

Special Issue: Feeding the World: The Future of Plant Breeding

## Review

## Domestication of Crop Metabolomes: Desired and Unintended Consequences

Saleh Alseekh,<sup>1,2</sup> Federico Scossa,<sup>1,3</sup> Weiwei Wen,<sup>4</sup> Jie Luo,<sup>5,6</sup> Jianbing Yan,<sup>5</sup> Romina Beleggia,<sup>7</sup> Harry J. Klee,<sup>8</sup> Sanwen Huang,<sup>9</sup> Roberto Papa,<sup>10,\*</sup> and Alisdair R. Fernie<sup>1,2,\*</sup>

The majority of the crops and vegetables of today were domesticated from their wild progenitors within the past 12 000 years. Considerable research effort has been expended on characterizing the genes undergoing positive and negative selection during the processes of crop domestication and improvement. Many studies have also documented how the contents of a handful of metabolites have been altered during human selection, but we are only beginning to unravel the true extent of the metabolic consequences of breeding. We highlight how crop metabolomes have been wittingly or unwittingly shaped by the processes of domestication, and highlight how we can identify new targets for metabolite engineering for the purpose of *de novo* domestication of crop wild relatives.

## Domestication and Crop Improvement

Human-guided domestication began ~12 000 years ago in the Middle East and Fertile Crescent, and subsequently occurred across the world including events in China, Mesoamerica and the Andes, Near Oceania, Sub-Saharan Africa, and eastern North America [1–3]. Despite our simplified title, we distinguish here between domestication, diversification, and crop improvement events, wherever possible, given that in both evolutionary and phenotypic terms they are clearly distinct processes [4]. Large surveys have revealed that domesticated plant species span some 160 taxonomic families with >2500 species having undergone some extent of the process and ~300 being fully domesticated [2,3,5]. Current models integrating archaeological, genetic, and genomic evidence suggest that domestication is a multistage process consisting of (i) the onset of cultivation, (ii) an increased frequency of desirable alleles, (iii) the formation of domesticated populations, and finally (iv) deliberate breeding. That said, delineating the history of domestication is highly complex in many species because of the presence of multiple domestication events and frequent post-domestication exchanges with the progenitor species [6–8]. Moreover, it is important to note that some species such as *Oryza nivara* and Brazil nut are cultivated without domestication, and there has been insightful analysis regarding the genetic bottlenecks associated with the initial selection [9]. Taken together, these studies have greatly enhanced our understanding of trait evolution and have provided considerable insight into both convergent and parallel evolution during domestication [10]. For example, the stay green gene, *SGR*, underpins seed dormancy in a range of species [11], a subset of fruit weight quantitative trait loci map to the same genomic region in tomato (*Solanum lycopersicum*) and pepper (*Capsicum annuum*) [12], and the glutinous grain improvement traits of rice (*Oryza sativa*), sorghum (*Sorghum bicolor*), barley (*Hordeum vulgare*), and millet (*Pennisetum glaucum*) are all defined by different mutations in orthologs of the *Waxy* gene [2]. By contrast, although initial views on the appearance of the classical traits of the domestication syndromes, such as the loss of seed shattering in cereals, were considered to be cases of parallel evolution [13], genetic mapping studies have recently revealed that multiple traits are often associated with non-homologous genes [14]. For example, the canonical domestication gene, *TEOSINTE BRANCHED 1 (tb1)* of maize (*Zea mays*) [15] has minor effects on branching in foxtail millet [16], and even within separate barley lineages different

## Highlights

Next-generation sequencing has dramatically boosted our ability to study selection during domestication and crop improvement.

Domestication of our crop species is characterized by a reduction in allelic diversity and massive changes in both their gene expression and visible phenotype.

An increasing number of studies suggest that metabolism is also considerably affected.

However, in contrast to gene expression changes which are largely conserved, changes at the level of the metabolome appear to be species-specific.

<sup>1</sup>Max-Planck-Institut für Molekulare Pflanzenphysiologie, 14476 Potsdam-Golm, Germany

<sup>2</sup>Center of Plant Systems Biology and Biotechnology, Plovdiv 4000, Bulgaria

<sup>3</sup>Council for Agricultural Research and Economics (CREA), Research Centre for Genomics and Bioinformatics (CREA-GB), 00178 Rome, Italy

<sup>4</sup>Key Laboratory of Horticultural Plant Biology (MOE), College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Hubei, Wuhan 430070, China

<sup>5</sup>National Key Laboratory of Crop Genetic Improvement and National Centre of Plant Gene Research, Huazhong Agricultural University Hubei, Wuhan 430070, China

<sup>6</sup>College of Tropical Crops, Hainan University, Haikou, Hainan, China

<sup>7</sup>Council for Agricultural Research and Economics (CREA), Research Centre for Cereal and Industrial Crops (CREA-CI), 71122 Foggia, Italy

<sup>8</sup>Horticultural Sciences, University of Florida, Gainesville, FL, USA

genes are causal for shattering [17]. These and many other examples prove that both common and distinct mechanisms underlie the phenotypic convergence of traits in the domestication syndrome. An excellent review on the role of the domestication of vegetatively propagated crops is provided by Denham *et al.* [18]. The majority of crop domestication genes reported to date are involved in diverse developmental processes that largely produced changes in morphological phenotypes [10,19,20]. Prominent amongst these genes are transcription factors, and many studies have shown that domestication has led to major transcriptional reprogramming (Table 1).

