The Diversity of Nutritional Metabolites: Origin, Dissection, and Application in Crop Breeding

Chuanying Fang[1], Jie Luo[1,2], and Shouchuang Wang[1]*

[1] Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresource, College of Tropical Crops, Hainan University, Haikou, China, [2] National Key Laboratory of Crop Genetic Improvement and National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan, China

The chemical diversity of plants is very high, and plant-based foods provide almost all the nutrients necessary for human health, either directly or indirectly. With advancements in plant metabolomics studies, the concept of nutritional metabolites has been expanded and updated. Because the concentration of many nutrients is usually low in plant-based foods, especially those from crops, metabolome-assisted breeding techniques using molecular markers associated with the synthesis of nutritional metabolites have been developed and used to improve nutritional quality of crops. Here, we review the origins of the diversity of nutrient metabolites from a genomic perspective and the role of gene duplication and divergence. In addition, we systematically review recent advances in the metabolomic and genetic basis of metabolite production in major crops. With the development of genome sequencing and metabolic detection technologies, multi-omic integrative analysis of genomes, transcriptomes, and metabolomes has greatly facilitated the deciphering of the genetic basis of metabolic pathways and the diversity of nutrient metabolites. Finally, we summarize the application of nutrient diversity in crop breeding and discuss the future development of a viable alternative to metabolome-assisted breeding techniques that can be used to improve crop nutrient quality.

Keywords: nutritional metabolites, metabolic diversity, crops, genetic bases, breeding
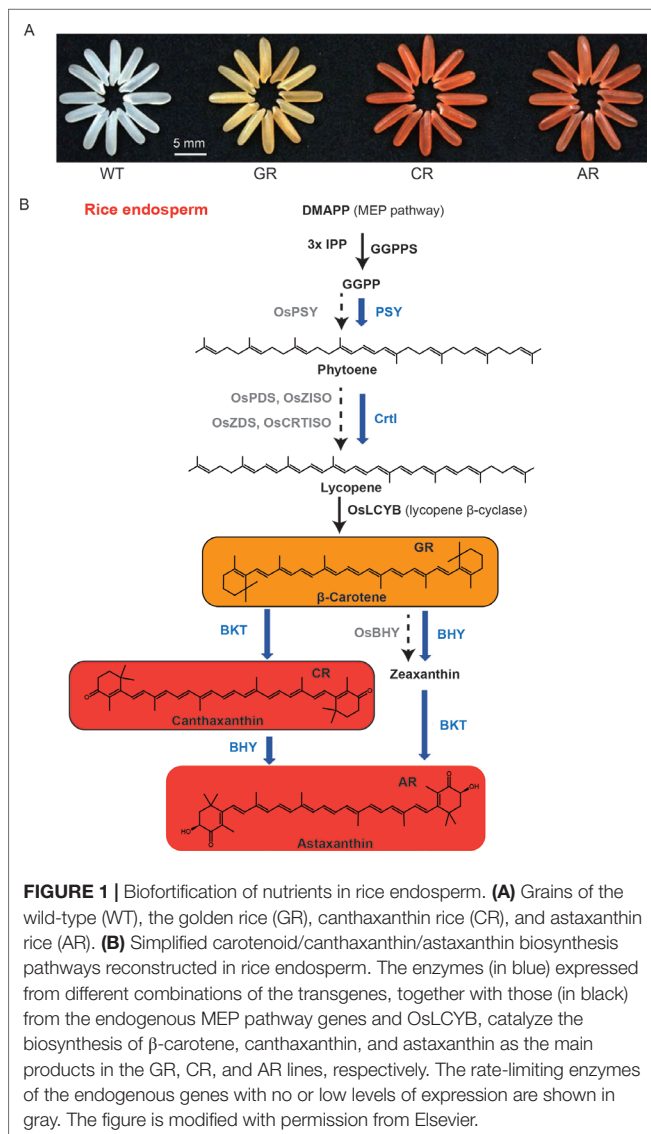
INTRODUCTION

The nutritional metabolites needed for humans to maintain health are mainly derived from plants, either directly or indirectly when plants are consumed by animals (Dellapenna, 1999). Plant-derived foods, especially crops, provide almost all essential human nutrients such as amino acids, vitamins (tocopherol, ascorbic acid, folic acid), sugars (sucrose, glucose), as well as other health-promoting phytochemicals (Hall et al., 2008). Traditionally, nutritional metabolites are generally not considered to be directly synthesized in the human body, or the specific factors required in their synthetic pathways are lacking or insufficient under certain conditions, and humans must obtain these components from exogenous food (e.g., some amino acids, fatty acids, vitamins) (Hounsome et al., 2008). Golden rice is an important achievement in the improvement of crop nutritional quality through genetically modified technology. Paine et al. developed "Golden rice 2," introducing *psy* from maize in combination with the *Erwinia uredovora* carotene desaturase (*crtl*) from *Erwinia*

uredovora that was used to generate the original golden rice. The β-carotene (provitamin A) content in golden rice is significantly improved, which is helpful in fighting against vitamin A deficiency (Paine et al., 2005). Recently, Zhu et al. introduced four synthetic genes in rice endosperm to achieve astaxanthin biosynthesis, and these four genes are *sZmPSY1*, *sPaCrtl*, *sCrBKT*, and *sHpBHY*, which encode the enzymes phytoene synthase, phytoene desaturase, β-carotene ketolase (BKT), and β-carotene hydroxylase, respectively (**Figure 1**) (Zhu et al., 2018b). However, many other phytochemicals are also nutrients that have a positive effect on human health, such as flavonoids, phytosterols, phenolic acids, carotenoids, polyunsaturated fatty acids, and glucosinolates, which are effective in preventing the occurrence of clinical disease risk (Appleby et al., 2014; Wang et al., 2014). In addition to crop yields and stress resistance, more researches have begun to focus on nutritional quality and how to improve novel nutrients such as anthocyanins, carotenoids, and resveratrol in major crops (Fernie et al., 2006). The anthocyanin

content of purple tomato and purple rice has been significantly improved by transgenic technology, and it is considered to be an important manifestation of crop nutrient quality improvement, achieving the goal of crop nutrient biofortification (Butelli et al., 2008; Zhu et al., 2017).

The diversity of nutrients in crops is very complex and difficult to assess. The amount and type of nutritional metabolites are strongly affected by genetic and environmental factors, which are major contributors to nutrient diversity (Hounsome et al., 2008; Ahuja et al., 2010). Because of the inability to move during growth, plants must evolve a series of protective mechanisms to counter the unfavorable environment in order to maintain normal life activities. Environmental factors, including light intensity, temperature, drought, UV radiation, salinity, toxic heavy metals, and so on, induce plants to produce rich metabolic diversity, which provides the possibility of screening and utilization of nutrients in plant (Orcutt and Nilsen, 2000; Hounsome et al., 2008). For example, UV-B radiation induces the accumulation of the multi-functional active flavonoids in the corresponding tissues (Treutter, 2006); under drought conditions, the metabolites, such as carbohydrate metabolites, glycine betaine, proline, ectoine, can be increased to alleviate the damage of plant cells caused by water shortage (Mahajan and Tuteja, 2005). Both biotic and abiotic stresses induce the diversity of metabolites and also affect the nutritional quality of crops. Most metabolites such as sugars, organic acids, amino acids, vitamins, hormones, flavonoids, phenolics, and glucosinolates are essential for plant growth, development, stress adaptation, and defense, and the diversity of these metabolites also determines the nutritional quality, color, taste, and smell as well as antioxidative, anticarcinogenic, anti-inflammatory, antimicrobial, and cholesterol-lowering properties of food (Hounsome et al., 2008). Therefore, deciphering the metabolic diversity and genetic regulation of nutrients makes it possible to develop metabolic markers and genetic loci for metabolome-assisted breeding and biofortification (Luo, 2015; Martin and Li, 2017). Taking advantage of high-throughput metabolic profiling and genome sequencing, a series of advances have been made in the structural identification, biochemical characterization, genetic basis of synthesis, localization, and health benefits of crop nutrient metabolites (Fang et al., 2019).

