primarily focus on additive genetic variance, while dominance, epistasis, and their interactions with environmental factors are often not explicitly modeled (Abney et al., 2001; Chen et al., 2015; Zhu et al., 2015). These effects may be partially absorbed into genetic variance estimates in pedigree-based analyses, but are difficult to identify in SNP-based analyses using unrelated individuals.

In summary, "missing heritability" is more appropriately understood as a difference in the identifiability of genetic variance under different statistical frameworks, rather than as an actual absence of genetic mechanisms. Pedigree-based and SNP-based heritability estimates reflect different aspects of genetic architecture; their discrepancy provides important insights into the multi-layered genetic basis of complex traits, rather than constituting mutually contradictory evidence. To facilitate a systematic comparison between traditional marker-assisted approaches and genome-wide statistical genetic methods in terms of research objectives, statistical assumptions, and application scenarios, representative methods-including linkage analysis, candidate gene approaches, and GWAS/GCTA-GREML-are summarized in Table 1.

Table 1 Comparison between traditional marker-assisted approaches and genome-wide statistical genetic methods

Comparison dimension	Traditional approaches (Linkage/Candidate gene)	Genome-wide approaches (GWAS/GCTA-GREML)
Research starting point	Hypothesis-driven candidate regions or genes	Genome-wide, hypothesis-free scanning
Primary data type	A limited number of molecular markers (e.g., RFLP, SSR)	High-density SNPs or whole-genome sequencing data
Study population	Structured populations or pedigrees	Natural populations or breeding populations
Scale of genetic signal	Single loci or local linkage intervals	Genome-wide, multi-locus signals
Core statistical assumptions	Strong prior assumptions with limited multiple testing	Explicit modeling of population structure and multiple testing
Main analytical objective	Identification of QTLs or candidate genes	Estimation of heritability and genetic architecture
Interpretation of results	Locus-specific effects and biological interpretation	Variance decomposition and predictability assessment
Suitability for complex traits	Limited power for highly polygenic traits	Well suited for highly polygenic traits
Role in breeding	Marker-assisted selection and locus validation	Guiding genomic selection and breeding strategy design
Representative methods	Linkage analysis, candidate gene analysis	GWAS、GCTA、GREMLGWAS, GCTA, GREML
Methodological limitations	Limited resolution, power depends on population design	Sample-size dependent, limited causal interpretation
Comparison dimension	Traditional approaches (Linkage/Candidate gene)	Genome-wide approaches(GWAS/GCTA-GREML)

Note: Traditional marker-assisted approaches rely mainly on linkage analysis and candidate gene strategies to identify QTLs or functional loci using a limited number of molecular markers in structured populations (Fang et al., 2001). Genome-wide methods, represented by GWAS and GCTA/GREML, use dense genome-wide markers to build statistical models for estimating heritability and dissecting the genetic architecture of complex traits. Although these approaches differ substantially in statistical assumptions and analytical scale, they are historically and conceptually connected in crop genetic improvement (Fang and Wu, 2026).

3 Principles of Constructing the Genome-wide Relationship Matrix (GRM)

3.1 Standardized genotype matrix

The construction of the genome-wide relationship matrix (GRM) is fundamentally based on a standardized genotype matrix. For each SNP locus in diploid species, genotypes are typically encoded as 0, 1, or 2, representing the number of copies of the reference allele carried by an individual. However, directly using these raw genotype encodings may introduce bias, because differences in allele frequencies across loci can lead to heterogeneity in variance (Forni et al., 2011; Wang et al., 2025).

To avoid such bias, genotype data must be standardized. Let the population frequency of the reference allele at a given locus be p. The observed genotype $x$for an individual at that locus is standardized as:

$$z=\frac{x-2p}{\sqrt{2p(1-p)}}$$