Advances in Computational Biology Methods and Applications for Nutritional Metabolomics of Vegetable Crops

Qiuhui Li *

Jilin Agricultural University College of Hortculture, Changchun, China

* Corresponding Author Email: liqiuhui@mails.jlau.edu.cn

Abstract. As an essential component of systems biology, metabolomics provides novel technological approaches and theoretical frameworks for studying nutritional quality in vegetable crops. This review summarizes recent advances in computational biology methods for vegetable crop nutritional metabolomics, including key technologies such as data preprocessing, statistical analysis, machine learning, network analysis, and pathway enrichment analysis. We focus on the applications of these methods in vegetable nutritional composition analysis, quality evaluation, stress response mechanism elucidation, and breeding-assisted selection. The current challenges are analyzed, including data standardization, method standardization, and multi-omics data integration. Future development trends are discussed to provide references for vegetable crop nutritional metabolomics research.

Keywords: *Vegetable crops; Nutritional metabolomics; Computational biology; Machine learning; Systems biology.*

1. Introduction

The nutritional quality of vegetable crops directly impacts human health[1,2]. Metabolomics, as a systems biology approach, enables large-scale detection of small molecule metabolites, offering new insights into vegetable nutritional metabolism, stress responses, and breeding improvement[3,4]. However, metabolomic data's high dimensionality, noise, and nonlinearity pose challenges for traditional statistical methods. Computational biology methods—including machine learning, network analysis, and pathway analysis—have significantly enhanced metabolomic data interpretation. This paper systematically reviews computational approaches in vegetable nutritional metabolomics, focusing on applications in nutritional analysis, variety identification, stress response, and breeding, while prospecting future development trends.

2. Computational Biology Methods for Vegetable Crop Nutritional Metabolomics

2.1. Data Preprocessing Methods

Data preprocessing is critical for ensuring metabolomics analysis quality, encompassing mass spectrometry data processing and standardization. For peak detection and alignment, software such as XCMS[5], MZmine [6], and MS-DIAL[7] employ dynamic time warping (DTW) algorithms to address chromatographic drift. Metabolite identification relies on accurate mass, isotope patterns, and MS/MS spectra, combined with databases like METLIN, HMDB, and KEGG for annotation. Quality control typically uses QC samples to evaluate stability, removing variables with RSD >30% to ensure data reliability.