In the current article, we review the diversity of nutrients in crops, ranging from traditional basic types to the novel types of nutrients that are currently receiving attention. We will summarize the origins of nutrient metabolite diversity from the perspective of the genome, focusing on the essential factors that determine metabolite diversity, including gene duplication and divergence. In addition, we will systematically review the latest advances in studies on the genetic basis of nutrient diversity, especially the forward genetics approach of metabolite-based genome-wide association study (mGWAS), which greatly promotes analyses of the genetic basis of metabolic pathways and diversity. Finally, we will summarize the application of molecular markers related to nutrients and their metabolic biosynthesis in crop breeding and discuss the future impact of metabolome-assisted breeding techniques on crop nutrient quality improvement.

FIGURE 1 | Biofortification of nutrients in rice endosperm. **(A)** Grains of the wild-type (WT), the golden rice (GR), canthaxanthin rice (CR), and astaxanthin rice (AR). **(B)** Simplified carotenoid/canthaxanthin/astaxanthin biosynthesis pathways reconstructed in rice endosperm. The enzymes (in blue) expressed from different combinations of the transgenes, together with those (in black) from the endogenous MEP pathway genes and OsLCYB, catalyze the biosynthesis of β-carotene, canthaxanthin, and astaxanthin as the main products in the GR, CR, and AR lines, respectively. The rate-limiting enzymes of the endogenous genes with no or low levels of expression are shown in gray. The figure is modified with permission from Elsevier.

THE ORIGIN OF THE DIVERSITY OF NUTRITIONAL METABOLITES IN THE CONTEXT OF GENOME EVOLUTION

Primordial metabolism is generally regarded as chemical intermediates interconnected by a smaller number of ancestral enzymes with multifunctionality (Croteau et al., 2006; Fani and Fondi, 2009). Necessary metabolic processes became established since the appearance of plants in the land (Austin et al., 2004). Currently, plants produce a repository of structurally diverse compounds, including those vital for growth and development and for interactions of plants with environment (Weng et al., 2012). With the development of cross-species metabolic profiling strategies, convergent and divergent evolution of metabolites have been identified in several species (Chen et al., 2016; Tohge et al., 2016; Zhou et al., 2016).

Among the demonstrated mechanisms of the evolution of metabolism, gene duplication and divergence have been documented to be vital sources of the raw material for such evolution (Moghe et al., 2017). There are at least four mechanistic categories of gene duplication, including i) tandem duplication, ii) polyploidy, iii) chromosomal segment duplication, and iv) single-gene transposition–duplication (Freeling, 2009). The cytochrome P450 monooxygenase (P450) family is found to catalyze NADPH- and O_2-dependent hydroxylation reactions, generally located at the cytoplasmic surface of the endoplasmic reticulum (Werck-Reichhart and Feyereisen, 2000). P450 proteins are involved in the biosynthesis of various primary and secondary metabolites. In plants, P450s are generally categorized into the plant-specific A type and non-plant-specific non-A type, according to the evolutionary relationship (Durst and Nelson, 1995). There were three rounds (i.e., γ, β, and α) of polyploidization in *A. thaliana* and all other Brassicaceae taxa (Bowers et al., 2003). The level of the cytochrome P450 gene evolutionary group was highly enriched throughout evolutionary history. In addition, tandem duplication is also important for the evolution of the cytochrome P450 supergene family (Yu et al., 2017). Furthermore, the evolution of genes encoding homospermidine synthase is also identified to be important for the biosynthesis of pyrrolizidine alkaloid (Ober et al., 2003). Family of plant transcription factors, such as the MADS-box family, and the diterpene synthases, such as isopimaradiene synthase and levopimaradiene/abietadiene synthase, undergo multiple gene duplications that increase secondary metabolites diversity (Keeling et al., 2008; Flagel and Wendel, 2009).

Gene duplications immediately lead to the presence of two identical gene copies. Both copies may remain almost unchanged or diverge functionally. Alternatively, one of the duplicates serves as a pseudogene. To explain the development of new enzymatic functions, at least two hypotheses have been proposed: the neofunctionalization hypothesis and the subfunctionalization hypothesis. According to the first hypothesis, one duplicate's function resembles that of the ancestral gene, while mutations accumulate in the other gene during evolution, leading to a loss (nonfunctionalization) or a gain (neofunctionalization) of function (Ono, 1973; Rodin and Riggs, 2003). In the latter hypothesis, the functions of the ancestral gene are divided between the daughter genes (Hughes, 1994). Alternatively, duplicated genes may also undergo intraspecific partitioning of functions. Because the divergent evolution of genes involved in plant metabolite synthesis and regulation is too large to be fully reviewed, we will mainly focus on the variation in expression. The subdivision of functions across duplicate genes may manifest as differential expression patterns across multiple genotypes or as differential expression patterns within a single genotype.

Kliebenstein found that duplicated genes have more variable transcript accumulation than the average gene. In addition, the expression of tandem duplicated genes are significantly variable than that of segmental duplications (Kliebenstein, 2008). Moreover, epigenetic alleles are also critical for the determination of nutritional compounds accumulation.

Vitamin E consisted of tocopherol and its derivates. The first step of tocopherol synthesis is catalyzed by homogentisate phytyl transferase, producing 2-methyl-6-phytylquinol. This precursor is further catalyzed by dimethyl-phytylquinol methyl transferase to synthesize γ- and a-tocopherol (Almeida et al., 2011). Quadrana et al. identified an expression quantitative trait locus (QTL) for vitamin E content in tomato fruits. A retrotransposon was found to be located in the promoter of the methyltransferase-encoded VTE3(1), whose methylation affects the expression of VTE3(1) (Quadrana et al., 2014).

DISSECTION OF THE GENETIC BASES OF NUTRITIONAL QUALITY IN CROPS

Plants produce structurally diverse chemicals in order to maintain normal life activities and adapt to ecological environments. Plant metabolites generally consist of primary and secondary metabolites (Luo, 2015). Primary metabolites are thought to be essential for growth and development and play an important role in maintaining the normal life activities of plants, while secondary metabolites are regarded as more closely related to stress responses, helping plants cope with biotic and abiotic stresses in a constantly changing environment (Weng et al., 2012; Wurtzel and Kutchan, 2016). The considerable chemical diversity of plants is the source of nutrients required for human health, and the nutritional status of crops is ultimately dependent on their metabolic composition and content (Memelink, 2005). Metabolomic approaches enable parallel assessment of the levels of a broad range of plant metabolites, providing the possibility to study the diversity of crop nutrient metabolites (Fernie and Schauer, 2009; Samota et al., 2017). The recently developed widely targeted metabolomic approach based on liquid chromatography–mass spectrometry enables high-throughput detection of metabolite content (Chen et al., 2013) and has been used in several species, including rice (Chen et al., 2014; Chen et al., 2016), maize (Wen et al., 2014), citrus (Wang et al., 2016; Wang et al., 2017), and tomato (Zhu et al., 2018a). Comprehensive metabolic profiling and natural variation analysis of flavonoids were carried out in rice, and a total of 91 flavonoids were identified and quantified (Dong et al., 2014). Many advances have been made in the study of plant nutrient biosynthesis, such as vitamin A and oil in maize (Harjes et al., 2008; Li et al., 2013), carotenoids, sugars, and organic acids in tomato

(Lu and Li, 2008; Bursac et al., 2017; Tieman et al., 2017), and isoflavones in soybeans (Bursac et al., 2017).

