | 752 / 40,639 | 2.32% | 2.37% |
| | Body-length | 752 / 40,639 | 1.76% | 1.85% |
| | Chest length | 752 / 40,639 | 0.23% | 0.23% |
| | Heart-girth | 752 / 40,639 | 2.60% | 2.75% |
| | Pubic bone-length | 752 / 40,639 | 0.00% | 0.03% |
| | Weight | 752 / 40,639 | 4.86% | 5.03% |
| **Wild Avians** | Clutch-size | 314 / 40,766 | 52.92% | 57.16% |
| **Average** | | | 30.27% | 30.68% |

[a]The "inbred" algorithm of GCTA was used to analyze the tomato and foxtail millet genotypes. The average inbreeding coefficients (Fhat) for these gentypes were > 0.8.

**Supplementary Table 4: BayesR SNP heritability estimates for simulated quantitative-phenotypes using the 200K simulated SNPs in LE and LD.**

| | The introduced non-random (associated) SNP position | | | | On a 200kb chromosome (Linkage Disequilibrium) | | | 300kb distance between two adjacent SNPs (Linkage Equilibrium) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Ref/Ref Genotype | Alt/Alt Genotype | | | BayesR | | Process time (BayesR/SAFE-$h^2$) | BayesR | | Process time (BayesR/SAFE-$h^2$) |
| Phenotypes | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | $h^2$ (%) | Process time | | $h^2$ (%) | Process time | |
| 1st | 40-50 (900) | 80-100 (100) | 6.995 | 93.005 | na | 150 min | 10.7 | na | 150 min | 10.7 |
| 2nd | 40-60 (900) | 75-100 (100) | 22.389 | 77.611 | na | 150 min | 10.7 | na | 150 min | 10.7 |
| 3rd | 40-70 (900) | 70-100 (100) | 50.267 | 49.733 | na | 150 min | 10.7 | 34.67 | 150 min | 10.7 |
| 4th | 40-80 (900) | 60-100 (100) | 80.69 | 19.31 | 6.04 | 150 min | 10.7 | 8.33 | 150 min | 10.7 |
| 5th | 40-90 (900) | 50-100 (100) | 97.17 | 2.83 | 3.47 | 150 min | 10.7 | 1.07 | 150 min | 10.7 |

Ref/Ref and Alt/Alt for the introduced associated single SNP, N: Number of genotypes (individuals)

$\sigma_g^2$: Inter-genotype variances of the phenotypes

$\sigma_e^2$: Intra-genotype variations

**Supplementary Table 5: BayesR SNP heritability estimates for simulated quantitative-phenotypes using the chr. 22 SNPs from human HapMap phase 1 genotypes.**

| | The introduced non-random (associated) SNP position | | | | | Real genotypes based on Chr22 + introduced single associated SNP | | |
|---|---|---|---|---|---|---|---|---|
| | Ref/Ref Genotype | Ref/Alt Genotype | Alt/Alt Genotype | | | BayesR | | Process time (BayesR/SAFE-$h^2$) |
| Phenotypes | range (N) | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | $h^2$ (%) | Process time | |
| 1st | 40-50 (982) | 80-100 (103) | 120-150 (7) | 5.704 | 94.296 | na | 120 min | 8.6 |
| 2nd | 40-60 (982) | 75-100 (103) | 112-150 (7) | 21.037 | 78.963 | na | 120 min | 8.6 |
| 3rd | 40-70 (982) | 70-100 (103) | 105-150 (7) | 41.501 | 58.499 | na | 120 min | 8.6 |
| 4th | 40-80 (982) | 60-100 (103) | 90-150 (7) | 72.099 | 27.901 | 22.21 | 120 min | 8.6 |
| 5th | 40-90 (982) | 50-100 (103) | 75-150 (7) | 91.416 | 8.584 | 6.17 | 120 min | 8.6 |

