5.2 Phenotypic data collection techniques: imaging measurement, metabolomics, and behavioral analysis

Phenotypic data collection is rapidly evolving from traditional manual measurement to high-throughput, automated, non-invasive, and multimodal integration approaches. This transformation has greatly enhanced data accuracy, efficiency, and reproducibility, providing a solid foundation for the systematic phenotypic characterization of transgenic livestock.

In imaging measurement, RGB and depth cameras, laser scanning, structured light, and 3D reconstruction technologies are widely used for body size measurement, weight estimation, and surface or muscle thickness analysis. CT and MRI can analyze tissue distribution and fat deposition patterns. When combined with computer vision algorithms such as convolutional neural networks (CNNs), these technologies enable automatic landmark recognition and phenotypic feature quantification, reducing human error and improving assessment efficiency.

In metabolomics and biochemical detection, gas chromatography-mass spectrometry (GC-MS) and liquid chromatography-mass spectrometry (LC-MS) can be used to analyze dynamic changes in energy metabolism, lipid synthesis, and amino acid utilization in samples such as blood, urine, and milk. Combined with transcriptomic and proteomic data, these methods can construct metabolic pathway networks, revealing potential systemic metabolic effects induced by genetic modifications.

In behavioral and physiological monitoring, video tracking systems and wearable sensors enable continuous recording of feeding, rumination, activity rhythms, social behavior, and stress responses. Monitoring heart rate, body temperature, and respiration rate provides quantifiable indicators of livestock adaptability and recovery speed under environmental changes or pathogen exposure, offering multidimensional data support for health and welfare assessment.

5.3 Data analysis and statistical models: multivariate regression, genetic and environmental interaction analysis

The analysis of multidimensional high-throughput data requires an integrated statistical framework to effectively extract genetic signals and reveal the patterns of gene–environment interaction (G×E).

In multivariate analysis, when multiple correlated phenotypes are studied, multivariate regression, canonical discriminant analysis, or stepwise discriminant analysis can be applied for joint modeling. These methods enhance the detection power of quantitative trait loci (QTLs) and gene effects by accounting for trait correlations. For datasets with nonlinear relationships or strong collinearity, partial least squares regression (PLSR) or principal component regression (PCR) can be used for dimensionality reduction before further analysis to reduce noise and improve model stability.

In gene–environment interaction studies, linear mixed models (LMMs) and random regression models (RRMs) are widely used to analyze longitudinal data and multi-environment experimental results. These models describe reaction norms of gene effects across environmental gradients or time, thus revealing genotype performance differences and adaptive traits under varying environments.

In multilevel and machine learning modeling, for complex high-dimensional data such as imaging, metabolomics, and behavioral datasets, algorithms such as random forest (RF), support vector machine (SVM), and deep learning can be applied to capture nonlinear patterns and higher-order interactions. These models enable individual-level trait prediction and cluster analysis. To prevent overfitting, cross-validation and independent external validation sets should be incorporated to evaluate model performance and ensure robustness.

6 Correlation Analysis Between Genetic Stability and Phenotypic Consistency

6.1 Relationship models between stability and phenotypic variation

Genetic stability — the structural integrity and consistent expression of exogenous genes across generations — is the foundation for achieving phenotypic consistency. When copy number and integration sites remain stable over generations and expression variance is low, the intergroup dispersion of target traits decreases significantly; conversely, rearrangements, copy number drift, or epigenetic silencing can amplify phenotypic fluctuations (Van Cott et al., 1997).