In the process of studying the diversity of crop nutrients, it is critical to clarify how each metabolite is synthesized, transported, and degraded and how the metabolic pathway is regulated (Fang et al., 2019). Advances in different omic technologies, such as genomics, transcriptomics, and metabolomics, have facilitated the qualitative and quantitative analysis of plant metabolites, as well as the detection of candidate genes involved in metabolic synthesis and regulation, which contribute to the diversity of plant metabolite modifications (Oksman-Caldentey and Saito, 2005; Urano et al., 2010; Wang et al., 2019). This strategy to decipher the genetic basis of nutrients is further facilitated by recent advances in next-generation sequencing technology. For example, Sadre et al. recently identified two key genes for camptothecin biosynthesis, namely, TDC1 and TDC2, by analyzing transcriptome and metabolome data. They also found that CYCLASE1 (CYC1) is coexpressed with TDC1, suggesting that it may also be involved in camptothecin biosynthesis (Sadre et al., 2016). Polturak et al. performed multi-species transcriptomic and metabolomic analyses in *Mirabilis jalapa* and additional betalain-producing species to identify candidate genes possibly involved in betalain biosynthesis. Among the identified candidate genes, the betalain-related cytochrome P450 and glucosyltransferase-type genes that catalyze tyrosine hydroxylation and cinnamoyl-glucose formation were further functionally characterized (Polturak and Aharoni, 2018; Polturak et al., 2018). Integration analysis of transcriptome and metabolome data is a powerful tool for deciphering the genetic determinants of metabolic pathways, yet it lacks the ability to unravel the genetic basis of natural variation in the plant metabolome (Fang et al., 2019).

To explore the genetic basis of the crop metabolome, forward genetics based on genomics and metabolomics is being widely used, for example, using the biparental populations to determine QTL mapping and using natural populations for genome-wide association studies (GWASs) (Kliebenstein, 2009; Zhao et al., 2011; Riedelsheimer et al., 2012a; Routaboul et al., 2012; Chen et al., 2014; Chen et al., 2016; Wang et al., 2019). A number of advances have been made in the identification of metabolic quantitative trait locus (mQTL) using ultra-high density maps constructed using next-generation sequencing technologies, for example, integrating ultra-high-density maps of rice and metabolic profiles of seeds for mQTL mapping to analyze the genetic basis of the rice metabolomes and identifying hundreds of mQTLs in flag leaves or germinating seeds (Gong et al., 2013). QTL analysis in tomato seeds revealed colocalization of six amino acids on chromosomes 2, 4, and 10, of which 10 candidate genes related to amino acid metabolism were screened on chromosome 2 (Toubiana et al., 2015). To gain insight into the genetic factors controlling seed metabolism, QTL mapping was performed using the relative content of 311 primary metabolites. A total of 786 mQTLs were unevenly distributed in the genome, forming multiple hotspots. A series of candidate genes, including bZIP10, were identified to provide a basis for further study of the natural variation of Arabidopsis seed metabolism-related genes (Knoch et al., 2017).

Metabolic GWAS (mGWAS), which is used to decipher the genetic basis of plant metabolite biosynthesis and regulation, has made many advances in Arabidopsis (Chan et al., 2010; Wu et al., 2018), maize (Wen et al., 2014; Jin et al., 2017), rice (Luo, 2015; Chen et al., 2016), tomato (Bauchet et al., 2017; Ye et al., 2017), and wheat (Peng et al., 2018). Angelovici et al. performed a non-targeted liquid chromatography–mass spectrometry-based metabolome profiling of 309 Arabidopsis germplasms grown in two separate environments and performed mGWAS analysis to determine 70 significant associations between candidated genes and metabolites (Angelovici et al., 2013). Riedelsheimer et al. performed mGWAS analysis with 56,110 single nucleotide polymorphisms (SNPs) and 118 metabolites in maize inbred lines, identifying 26 different metabolites closely related to maize SNPs, of which p-coumaric acid and caffeic acid are closely related to the chromosome 9 region, which contains a gene encoding the key enzyme cinnamoyl-CoA reductase in the synthesis of lignin monomers (Riedelsheimer et al., 2012b). In rice, mGWAS analysis using 175 rice germplasms successfully identified 323 associations between 143 SNPs and 89 secondary metabolites, revealing the genetic mechanism of natural variation in rice secondary metabolite composition (Matsuda et al., 2015). Futhermore, Chen et al. performed quantitative analysis of 840 metabolites on 524 natural rice populations and used mGWAS to identify many important genetic loci associated with different metabolites (Chen et al., 2014). A systematic study of the genetic and biochemical bases of natural variation in flavonoids and polyamines in rice led to the identification of candidate genes related to their biosynthesis by mQTL and mGWAS methods (Peng et al., 2016; Peng et al., 2017). mGWAS can also identify genetic loci that affect most of the target flavor chemicals in tomato. Tieman et al. identified 2,014,488 common SNPs in 398 tomato germplasm genomes and identified 251 flavor-related signals using mGWAS (Tieman et al., 2017). Peng et al., based on six multi-locus GWAS models of 14,646 SNPs, found that 15 candidate genes are involved in free amino acid biosynthesis in wheat and functionally identified the candidate gene *TraesCS1D01G052500* encoding tryptophan decarboxylase, which provides new insights into understanding the biosynthesis of free amino acid in wheat (Peng et al., 2018).

Multi-omics integration analysis and multiple stages of development and different organizational analyses have been increasingly used to provide insight into biological mechanisms since combining multiple different types of datasets can compensate for missing or unreliable information in any single data type (Fang et al., 2019). Metabolic profiling combined with transcriptome analysis has been used to identify new gene clusters and GAME9 transcription factors involved in steroidal glycoalkaloid biosynthesis (Itkin et al., 2013; Cardenas et al., 2016). Joint metabolomic and genomic data subsequently allowed comprehensive refinement of steroidal glycoalkaloid biosynthesis (Schwahn et al., 2014). We recently performed multi-omics analysis of 610 tomato varieties, including genomes, transcriptomes, and metabolomes, to explore changes in fruit metabolomes during human-directed breeding (Zhu et al., 2018a). A total of 13,361 triple relationships (metabolite–SNP–gene), including 371 metabolites, 970 SNPs, and 535 genes, were constructed by mGWAS and eQTL analysis, which facilitated the identification of candidate genes and the clarification of metabolic pathways. For example, the SNP of *SlMYB12* (SNPy) discovered by excavating the abovementioned triple relationships was correlated with 69 metabolites and 69 genes in the mGWAS and eQTL analysis, and mutation of the

