Tree Genetics and Molecular Breeding 2025, Vol.15, No.5, 192-201
http://genbreedpublisher.com/index.php/tgmb

Feature Review **Open Access**

Metabolomic and Genomic Analysis of Theobromine Biosynthesis in *Theobroma cacao*

Xingzhu Feng, Shaomin Yang, Jianli Zhong
Hainan Institute of Biotechnology, Haikou, 570206, Hainan, China
Corresponding email: jianli.zhong@hibio.org

Tree Genetics and Molecular Breeding, 2025, Vol.15, No.5 doi: 10.5376/tgmb.2025.15.0023
Received: 01 Sep., 2025
Accepted: 07 Oct., 2025
Published: 14 Oct., 2025

Preferred citation for this article:
Feng X.Z., Yang S.M., and Zhong J.L., 2025, Metabolomic and genomic analysis of theobromine biosynthesis in *Theobroma cacao*, Tree Genetics and Molecular Breeding, 15(5): 192-201 (doi: 10.5376/tgmb.2025.15.0023)

Abstract This study combined metabolomics and genomics data to analyze the biosynthesis of theobromine in *Theobroma cacao*. The results show that theobromine is mainly synthesized in the young peels and cotyledons, and gradually accumulates in the seeds as the tissues mature. Metabolic pathway analysis revealed that the synthesis of theobromine is related to the transformation of purine nucleotides such as AMP and GMP, with key intermediates including 7-methylxanthosine and 3-methylxanthine. Genomic research has found that during the domestication process, genes related to theobromine metabolism were significantly selected, thereby affecting the theobromine content in different varieties. Multi-omics data also indicate that post-processing steps such as fermentation can lead to a decrease in theobromine levels. This study aims to provide molecular targets and theoretical references for improving cocoa quality and developing functional components.

Keywords *Theobroma cacao*; Theobromine; Metabolomics; Genomics; Biosynthetic pathway

1 Introduction

Theobroma cacao is an important cash crop in tropical regions. Its seeds-cocoa beans-are the main raw materials for chocolate and a variety of foods, beverages and cosmetics. The cocoa industry has brought income to millions of smallholder farmers in West Africa, Central and South America, and Asia, and holds an important position in global agricultural trade, with an annual output of several million tons and a value of approximately 12 billion US dollars (González et al., 2019). Furthermore, cocoa bean by-products such as bean shells and peels are also rich in bioactive substances and are regarded as new resources for sustainable agriculture and the development of the bioeconomy.

Theobromine is the most abundant methylxanthine alkaloid in cocoa beans. It not only endows chocolate with a unique bitterness and flavor, but also has multiple physiological functions. Studies have shown that theobromine has health benefits such as antioxidation, anti-inflammation and cardiovascular protection (Sugimoto et al., 2014). Theobromine also plays an important role in the growth process of *Theobroma cacao* itself, which can help the plant resist pathogens and environmental stress (Pereira-Caro et al., 2013; Cornejo et al., 2018). Its content and distribution are affected by variety, developmental stage and processing procedure, and thereby determine the quality and functional characteristics of cocoa products (Zheng et al., 2004; Cortez et al., 2023; Ruiz-Santiago et al., 2024).

This study combines metabolomics and genomics to elucidate the biosynthetic pathways and regulatory mechanisms of theobromine, reveal its accumulation changes at different tissues and developmental stages, analyze related metabolic pathways and key enzyme genes, and explore the selection pressure and genetic variation of these genes in the domestication and variety improvement of *Theobroma cacao*. This study aims to provide theoretical basis and molecular tools for improving cocoa quality, developing functional foods, and promoting the sustainable utilization of cocoa resources.

2 Biochemical Pathway of Theobromine Biosynthesis

2.1 Overview of purine alkaloid biosynthesis

Theobromine and caffeine both belong to purine alkaloids and are the main active components in *Theobroma* *cacao* seeds. Their synthetic starting points are nucleotides such as adenosine monophosphate (AMP) and guanosine monophosphate (GMP). After a series of transformations, it is first converted into xanthosine, and then different methyl xanthine compounds are formed through methylation reactions. Theobromine is the most abundant purine alkaloid in the cacao tree, especially accumulating in large quantities in young tissues and the later stage of seed development (Koyama et al., 2003; Zheng et al., 2004; Pereira-Caro et al., 2013).

2.2 Enzymatic steps from xanthosine to theobromine (XMT, MXMT, DXMT enzymes)

The synthesis process of theobromine mainly consists of several steps. Firstly, xanthosine is converted into 7-methylxanthosine under the action of 7-methylxanthosine methyltransferase (XMT). Then, 7-methylxanthine nucleoside hydrolyzes to form 7-methylxanthine. Then, under the action of 3-methylxanthine methyltransferase (MXMT), it is further methylated to form theobromine (3,7-dimethylxanthine). In some tissues, theobromine can still be converted into caffeine (1,3,7-trimethylxanthine) under the action of 1-methylxanthine methyltransferase, but this process is relatively slow in *Theobroma cacao*, so the content of theobromine is much higher than that of caffeine (Koyama et al., 2003; Pereira-Caro et al., 2013).