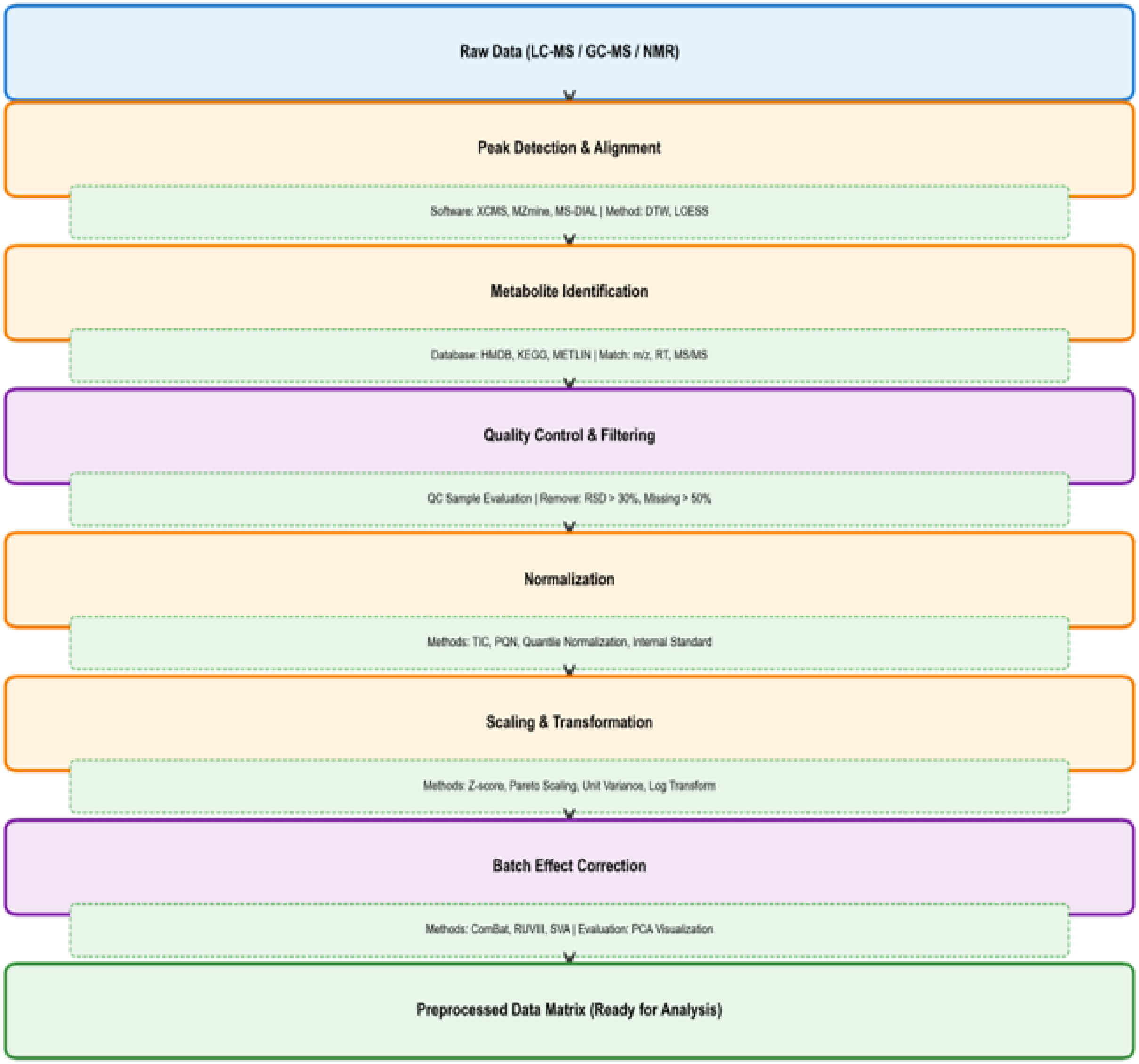

Figure 1. Detailed workflow of metabolomics data preprocessing.

Data standardization is an important step in eliminating systematic errors and improving data comparability. Normalization methods are mainly used to correct overall concentration differences between samples. Common methods include total ion chromatogram (TIC) normalization, probabilistic quotient normalization (PQN), and quantile normalization[8]. Scaling methods are used to adjust scale differences between different metabolites. Z-score scaling, Pareto scaling, and unit variance scaling are currently the three most widely used methods[9]. When batch effects exist in the data, ComBat algorithm or RUVIII method is needed for correction to eliminate systematic differences between technical replicates and batches[10] in Figure 1.

2.2. Multivariate Statistical Analysis Methods

After completing data preprocessing, multivariate statistical analysis becomes the core method for mining biological information from metabolomic data, extracting key information from high-dimensional data and revealing differential patterns between samples. Research shows that integrating genetic diversity and metabolomics information helps with gene function annotation[11]. Table 1 provides a comprehensive comparison of the main computational methods used in metabolomic data analysis.

Table 1. Comparison of Main Computational Methods for Metabolomic Data Analysis.

Method Type	Representative Algorithm	Main Function	Advantages	Limitations	Typical Application Scenarios
Unsupervised Learning	PCA	Dimensionality reduction visualization	No labels needed, discovers overall patterns	Does not consider grouping information	Preliminary data exploration, outlier detection
	HCA	Sample clustering	Intuitively shows similarity	Sensitive to distance metrics	Sample grouping, metabolic pattern recognition
	ICA	Independent component extraction	Separates statistically independent signals	Poor interpretability	Mixed signal separation, feature extraction
Supervised Learning	PLS-DA	Classification discrimination	Considers grouping information	Easy to overfit	Differential metabolite screening
	OPLS-DA	Orthogonal separation	Removes irrelevant variation	Requires validation	Biomarker discovery
	sPLS-DA	Sparse discrimination	Automatic feature selection	Complex parameter selection	High-dimensional data classification
Traditional Machine Learning	SVM	Nonlinear classification	Handles high-dimensional small samples	Difficult parameter tuning	Variety identification, quality prediction
	Random Forest	Ensemble classification	Feature importance scoring	Complex model	Metabolite importance ranking
	XGBoost	Gradient boosting	High accuracy prediction	High computational complexity	Nutritional value prediction
Deep Learning	DNN	Deep representation	Learns complex patterns	Requires large samples	Complex metabolic pattern recognition
	CNN	Spectral feature extraction	Automatic feature learning	Black box characteristics	Mass spectrometry data analysis
	Autoencoder	Dimensionality reduction and reconstruction	Unsupervised feature learning	Difficult interpretation	Data dimensionality reduction, anomaly detection
Network Analysis	Correlation	Association network	Simple and intuitive	Contains indirect correlations	Metabolite association analysis
	Partial Corr	Partial correlation network	Direct regulatory relationships	Complex calculation	Regulatory network construction
	WGCNA	Module identification	Discovers functional modules	Strong parameter dependence	Co-regulated module mining

This table provides a comprehensive comparison of the main computational methods used in metabolomic data analysis, including unsupervised learning (PCA, HCA, ICA), supervised learning (PLS-DA, OPLS-DA, sPLS-DA), traditional machine learning (SVM, Random Forest, XGBoost), deep learning (DNN, CNN, Autoencoder), and network analysis (Correlation, Partial Correlation, WGCNA). Each method is evaluated based on its representative algorithms, main functions, advantages, limitations, and typical application scenarios. Selecting appropriate analysis methods requires comprehensive consideration of data characteristics, sample size, research objectives, and computational resources. In practice, multiple methods are often used in combination to obtain more reliable results.

