

6.3 Integrating multi-omics and phenotyping data to explain seed quality formation

Integrated omics studies now explain varietal differences in seed quality formation by linking developmental phenotypes to coordinated changes in transcripts, metabolites, and candidate regulators. In two contrasting cultivars, combined transcriptomics and metabolomics identified more than 12,000 differentially expressed genes and 315 differential metabolites across seed development, and the authors proposed that high-protein varieties differ from high-oil types through altered sensitivity to desiccation, photomorphogenesis, and senescence timing; in a separate nine-cultivar multi-omics study, more than 1,000 differential metabolites and 7 000 differentially expressed genes were used to define 10 core genes associated with oil content, with *GmADH1* and *GmCrRLK1L34* emerging as hub regulators (Xu et al., 2022). These integrative datasets also converge on central metabolism, because metabolomic analysis of extreme high- and low-protein/oil lines showed enhanced Calvin cycle, TCA cycle, and glycolytic activity that supports carbon entry into amino acid and fatty acid synthesis, helping explain how developmental resource partitioning generates contrasting seed compositions (Cui et al., 2025).

High-resolution phenotyping adds spatial and quantitative context that bulk omics alone cannot provide. In wild soybean, spatial transcriptomics, spatial metabolomics, and single-cell RNA sequencing at mid-maturity showed tissue-level separation between protein- and lipid-associated metabolism and identified *GsMAPK23-4* as a candidate regulator of seed quality, while FT-NIR phenotyping across 191 diverse accessions produced highly accurate protein and oil prediction models with R^2 values above 95%, enabling rapid compositional screening across maturity groups and seed phenotypes. Genetic integration strengthens this framework further: QTL mapping, BSA-seq, and RNA-seq identified 37 QTLs and 12 preliminarily validated candidate genes for protein and oil, showing that multi-omics and phenotyping together can move from descriptive developmental patterns to testable loci and breeding targets (Fang et al., 2025).

7 Breeding Strategies and Future Perspectives for Improving Soybean Protein and Oil Quality

7.1 Genetic improvement of soybean protein and oil traits

Genetic improvement of soybean protein and oil traits still depends on treating these characters as complex quantitative phenotypes shaped by many loci and strong genotype $\times$ environment interaction. QTL mapping, GWAS, and meta-QTL analysis have repeatedly identified stable genomic regions for both traits, and the narrower intervals produced by GWAS now make marker-assisted allele selection more practical for predictable compositional improvement (Kumar et al., 2021).

Breeding strategies are also shifting from locus discovery alone to predictive and design-based selection. Genomic selection achieved cross-validation accuracies of 0.68 for protein and 0.64 for oil, indicating that breeders can identify many top-quartile composition lines before extensive phenotyping (Miller et al., 2023). At the same time, integrative genetics shows that some quality traits can be improved together rather than traded off absolutely, because seed weight and oil content share a positive genetic correlation and candidate genes such as *GmRWOS1* have already been functionally validated for coordinated trait regulation (Yuan et al., 2024).

7.2 Optimizing agronomic management for enhanced seed quality

Agronomic management can partly buffer quality losses, but environment remains the dominant driver of soybean seed composition across production systems. In a synthesis of 13 574 U.S. data points, site-year explained more than 70% of the variation in protein, oil, and yield, which means management usually works by shifting how crops experience temperature, radiation, water, and nitrogen during seed filling. Within that constraint, delayed planting consistently reduced oil concentration, whereas corn-soybean rotation improved both composition and yield, showing that timing and system context can move quality in useful directions.

More targeted interventions can improve seed quality without necessarily imposing the usual yield penalty. Lower fertilizer N rates increased both oil and protein concentration in the U.S. synthesis, and irrigation in water-limited Nebraska fields increased yield and protein concentration simultaneously in about two-thirds of fields, especially where conditions were not favorable for oil synthesis (Carciochi et al., 2023). Regional field studies further show