Ref/Ref, Ref/Alt, and Alt/Alt for the introduced associated single SNP

$\sigma_g^2$: Inter-genotype variances of the phenotypes

$\sigma_e^2$: Intra-genotype variations

N: number of genotypes (individuals)

**Supplementary Table 6: The SNP heritability estimates for real phenotypes using BayesR and SAFE-$h^2$.**

| | | Heritability estimates ($h^2$ %) | | |
|---|---|---|---|---|
| | Phenotype | BayesR | SAFE-$h^2$ by EMMAX | SAFE-$h^2$ by GCTA_GREML |
| **Foxtail millet** | | | | |
| 827 genotypes & 161k SNPs | OUT_25 | 27.84 | 71.73 | 72.05 |
| 827 genotypes & 161k SNPs | OUT_312 | 9.47 | 54.05 | 54.01 |
| **Tomato** | | | | |
| 411 genotypes & 308k SNPs | Betaine | 17.4 | 68.08 | 70.02 |
| 411 genotypes & 308k SNPs | Hesperetin | 7.37 | 62.85 | 66.41 |
| **Arabidopsis** | | | | |
| 345 genotypes & 2.8M SNPs | Root-length | 6.3 | 71.5 | 71.4 |
| 936 genotypes & 2.8M SNPs | FBE | 69.41 | 93.1 | 93.1 |
| **Apple** | | | | |
| 136 genotypes & 98k SNPs | Avicularin | 24.2 | 72.63 | 77.08 |
| 136 genotypes & 98k SNPs | Catechin | 35.48 | 55.73 | 53.61 |

FBE stands for flowering bud emergence.

**Supplementary Table 7. SNP heritability estimations on additive or additive + non-additive effects of a biallelic locus.**