2.2.1. Unsupervised Learning Methods

Unsupervised learning methods do not require prior knowledge of sample grouping information and can discover intrinsic differential patterns between samples from the data itself. Principal component analysis (PCA) is the most commonly used unsupervised method. Through linear combination, it projects high-dimensional metabolomic data onto a low-dimensional space. While achieving dimensionality reduction visualization, it can also identify major sources of variation and outlier samples. Statistical hypothesis testing methods for principal component analysis have been applied to metabolite set enrichment analysis[15]. Hierarchical clustering analysis (HCA) groups samples based on the similarity between metabolites, intuitively displaying the hierarchical relationships between samples with dendrograms, helping to reveal metabolic patterns of different samples. Independent component analysis (ICA) assumes that observed data is linearly mixed from several statistically independent components. By maximizing statistical independence between components, potential biological signals can be separated to discover hidden metabolic processes. These methods have been extensively validated in plant metabolomics studies[12,13,14].

2.2.2. Supervised Learning Methods

Supervised learning methods can more specifically conduct classification analysis and differential metabolite screening by introducing sample grouping information. Partial least squares discriminant analysis (PLS-DA) combines the ideas of partial least squares regression and discriminant analysis, achieving classification prediction by maximizing inter-group differences while screening differential metabolites that contribute significantly to classification. Orthogonal partial least squares discriminant analysis (OPLS-DA), based on PLS-DA, decomposes data into two parts: related and unrelated to grouping, removing systematic variation unrelated to grouping, making the biological interpretation of the model clearer. For high-dimensional data, sparse PLS-DA (sPLS-DA) achieves automatic feature selection by introducing sparsity constraints, reducing model complexity while ensuring classification accuracy and improving interpretability. The effectiveness of these methods has been demonstrated in numerous vegetable metabolomics studies[16,17].

2.3. Machine Learning Methods

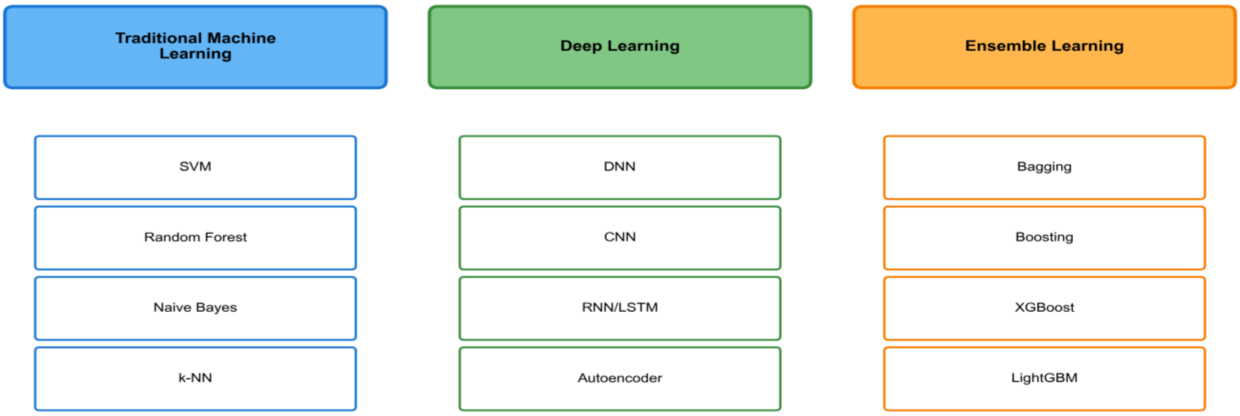

Figure 2. Machine Learning Method Classification System.

This figure presents a comprehensive classification of machine learning methods used in metabolomics data analysis, organized into three major categories: Traditional Machine Learning (blue, including SVM, Random Forest, Naive Bayes, and k-NN), Deep Learning (green, including DNN, CNN, RNN/LSTM, and Autoencoder), and Ensemble Learning (orange, including Bagging, Boosting, XGBoost, and LightGBM). Each category represents different algorithmic approaches with distinct advantages and application scenarios in vegetable crop nutritional metabolomics research.

Machine learning has been widely applied in metabolomic data analysis (Figure 2). Traditional algorithms include: support vector machines (SVM), which excel at high-dimensional small sample classification; random forest (RF), which provides both classification and feature importance evaluation for identifying key metabolite markers; naive Bayes and k-nearest neighbors (k-NN), which offer computational simplicity for multi-class problems.

