

appropriate growth periods according to target latitude and cropping system (Ravelombola et al., 2021; Gai et al., 2025). For yield-related traits, although total yield is a typical polygenic trait, component traits such as seed weight, pod number, and per-plant yield still contain QTLs with moderate to large effects. For example, an approximately 11.5 Mb region on chromosome 10 is significantly associated with both seed weight and yield, showing high breeding value (Ravelombola et al., 2021). Therefore, MAS can be preferentially applied for precise manipulation of major-effect agronomic traits such as plant architecture and growth period, and can be combined with genomic selection (GS) to simultaneously capture the contributions of both major- and minor-effect loci to yield potential (Ravelombola et al., 2021; Gai et al., 2025).

3 Application of MAS in Quality Improvement

3.1 Protein and oil content

Soybean seed protein and oil contents are core traits determining nutritional quality and processing value. Both traits have relatively high heritability, but they are usually negatively correlated with each other and show certain trade-offs with yield, making their coordinated improvement difficult (Lee et al., 2019). Existing GWAS, linkage analyses, and multi-environment trials have shown that although protein and oil contents are mainly controlled by multiple genes, relatively stable large-effect QTLs exist in regions such as chromosomes 15 and 20. In addition, several regions on chromosomes 2, 8, and 14 can also explain a relatively high proportion of phenotypic variation, and therefore have become priority targets for MAS utilization (Jin et al., 2023). These results indicate that, in quality breeding, MAS is particularly suitable for early-generation screening targeting stable major-effect loci, thereby improving the efficiency of identifying high-protein or high-oil materials.

In terms of oil improvement, high-resolution GWAS has identified multiple SNP-enriched regions associated with oil content, some of which contain candidate genes related to lipid metabolism and transport. Meta-QTLs formed through meta-analysis of previous QTLs further improve the stability and transferability of markers (Kumar et al., 2022). Among them, *GmSWEET39* on chromosome 15 has been identified as an important causal gene simultaneously affecting the protein-oil balance. Its different haplotypes correspond to “high-oil and low-protein” or “relatively high-protein” phenotypes, providing a functional marker basis for the targeted design of quality combinations (Figure 2) (Zhang et al., 2020). Therefore, by jointly tracking major-effect QTLs and key genes through functional markers such as KASP, MAS can, to a certain extent, alleviate the negative correlation between protein and oil and achieve targeted optimization of seed composition (Patel et al., 2025).

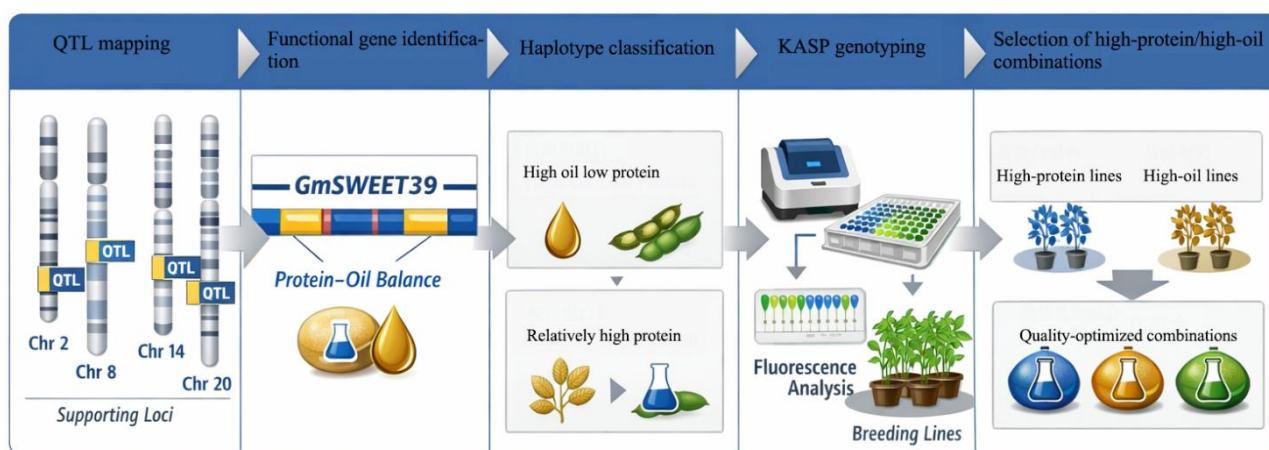


Figure 2 Functional marker-assisted optimization of soybean seed protein–oil balance based on *GmSWEET39* and major QTL

3.2 Specialized metabolites

In addition to protein and oil, soybean seeds are also rich in various specialized metabolites with nutritional and functional value, among which isoflavones and specific fatty acid compositions are the most representative. Isoflavones, such as genistein and daidzein, are synthesized through the phenylpropanoid-flavonoid pathway, and their natural variation is closely related to differences in the expression of key enzyme genes such as CHS, CHI,