

5 Methodological Extensions and Variants

To facilitate comparison of different extensions in terms of statistical objectives and applicability, this study adopts a unified analytical framework for commonly used GREML-based methods (Table S2).

5.1 LOCO (leave-one-chromosome-out) strategy

When estimating heritability and performing genomic prediction within the GREML framework, sources of bias embedded in model specification are not always immediately apparent. Among these, proximal contamination represents a typical issue with important methodological implications. Specifically, when analysis focuses on a particular chromosome or a localized genomic region, if markers from that same region are simultaneously used to construct the genomic relationship matrix (GRM), the linkage disequilibrium (LD) information they carry can “feed back” into the model through the relationship structure. This feedback mechanism leads to a systematic inflation of the estimated contribution of the focal region, fundamentally reflecting a lack of identifiability in parameter decomposition and the resulting estimation bias (Yang et al., 2011; Van den Berg et al., 2019).

The LOCO (leave-one-chromosome-out) approach offers a targeted correction strategy for this problem. Rather than restructuring the model in a complex way, its core logic is to deliberately exclude all markers from the chromosome of interest when constructing the GRM used to estimate that chromosome’s genetic effect. In doing so, it effectively blocks the indirect feedback pathway through which local LD information influences its own effect estimate (Yang et al., 2011). This strategy implicitly relies on the assumption that genetic contributions from different chromosomes can be treated as approximately independent in a statistical sense, such that removing information from the target chromosome does not substantially impair the modeling of the remaining genomic background. Under this condition, LOCO can mitigate endogenous bias in local effect estimation without altering the overall modeling framework.

From the perspective of genomic architecture, the advantages of LOCO become particularly evident under certain conditions. In organisms with a relatively small number of chromosomes, extended LD blocks, or phenotypic variation driven by a limited number of large-effect loci, local LD structures are more likely to generate strong signal coupling within the GRM, thereby amplifying the impact of proximal contamination. This feature is especially pronounced in many crop genomes, making the LOCO strategy highly compatible with studies in agricultural genetics and breeding. In contrast, for species with a larger number of chromosomes and rapid LD decay, the severity of proximal contamination is often reduced, and the marginal benefit of applying LOCO correspondingly diminishes.

It is important to note that LOCO is not a universal solution for bias correction. Its utility is primarily confined to addressing proximal contamination and does not extend to systematic control of population structure, long-range LD heterogeneity, or other complex confounding factors. Therefore, its application should be guided by empirical evaluation rather than assumed necessity. In practice, researchers may compare heritability estimates or marker effect sizes obtained from standard GRM-based models and LOCO-adjusted models to assess the extent of proximal contamination (Van den Berg et al., 2019). If results are highly consistent across the two settings, the additional computational burden and model partitioning introduced by LOCO may not yield substantial benefits. Conversely, pronounced discrepancies indicate that local LD “feedback” is indeed influencing parameter estimation, in which case the use of a leave-one-chromosome-out strategy is both statistically justified and practically valuable.

5.2 Partitioning heritability by functional categories

In the traditional GREML framework, all SNPs are assumed to have equal prior weights by default; that is, their contributions to the overall genetic variance are treated as statistically homogeneous. However, this assumption is often difficult to sustain for complex traits, because different functional regions of the genome vary substantially in their biological mechanisms and evolutionary constraints, which in turn leads to spatial heterogeneity in the distribution of genetic effects. Against this background, research approaches that partition heritability by functional category have gradually developed. Their core objective is to reveal how genetic variance is distributed