Deep learning methods construct multi-layer neural networks for automatic feature learning. Deep neural networks (DNN) capture complex nonlinear metabolic patterns; convolutional neural networks (CNN) automatically extract spectral features; recurrent neural networks (RNN/LSTM) model temporal metabolic dynamics across growth stages or stress conditions; autoencoders enable unsupervised dimensionality reduction and outlier detection. However, deep learning typically requires large training samples, limiting its application in vegetable metabolomics where sample sizes are often restricted. Recent advances in deep learning for metabolomics have been reviewed extensively[18,19].

Ensemble learning combines multiple base learners for improved prediction. Gradient boosting methods—including GBDT, XGBoost, LightGBM, and CatBoost—have become popular choices, offering advantages in handling imbalanced data, computational efficiency, and robustness while preventing overfitting.

2.4. Network Analysis Methods

Network analysis provides a systems biology perspective for understanding complex metabolite regulatory relationships. Correlation network analysis (Pearson/Spearman) reveals metabolite co-change patterns, while partial correlation networks distinguish direct from indirect associations by removing confounding variables. Gaussian graphical models (GGM) infer conditional dependencies, and Bayesian networks represent probabilistic causal relationships between metabolites. Network-based approaches have proven valuable for understanding metabolic regulation[20].

Weighted gene co-expression network analysis (WGCNA), originally developed for transcriptomics, has been successfully applied to metabolomics. This method constructs scale-free networks and identifies co-expression modules through hierarchical clustering—metabolites within the same module share similar variation patterns and likely participate in common biological processes. By associating module eigengenes with phenotypic traits, hub metabolites can be identified for in-depth mechanistic studies. In vegetable research, WGCNA has been applied to analyze tomato fruit ripening and cucumber stress response networks. WGCNA has been successfully applied to various plant species[24, 25, 26].

2.5. Pathway Analysis Methods

Pathway analysis maps metabolites to known metabolic pathways to reveal the biological significance of differential metabolites. Enrichment analysis evaluates whether differential metabolites are significantly enriched in specific metabolic pathways through statistical testing methods such as hypergeometric tests or Fisher's exact tests. Commonly used pathway databases include KEGG[21], PlantCyc, MetaCyc, etc. Metabolite set enrichment analysis (MSEA) considers the correlation structure between metabolites, using methods similar to gene set enrichment analysis (GSEA) to identify significantly changed metabolic pathways, improving the sensitivity and accuracy of pathway discovery[22].

Topological analysis considers the topological structure of metabolic pathways, evaluating pathway importance by calculating centrality indices of metabolites in pathways (such as betweenness centrality, degree centrality). Metabolites at key positions may play important regulatory roles. Impact factor analysis comprehensively considers metabolite changes and topological positions, providing a more accurate assessment of pathway perturbation degree. Visualization tools such as MetaboAnalyst[23], Cytoscape, etc., can intuitively display pathway analysis results, helping researchers understand the biological significance of metabolic changes. Figure 3 illustrates the typical analytical workflow for vegetable crop nutritional metabolomics.

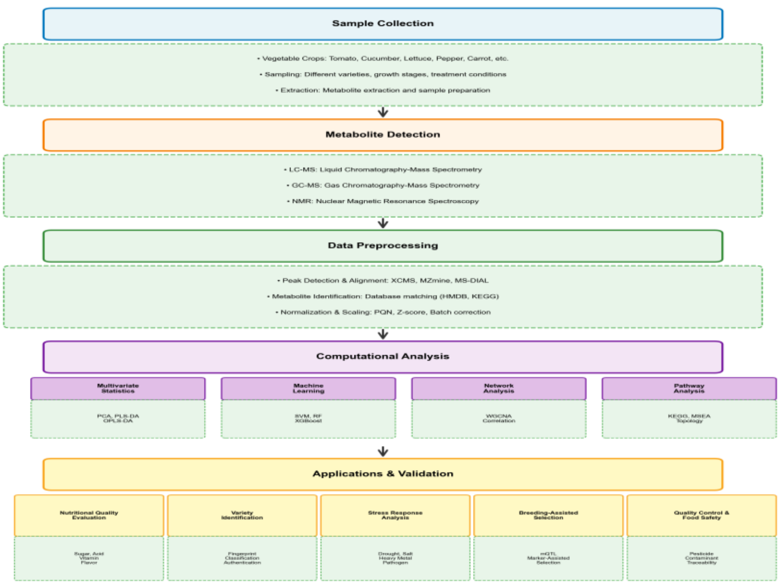

Figure 3. Typical analytical workflow for vegetable crop nutritional metabolomics.

This figure illustrates the complete analytical workflow from sample collection to biological interpretation in vegetable crop nutritional metabolomics research. The workflow begins with sample collection and metabolite extraction from vegetable crops (such as tomatoes, cucumbers, lettuce, peppers, carrots, etc.), followed by LC-MS or GC-MS detection platforms for metabolite detection. Raw data undergoes preprocessing (including peak detection, metabolite identification, data normalization, and batch effect correction) and subsequent multi-level computational analyses (including multivariate statistics, machine learning classification, network analysis, and pathway analysis). Finally, comprehensive interpretation of biological significance is performed based on analysis results, revealing nutritional quality mechanisms, stress response mechanisms, and providing theoretical support for breeding improvement. This integrated analytical strategy fully leverages the advantages of computational biology methods, extracting maximum biological information from metabolomic data to promote in-depth research on vegetable crop nutritional metabolism.