2.3 Regulation of metabolic flux in caffeine and theobromine pathways

The accumulation of theobromine and caffeine can be influenced by tissue type, developmental stage and enzyme activity. Studies have found that theobromine has the highest content in young fruits and leaves, gradually decreases with maturity, but increases significantly in the later stage of seed development (Zheng et al., 2004; Pereira-Caro et al., 2013). The synthesis rate of caffeine is relatively low, mainly limited by the activity of the DXMT enzyme. Therefore, theobromine is the main end product in *Theobroma cacao*. In addition, metabolic fluxes are also regulated by gene expression and substrate specificity. The substrate specificity of N-methyltransferase determines the relative ratio of theobromine to caffeine (Koyama et al., 2003; González et al., 2019). Genomic studies have also shown that theobromine metabolism-related genes were subjected to selective pressure during domestication, which affected the accumulation pattern of alkaloids in different varieties (Cornejo et al., 2018).

3 Metabolomic Profiling of Theobromine in Cacao

3.1 Analytical platforms: LC- MS, GC- MS, and NMR for alkaloid profiling

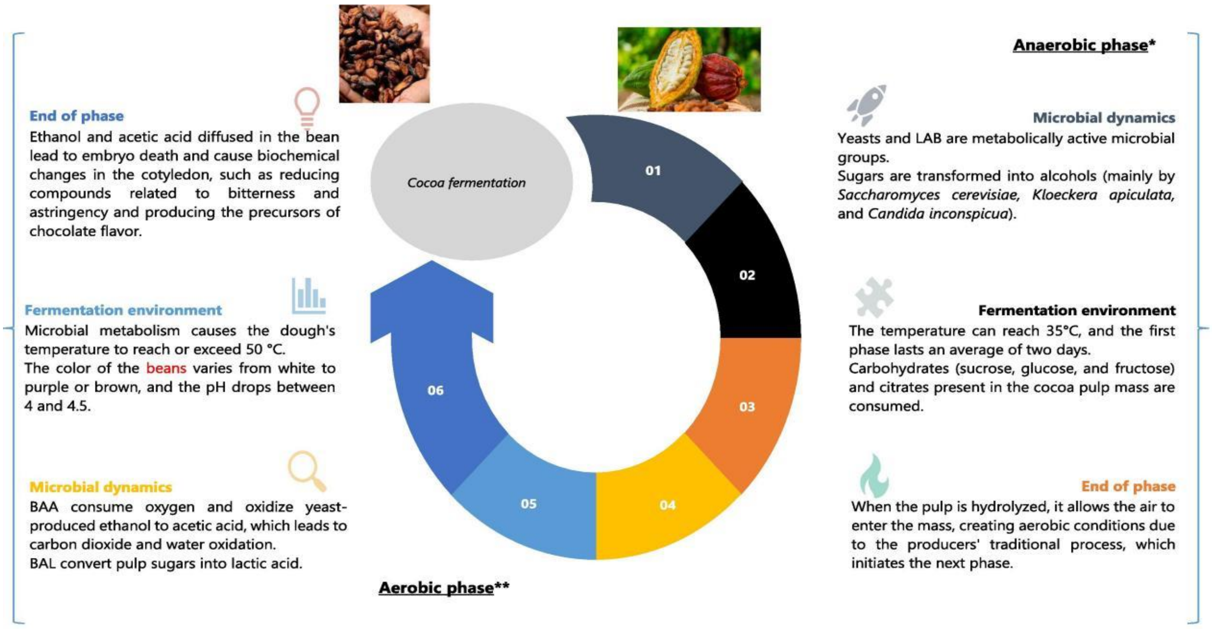
Theobromine is the main methyl xanthine alkaloid in *Theobroma cacao*. Metabolomics analysis of it relies on multiple highly sensitive platforms. Common methods include high performance liquid chromatography-mass spectrometry (LC-MS) and gas chromatography-mass spectrometry (GC-MS), which can effectively separate and detect methyl xanthine compounds in complex matrices (Pérez et al., 2018; González et al., 2019; Llerena et al., 2023). In addition, nuclear magnetic resonance (NMR) is also frequently used for the overall metabolite analysis of cocoa beans and their fermentation products. NMR can simultaneously detect amino acids, organic acids, sugars and methyl xanthine, etc., providing metabolic fingerprints for cocoa beans of different varieties and origins (Caligiani et al., 2010; Boutchouang et al., 2024). Overall, LC-MS and GC-MS stand out more in terms of sensitivity and specificity, while NMR is more suitable for high-throughput, non-destructive testing and multi-component quantification.

3.2 Quantitative variation of theobromine across tissues, developmental stages, and cultivars

The content of theobromine varies significantly among different tissues, developmental stages and varieties. Studies have found that theobromine levels are relatively high in young leaves and exocarp of fruits, gradually decreasing as tissues mature. However, in mature seeds, especially cotyledon, they accumulate in large quantities (Koyama et al., 2003; Zheng et al., 2004). There are also differences among different varieties. For example, the theobromine content in Nacional × Trinitario type bean shells is higher than that in CCN-51 type (Llerena et al., 2023). In addition, the theobromine/caffeine ratio of Forastero and Trinitario varieties can be used as a chemical marker to distinguish varieties (Collazos-Escobar et al., 2024). Metabolomics studies have further revealed the dynamic distribution patterns of theobromine at different developmental stages and among tissues (Koyama et al., 2003; Zheng et al., 2004).

3.3 Environmental and postharvest factors affecting theobromine levels

Environmental and processing conditions can also affect theobromine levels. Water stress can alter the accumulation of theobromine and related metabolites by regulating metabolic pathways (Boutchouang et al., 2024). During the fermentation process, the contents of methylxanthines such as theobromine and caffeine gradually decrease over time, and the variation range varies among different varieties (Figure 1) (Balcázar-Zumaeta et al., 2022; Cortez et al., 2023). Processes such as baking and drying can also affect the final content and distribution of theobromine (Figure 2) (Mladenović et al., 2018; Collazos-Escobar et al., 2024). Therefore, the environment and processing methods are important factors determining the metabolic characteristics of theobromine.