SNP resulted in a decrease in nutrient flavonoid content, resulting in the formation of pink tomato (Zhu et al., 2018a). Multi-omics integrative analysis has also made breakthroughs in the study of tomato flavor and cucumber bitterness. Thirty-seven metabolites, including total soluble solids, glucose, fructose, citric acid, and malic acids, were found to affect tomato flavor, and a total of 251 association signals were detected for 20 traits, including four nonvolatile and 15 volatile flavor chemicals (Tieman et al., 2017). Shang et al. discovered that two TFs regulate nine genes in the cucurbitacin C biosynthetic pathway and proposed a model for how extremely bitter wild cucumber was domesticated into nonbitter cultivars (Shang et al., 2014). These examples show that exploring the biochemical and genetic bases of nutrient diversity can provide new opportunities to increase the level of nutrient biofortification or to change the flavor characteristics that are beneficial to human health (Dixon, 1999).
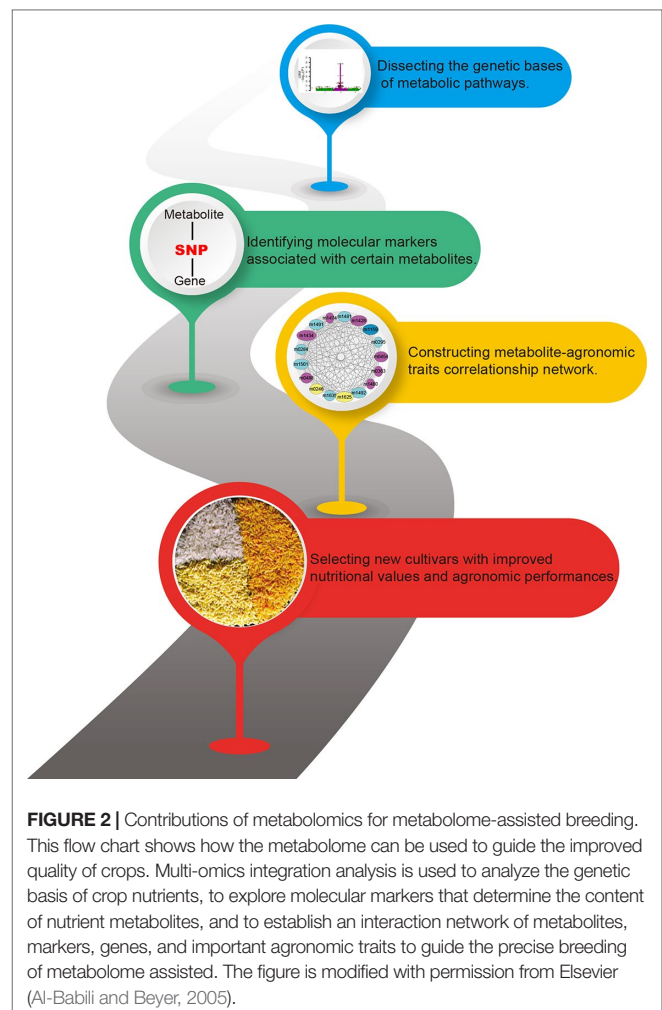
THE APPLICATION OF METABOLIC DIVERSITY IN CROP BREEDING

As described by the adage "health comes from the farm, not the pharmacy," crops serve as sources of metabolites essential for the nutrition and health of humans. Ongoing international biofortification research and breeding programs strive to improve life and well-being (Riezzo et al., 2005). Vitamins and *anthocyanins* are important targets for biofortification because they are sourced primarily from food.

Vitamins

Structural genes and their origination are essential for biofortification. A group of fat-soluble C20 carotenoid derivatives are denoted as vitamin A, including retinal, retinol and its esters, and retinoic acid. Certain carotenoids, referred to as provitamin A, are cleaved to form vitamin A within the body (Yeum and Russell, 2002). Vitamin A is essential for human health and development (West et al., 2002). Golden rice was developed to deliver provitamin A to ease the global deficiency of vitamin A; in this rice, the carotenoid biosynthetic pathway is reconstituted in the endosperm. Although the endosperm of rice cultivars does not accumulate provitamin A, the earlier intermediate geranylgeranyl diphosphate is present in rice endosperm, which can produce the phytoene under the catalization of plant phytoene synthase (PSY). A transgenic approach was adopted to accumulate provitamin A in the endosperm of rice by expressing a *PSY* and a bacterial phytoene desaturase (*CrtI*) (Ye et al., 2000). Genetically modified golden rice produces as much as 1.6 mg/g total carotenoids in the endosperm, leading to its characteristic yellow color. The limiting and major regulatory step for carotenoid biosynthesis is thought to be phytoene synthase (Fraser et al., 1994; Ronen et al., 1999; Fraser et al., 2002). To increase the carotenoid content of golden rice, systematic tests of psy genes from different plant species were carried out. Hence, "Golden Rice 2" was created by expressing the maize-originated psy gene and the *CrtI* gene from *Erwinia uredovora*, leading to the accumulation of up to 37 ug/g total carotenoids in the endosperm and preferential production of β-carotene (Paine et al., 2005).

Biofortification can be carried out by ectopic expression of metabolic pathways in crops. Astaxanthin, a red ketocarotenoid synthesized from β-carotene, is used in feedstuffs as a supplement. BKT and β-carotene hydroxylase are essential for the producing of astaxanthin (Higuera-Ciapara et al., 2006). Although different hydroxylated carotenoids pile up in the majority of higher plants, the biosynthesis of ketocarotenoids is impaired due to the absence of BKT genes (Cunningham and Gantt, 2005; Zhu et al., 2009). Astaxanthin has been successfully ectopically expressed in several species with the presence of native β-carotene by introducing two (β-carotene hydroxylase and BKT) transgenes or a single (BKT) transgene (Hasunuma et al., 2008; Jayaraj et al., 2008; Huang et al., 2013; Harada et al., 2014; Campbell et al., 2015; Farre et al., 2016). However, rice endosperm does not accumulate β-carotene, which can be preferentially produced by overexpressing *ZmPSY1* and *PaCrtl* (Paine et al., 2005). Zhu et al. developed canthaxanthin rice, which has a high ketocarotenoid content, by expressing *ZmPSY1*, *PaCrtl*, and *CrBKT* in the rice endosperm (**Figure 1**) (Zhu et al., 2018b).

Folate, only synthesized *de novo* in plants and microorganisms, decreases the risk of several diseases (Iyer and Tomar, 2009). Folate biofortification was performed in both tomato fruits and rice seeds (Diaz De La Garza et al., 2007; Storozhenko et al., 2007). To increase the folate content in tomato and rice, two genes encoding

FIGURE 2 | Contributions of metabolomics for metabolome-assisted breeding. This flow chart shows how the metabolome can be used to guide the improved quality of crops. Multi-omics integration analysis is used to analyze the genetic basis of crop nutrients, to explore molecular markers that determine the content of nutrient metabolites, and to establish an interaction network of metabolites, markers, genes, and important agronomic traits to guide the precise breeding of metabolome assisted. The figure is modified with permission from Elsevier (Al-Babili and Beyer, 2005).

enzymes in folate biosynthesis were overexpressed, including an aminodeoxychorismate synthase and a GTP cyclohydrolase I. However, folate is prone to degradation upon storage (Blancquaert et al., 2015) since it is sensitive to oxidation, pondus hydrogenii, and temperature (Scott et al., 2000; De Brouwer et al., 2007). Approximately 40% of the folate remained in the rice grain after storage for 8 months, indicating that the folate may be protected from degradation since folate polyglutamylation by folylpolyglutamate synthetase (FPGS) or complexation with folate-binding proteins (FBP) can improve folate stability (Blancquaert et al., 2010; Blancquaert et al., 2014). Blancquaert et al. aimed to enhance folate concentration and stability by overexpressing the *A. thaliana* cDNA encoding FPGS (ctF) or a synthetic soluble FBP. The transgenic rice supplies 150 times more folate than the wild type (Blancquaert et al., 2015).