3. Applications in Vegetable Crop Nutritional Metabolomics Figure 4 shows the distribution of application fields in vegetable crop metabolomics research from 2019 to 2024.

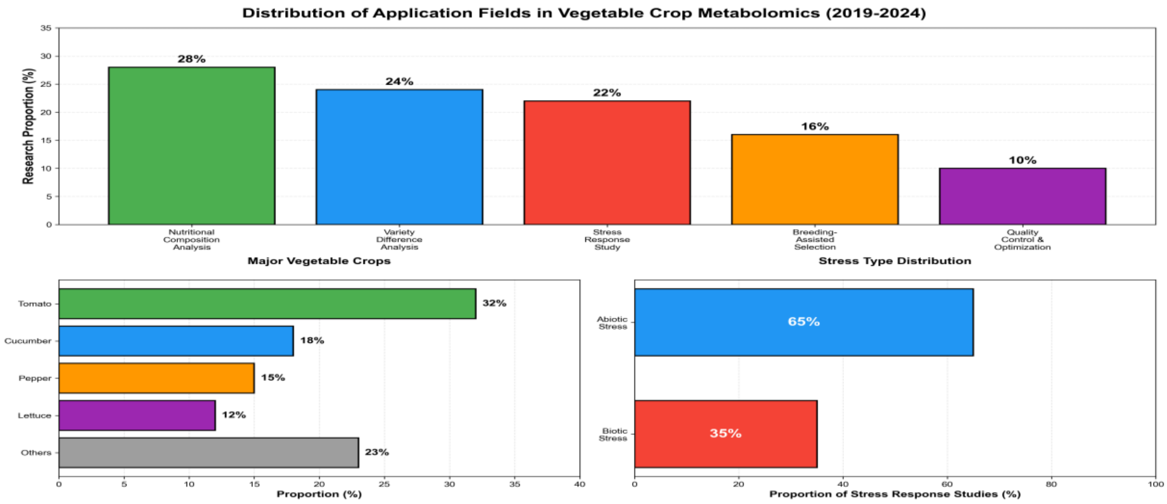

Figure 4. Distribution of Application Fields in Vegetable Crop Metabolomics (2019-2024).

This figure illustrates the distribution of application fields in vegetable crop metabolomics research from 2019 to 2024. The top panel shows the five main research areas: Nutritional Composition Analysis (28%), Variety Difference Analysis (24%), Stress Response Study (22%), Breeding-Assisted Selection (16%), and Quality Control & Optimization (10%). The bottom left panel displays the distribution of major vegetable crops studied, with tomato being the most prevalent (32%), followed by cucumber (18%), pepper (15%), lettuce (12%), and others (23%). The bottom right panel demonstrates the distribution of stress response studies, showing that abiotic stress (drought, salt) accounts for 65% while biotic stress (pathogen) represents 35% of the research focus.

3.1. Nutritional Quality Evaluation and Analysis

Metabolomics enables comprehensive vegetable nutritional quality evaluation by detecting hundreds to thousands of metabolites. Machine learning models accurately distinguish varieties based on metabolite fingerprints—for example, GC-MS with PLS-DA identifies tomato varieties differing in sugar, organic acid, and amino acid contents. Network analysis reveals co-variation patterns, showing how sugars, organic acids, and amino acids collectively determine taste quality. Pathway analysis maps metabolites to nutritional pathways, revealing quality formation mechanisms—cultivated tomatoes differ from wild varieties in flavonoid and terpenoid biosynthesis pathways affecting nutrient content and flavor. The comprehensive nutritional quality evaluation system is presented in Figure 5.

Figure description: The comprehensive nutritional quality evaluation system comprises four steps: ① Sample Detection—metabolomics detection of vitamins, polyphenols, and secondary metabolites; ② Data Processing—PCA, cluster analysis, and machine learning algorithms; ③Index Construction—comprehensive nutritional indices, antioxidant activity evaluation, and quality grading; ④ Practical Application—variety screening, breeding improvement, and quality control.

Figure 5. Nutritional Quality Evaluation System.

3.2. Variety Identification and Classification

Metabolomic fingerprinting combined with machine learning methods provides new technical means for vegetable variety identification. Different varieties have unique metabolic characteristics that can be used for variety classification and authentication. Support vector machines, random forests, and other algorithms can establish high-precision classification models based on metabolomic data to distinguish different varieties. For example, based on metabolomic data and SVM models, cucumber variety classification accuracy can reach over 95%[27]. Deep learning methods further improve classification accuracy; convolutional neural networks can automatically extract effective features from spectral data for variety identification[28].

