This transformation centers the genotype (by subtracting its expectation, $2p$) and scales it (by dividing by its standard deviation, $\sqrt{2p(1-p)}$), ensuring that all loci contribute comparably to the matrix calculation (Forni et al., 2011; Granato et al., 2018).

This standardization has important statistical implications. On the one hand, it removes variance heterogeneity caused by allele frequency differences across loci, making the GRM estimation more reflective of true genetic similarity (Wang et al., 2025). On the other hand, it effectively distinguishes between allele frequency differences arising from random genetic drift and those reflecting genuine shared genetic background, thereby enabling the construction of a robust relationship matrix at the genome-wide level. This approach has been widely applied in genomic prediction, heritability estimation, and association studies, and has been integrated into various molecular breeding tools (Forni et al., 2011; Granato et al., 2018).

3.2 GRM formula and intuitive interpretation

After constructing the standardized genotype matrix $\mathbf{Z}$, the GRM can be expressed as:

$$\mathbf{G}=\frac{1}{M}\mathbf{Z}\mathbf{Z}^{T}$$

where M denotes the total number of SNPs across the genome, and each matrix element G_{ij}represents the genomic similarity between individuals i and j (Forni et al., 2011; Wang et al., 2025).

Intuitively, the GRM measures the similarity between two individuals based on their standardized genotypes across all marker loci, and its values reflect their additive genetic relatedness at the population level. The diagonal elements represent self-relatedness (or inbreeding), with an expected value close to 1, while off-diagonal elements quantify pairwise relatedness between individuals. Values approaching 1 indicate high genetic similarity, whereas values close to 0 suggest near independence.

From a statistical perspective, the GRM can be interpreted as a genome-wide weighted average of identity-by-state (IBS) (Forni et al., 2011). Unlike traditional pedigree-based relationship matrices, the GRM does not rely on prior pedigree information but is constructed directly from molecular data, enabling it to capture realized genetic similarity. This property allows the GRM to be applied not only to large-scale natural populations without pedigree records, but also to more accurately characterize complex population structures and latent genetic diversity (Bilton et al., 2024; Wang et al., 2025).

3.3 Example: visualization and comparison of GRM structures in human and crop populations

In high-level human genetics studies, the GRM is often visualized using heatmaps or distributions of pairwise relatedness to intuitively illustrate additive genetic similarity among individuals (Figure 1). For example, in studies based on the UK Biobank (Yang et al., 2015; Speed et al., 2016; Hou et al., 2019), GRM heatmaps typically exhibit a highly sparse structure centered along the diagonal: diagonal elements are close to 1, reflecting the standardized genetic variance of individuals themselves, while off-diagonal elements are mostly concentrated near zero, with weak clustering patterns appearing only in the presence of subtle population structure or residual relatedness. This structural feature indicates that, after stringent quality control (QC) and removal of close relatives, the GRM can stably capture SNP-derived additive genetic similarity among unrelated individuals.

Similar structural patterns can also be observed in crop populations, but their manifestation is strongly influenced by population composition and linkage disequilibrium (LD) structure. In inbred populations such as rice or maize, where the number of chromosomes is limited, LD blocks are relatively large, and subpopulation differentiation is pronounced, GRM heatmaps often display clearer block-like structures corresponding to different genetic backgrounds or breeding origins (Granato et al., 2018). This comparison highlights that, although the statistical definition of the GRM remains consistent across species, its empirical structure is highly dependent on population history, LD architecture, and sampling design.

It is important to note that the elements of the GRM represent standardized additive genetic covariances, rather than correlation coefficients. Therefore, when the number of markers is limited and allele frequencies are estimated from the sample, diagonal elements or values for highly related individuals may slightly exceed 1.