

U.S. synthesis, low to moderate fertilizer N inputs increased both protein and oil concentration, although environmental variation still explained most of the total variation in composition. Field experiments in high-yield environments similarly showed that full-season nitrogen supply increased seed protein concentration without increasing oil concentration, while raising both protein and oil yields through higher seed production, indicating that N limitation can restrict both composition and total reserve output.

Management effects beyond nitrogen are real but less uniform, and they often interact with water regime, latitude, and cropping system. Recent field studies showed that late-season N application during seed filling increased protein concentration by 1.2-2.8% with little or no effect on oil concentration, supporting targeted N management as a practical strategy for improving meal quality (Khatri et al., 2026). By contrast, broader management syntheses indicate that delayed planting decreases oil concentration, corn-soybean rotation tends to improve composition and yield, and practices such as no-till, seed treatment, foliar nutrient application, and fungicide produce mixed responses, showing that cultivation practices mainly modify composition indirectly through their effects on crop growth environment and stress exposure.

5 Advanced Technologies for Monitoring and Predicting Protein and Oil Dynamics

5.1 Biochemical and omics approaches for understanding seed composition

Biochemical and omics approaches have become the main tools for dissecting how protein and oil contents change during soybean seed development because they capture regulatory variation across transcripts, proteins, and metabolites rather than only final composition. Integrated transcriptomic, proteomic, and metabolomic studies show that the pathways most consistently linked to protein and oil divergence involve glycolysis and carbon metabolism, while systems-level analyses further indicate that metabolic flux mapping during seed fill is especially useful for connecting developmental gene expression with the biochemical networks that determine mature seed composition (Mo et al., 2024). These datasets also reveal that different omics layers contribute distinct kinds of information, which is important for interpreting developmental dynamics. Joint transcriptome-proteome analysis identified generally poor correspondence between mRNA and protein abundance, while metabolomics studies in contrasting high-protein and high-oil lines detected coordinated shifts in the Calvin cycle, TCA cycle, and glycolysis that favor routing carbon into amino acid and fatty acid synthesis (Mo et al., 2024; Cui et al., 2025).

Proteomics has been particularly valuable for defining the temporal sequence of storage reserve accumulation during seed filling. High-resolution proteome mapping across 2 to 6 weeks after flowering showed a developmental decrease in metabolism-related proteins together with an increase in proteins associated with destination and storage, and more recent TMT-based proteomics further demonstrated that major 7S and 11S storage proteins accumulate steadily from early to late seed filling (Islam et al., 2021). At a broader scale, integrative omics has strengthened gene discovery for seed quality improvement by linking developmental expression patterns to inherited composition loci. Meta-analysis and time-course transcriptomics identified 11 shared meta-QTL hot regions for oil and protein and seven hub genes associated with storage accumulation, while multi-omics comparison among contrasting cultivars detected 22 candidate genes with potential simultaneous negative regulation of protein and oil content (Mo et al., 2024).

5.2 Imaging and phenotyping technologies for seed development monitoring

Non-destructive imaging technologies now allow soybean seed composition and developmental status to be monitored at the single-seed level, which is a major advance over destructive wet-chemistry assays. Near-infrared hyperspectral imaging predicted single-seed protein content with $R^2 = 0.92$ and RMSE of 1.08%, and it also generated chemical maps that visualized within-seed protein distribution for rapid screening of large numbers of samples (Aulia et al., 2022). Hyperspectral approaches are also effective for monitoring oil quality traits that contribute to overall lipid dynamics during development and selection. Reflective hyperspectral imaging classified oleic and linoleic acid contents of single seeds with validation accuracies of 90% and 93.3%, and the study emphasized that single-seed measurement is critical because bulk-seed spectroscopy cannot resolve the composition of individual seeds needed for precision breeding (Fu et al., 2021).