Metabolic markers specific .to varieties can be discovered through feature selection algorithms. These marker metabolites not only help with variety identification but may also be related to variety-specific traits (such as disease resistance, stress tolerance, nutritional quality, etc.). Network analysis reveals metabolic network differences between different varieties, helping to understand the metabolic basis of variety trait differences. In breeding, metabolic marker-assisted selection can accelerate the breeding process and improve breeding efficiency. These approaches have been validated in multiple vegetable crops[29,30].

3.3. Stress Response Mechanism Analysis

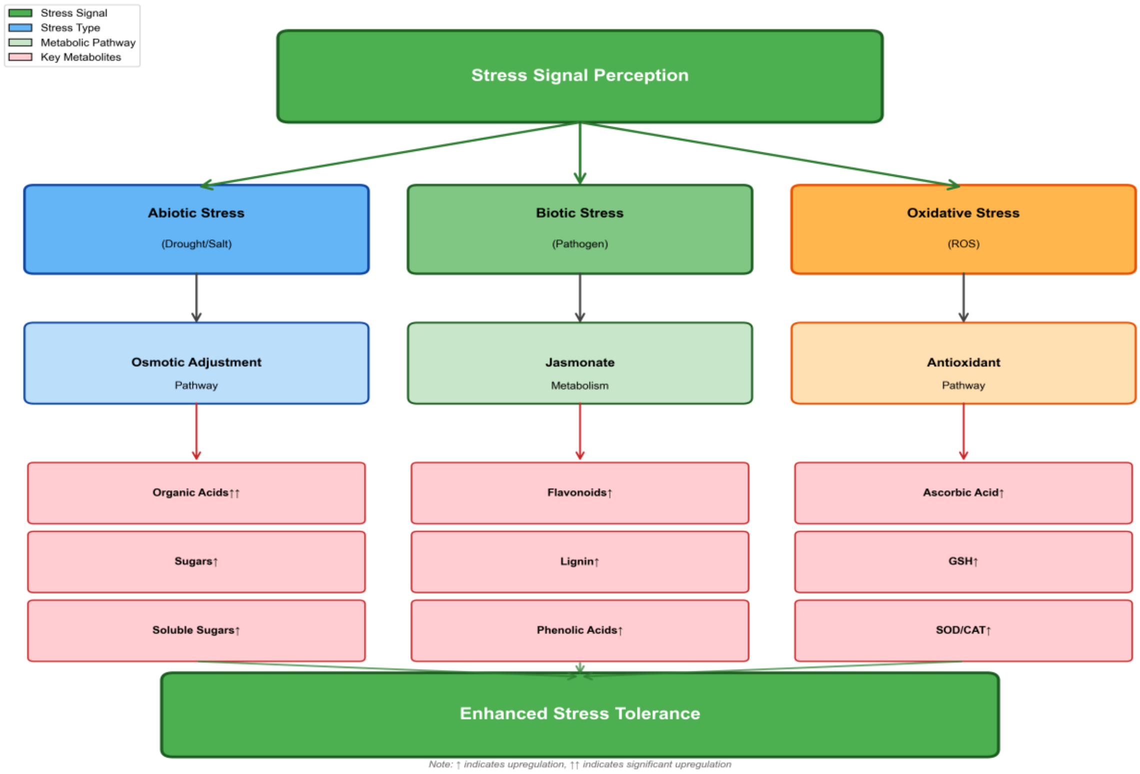

Figure 6. Stress Response Network.

Figure description: This metabolic network illustrates vegetable stress responses. Stress signal perception branches into three types: abiotic (drought/salt), biotic (pathogen), and oxidative (ROS) stress, activating osmotic adjustment, jasmonate metabolism, and antioxidant pathways, respectively. Key upregulated metabolites include organic acids and sugars for osmotic adjustment, flavonoids and phenolic acids for pathogen defense, and ascorbic acid/GSH for oxidative protection, ultimately converging to enhanced stress tolerance.

Applications: Metabolomics reveals stress-induced metabolic reprogramming in vegetables (Figure 6). Time-series analysis combined with machine learning captures dynamic stress responses—for example, identifying amino acid and carbohydrate metabolism changes in lettuce under nitrogen stress. Network analysis (e.g., WGCNA) identifies key regulatory nodes and hub metabolites associated with drought stress in tomatoes. Pathway analysis reveals adaptation mechanisms including osmoprotectant accumulation (proline, betaine) and enhanced antioxidant synthesis. Deep learning models can predict plant stress states from metabolic profiles, supporting precision agricultural management.

3.4. Breeding-Assisted Selection

Metabolomics provides new selection tools for vegetable breeding. Traditional breeding mainly relies on phenotypic selection, which is time-consuming and inefficient. Metabolic marker-assisted selection can identify superior individuals at early stages, shortening the breeding cycle. By integrating metabolomic data with genomic and phenotypic data, quantitative trait loci (mQTL) related to important traits can be identified, providing molecular markers for breeding. For example, combining metabolomics and genomics, mQTLs controlling flavonoid content in peppers have been identified, providing targets for nutritional quality improvement breeding[31].