| Phenotypes | | | | | | SAFE-$h^2$ estimates of SNP-heritability | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Ref/Ref Genotype | Ref/Alt Genotype | Alt/Alt Genotype | | | using EMMAX model | | | | using GEMMA model | | | |
| Data_Genetic effect | range (N) | range (N) | range (N) | $\sigma_e^2$ (%) | $\sigma_g^2$ (%) | A[a] | AD[b] | ADO[c] | ADOH[d] | A[a] | AD[b] | ADO[c] | ADOH[d] |
| 1st_Additice | 40-50 (982) | 80-100 (103) | 120-150 (7) | 5.7 | 94.3 | 94.29 | 94.29 | 94.29 | 94.29 | 94.28 | 94.28 | 94.28 | 94.28 |
| 1st_Dominance | 40-50 (982) | 120-150 (103) | 120-150 (7) | 2.16 | 97.84 | 92.8 | ▲ 97.83 | ▲ 97.83 | ▲ 97.83 | 92.8 | ▲ 97.83 | ▲ 97.83 | ▲ 97.83 |
| 1st_Overdominance | 40-50 (982) | 120-150 (103) | 80-100 (7) | 2.35 | 97.65 | 84.77 | ▲ 95.9 | ▲ 97.65 | ▲ 97.65 | 84.76 | ▲ 95.9 | ▲ 97.64 | ▲ 97.64 |
| 1st_Heterosis-like | 40-50 (982) | 120-150 (103) | 40-50 (7) | 6.11 | 93.89 | 70.51 | ▲ 86.47 | ▲ 91.72 | ▲ 93.86 | 70.49 | ▲ 86.46 | ▲ 91.71 | ▲ 93.86 |
| 2nd_Additive | 40-60 (982) | 75-100 (103) | 112-150 (7) | 21.04 | 78.96 | 78.86 | 78.86 | 78.86 | 78.86 | 78.85 | 78.85 | 78.85 | 78.85 |
| 2nd_Dominance | 40-60 (982) | 112-150 (103) | 112-150 (7) | 12.75 | 87.25 | 82.1 | ▲ 87.23 | ▲ 87.23 | ▲ 87.23 | 82.1 | ▲ 87.22 | ▲ 87.22 | ▲ 87.22 |
| 2nd_Overdominance | 40-60 (982) | 112-150 (103) | 75-100 (7) | 14.7 | 85.3 | 71.58 | ▲ 82.7 | ▲ 85.13 | ▲ 85.13 | 71.56 | ▲ 82.69 | ▲ 85.12 | ▲ 85.12 |
| 2nd_Heterosis-like | 40-60 (982) | 112-150 (103) | 40-60 (7) | 8.1 | 91.9 | 70.56 | ▲ 85.58 | ▲ 90.28 | ▲ 91.89 | 70.54 | ▲ 85.57 | ▲ 90.27 | 91.89 |
| 3rd_Additive | 40-70 (982) | 70-100 (103) | 105-150 (7) | 41.5 | 58.5 | 58.34 | 58.34 | 58.34 | 58.34 | 58.31 | 58.31 | 58.31 | 58.31 |
| 3rd_Dominance | 40-70 (982) | 105-150 (103) | 105-150 (7) | 15.47 | 84.53 | 79.48 | ▲ 84.51 | ▲ 84.51 | ▲ 84.51 | 79.47 | ▲ 84.5 | ▲ 84.5 | ▲ 84.5 |
| 3rd_Overdominance | 40-70 (982) | 105-150 (103) | 70-100 (7) | 15.16 | 84.84 | 72.05 | ▲ 82.65 | ▲ 84.76 | ▲ 84.76 | 72.04 | ▲ 82.64 | ▲ 84.74 | ▲ 84.74 |
| 3rd_Heterosis-like | 40-70 (982) | 105-150 (103) | 40-70 (7) | 16.44 | 83.56 | 62.59 | ▲ 76.85 | ▲ 81.56 | ▲ 83.52 | 62.57 | ▲ 76.84 | ▲ 81.55 | ▲ 83.5 |
| 4th_Additive | 40-80 (982) | 60-100 (103) | 90-150 (7) | 72.1 | 27.9 | 27.34 | 27.34 | 27.34 | 27.34 | 27.33 | 27.33 | 27.33 | 27.33 |
| 4th_Dominance | 40-80 (982) | 90-150 (103) | 90-150 (7) | 33.36 | 66.64 | 62.6 | ▲ 66.6 | ▲ 66.6 | ▲ 66.6 | 62.58 | ▲ 66.59 | ▲ 66.59 | ▲ 66.59 |
| 4th_Overdominance | 40-80 (982) | 90-150 (103) | 60-100 (7) | 32.11 | 67.89 | 56.78 | ▲ 65.73 | ▲ 67.72 | ▲ 67.72 | 56.76 | ▲ 65.71 | ▲ 67.7 | ▲ 67.7 |
| 4th_Heterosis | 40-80 (982) | 90-150 (103) | 40-80 (7) | 35.4 | 64.6 | 51.1 | ▲ 60.96 | ▲ 63.84 | ▲ 64.5 | 50.98 | ▲ 60.94 | ▲ 63.82 | ▲ 64.49 |
| 5th_Additive | 40-90 (982) | 50-100 (103) | 75-150 (7) | 91.42 | 8.58 | 7.7 | 7.7 | 7.7 | 7.7 | 7.69 | 7.69 | 7.69 | 7.69 |
| 5th_Dominance | 40-90 (982) | 75-150 (103) | 75-150 (7) | 61.67 | 38.33 | 36.61 | ▲ 38.24 | ▲ 38.24 | ▲ 38.24 | 36.59 | ▲ 38.22 | ▲ 38.22 | ▲ 38.22 |
| 5th_Overdominance | 40-90 (982) | 75-150 (103) | 50-100 (7) | 55.12 | 44.88 | 34.61 | ▲ 41.87 | ▲ 44.1 | ▲ 44.82 | 34.59 | ▲ 41.84 | ▲ 44.08 | ▲ 44.8 |
| 5th_Heterosis-like | 40-90 (982) | 75-150 (103) | 40-90 (7) | 56 | 44 | 34.25 | ▲ 41.23 | ▲ 43.33 | ▲ 43.93 | 34.23 | ▲ 41.21 | ▲ 43.3 | ▲ 43.91 |
| 6th_Additive | 40-95 (982) | 45-100 (103) | 50-150 (7) | 97.56 | 2.44 | 2.3 | 2.3 | 2.3 | 2.3 | 2.29 | 2.29 | 2.29 | 2.29 |
| 6th_Dominance | 40-95 (982) | 50-150 (103) | 50-150 (7) | 76.6 | 23.4 | 20.67 | ▲ 23.03 | ▲ 23.31 | ▲ 23.31 | 20.65 | ▲ 23.05 | ▲ 23.29 | ▲ 23.29 |
| 6th_Overdominance | 40-95 (982) | 50-150 (103) | 45-100 (7) | 76.64 | 23.36 | 19.5 | ▲ 22.56 | ▲ 23.24 | ▲ 23.24 | 19.49 | ▲ 22.55 | ▲ 23.22 | ▲ 23.22 |
| 6th_Heterosis-like | 40-95 (982) | 50-150 (103) | 40-95 (7) | 76.57 | 23.43 | 16.19 | ▲ 20.62 | ▲ 22.57 | ▲ 23.21 | 16.17 | ▲ 20.61 | ▲ 22.56 | ▲ 23.19 |