End of phase
Ethanol and acetic acid diffused in the bean lead to embryo death and cause biochemical changes in the cotyledon, such as reducing compounds related to bitterness and astringency and producing the precursors of chocolate flavor.

Fermentation environment
Microbial metabolism causes the dough's temperature to reach or exceed 50 °C.
The color of the beans varies from white to purple or brown, and the pH drops between 4 and 4.5.

Microbial dynamics
BAA consume oxygen and oxidize yeast-produced ethanol to acetic acid, which leads to carbon dioxide and water oxidation.
BAL convert pulp sugars into lactic acid.

Aerobic phase**

Anaerobic phase*

Microbial dynamics
Yeasts and LAB are metabolically active microbial groups.
Sugars are transformed into alcohols (mainly by *Saccharomyces cerevisiae*, *Kloeckera apiculata*, and *Candida inconspicua*).

Fermentation environment
The temperature can reach 35°C, and the first phase lasts an average of two days.
Carbohydrates (sucrose, glucose, and fructose) and citrates present in the cocoa pulp mass are consumed.

End of phase
When the pulp is hydrolyzed, it allows the air to enter the mass, creating aerobic conditions due to the producers' traditional process, which initiates the next phase.

Figure 1 Cocoa fermentation phases (Adopted from Balcázar-Zumaeta et al., 2022)

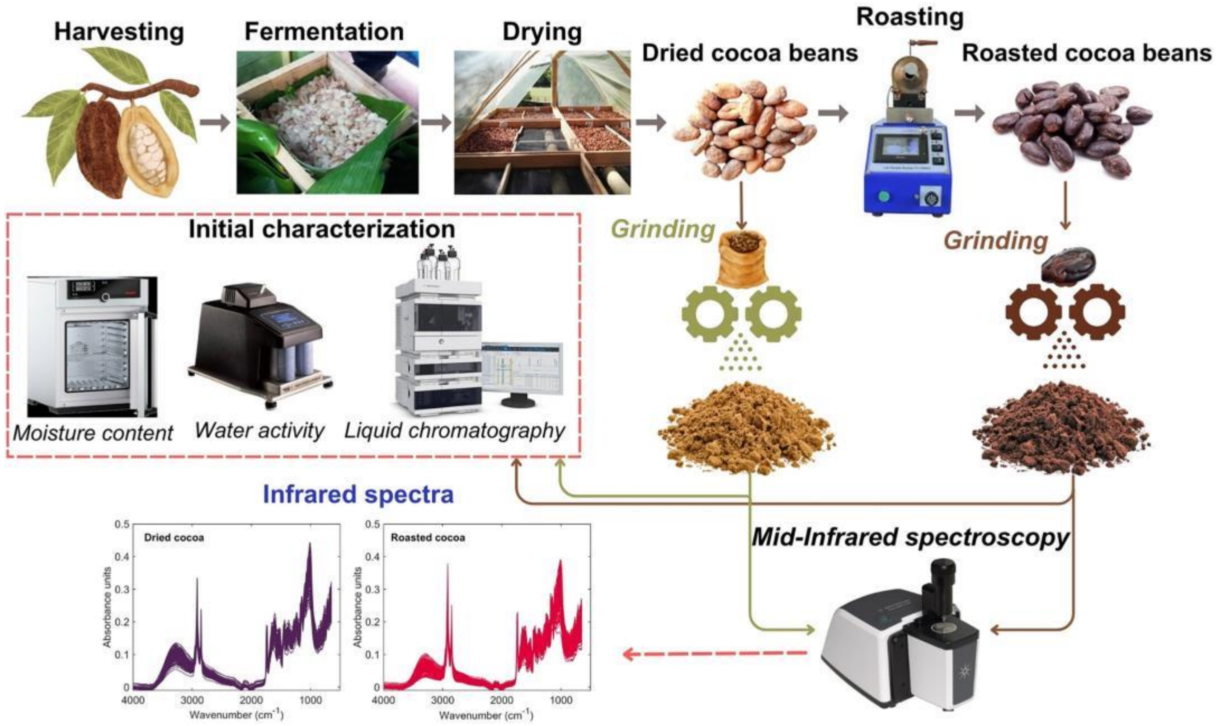


Figure 2 Flow chart illustrating the experimental procedure used to determine mid-infrared spectra and theobromine/caffeine content of cocoa samples for obtaining the dataset (Adopted from Collazos-Escobar et al., 2024)

4 Genomic Insights into Theobromine Biosynthesis

4.1 Genome sequencing and annotation of *Theobroma cacao*

High-quality genomic sequencing and annotation of *Theobroma cacao* have provided a basis for studying the synthetic mechanisms of metabolites such as theobromine. Current research has covered approximately 80% of *Theobroma cacao* genome, and most genes are anchored on 10 chromosomes. The results showed that some gene families of *Theobroma cacao* had specific expansions, especially those related to secondary metabolites, such as flavonoids and methylxanthines (Argout et al., 2011). Meanwhile, the study also identified and annotated over 7 million SNPs. These data provided rich resources for functional gene localization and trait improvement (Cornejo et al., 2018). These genomic achievements not only contribute to the discovery of genes related to theobromine synthesis, but also lay a solid foundation for molecular breeding and functional gene research (Allègre et al., 2011).