Anthocyanins

Transcriptional regulation of the genes in entire metabolic pathways provide effective tools for metabolic engineering. Anthocyanins are viewed as compounds beneficial for human health, which decrease the risk of certain cancers and other diseases (Wang and Stoner, 2008; Deng et al., 2013; Zhang et al., 2014). Butelli et al. overexpressed the Delila (*Del*) and Rosea1 (*Ros1*) genes from the snapdragon *Antirrhinum majus* in tomato, which encode a basic helix-loop-helix transcription factor and a MYB-related transcription factor, respectively. The transgenic tomatoes exhibited significantly activated transcription levels of key genes, including almost all of the genes required for anthocyanin biosynthesis and genes essential for side-chain modification. Consequently, overexpression of Del/Ros1 activated the production of anthocyanins in tomatoes, resulting in a purple color (Butelli et al., 2008). AtMYB12 driven by the fruit-specific E8 promoter increases the expression levels of genes in primary metabolism and flavonol and hydroxycinnamic ester biosynthesis in tomato and activates the accumulation of both flavonols and caffeoyl quinic acids. Indigo tomato was developed by crossing AtMYB12 tomato with the purple Del/Ros1 tomato line, which accumulates even greater amounts of chlorogenic acid, flavonols, and anthocyanins (Zhang et al., 2015). Although anthocyanins accumulate in several tissues of plants, the endosperm of cereals lacks anthocyanins. The pericarp of some special varieties of rice accumulates anthocyanins and proanthocyanidins. Many efforts have been made to decode

the sophisticated anthocyanin biosynthesis pathway in plants, leading to the identification of conserved enzymes, as well as several regulatory proteins (Hichri et al., 2011; Dixon et al., 2013; Zhang et al., 2014; Yuan and Grotewold, 2015). To develop rice with a high anthocyanin content in the endosperm, eight anthocyanin pathway genes were transferred into rice calli, including six structural genes for anthocyanin biosynthesis from Coleus and two regulatory genes. The transgenic plants displayed purple endosperm due to the activated accumulation of anthocyanins and were renamed Zijingmi in Chinese (Zhu et al., 2017).

FUTURE PERSPECTIVES

Past research has focused largely on annotating more metabolites and decoding metabolic pathways. However, a far more exciting research front in crop breeding has been produced by the multi-omics studies. Here, we have reviewed recent advances, focusing on the diversity of phytonutrients and its genetic bases. Our knowledge on the biosynthesis and the diversity of plant metabolites will be enhanced by studies with multi-omics data. The metabolome-assistant breeding will contribute greatly to the improvement of crops with additional nutritional value (**Figure 2**) and that natural and artificial populations of crops will provide vast gene resources and parental materials.

AUTHOR CONTRIBUTIONS

SW and CF wrote the manuscript. SW, CF and JL revised the manuscript.

FUNDING

REFERENCES

Ahuja, I., De Vos, R. C. H., Bones, A. M., and Hall, R. D. (2010). Plant molecular stress responses face climate change. *Trends Plant Sci.* 15, 664–674. doi: 10.1016/j.tplants.2010.08.002

Al-Babili, S., and Beyer, P. (2005). Golden Rice – five years on the road – five years to go? *Trends Plant Sci.* 10, 565–573. doi: 10.1016/j.tplants.2005.10.006

Almeida, J., Quadrana, L., Asis, R., Setta, N., De Godoy, F., Bermudez, L., et al. (2011). Genetic dissection of vitamin E biosynthesis in tomato. *J. Exp. Bot.* 62, 3781–3798. doi: 10.1093/jxb/err055

Angelovici, R., Lipka, A. E., Deason, N., Gonzalez-Jorge, S., Lin, H., Cepela, J., et al. (2013). Genome-wide analysis of branched-chain amino acid levels in Arabidopsis seeds. *Plant Cell* 25, 4827–4843. doi: 10.1105/tpc.113.119370

Appleby, P. N., Key, T. J., and Bradbury, K. E. (2014). Fruit, vegetable, and fiber intake in relation to cancer risk: findings from the European prospective investigation into cancer and nutrition (EPIC). *Am. J. Clin. Nutr.* 100, 394S–398S. doi: 10.3945/ajcn.113.071357

Austin, M. B., Bowman, M. E., Ferrer, J. L., Schroder, J., and Noel, J. P. (2004). An aldol switch discovered in stilbene synthases mediates cyclization specificity of type III polyketide synthases. *Chem. Biol.* 11, 1179–1194. doi: 10.1016/j.chembiol.2004.05.024

Bauchet, G., Grenier, S., Samson, N., Segura, V., Kende, A., Beekwilder, J., et al. (2017). Identification of major loci and genomic regions controlling acid and volatile content in tomato fruit: implications for flavor improvement. *New Phytol.* 215, 624–641. doi: 10.1111/nph.14615

Blancquaert, D., De Steur, H., Gellynck, X., and Van Der Straeten, D. (2014). Present and future of folate biofortification of crop plants. *J. Exp. Bot.* 65, 895–906. doi: 10.1093/jxb/ert483

Blancquaert, D., Storozhenko, S., Loizeau, K., De Steur, H., De Brouwer, V., Viaene, J., et al. (2010). Folates and folic acid: from fundamental research

toward sustainable health. *Crit. Rev. Plant Sci.* 29, 14–35. doi: 10.1080/07352680903436283

Blancquaert, D., Van Daele, J., Strobbe, S., Kiekens, F., Storozhenko, S., De Steur, H., et al. (2015). Improving folate (vitamin B9) stability in biofortified rice through metabolic engineering. *Nat. Biotechnol.* 33, 1076–1078. doi: 10.1038/nbt.3358

Bowers, J. E., Chapman, B. A., Rong, J., and Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422, 433–438. doi: 10.1038/nature01521

Bursac, M., Krstonosic, M. A., Miladinovic, J., Malencic, D., Gvozdenovic, L., and Cvejic, J. H. (2017). Isoflavone composition, total phenolic content and antioxidant capacity of soybeans with colored seed coat. *Nat. Prod. Commun.* 12, 527–532. doi: 10.1177/1934578X1701200417

Butelli, E., Titta, L., Giorgio, M., Mock, H.-P., Matros, A., Peterek, S., et al. (2008). Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat. Biotechnol.* 26, 1301–1308. doi: 10.1038/nbt.1506

Campbell, R., Morris, W. L., Mortimer, C. L., Misawa, N., Ducreux, L. J., Morris, J. A., et al. (2015). Optimising ketocarotenoid production in potato tubers: effect of genetic background, transgene combinations and environment. *Plant Sci.* 234, 27–37. doi: 10.1016/j.plantsci.2015.01.014

Cardenas, P. D., Sonawane, P. D., Pollier, J., Vanden Bossche, R., Dewangan, V., Weithorn, E., et al. (2016). GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. *Nat. Commun.* 7, 10654. doi: 10.1038/ncomms10654

Chan, E. K., Rowe, H. C., Hansen, B. G., and Kliebenstein, D. J. (2010). The complex genetic architecture of the metabolome. *PLoS Genet.* 6, e1001198. doi: 10.1371/journal.pgen.1001198

Chen, W., Gao, Y., Xie, W., Gong, L., Lu, K., Wang, W., et al. (2014). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* 46, 714–721. doi: 10.1038/ng.3007

Chen, W., Gong, L., Guo, Z., Wang, W., Zhang, H., Liu, X., et al. (2013). A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol. Plant* 6, 1769–1780. doi: 10.1093/mp/sst080

Chen, W., Wang, W., Peng, M., Gong, L., Gao, Y., Wan, J., et al. (2016). Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nat. Commun.* 7, 12767. doi: 10.1038/ncomms12767

Croteau, R., Ketchum, R. E., Long, R. M., Kaspera, R., and Wildung, M. R. (2006). Taxol biosynthesis and molecular genetics. *Phytochem. Rev.* 5, 75–97. doi: 10.1007/s11101-005-3748-2

Cunningham, F. X., Jr., and Gantt, E. (2005). A study in scarlet: enzymes of ketocarotenoid biosynthesis in the flowers of Adonis aestivalis. *Plant J.* 41, 478–492. doi: 10.1111/j.1365-313X.2004.02309.x

De Brouwer, V., Zhang, G. F., Storozhenko, S., Straeten, D. V., and Lambert, W. E. (2007). pH stability of individual folates during critical sample preparation steps in prevision of the analysis of plant folates. *Phytochem Anal.* 18, 496–508. doi: 10.1002/pca.1006

Dellapenna, D. (1999). Nutritional genomics: manipulating plant micronutrients to improve human health. *Science* 285, 375–379. doi: 10.1126/science.285.5426.375