Green triangles indicate improved estimation when compared to the analysis on the original genotypes (a). Red triangles indicate underestimations when compared to the analysis on the original genotypes (a). $\sigma_e^2$ & $\sigma_g^2$ represent the the enviromental and genetic variances, respectively. $\sigma_g^2$ is equivalent to the broad-sense heritability. [a]On original genotypes to capture additive effects; [b]On adjusted genotypes for dominance effects to capture additive and dominance effects; [c]On adjusted genotypes for dominance and overdominance effects to capture additive, dominance, and overdominace effects; [d]On adjusted genotypes for dominance,overdominance, and heterosis-like effects to capture additive, dominance, overdominace, and heterosis-like effects.

**Supplementary Table 8. Heritability estimates based on *p*-values, calculated using EMMAX and GEMMA.**

| | | Heritability estimates ($h^2$ %) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Association analysis using EMMAX (*p*-value source) | | | | | | Association analysis using GEMMA (*p*-value source) | | | | | |
| | Phenotype | EMMAX (on all SNPs) | SAFE-$h^2$ using EMMAX | GCTA_GREML (on all SNPs) | SAFE-$h^2$ using GCTA_GREML | GEMMA (on all SNPs) | SAFE-$h^2$ using GEMMA | EMMAX (on all SNPs) | SAFE-$h^2$ using EMMAX | GCTA_GREML (on all SNPs) | SAFE-$h^2$ using GCTA_GREML | GEMMA (on all SNPs) | SAFE-$h^2$ using GEMMA |
| **Foxtail millet** | | | | | | | | | | | | | |
| 827 genotypes & 161k SNPs | OUT_25 | 32.42 | 71.73 | 34.26 | 72.05 | 32.94 | 74.41 | 32.43 | 94.97 | 34.26 | 95.17 | 32.94 | 97.97 |
| 827 genotypes & 161k SNPs | OUT_312 | 17.48 | 54.05 | 18.68 | 54.01 | 17.86 | 57.37 | 17.49 | 85.01 | 18.68 | 86.02 | 17.86 | 88.567 |
| **Tomato** | | | | | | | | | | | | | |
| 411 genotypes & 308k SNPs | Betaine | 18.43 | 68.08 | 19.99 | 70.02 | 18.02 | 67.35 | 18.43 | 73.23 | 19.99 | 74.85 | 18.02 | 74.46 |
| 411 genotypes & 308k SNPs | Hesperetin | 0.21 | 63.39 | 0.23 | 66.84 | 0.26 | 62.5 | 0.21 | 62.04 | 0.23 | 65.58 | 0.26 | 61.18 |
| **Arabidopsis** | | | | | | | | | | | | | |
| 345 genotypes & 2.8M SNPs | Root-length | 0.01 | 71.47 | 0.01 | 71.42 | 0 | 71.41 | 0.01 | 71.58 | 0 | 71.52 | 0 | 71.52 |
| 936 genotypes & 2.8M SNPs | FBE | 51.03 | 93.14 | 51 | 93.13 | 51 | 93.13 | 51.03 | 93.58 | 51 | 93.58 | 51 | 93.58 |
| **Apple** | | | | | | | | | | | | | |
| 136 genotypes & 98k SNPs | Avicularin | 0.09 | 72.63 | 0.09 | 72.08 | 0.01 | 80.34 | 0.09 | 79.74 | 0.09 | 79.26 | 1.47 | 86.54 |
| 136 genotypes & 98k SNPs | Catechin | 42.69 | 55.73 | 41.61 | 53.61 | 45.05 | 58.88 | 42.69 | 95.77 | 41.61 | 95.65 | 45.05 | 97.09 |