Machine learning models can predict phenotypes based on metabolic profiles, such as predicting fruit quality, yield potential, stress resistance, etc. These prediction models can assist breeders in early selection, improving breeding efficiency. Network analysis reveals metabolic regulatory networks of complex traits, helping to understand the genetic basis of traits and providing theoretical guidance for breeding design. With the development of genome editing technologies, key metabolic regulatory genes can be precisely modified to achieve directional improvement of target traits.

3.5. Quality Control and Food Safety

Metabolomics plays an important role in vegetable quality control and food safety assessment. By detecting harmful substances (pesticide residues, heavy metals, mycotoxins, etc.) metabolites and their metabolic products, vegetable safety can be evaluated. Machine learning classification models can distinguish safe products from contaminated products based on metabolic profiles. For example, based on metabolomic data and random forest algorithms, heavy metal-contaminated vegetables can be quickly identified[32].

Non-targeted metabolomics can discover unknown harmful substances or adulterants. Anomaly detection algorithms (such as isolation forest, one-class SVM, etc.) can identify abnormal samples, prompting potential quality issues. Metabolic fingerprinting can be used for vegetable origin tracing and authenticity identification, preventing fraud. Network analysis reveals metabolic network changes in vegetables under different storage and processing conditions, providing a basis for optimizing preservation and processing techniques.

4. Challenges and Prospects

4.1. Current Challenges

Despite significant progress in computational biology methods for vegetable crop nutritional metabolomics, many challenges remain. First, data standardization issues are prominent. Different platforms, different laboratories generate data with large systematic differences, lacking unified data standards and quality control specifications, limiting data comparability and sharing. Second, method standardization is insufficient. There are many types of analytical methods, with large parameter setting differences, lacking authoritative method evaluation and selection guidelines, making it difficult for researchers to choose appropriate methods. Third, multi-omics data integration is difficult. Although integration of metabolomics with genomics, transcriptomics, proteomics, etc., can provide more comprehensive biological information, data heterogeneity and scale differences between different omics make integration analysis face technical challenges.

Fourth, biological interpretation capabilities need improvement. Many machine learning and deep learning models have "black box" characteristics, with unclear biological significance of results, requiring the development of more interpretable analytical methods. Fifth, sample sizes are generally small. Deep learning methods require large samples, but most vegetable crop metabolomics studies have limited sample sizes, restricting the application of advanced algorithms. Sixth, computational resource requirements are high. Network analysis, deep learning, and other methods require powerful computing power, and many grassroots research institutions lack sufficient computational resources.

4.2. Future Development Directions

Future development of vegetable nutritional metabolomics will focus on several key areas:

1.Standardization: Establish standardized data generation, analysis processes, and quality control specifications; build public repositories for data sharing.

2.Interpretable AI: Develop machine learning and deep learning methods integrating biological knowledge bases (metabolic pathways, regulatory networks) to enhance model interpretability.

3.Multi-omics integration: Develop efficient algorithms combining genomics, transcriptomics, proteomics, and metabolomics for comprehensive biological analysis.

4.Specialized tools: Create user-friendly analytical platforms tailored for vegetable metabolomics to lower technical barriers.

5.Advanced AI applications: Apply transfer learning, few-shot learning, and AutoML to address small sample issues and simplify model construction.

6.Research translation: Bridge computational methods with practical applications in breeding, quality improvement, and precision agriculture.

7.Crop diversit: Expand research beyond major crops to reveal nutritional metabolic characteristics across diverse vegetable species.

8.Dynamic analysis: Develop time-series and spatial metabolomics methods to reveal metabolic dynamics and spatial distribution patterns.

5. Conclusion

Computational biology methods provide powerful tools for vegetable crop nutritional metabolomics research. From data preprocessing, statistical analysis, machine learning, to network and pathway analysis, a complete methodological system has been formed. These methods have demonstrated important application value in nutritional quality evaluation, variety identification, stress response mechanism analysis, breeding-assisted selection, and other aspects, providing new technical means for vegetable science research and industry development. However, challenges such as data standardization, method standardization, multi-omics data integration, and biological interpretation still exist. Future research needs to strengthen methodological innovation and integration, promote technology popularization and application, and contribute to ensuring vegetable nutritional quality and food safety. The integration of these methods continues to advance our understanding of plant metabolism[33,34,35].

References

[1] Fiehn O. 2002. Metabolomics-the link between genotypes and phenotypes. Plant Molecular Biology, 48(1-2): 155-171.

[2] Saito K, Matsuda F. 2010. Metabolomics for functional genomics, systems biology, and biotechnology. Annual Review of Plant Biology, 61: 463-489.