4.2 Identification of candidate genes and regulatory networks

Genome-wide annotation and functional analysis have identified a number of candidate genes and regulatory networks closely related to theobromine synthesis. Theobromine is a purine alkaloid, and its synthesis depends on multiple methyltransferase and other enzyme genes. Population genome studies have found that during domestication, theobromine metabolism-related genes were subjected to significant selective pressure (Cornejo et al., 2018). In addition, some gene families (such as GASA, BAHD, MADS-box, etc.) have been systematically identified in *Theobroma cacao*. Expression profiling analysis indicated that these genes were involved in seed development, disease resistance and metabolic regulation (Abdullah et al., 2021a; Abdullah et al., 2021b; Zhang et al., 2021). High-density genetic mapping and QTL mapping also helped achieve fine mapping of candidate genes related to theobromine content (Allègre et al., 2011).

4.3 Comparative genomics with other Theobromine/caffeine-producing species

Comparative genomics shows that there are significant differences in genomic structure and functional genes between *Theobroma cacao* and other species that produce methyl xanthine, such as coffee and tea. *Theobroma cacao* mainly accumulates theobromine rather than caffeine, which is related to the expression and regulation mode of its methylase gene (Zheng et al., 2004; Agyirifo et al., 2019). Comparisons with the closely related species *Theobroma grandiflorum* also revealed differences in genomic structure, genetic content and evolutionary pathways between the two, providing evidence for understanding the evolutionary basis of theobromine synthesis (Kuhn et al., 2010). In addition, the comparison of chloroplast and mitochondrial genomes also provides new ideas for understanding the phylogenetic and metabolic diversity of *Theobroma cacao* and related species (De Abreu et al., 2022; Tineo et al., 2025).

5 Transcriptomic and Functional Genomics Approaches

5.1 Expression profiling of alkaloid biosynthesis genes

Through high-throughput technologies such as RNA sequencing (RNA-seq), researchers can analyze the expression of alkaloid synthesis genes in *Theobroma cacao*. For instance, co-expression analysis has identified some gene modules related to protein modification, flowering and water transport. These modules showed obvious polygenic differentiation in different *Theobroma cacao* populations, providing clues for understanding the regulation of alkaloid synthesis genes (Hämälä et al., 2019). In addition, genome-wide analyses of some specific gene families (such as BAHD, GASA, etc.) revealed that some of their members were upregulated in expression during embryonic development and stress conditions (such as fungal infection), suggesting that they might be related to alkaloid metabolism and stress resistance (Abdullah et al., 2021a; Abdullah et al., 2021b; Nguyen et al., 2025).

5.2 Gene editing and functional validation

In functional gene research, transient expression systems and gene editing techniques (such as CRISPR and RNAi) are often used to verify the functions of candidate genes. For instance, Agrobacterium-mediated transient transformation can enable the expression of exogenous genes in *Theobroma cacao* leaves within a short period of time. Researchers can detect the expression changes of the target gene through qRT-PCR and observe whether it affects metabolites or disease resistance (Fister et al., 2016). At present, there are not many studies directly verifying theobromine synthesis genes using CRISPR or RNAi. However, these techniques have been applied in some defense genes (such as chitinase), and their functions have been confirmed, laying the foundation for subsequent studies of theobrosin-related genes (Fister et al., 2016).

5.3 Integration of transcriptomics with metabolomic data

Combining transcriptome and metabolome data for analysis can provide a clearer view of the relationship between gene expression and metabolites. Research has found that the accumulation of theobromine at different developmental stages is closely related to the expression of related synthetic genes. Meanwhile, genotype differences among different populations can affect the efficiency of alkaloid metabolic pathways (Koyama et al., 2003; Zheng et al., 2004). Genomic selection signal analysis also revealed that during domestication, theobromine synthesis-related genes were strongly selected, further indicating that these genes are important in *Theobroma cacao* quality formation (Cornejo et al., 2018).

6 Case Study: Metabolomic and Genomic Analysis in Cacao Cultivars

6.1 Selection of cultivars with contrasting theobromine content

The varieties of *Theobroma cacao* vary greatly in the content of theobromine. Traditionally, the main cultivation types are Criollo, Forastero and Trinitario. Modern genomic research has further classified *Theobroma cacao* germplasm into ten genetic populations, including Amelonado, Contamana, Criollo, Curaray, Guianna, Iquitos, Marañon, Nacional, Nanay and Purús. These groups showed significant differences in genetic structure and metabolite accumulation, providing a basis for comparing high and low theobromine varieties (Zheng et al., 2004; Cornejo et al., 2018).

6.2 Methodology: integrated metabolomic- genomic analysis pipeline

The research process is generally divided into several steps. First, collect leaf or fruit samples from different varieties. Secondly, whole-genome resequencing was conducted using high-throughput sequencing technologies (such as the Illumina platform) to obtain a large amount of SNP and gene annotation information (Argout et al., 2011; Cornejo et al., 2018). Thirdly, the contents of theobromine and related metabolites were determined by metabolomics methods such as HPLC-MS/MS (Zheng et al., 2004). Finally, through GWAS and expression profiling analysis, the associations between metabolite levels and gene variations or gene expression patterns were identified (Zheng et al., 2004; Cornejo et al., 2018).