**Supplementary text 1: Independently simulated datasets cannot be used as replicates**.

When simulating datasets, every dataset shows the similar downward bias in heritability estimates by adding unassociated SNPs, i.e., with large association *p*-values (see Fig. 2b and Suppl. Fig. 2-6). This downward bias is also observed in real-datasets (see Fig 2c, Suppl. Fig. 7, Suppl. Fig. 9, Suppl. Table 2 and 3).

We can use *p*-values from one simulated dataset to make heritability estimates in another simulated dataset (as a replicated study), but only if the second dataset is not independently simulated from scratch. Under these conditions, we will observe the same pattern of downward bias shown in Fig. 2b and Suppl. Figs. 2-6.

An example of a proper replication:

**Original study**

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Individual/Genotypes: | 1, | 2, | 3, | 4, | 5, | 6, | 7, | 8, | 9, | 10, | 11, | 12 |
| Phenotype A: | 1, | 2, | 5, | 3, | 2, | 3, | 20, | 22, | 28, | 20, | 25, | 22 |

**Replicated study**

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Individual/Genotypes: | 1, | 2, | 3, | 4, | 5, | 6, | 7, | 8, | 9, | 10, | 11, | 12 |
| Phenotype A: | 1±0.2 | 2±0.2 | 5±0.2 | 3±0.2 | 2±0.2 | 3±0.2 | 20±0.2 | 22±0.2 | 28±0.2 | 20±0.2 | 25±0.2 | 22±0.2 |

(levels of sampling noise, e.g., 0.2)

In contrast, it is not possible to have replication through full re-simulation of either of genotypes or phenotypes. When re-simulating phenotype from scratch, we generate new independent phenotypes, for example the original was phenotype A and now we have observation of phenotype B.

An example of a wrong replication:

**Original study**

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Individual/Genotypes: | 1, | 2, | 3, | 4, | 5, | 6, | 7, | 8, | 9, | 10, | 11, | 12 |
| Phenotype A: | 1, | 2, | 5, | 3, | 2, | 3, | 20, | 22, | 28, | 20, | 25, | 22 |

**Wrong replicated study**

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Individual/Genotypes: | 1, | 2, | 3, | 4, | 5, | 6, | 7, | 8, | 9, | 10, | 11, | 12 |
| Phenotype B: | 26, | 4, | 2, | 27, | 1, | 26, | 4, | 2, | 21, | 5, | 23, | 25 |

Here, we cannot use the *p*-values from the original dataset to make heritability estimation in the second dataset which was simulated independently from scratch (i.e., wrong replicated study).

When re-simulating genotypes, it also means a complete new cohort, for example the original was a cohort on disease A and the re-simulated is a cohort on disease B. So, that is why *p*-values obtain from one simulation, so-called “original study/cohort”, cannot be used in another dataset which is simulated from scratch.