Deng, G. F., Xu, X. R., Zhang, Y., Li, D., Gan, R. Y., and Li, H. B. (2013). Phenolic compounds and bioactivities of pigmented rice. *Crit. Rev. Food Sci. Nutr.* 53, 296–306. doi: 10.1080/10408398.2010.529624

Diaz De La Garza, R. I., Gregory, J. F., 3rd, and Hanson, A. D. (2007). Folate biofortification of tomato fruit. *Proc. Natl. Acad. Sci. U. S. A.* 104, 4218–4222. doi: 10.1073/pnas.0700409104

Dixon, R. A. (1999). Plant natural products: the molecular genetic basis of biosynthetic diversity. *Curr. Opin. Biotechnol.* 10, 192–197. doi: 10.1016/S0958-1669(99)80034-2

Dixon, R. A., Liu, C., and Jun, J. H. (2013). Metabolic engineering of anthocyanins and condensed tannins in plants. *Curr. Opin. Biotechnol.* 24, 329–335. doi: 10.1016/j.copbio.2012.07.004

Dong, X., Chen, W., Wang, W., Zhang, H., Liu, X., and Luo, J. (2014). Comprehensive profiling and natural variation of flavonoids in rice. *J. Integr. Plant Biol.* 56, 876–886. doi: 10.1111/jipb.12204

Durst, F., and Nelson, D. R. (1995). Diversity and evolution of plant P450 and P450-reductases. *Drug Metabol. Drug Interact.* 12, 189–206. doi: 10.1515/DMDI.1995.12.3-4.189

Fang, C. Y., Fernie, A. R., and Luo, J. (2019). Exploring the diversity of plant metabolism. *Trends Plant Sci.* 24, 83–98. doi: 10.1016/j.tplants.2018.09.006

Fani, R., and Fondi, M. (2009). Origin and evolution of metabolic pathways. *Phys. Life Rev.* 6, 23–52. doi: 10.1016/j.plrev.2008.12.003

Farre, G., Perez-Fons, L., Decourcelle, M., Breitenbach, J., Hem, S., Zhu, C., et al. (2016). Metabolic engineering of astaxanthin biosynthesis in maize endosperm and characterization of a prototype high oil hybrid. *Transgenic Res.* 25, 477–489. doi: 10.1007/s11248-016-9943-7

Fernie, A. R., and Schauer, N. (2009). Metabolomics-assisted breeding: a viable option for crop improvement? *Trends Genet.* 25, 39–48. doi: 10.1016/j.tig.2008.10.010

Fernie, A. R., Tadmor, Y., and Zamir, D. (2006). Natural genetic variation for improving crop quality. *Curr. Opin. Plant Biol.* 9, 196–202. doi: 10.1016/j.pbi.2006.01.010

Flagel, L. E., and Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytol.* 183, 557–564. doi: 10.1111/j.1469-8137.2009.02923.x

Fraser, P. D., Romer, S., Shipton, C. A., Mills, P. B., Kiano, J. W., Misawa, N., et al. (2002). Evaluation of transgenic tomato plants expressing an additional phytoene synthase in a fruit-specific manner. *Proc. Natl. Acad. Sci. USA* 99, 1092–1097. doi: 10.1073/pnas.241374598

Fraser, P. D., Truesdale, M. R., Bird, C. R., Schuch, W., and Bramley, P. M. (1994). Carotenoid biosynthesis during tomato fruit development (evidence for tissue-specific gene expression). *Plant Physiol.* 105, 405–413. doi: 10.1104/pp.105.1.405

Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* 60, 433–453. doi: 10.1146/annurev.arplant.043008.092122

Gong, L., Chen, W., Gao, Y., Liu, X., Zhang, H., Xu, C., et al. (2013). Genetic analysis of the metabolome exemplified using a rice population. *Proc. Natl. Acad. Sci. U. S. A.* 110, 20320–20325. doi: 10.1073/pnas.1319681110

Hall, R. D., Brouwer, I. D., and Fitzgerald, M. A. (2008). Plant metabolomics and its potential application for human nutrition. *Physiol. Plantarum* 132, 162–175. doi: 10.1111/j.1399-3054.2007.00989.x

Harada, H., Maoka, T., Osawa, A., Hattan, J., Kanamoto, H., Shindo, K., et al. (2014). Construction of transplastomic lettuce (*Lactuca sativa*) dominantly producing astaxanthin fatty acid esters and detailed chemical analysis of generated carotenoids. *Transgenic Res.* 23, 303–315. doi: 10.1007/s11248-013-9750-3

Harjes, C. E., Rocheford, T. R., Bai, L., Brutnell, T. P., Kandianis, C. B., Sowinski, S. G., et al. (2008). Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science* 319, 330–333. doi: 10.1126/science.1150255

Hasunuma, T., Miyazawa, S., Yoshimura, S., Shinzaki, Y., Tomizawa, K., Shindo, K., et al. (2008). Biosynthesis of astaxanthin in tobacco leaves by transplastomic engineering. *Plant J.* 55, 857–868. doi: 10.1111/j.1365-313X.2008.03559.x

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., and Lauvergeat, V. (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *J. Exp. Bot.* 62, 2465–2483. doi: 10.1093/jxb/erq442

Higuera-Ciapara, I., Felix-Valenzuela, L., and Goycoolea, F. M. (2006). Astaxanthin: a review of its chemistry and applications. *Crit. Rev. Food Sci. Nutr.* 46, 185–196. doi: 10.1080/10408690590957188

Hounsome, N., Hounsome, B., Tomos, D., and Edwards-Jones, G. (2008). Plant metabolites and nutritional quality of vegetables. *J. Food Sci.* 73, R48–R65. doi: 10.1111/j.1750-3841.2008.00716.x

Huang, J. C., Zhong, Y. J., Liu, J., Sandmann, G., and Chen, F. (2013). Metabolic engineering of tomato for high-yield production of astaxanthin. *Metab. Eng.* 17, 59–67. doi: 10.1016/j.ymben.2013.02.005

Hughes, A. L. (1994). The evolution of functionally novel proteins after gene duplication. *Proc. Biol. Sci.* 256, 119–124. doi: 10.1098/rspb.1994.0058

Itkin, M., Heinig, U., Tzfadia, O., Bhide, A. J., Shinde, B., Cardenas, P. D., et al. (2013). Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science* 341, 175–179. doi: 10.1126/science.1240230

Iyer, R., and Tomar, S. K. (2009). Folate: a functional food constituent. *J. Food Sci.* 74, R114–R122. doi: 10.1111/j.1750-3841.2009.01359.x

Jayaraj, J., Devlin, R., and Punja, Z. (2008). Metabolic engineering of novel ketocarotenoid production in carrot plants. *Transgenic Res.* 17, 489–501. doi: 10.1007/s11248-007-9120-0

Jin, M., Zhang, X., Zhao, M., Deng, M., Du, Y., Zhou, Y., et al. (2017). Integrated genomics-based mapping reveals the genetics underlying maize flavonoid biosynthesis. *BMC Plant Biol.* 17, 17. doi: 10.1186/s12870-017-0972-z

Keeling, C. I., Weisshaar, S., Lin, R. P., and Bohlmann, J. (2008). Functional plasticity of paralogous diterpene synthases involved in conifer defense. *Proc. Natl. Acad. Sci. U. S. A.* 105, 1085–1090. doi: 10.1073/pnas.0709466105

Kliebenstein, D. (2009). Advancing genetic theory and application by metabolic quantitative trait loci analysis. *Plant Cell* 21, 1637–1646. doi: 10.1105/tpc.109.067611

Kliebenstein, D. J. (2008). A role for gene duplication and natural variation of gene expression in the evolution of metabolism. *PLoS One* 3, e1838. doi: 10.1371/journal.pone.0001838