6.3 Findings: key genes, metabolite variations, and breeding implications

The results indicated that domesticated varieties such as Criollo underwent intense selection in genes related to theobromine metabolism. Some genes are closely related to the synthesis of flavor substances such as theobromine and anthocyanins (Cornejo et al., 2018). Theobromine is mainly synthesized in young fruits and cotylodes. The pathway is that AMP/IMP is first converted to 7-methylxanthine nucleoside, and then theobromine is generated through multiple steps of reaction (Koyama et al., 2003). There are significant differences in the ability of theobromine synthesis and degradation among different varieties and tissues. Moreover, most theobromine is directly synthesized in the seeds rather than transported from the pericarp (Zheng et al., 2004). Genomic studies have also found that beneficial genes related to flavors such as theobromine were selected during domestication, but some harmful mutations were also accumulated at the same time, which may affect the adaptability and yield of plants (Cornejo et al., 2018). These achievements provide important genetic resources for molecular breeding and contribute to the cultivation of new varieties with high theobromine content, excellent flavor and stress resistance (Argout et al., 2011; Cornejo et al., 2018).

7 Applications for Breeding and Industry

7.1 Marker-assisted breeding for flavor and health-enhancing alkaloid content

Studies have shown that the biosynthetic pathway and genetic basis of theobromine are relatively clear, which provides strong support for molecular marker-assisted breeding (MAS). The results of population genomics show that during the domestication of *Theobroma cacao*, theobromine metabolism-related genes were strongly selected, especially in the Criollo variety, where the variations of related genes were closely related to flavor and functional components (Cornejo et al., 2018). At present, there are a large number of SNP resources and genomic annotation data, which provide a basis for the development of molecular markers related to high theobromine content, superior flavor and disease resistance (Cornejo et al., 2018). Through these markers, breeders can more quickly screen out new materials that have both good flavor and health value to meet the demands of high-end chocolate and functional foods (Wickramasuriya and Dunwell, 2017).

7.2 Biotechnological strategies for metabolic engineering of theobromine levels

With the development of genomics and metabolomics, metabolic engineering has become another means to regulate theobromine levels. By using gene editing and transgenic techniques, key enzyme genes (such as 7-methylxanthine synthase) can be targeted and regulated, thereby increasing or decreasing theobromine content (Koyama et al., 2003). In addition, techniques such as in vitro embryogenesis and genetic transformation have been applied to the genetic improvement of *Theobroma cacao*, providing a platform for the precise regulation of alkaloid metabolism. These methods can not only help cultivate specialized varieties with high or low theobromine, but also carry out multi-trait improvement in combination with goals such as disease resistance and stress tolerance (Wickramasuriya and Dunwell, 2017).

7.3 Industrial implications for chocolate quality, pharmacology, and sustainability

The level of theobromine directly affects the flavor, bitterness and functionality of chocolate, thereby determining the market positioning and consumer experience of the product. Cocoa beans with a high theobromine content can produce chocolate with a stronger flavor and higher health value, meeting the demands of high-end and functional markets (Cortez et al., 2023). Meanwhile, theobromine and its derivatives have antioxidant and cardiovascular protective effects in the fields of health care and pharmacology, and have also expanded the application prospects of *Theobroma cacao* industry (Wickramasuriya and Dunwell, 2017). Increasing theobromine levels through genomics and metabolomics helps improve raw material utilization and product added value, and promotes the sustainable development of the industry (Wickramasuriya and Dunwell, 2017; Cornejo et al., 2018). Meanwhile, precision breeding and biotechnology improvement can also reduce dependence on the environment, enhance production efficiency, increase the income of small-scale farmers, and promote the green transformation of the global *Theobroma cacao* industry (Wickramasuriya and Dunwell, 2017).

8 Challenges and Research Gaps

8.1 Complexity of alkaloid regulation and redundancy in gene families

The synthesis of theobromine involves multiple purine metabolic pathways and also requires the joint action of various enzymes. This regulatory network is very complex. Research has found that there are often multiple homologous genes and repetitive enzyme activities in related pathways, so the metabolic patterns of different tissues and developmental stages vary greatly. For instance, the theobromine content in young fruits and tender leaves is relatively high, but it will significantly decrease as tissues mature. Moreover, the activities of synthetic and degradation pathways in different tissues also vary (Koyama et al., 2003; Zheng et al., 2004). Transcriptome analysis also revealed that multiple transcription factors and signaling pathways interact, which makes it more difficult to identify key regulatory nodes. The redundancy of gene families also indicates that it is difficult to precisely regulate theobromine content through single gene editing (Gallego et al., 2021).

8.2 Integration difficulties between metabolomic and genomic datasets

Although metabolomics and genomic technologies have provided a wealth of data, the integration of the two remains highly challenging. The dynamic changes of metabolite abundance are often inconsistent with the expression of genes at different times and in different tissues, so it is difficult to directly establish the correspondence between genes and metabolites (Gallego et al., 2021). Environmental factors, such as light exposure and fermentation treatment, can also simultaneously affect metabolites and gene expression, increasing the complexity of integration (Cortez et al., 2023). At present, there is a lack of a unified analytical framework and standardized methods, which limits the in-depth understanding of the regulatory mechanism of theobromine synthesis (Gallego et al., 2021).

8.3 Limited field-level validation of laboratory findings

Most research on theobromine is conducted in laboratories, mainly at the cellular, tissue or organ level. The drawback is the lack of validation in the field environment (Koyama et al., 2003; Zheng et al., 2004). The metabolic and gene expression patterns in the laboratory often differ from those in the field environment. Especially when facing environmental stress, pests and diseases or agronomic management, the differences are more obvious (Gallego et al., 2021). In addition, post-harvest treatments, such as fermentation and baking, can also significantly affect the content of theobromine and related alkaloids, but the molecular mechanism is not fully understood (Febrianto and Zhu, 2022; Cortez et al., 2023). Therefore, for laboratory achievements to be truly applied to field production, more research is still needed.

