In this context, GRM heatmaps serve not only as a visualization tool, but also as an important diagnostic instrument for understanding population structure, assessing potential confounding factors, and interpreting subsequent GREML-based heritability estimates.

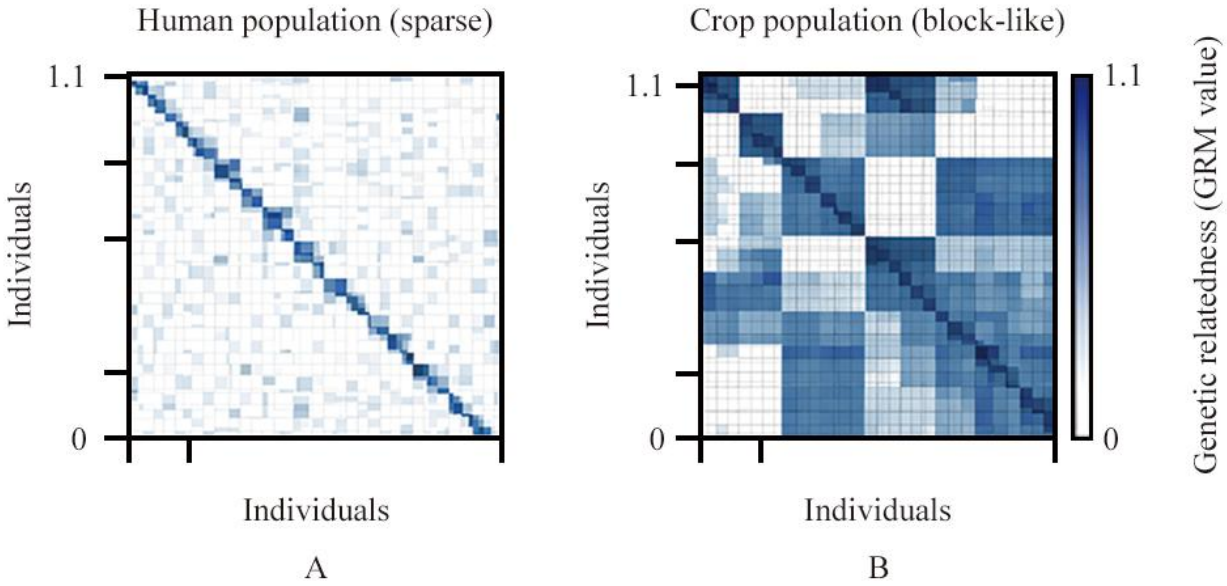

Figure 1 Illustrative comparison of GRM structures in human and crop populations

Caption: Illustrative schematic based on published studies (Yang et al., 2015; Speed et al., 2016). Panel A shows a schematic GRM heatmap representative of large human cohorts after standard quality control and removal of close relatives, as commonly observed in studies such as UK Biobank-based analyses. The matrix is characterized by strong diagonal elements (self-relatedness) and sparse off-diagonal values centered near zero, reflecting weak pairwise genetic relatedness among largely unrelated individuals. Panel B illustrates a typical GRM structure for crop populations, where pronounced block-like patterns arise due to strong population structure, limited numbers of chromosomes, extended linkage disequilibrium, and shared breeding history. These contrasting patterns highlight that, although the statistical definition of the GRM is consistent across species, its empirical structure is highly dependent on population history, LD architecture, and sampling design. The figure is schematic and intended for diagnostic illustration rather than representation of a specific dataset. Note that GRM values are not constrained to the interval [−1, 1]; diagonal elements and highly related pairs may slightly exceed 1 due to finite marker density and allele-frequency estimation.

4 GREML and REML Estimation

4.1 Model derivation

Heritability estimation based on the genome-wide relationship matrix (GRM) is typically conducted within the framework of a linear mixed model (LMM) (Da et al., 2014; Yang et al., 2016; Zhou et al., 2020). Its general form can be written as:

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{g}+\mathbf{e}$$

where $\mathbf{y}$denotes the vector of phenotypic observations, $\mathbf{X}\boldsymbol{\beta}$ represents fixed effects (e.g., population structure, environmental factors, or other covariates), $\mathbf{g}$ denotes the random additive genetic effects, and $\mathbf{e}$ is the independent residual error term. Unlike traditional heritability estimation approaches, the LMM framework allows for simultaneous control of systematic confounding and estimation of genotype-related variance components within a unified model.

In variance decomposition, the random genetic effects are assumed to follow a multivariate normal distribution with mean zero and a covariance structure proportional to the GRM:

$$\mathbf{g}\sim N(0,\sigma_g^2\mathbf{G})$$

where σ_g^2denotes the additive genetic variance and $\mathbf{G}$is the GRM. The environmental residuals are assumed to follow:

$$\mathbf{e}\sim N(0,\sigma_e^2\mathbf{I})$$