[3] Fernie AR, Schauer N. 2009. Metabolomics-assisted breeding: a viable option for crop improvement? Trends in Genetics, 25(1): 39-48.

[4] Fukushima A, Kusano M. 2013. Recent progress in the development of metabolome databases for plant systems biology. Frontiers in Plant Science, 4: 73.

[5] Smith CA, Want EJ, O'Maille G, et al. 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Analytical Chemistry, 78(3): 779-787.

[6] Pluskal T, Castillo S, Villar-Briones A, et al. 2010. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics, 11: 395.

[7] Tsugawa H, Cajka T, Kind T, et al. 2015. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nature Methods, 12(6): 523-526.

[8] Wishart DS, Feunang YD, Marcu A, et al. 2018. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Research, 46(D1): D608-D617.

[9] Dunn WB, Broadhurst D, Begley P, et al. 2011. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. Nature Protocols, 6(7): 1060-1083.

[10] Dieterle F, Ross A, Schlotterbeck G, et al. 2006. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Analytical Chemistry, 78(13): 4281-4290.

[11] van den Berg RA, Hoefsloot HC, Westerhuis JA, et al. 2006. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics, 7: 142.

[12] Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics, 8(1): 118-127.

[13] Keurentjes JJ, Fu J, de Vos CH, et al. 2006. The genetics of plant metabolism. Nature Genetics, 38(7): 842-849.

[14] Tohge T, Fernie AR. 2010. Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. Nature Protocols, 5(6): 1210-1227.

[15] Yamamoto H, Fujimori T, Sato H, et al. 2014. Statistical hypothesis testing of factor loading in principal component analysis and its application to metabolite set enrichment analysis. BMC Bioinformatics, 15: 51.

[16] Pomyen Y, Wanichthanarak K, Poungsombat P, et al. 2020. Deep metabolome: applications of deep learning in metabolomics. Computational and Structural Biotechnology Journal, 18: 2818-2825.

[17] Wen B, Mei Z, Zeng C, et al. 2017. metaX: a flexible and comprehensive software for processing metabolomics data. BMC Bioinformatics, 18(1): 183.

[18] Kopka J, Schauer N, Krueger S, et al. 2005. GMD@CSB.DB: the Golm Metabolome Database. Bioinformatics, 21(8): 1635-1638.

[19] Saito K, Hirai MY, Yonekura-Sakakibara K. 2008. Decoding genes with coexpression networks and metabolomics - 'majority report by precogs'. Trends in Plant Science, 13(1): 36-43.

[20] Keurentjes JJ, Fu J, Terpstra IR, et al. 2007. Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. Proceedings of the National Academy of Sciences, 104(5): 1708-1713.

[21] Krumsiek J, Suhre K, Illig T, et al. 2011. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. BMC Systems Biology, 5: 21.

[22] Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, 9: 559.

[23] Kanehisa M, Goto S, Sato Y, et al. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Research, 42(D1): D199-D205.

[24] Xia J, Sinelnikov IV, Han B, et al. 2015. MetaboAnalyst 3.0—making metabolomics more meaningful. Nucleic Acids Research, 43(W1): W251-W257.

[25] Schauer N, Fernie AR. 2006. Plant metabolomics: towards biological function and mechanism. Trends in Plant Science, 11(10): 508-516.

[26] Fiehn O, Kopka J, Dörmann P, et al. 2000. Metabolite profiling for plant functional genomics. Nature Biotechnology, 18(11): 1157-1161.

[27] Roessner U, Luedemann A, Brust D, et al. 2001. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. The Plant Cell, 13(1): 11-29.

[28] Roessner U, Wagner C, Kopka J, et al. 2000. Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. The Plant Journal, 23(1): 131-142.

[29] Sumner LW, Mendes P, Dixon RA. 2003. Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry, 62(6): 817-836.

[30] Hall RD. 2006. Plant metabolomics: from holistic hope, to hype, to hot topic. New Phytologist, 169(3): 453-468.

[31] Kusano M, Redestig H, Hirai T, et al. 2011. Covering chemical diversity of genetically-modified tomatoes using metabolomics for objective substantial equivalence assessment. PLoS One, 6(2): e16989.

[32] Kusano M, Fukushima A, Kobayashi M, et al. 2007. Application of a metabolomic method combining one-dimensional and two-dimensional gas chromatography-time-of-flight/mass spectrometry to metabolic phenotyping of natural variants in rice. Journal of Chromatography B, 855(1): 71-79.

[33] Schauer N, Semel Y, Roessner U, et al. 2006. Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nature Biotechnology, 24(4): 447-454.

[34] Tohge T, Nishiyama Y, Hirai MY, et al. 2005. Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. The Plant Journal, 42(2): 218-235.

[35] Liu CW, Murray JD. 2016. The role of flavonoids in nodulation host-range specificity: an update. Plants, 5(3): 33.