9 Future Perspectives

9.1 Multi-omics integration: metabolomics, genomics, transcriptomics, proteomics

Multi-omics integration has become an important method for studying complex biological processes. Analyzing the genome, transcriptome, proteome and metabolome together can provide a more comprehensive understanding of the synthesis process and regulatory network of theobromine. This method can link genes and phenotypes, and also identify key regulatory factors and metabolic pathways, providing a basis for functional gene research and molecular marker development (Picard et al., 2021; Wörheide et al., 2021; Ramos-López et al., 2022). In the study of cocoa processing and flavor, multi-omics analysis also revealed metabolite differences in different regions, varieties and processing stages, and more molecules related to flavor were understood through proteomic and peptidomic studies (Agyirifo et al., 2019; Herrera-Rocha et al., 2023). In the future, with the development of high-throughput technologies and data analysis tools, multi-omics integration will play a greater role in cocoa quality improvement and functional component development (Mahmood et al., 2022; Shankar and Sharma, 2022; Cembrowska-Lech et al., 2023).

9.2 AI-driven metabolite prediction and pathway modeling

Artificial intelligence (AI) and machine learning (ML) provide new means for the integration of multi-omics data and the analysis of complex networks. AI can help predict the relationship among genotypes, environment and phenotypes through deep learning, knowledge graphs and generative models, and can also discover new molecular targets and biomarkers (Picard et al., 2021; Mahmood et al., 2022; Yan and Wang, 2022). In metabolite prediction, AI can handle complex big data and also improve the training effect of models through synthetic data, thereby more accurately identifying new metabolic pathways and key enzymes (Shankar and Sharma, 2022; Wu and Xie, 2024). These technologies will contribute to more precise regulation of theobromine and related flavors and nutritional components (Yan and Wang, 2022; Cembrowska-Lech et al., 2023).

9.3 Toward precision breeding for optimized flavor and nutritional value

Combining multi-omics with AI provides a new direction for the precise breeding of *Theobroma cacao*. By integrating genomic, transcriptomic, proteomic and metabolomic data and combining them with AI models, complex traits such as flavor and nutrition can be predicted more quickly and accurately (Shankar and Sharma, 2022; Yan and Wang, 2022). Precision breeding can accelerate the selection and breeding of superior varieties, and also improve the flavor and nutrition of cocoa in a targeted manner according to different market demands, achieving optimization from genes to products (Mahmood et al., 2022). In the future, with the accumulation of more data and the improvement of AI algorithms, intelligent breeding relying on molecular markers and phenotypic prediction will become an important driving force for promoting the high-quality development of *Theobroma cacao* industry (Yan and Wang, 2022; Cembrowska-Lech et al., 2023).

Acknowledgments

The authors appreciate the comments from two anonymous peer reviewers on the manuscript of this study. The authors also thank the group members for helping to organize the research data.

Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Abdullah, Faraji S., Heidari P., and Poczai P., 2021a, The *BAHD* gene family in cacao (*Theobroma cacao*, Malvaceae): genome-wide identification and expression analysis, Front. Ecol. Evol.,, 9: 707708.
https://doi.org/10.3389/fevo.2021.707708

Abdullah, Faraji S., Mehmood F., Malik H., Ahmed I., Heidari P., and Poczai P., 2021b, The GASA gene family in cacao (*Theobroma cacao*, Malvaceae): genome wide identification and expression analysis, Agronomy, 11(7): 1425.
https://doi.org/10.3390/agronomy11071425

Agyirifo D., Wamalwa M., Otwe E., Galyuon I., Runo S., Takrama J., and Ngeranwa J., 2019, Metagenomics analysis of cocoa bean fermentation microbiome identifying species diversity and putative functional capabilities, Heliyon, 5(7): e02170.
https://doi.org/10.1016/j.heliyon.2019.e02170

Allègre M., Argout X., Boccara M., Fouet O., Roguet Y., Bérard A., Thévenin J., Chauveau A., Rivallan R., Clément D., Courtois B., Gramacho K., Boland-Augé A., Tahi M., Umaharan P., Brunel D., and Lanaud C., 2011, Discovery and mapping of a new expressed sequence tag-single nucleotide polymorphism and simple sequence repeat panel for large-scale genetic studies and breeding of *Theobroma cacao* L., DNA Research, 19(1): 23-35.
https://doi.org/10.1093/dnares/dsr039

Argout X., Salse J., Aury J., Guiltinan M., Droc G., Gouzy J., Allègre M., Chaparro C., Legavre T., Maximova S., Abrouk M., Murat F., Fouet O., Poulain J., Ruiz M., Roguet Y., Rodier-Goud M., Barbosa-Neto J., Sabot F., Kudrna D., Ammiraju J., Schuster S., Carlson J., Sallet E., Schiex T., Diévart A., Kramer M., Gelley L., Shi Z., Bérard A., Viot C., Boccara M., Risterucci A., Guignon V., Sabau X., Axtell M., Ma Z., Zhang Y., Brown S., Bourge M., Golser W., Song X., Clément D., Rivallan R., Tahi M., Akaza J., Pitollat B., Gramacho K., D'Hont A., Brunel D., Infante D., Kébé I., Costet P., Wing R., McCombie W., Guiderdoni E., Quétier F., Panaud O., Wincker P., Bocs S., and Lanaud C., 2011, The genome of *Theobroma cacao*, Nature Genetics, 43: 101-108.
https://doi.org/10.1038/ng.736