Knoch, D., Riewe, D., Meyer, R. C., Boudichevskaia, A., Schmidt, R., and Altmann, T. (2017). Genetic dissection of metabolite variation in Arabidopsis seeds: evidence for mQTL hotspots and a master regulatory locus of seed metabolism. *J. Exp. Bot.* 68, 1655–1667. doi: 10.1093/jxb/erx049

Li, H., Peng, Z. Y., Yang, X. H., Wang, W. D., Fu, J. J., Wang, J. H., et al. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* 45, 43–U72. doi: 10.1038/ng.2484

Lu, S., and Li, L. (2008). Carotenoid metabolism: biosynthesis, regulation, and beyond. *J. Integr. Plant Biol.* 50, 778–785. doi: 10.1111/j.1744-7909.2008.00708.x

Luo, J. (2015). Metabolite-based genome-wide association studies in plants. *Curr. Opin. Plant Biol.* 24, 31–38. doi: 10.1016/j.pbi.2015.01.006

Mahajan, S., and Tuteja, N. (2005). Cold, salinity and drought stresses: an overview. *Arch Biochem Bioph.* 444, 139–158. doi: 10.1016/j.abb.2005.10.018

Martin, C., and Li, J. (2017). Medicine is not health care, food is health care: plant metabolic engineering, diet and human health. *New Phytol.* 216, 699–719. doi: 10.1111/nph.14730

Matsuda, F., Nakabayashi, R., Yang, Z., Okazaki, Y., Yonemaru, J.-I., Ebana, K., et al. (2015). Metabolome-genome-wide association study dissects genetic architecture for generating natural variation in rice secondary metabolism. *Plant J.* 81, 13–23. doi: 10.1111/tpj.12681

Memelink, J. (2005). Tailoring the plant metabolome without a loose stitch. *Trends Plant Sci.* 10, 305–307. doi: 10.1016/j.tplants.2005.05.006

Moghe, G. D., Leong, B. J., Hurney, S. M., Daniel Jones, A., and Last, R. L. (2017). Evolutionary routes to biochemical innovation revealed by integrative analysis of a plant-defense related specialized metabolic pathway. *Elife* 6, e28468. doi: 10.7554/eLife.28468

Ober, D., Harms, R., Witte, L., and Hartmann, T. (2003). Molecular evolution by change of function: Alkaloid-specific homospermidine synthase retained all properties of deoxyhypusine synthase except binding the eIF5A precursor protein. *J. Biol. Chem.* 278, 2805–12812. doi: 10.1074/jbc.M207112200

Oksman-Caldentey, K. M., and Saito, K. (2005). Integrating genomics and metabolomics for engineering plant metabolic pathways. *Curr. Opin. Biotech.* 16, 174–179. doi: 10.1016/j.copbio.2005.02.007

Ono, S. (1973). Ancient linkage groups and frozen accidents. *Nature* 244, 259–262. doi: 10.1038/244259a0

Orcutt, D. M., and Nilsen, E. T. (2000). *The physiology of plants under stress soil and biotic factors*. New York: John Wiley and Sons Inc. p.680.

Paine, J. A., Shipton, C. A., Chaggar, S., Howells, R. M., Kennedy, M. J., Vernon, G., et al. (2005). Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat. Biotechnol.* 23, 482–487. doi: 10.1038/nbt1082

Peng, M., Gao, Y., Chen, W., Wang, W., Shen, S., Shi, J., et al. (2016). Evolutionarily distinct BAHD *N*-acyltransferases are responsible for natural variation of aromatic amine conjugates in rice. *Plant Cell* 28, 1533–1550. doi: 10.1105/tpc.16.00265

Peng, M., Shahzad, R., Gul, A., Subthain, H., Shen, S., Lei, L., et al. (2017). Differentially evolved glucosyltransferases determine natural variation of rice flavone accumulation and UV-tolerance. *Nat. Commun.* 8, 1975. doi: 10.1038/s41467-017-02168-x

Peng, Y., Liu, H., Chen, J., Shi, T., Zhang, C., Sun, D., et al. (2018). Genome-wide association studies of free amino acid levels by six multi-locus models in bread wheat. *Front. Plant Sci.* 9, 1196. doi: 10.3389/fpls.2018.01196

Polturak, G., and Aharoni, A. (2018). "La Vie en Rose": biosynthesis, sources, and applications of betalain pigments. *Mol. Plant* 11, 7–22. doi: 10.1016/j.molp.2017.10.008

Polturak, G., Heinig, U., Grossman, N., Battat, M., Leshkowitz, D., Malitsky, S., et al. (2018). Transcriptome and metabolic profiling provides insights into betalain biosynthesis and evolution in *Mirabilis jalapa*. *Mol. Plant* 11, 189–204. doi: 10.1016/j.molp.2017.12.002

Quadrana, L., Almeida, J., Asís, R., Duffy, T., Dominguez, P. G., Bermudez, L., et al. (2014). Natural occurring epialleles determine vitamin E accumulation in tomato fruits. *Nat. Commun.* 5, 4027. doi: 10.1038/ncomms5027

Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., Lisec, J., Technow, F., Sulpice, R., et al. (2012a). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat. Genet.* 44, 217–220. doi: 10.1038/ng.1033

Riedelsheimer, C., Lisec, J., Czedik-Eysenberg, A., Sulpice, R., Flis, A., Grieder, C., et al. (2012). Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proc. Natl. Acad. Sci. USA* 109, 8872–8877. doi: 10.1073/pnas.1120813109

Riezzo, G., Chiloiro, M., and Russo, F. (2005). Functional foods: salient features and clinical applications. *Curr. Drug Targets Immune Endocr. Metabol. Disord.* 5, 331–337. doi: 10.2174/1568008054863790

Rodin, S., and Riggs, A. (2003). Epigenetic silencing may aid evolution by gene duplication. *J. Mol. Evol.* 56, 718–729. doi: 10.1007/s00239-002-2446-6

Ronen, G., Cohen, M., Zamir, D., and Hirshberg, J. (1999). Regulation of carotenoid biosynthesis during tomato fruit development: expression of the gene for lycopene epsilon-cyclase is down-regulated during ripening and is elevated in the mutant Delta. *Plant J.* 17, 341–351. doi: 10.1046/j.1365-313X.1999.00381.x

Routaboul, J. M., Dubos, C., Beck, G., Marquis, C., Bidzinski, P., Loudet, O., et al. (2012). Metabolite profiling and quantitative genetics of natural variation for flavonoids in Arabidopsis. *J. Exp. Bot.* 63, 3749–3764. doi: 10.1093/jxb/ers067

Sadre, R., Magallanes-Lundback, M., Pradhan, S., Salim, V., Mesberg, A., Jones, A. D., et al. (2016). Metabolite diversity in alkaloid biosynthesis: a multilane (diastereomer) highway for camptothecin synthesis in *Camptotheca acuminata*. *Plant Cell* 28, 1926–1944. doi: 10.1105/tpc.16.00193

Samota, M. K., Bhatt, L., Yadav, D. K., Garg, N., and Bajiya, R. (2017). Metabolomics for functional genomics. *Int. J. Curr. Microbiol. Appl. Sci.* 6, 2531–2537. doi: 10.20546/ijcmas.2017.605.284

Schwahn, K., De Souza, L. P., Fernie, A. R., and Tohge, T. (2014). Metabolomics-assisted refinement of the pathways of steroidal glycoalkaloid biosynthesis in the tomato clade. *J. Integr. Plant Biol.* 56, 864–875. doi: 10.1111/jipb.12274