Balcázar-Zumaeta C., Castro-Alayo E., Cayo-Colca I., Idrogo-Vásquez G., and Muñoz-Astecker L., 2022, Metabolomics during the spontaneous fermentation in cocoa (*Theobroma cacao* L.): an exploraty review, Food Research International, 163: 112190.
https://doi.org/10.1016/j.foodres.2022.112190

Boutchouang R., Fliniaux O., Eyamo J., Djabou A., Fontaine J., Molinié R., Mesnard F., and Niemenak N., 2024, Metabolome profiling of cacao (*Theobroma cacao* L.) callus under drought stress conditions induced by polyethylene glycol (PEG) as osmoticant, Phytochemical Analysis, 35(4): 708-722.
https://doi.org/10.1002/pca.3323

Caligiani A., Acquotti D., Cirlini M., and Palla G., 2010, 1H NMR study of fermented cocoa (*Theobroma cacao* L.) beans, Journal of Agricultural and Food Chemistry, 58(23): 12105-12111.
https://doi.org/10.1021/jf102985w

Cembrowska-Lech D., Krzemińska A., Miller T., Nowakowska A., Adamski C., Radaczyńska M., Mikiciuk G., and Mikiciuk M., 2023, An integrated multi-omics and artificial intelligence framework for advance plant phenotyping in horticulture, Biology, 12(10): 1298.
https://doi.org/10.3390/biology12101298

Collazos-Escobar G., Bahamón-Monje A., and Gutiérrez-Gúzman N., 2024, Mid-infrared spectra of dried and roasted cocoa (*Theobroma cacao* L.): a dataset for machine learning-based classification of cocoa varieties and prediction of theobromine and caffeine content, Data in Brief, 58: 111243.
https://doi.org/10.1016/j.dib.2024.111243

Cornejo O., Yee M., Domínguez V., Andrews M., Sockell A., Strandberg E., Livingstone D., Stack C., Romero A., Umaharan P., Royaert S., Tawari N., Ng P., Gutiérrez O., Phillips W., Mockaitis K., Bustamante C., and Motamayor J., 2018, Population genomic analyses of the chocolate tree, *Theobroma cacao* L., provide insights into its domestication process, Communications Biology, 1: 167.
https://doi.org/10.1038/s42003-018-0168-6

Cortez D., Quispe-Sanchez L., Mestanza M., Oliva M., Yoplac I., Torres C., and Chávez S., 2023, Changes in bioactive compounds during fermentation of cocoa (*Theobroma cacao*) harvested in Amazonas-Peru, Current Research in Food Science, 6: 100494.
https://doi.org/10.1016/j.crfs.2023.100494

De Abreu V., Alves R., Silva S., Ferro J., Domingues D., Miranda V., and Varani A., 2022, Comparative analyses of *Theobroma cacao* and *T. grandiflorum* mitogenomes reveal conserved gene content embedded within complex and plastic structures, Gene, 849: 146904.
https://doi.org/10.1016/j.gene.2022.146904

Febrianto N., and Zhu F., 2022, Comparison of bioactive components and flavor volatiles of diverse cocoa genotypes of *Theobroma grandiflorum*, *Theobroma bicolor*, *Theobroma subincanum* and *Theobroma cacao*, Food Research International, 161: 111764.
https://doi.org/10.1016/j.foodres.2022.111764

Fister A., Shi Z., Zhang Y., Helliwell E., Maximova S., and Guiltinan M., 2016, Protocol: transient expression system for functional genomics in the tropical tree *Theobroma cacao* L., Plant Methods, 12: 19.
https://doi.org/10.1186/s13007-016-0119-5

Gallego A., Rojas L., Valencia W., Atehortúa L., Urrea A., Fister A., Guiltinan M., Maximova S., and Pabón-Mora N., 2021, Transcriptomic analyses of cacao flavonoids produced in photobioreactors, BMC Genomics, 22: 551.
https://doi.org/10.1186/s12864-021-07871-0

González J., Monan M., Pérez J., Gómez E., De La C. Salgado D., and Pérez D., 2019, Determination of theobromine and caffeine in *Theobroma cacao* husk from ethanolic extract by GC-MS after CC separation, OALib, 6: e5771.
https://doi.org/10.4236/oalib.1105771

Hämälä T., Guiltinan M., Marden J., Maximova S., dePamphilis C., and Tiffin P., 2019, Gene expression modularity reveals footprints of polygenic adaptation in *Theobroma cacao*, Molecular Biology and Evolution, 37: 110-123.
https://doi.org/10.1093/molbev/msz206

Herrera-Rocha F., Fernández-Niño M., Cala M., Duitama J., and Barrios A., 2023, Omics approaches to understand cocoa processing and chocolate flavor development: a review, Food Research International, 165: 112555.
https://doi.org/10.1016/j.foodres.2023.112555

Koyama Y., Tomoda Y., Kato M., and Ashihara H., 2003, Metabolism of purine bases, nucleosides and alkaloids in theobromine-forming *Theobroma cacao* leaves, Plant Physiology and Biochemistry, 41: 977-984.
https://doi.org/10.1016/j.plaphy.2003.07.002

Kuhn D., Figueira A., Lopes U., Motamayor J., Meerow A., Cariaga K., Freeman B., Livingstone D., and Schnell R., 2010, Evaluating *Theobroma grandiflorum* for comparative genomic studies with *Theobroma cacao*, Tree Genetics and Genomes, 6: 783-792.
https://doi.org/10.1007/s11295-010-0291-0

Llerena W., Samaniego I., Vallejo C., Arreaga A., Zhunio B., Coronel Z., Quiroz J., Angós I., and Carrillo W., 2023, Profile of bioactive components of cocoa (*Theobroma cacao* L.) by-products from ecuador and evaluation of their antioxidant activity, Foods, 12(13): 2583.
https://doi.org/10.3390/foods12132583

Mahmood U., Li X., Fan Y., Chang W., Niu Y., Li J., Qu C., and Lu K., 2022, Multi-omics revolution to promote plant breeding efficiency, Frontiers in Plant Science, 13: 1062952.
https://doi.org/10.3389/fpls.2022.1062952

Mladenović K., Root Y., and Ramanathan D., 2018, UHPLC-HRMS analysis of theobromine in *Theobroma cacao* and its products, Journal of Nutrition and Food Sciences, 8: 737.
https://doi.org/10.4172/2155-9600.1000737

Nguyen Q., Cao P., Chu N., Tran L., Le M., Luong H., Le Q., Dong G., and Chu H., 2025, The trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase in cocoa (*Theobroma cacao* L.): genome-wide identification and expression analysis, HAYATI Journal of Biosciences, 32(4): 940-949.
https://doi.org/10.4308/hjb.32.4.940-949

Pereira-Caro G., Borges G., Nagai C., Jackson M., Yokota T., Crozier A., and Ashihara H., 2013, Profiles of phenolic compounds and purine alkaloids during the development of seeds of *Theobroma cacao* cv. Trinitario, Journal of Agricultural and Food Chemistry, 61(2): 427-434.
https://doi.org/10.1021/jf304397m

Pérez D., González J., Gómez E., Pérez J., Rodríguez J., and Monan M., 2018, Chemical characterization of metabolites from the husk of *Theobroma cacao* by GC-MS in Cuba, Journal of Applied Biotechnology, 6(2): 39-48.
https://doi.org/10.5296/jab.v6i2.13263

Picard M., Scott-Boyer M., Bodein A., Périn O., and Droit A., 2021, Integration strategies of multi-omics data for machine learning analysis, Computational and Structural Biotechnology Journal, 19: 3735-3746.
https://doi.org/10.1016/j.csbj.2021.06.030

Ramos-López O., Martínez J., and Milagro F., 2022, Holistic integration of omics tools for precision nutrition in health and disease, Nutrients, 14(19): 4074.
https://doi.org/10.3390/nu14194074

Ruiz-Santiago F., Márquez-Rocha F., García-Alamilla P., Carrera-Lanestosa A., Ramírez-López C., Ocaranza-Sánchez E., and Jiménez-Rodríguez D., 2024, Physicochemical and biochemical changes in cocoa during the fermentation step, Fermentation, 10(8): 405.
https://doi.org/10.3390/fermentation10080405

Shankar A., and Sharma K., 2022, Fungal secondary metabolites in food and pharmaceuticals in the era of multi-omics, Applied Microbiology and Biotechnology, 106: 3465-3488.
https://doi.org/10.1007/s00253-022-11945-8

Sugimoto N., Miwa S., Hitomi Y., Nakamura H., Tsuchiya H., and Yachie A., 2014, Theobromine, the primary methylxanthine found in *Theobroma cacao*, prevents malignant glioblastoma proliferation by negatively regulating phosphodiesterase-4, extracellular signal-regulated kinase, Akt/mammalian target of rapamycin kinase, and nuclear factor-kappa B, Nutrition and Cancer, 66: 419-423.
https://doi.org/10.1080/01635581.2013.877497

Tineo D., Bustamante D., Calderon M., and Oliva M., 2025, Comparative analyses of chloroplast genomes of *Theobroma cacao* from northern Peru, PLoS One, 20(3): e0316148.
https://doi.org/10.1371/journal.pone.0316148

Wickramasuriya A., and Dunwell J., 2017, Cacao biotechnology: current status and future prospects, Plant Biotechnology Journal, 16: 4-17.
https://doi.org/10.1111/pbi.12848

Wörheide M., Krumsiek J., Kastenmüller G., and Arnold M., 2021, Multi-omics integration in biomedical research - a metabolomics-centric review, Analytica Chimica Acta, 1141: 144-162.
https://doi.org/10.1016/j.aca.2020.10.038

Wu Y., and Xie L., 2024, AI-driven multi-omics integration for multi-scale predictive modeling of genotype-environment-phenotype relationships, Computational and Structural Biotechnology Journal, 27: 265-277.
https://doi.org/10.1016/j.csbj.2024.12.030

Yan J., and Wang X., 2022, Machine learning bridges omics sciences and plant breeding, Trends in Plant Science, 28(2): 199-210.
https://doi.org/10.1016/j.tplants.2022.08.018

Zhang Q., Hou S., Sun Z., Chen J., Meng J., Liang D., Wu R., and Guo Y., 2021, Genome-wide identification and analysis of the MADS-box gene family in *Theobroma cacao*, Genes, 12(11): 1799.
https://doi.org/10.3390/genes12111799

Zheng X., Koyama Y., Nagai C., and Ashihara H., 2004, Biosynthesis, accumulation and degradation of theobromine in developing *Theobroma cacao* fruits, Journal of Plant Physiology, 161(4): 363-369.
https://doi.org/10.1078/0176-1617-01253