Scott, J., Rébeillé, F., and Fletcher, J. (2000). Folic acid and folates: the feasibility for nutritional enhancement in plant foods. *J. Sci. Food Agr.* 80, 795–824. doi: 10.1002/(SICI)1097-0010(20000515)80:7<795::AID-JSFA599>3.0.CO;2-K

Shang, Y., Ma, Y., Zhou, Y., Zhang, H., Duan, L., Chen, H., et al. (2014). Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* 346, 1084–1088. doi: 10.1126/science.1259215

Storozhenko, S., De Brouwer, V., Volckaert, M., Navarrete, O., Blancquaert, D., Zhang, G. F., et al. (2007). Folate fortification of rice by metabolic engineering. *Nat. Biotechnol.* 25, 1277–1279. doi: 10.1038/nbt1351

Tieman, D., Zhu, G., Resende, M. F., Jr., Lin, T., Nguyen, C., Bies, D., et al. (2017). A chemical genetic roadmap to improved tomato flavor. *Science* 355, 391–394. doi: 10.1126/science.aal1556

Tohge, T., Wendenburg, R., Ishihara, H., Nakabayashi, R., Watanabe, M., Sulpice, R., et al. (2016). Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in Brassicaceae. *Nat. Commun.* 7, 12399. doi: 10.1038/ncomms12399

Toubiana, D., Batushansky, A., Tzfadia, O., Scossa, F., Khan, A., Barak, S., et al. (2015). Combined correlation-based network and mQTL analyses efficiently identified loci for branched-chain amino acid, serine to threonine, and proline metabolism in tomato seeds. *Plant J.* 81, 121–133. doi: 10.1111/tpj.12717

Treutter, D. (2006). Significance of flavonoids in plant resistance: a review. *Environ. Chem. Lett.* 4, 147. doi: 10.1007/s10311-006-0068-8

Urano, K., Kurihara, Y., Seki, M., and Shinozaki, K. (2010). 'Omics' analyses of regulatory networks in plant abiotic stress responses. *Curr. Opin. Plant Biol.* 13, 132–138. doi: 10.1016/j.pbi.2009.12.006

Wang, L. S., and Stoner, G. D. (2008). Anthocyanins and their role in cancer prevention. *Cancer Lett.* 269, 281–290. doi: 10.1016/j.canlet.2008.05.020

Wang, S., Alseekh, S., Fernie, A. R., and Luo, J. (2019). The structure and function of major plant metabolite modifications. *Mol. Plant* 12, 899–919. doi: 10.1016/j.molp.2019.06.001

Wang, S., Moustaid-Moussa, N., Chen, L., Mo, H., Shastri, A., Su, R., et al. (2014). Novel insights of dietary polyphenols and obesity. *J. Nutr. Biochem.* 25, 1–18. doi: 10.1016/j.jnutbio.2013.09.001

Wang, S., Tu, H., Wan, J., Chen, W., Liu, X., Luo, J., et al. (2016). Spatio-temporal distribution and natural variation of metabolites in citrus fruits. *Food Chem.* 199, 8–17. doi: 10.1016/j.foodchem.2015.11.113

Wang, S., Yang, C., Tu, H., Zhou, J., Liu, X., Cheng, Y., et al. (2017). Characterization and metabolic diversity of flavonoids in citrus species. *Sci. Rep.* 7, 10549. doi: 10.1038/s41598-017-10970-2

Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., et al. (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat. Commun.* 5, 3438. doi: 10.1038/ncomms4438

Weng, J. K., Philippe, R. N., and Noel, J. P. (2012). The rise of chemodiversity in plants. *Science* 336, 1667–1670. doi: 10.1126/science.1217411

Werck-Reichhart, D., and Feyereisen, R. (2000). Cytochromes P450: a success story. *Genome Biol.* 1, 3003. doi: 10.1186/gb-2000-1-6-reviews3003

West, C. E., Eilander, A., and van Lieshout, M. (2002). Consequences of revised estimates of carotenoid bioefficacy for dietary control of vitamin A deficiency in developing countries. *J. Nutr.* 132, 2920S. doi: 10.1093/jn/132.9.2920S

Wu, S., Tohge, T., Cuadros-Inostroza, A., Tong, H., Tenenboim, H., Kooke, R., et al. (2018). Mapping the Arabidopsis metabolic landscape by untargeted metabolomics at different environmental conditions. *Mol. Plant* 11, 118–134. doi: 10.1016/j.molp.2017.08.012

Wurtzel, E. T., and Kutchan, T. M. (2016). Plant metabolism, the diverse chemistry set of the future. *Science* 353, 1232–1236. doi: 10.1126/science.aad2062

Ye, X., Al-Babili, S., Klöti, A., Zhang, J., Lucca, P., Beyer, P., et al. (2000). Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* 287, 303–305. doi: 10.1126/science.287.5451.303

Ye, J., Wang, X., Hu, T., Zhang, F., Wang, B., Li, C., et al. (2017). An InDel in the promoter of Al-ACTIVATED MALATE TRANSPORTER9 selected during tomato domestication determines fruit malate contents and aluminum tolerance. *Plant Cell* 29, 2249–2268. doi: 10.1105/tpc.17.00211

Yeum, K. J., and Russell, R. M. (2002). Carotenoid bioavailability and bioconversion. *Annu. Rev. Nutr.* 22, 483–504. doi: 10.1146/annurev.nutr.22.010402.102834

Yu, J., Tehrim, S., Wang, L., Dossa, K., Zhang, X., Ke, T., et al. (2017). Evolutionary history and functional divergence of the cytochrome P450 gene superfamily between Arabidopsis thaliana and Brassica species uncover effects of whole genome and tandem duplications. *BMC Genomics* 18, 733. doi: 10.1186/s12864-017-4094-7

Yuan, L., and Grotewold, E. (2015). Metabolic engineering to enhance the value of plants as green factories. *Metab. Eng.* 27, 83–91. doi: 10.1016/j.ymben.2014.11.005

Zhang, Y., Butelli, E., Alseekh, S., Tohge, T., Rallapalli, G., Luo, J., et al. (2015). Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nat. Commun.* 6, 8635. doi: 10.1038/ncomms9635

Zhang, Y., Butelli, E., and Martin, C. (2014). Engineering anthocyanin biosynthesis in plants. *Curr. Opin. Plant Biol.* 19, 81–90. doi: 10.1016/j.pbi.2014.05.011

Zhao, K., Tung, C. W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. *Nat Commun* 2, 467. doi: 10.1038/ncomms1467

Zhou, Y., Ma, Y., Zeng, J., Duan, L., Xue, X., Wang, H., et al. (2016). Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. *Nat. Plants* 2, 16183. doi: 10.1038/nplants.2016.183

Zhu, C., Naqvi, S., Capell, T., and Christou, P. (2009). Metabolic engineering of ketocarotenoid biosynthesis in higher plants. *Arch. Biochem. Biophys.* 483, 182–190. doi: 10.1016/j.abb.2008.10.029

Zhu, G., Wang, S., Huang, Z., Zhang, S., Liao, Q., Zhang, C., et al. (2018a). Rewiring of the fruit metabolome in tomato breeding. *Cell* 172, 249. doi: 10.1016/j.cell.2017.12.019

Zhu, Q., Zeng, D., Yu, S., Cui, C., Li, J., Li, H., et al. (2018b). From Golden Rice to aSTARice: bioengineering astaxanthin biosynthesis in rice endosperm. *Mol. Plant* 11, 1440–1448. doi: 10.1016/j.molp.2018.09.007

Zhu, Q. L., Yu, S. Z., Zeng, D. C., Liu, H. M., Wang, H. C., Yang, Z. F., et al. (2017). Development of "Purple Endosperm Rice" by engineering anthocyanin biosynthesis in the endosperm with a high-efficiency transgene stacking system. *Mol. Plant* 10, 918–929. doi: 10.1016/j.molp.2017.05.008.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.