Many studies have also shown how the contents of a handful of metabolites have been altered during the initial unintended and the protracted artificial selection phases during domestication. Indeed, a seemingly somewhat overlooked study on the domestication traits of 203 global food crops highlighted that changes in secondary metabolites are highly common domestication traits [21]. The emergence of a sweet, fleshy fruit, for example, can be considered to be an evolutionary adaptation to promote seed dispersal by animals; deliberate human selection subsequently transformed many fruit crops into obligate domesticates [22]. Most of the fruits traits shaped initially by zoochory

<sup>9</sup>Genome Analysis Laboratory of the Ministry of Agriculture – Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China  
<sup>10</sup>Department of Agricultural, Food, and Environmental Sciences, Università Politecnica delle Marche, 60131 Ancona, Italy

\*Correspondence: [r.papa@staff.univpm.it](mailto:r.papa@staff.univpm.it) (R. Papa) and [femie@mpimp-golm.mpg.de](mailto:femie@mpimp-golm.mpg.de) (A.R. Fernie).

Table 1. Genes Underlying Domestication That Are Directly Related to Metabolic Changes

Crop species	Gene	Function	Phenotypes	Refs
Apple	<i>Ma</i>	Transporter	Fruit acidity	[120,121]
Cucumber	<i>CmBr</i> , <i>CmBt</i>	Cytochrome P450s (CYPs)	Triterpenes/bitter taste	[79]
Melon	<i>CmPH</i>	Transporter	Acidity, sweetness	[80]
Watermelon	<i>QBRX2-1</i> , <i>CITST2</i> , <i>LCYB</i> , <i>CIAGA2</i>	Transporters, lycopene $\alpha$ -cyclase, $\alpha$ -galactosidase	Sweetness, sugars, carotenoids	[82,101,122]
Citrus sinensis	<i>Ruby</i> (MYB)	Transcription factor	Anthocyanin	[117]
Almond ( <i>Prunus dulcis</i> Miller)	<i>bHLH1</i> to <i>bHLH5</i>	Transcription factor	Bitter and toxic cyanogenic diglucoside amygdalin	[123]
Grape	<i>VvmybA1</i> (MYBA1)	Transcription factor	Anthocyanin biosynthesis/color	[124]
Grape	<i>VvMYBA2</i>	Transcription factor	Anthocyanin biosynthesis/color	[125]
Potato	<i>GLYCOALKALOID METABOLISM 4</i> (GAME4)	Cytochrome P450	Steroidal alkaloids	[106]
Tomato	<i>Solyc10 g085230</i>	UDP-glycosyltransferase	Steroidal alkaloids	[54]
Tomato	<i>Lin5</i>		Sugars	[126]
Tomato	<i>E8</i> (Solyc09g089580)	1-Aminocyclopropane-1-carboxylate oxidase	Volatiles	[126]
Tomato	<i>TomloxC</i>	Lipoxygenases	Lipids and volatiles	[39]
Tomato	<i>LIP1</i>	Lipase	Lipids and volatiles	[127]
<i>Ziziphus jujuba</i> Mill.	Currently still not resolved		Fruit sweetness	[67]
<i>Glycine max</i>	<i>R2R3 MYB</i>	Transcription factor	Anthocyanin (seed color)	[128]
Rice	<i>Black hull 4</i> ( <i>Bh4</i> )	Transporter	Color	[129]
Maize	<i>Teosinte branched 1</i> ( <i>tb1</i> )	Transcription factor	Regulating phytohormones such as gibberellins, abscisic acid, and jasmonic acid	[76]
<i>Chenopodium quinoa</i>	<i>TSARL1</i> ( <i>bHLH</i> )	Transcription factor	Saponins	[130]
Rice, potato	<i>Dihydroflavonol-4-reductase</i> ( <i>DFR</i> )	Dihydroflavonol reductase	Anthocyanin and proanthocyanidin	[131,132]
Tomato, maize	<i>MYB12</i>	Transcription factor	Flavonoids	[54,133]
Barley, rice	<i>sd1</i>	Enzyme (oxidation)	Hormone biosynthesis	[133,134]

(the dispersal of plant spores or seeds by animals), and later by human cultivation, involved color and flavor-related compounds, with a trend towards an increased sugar content and a reduction in bitterness and acidity [23]. Over the past few years such studies have, however, expanded to the metabolome level, considerably broadening the type of questions that can be addressed. These studies initially characterized broad patterns of change during domestication, for example revealing that wheat domestication was first characterized by a reduction in unsaturated fatty acids during the primary domestication event, and altered amino acid content then accompanied the secondary domestication event [24]. As we describe in the following text, in recent years genomics and transcriptomics have been used to study the effect of domestication on a wide range of species, including the major cereals and a range of other crops. We discuss aspects of these studies that seem most likely to impact on metabolism. We focus mainly on the application of metabolomics to unravel the effect of domestication on the metabolome in multiple important crop species, distinguishing, where possible, metabolic changes that are suggestive of being direct targets of selection from those where metabolic variation between wild and domesticated forms appears instead to be derived from processes other than selection (e.g., hitch-hiking, pleiotropy, or drift). In doing so, we use published data to highlight recent advances and insights into how crop metabolomes have been wittingly or unwittingly shaped by the processes of domestication and crop improvement. We suggest that metabolic change is a major feature of both processes and we discuss evidence supporting this conclusion derived from genomics and transcriptomics. Finally, we provide a roadmap to (i) gain a better understanding of the metabolic consequences of domestication and breeding, (ii) better characterize existing plant genetic resources, and (iii) identify new targets for metabolite engineering and gene editing approaches for the purpose of *de novo* domestication of crop wild relatives.

### Genome-Level Assessment of the Effects of Domestication

Essentially, the investigation of the genetic basis of phenotypic changes related to domestication has followed two main routes. In the first, classical quantitative genetic studies have been applied to the progeny of a controlled cross between a wild and a domesticated parent to identify associations between genetic polymorphisms and the phenotypic trait under investigation. This approach led to identification of genes controlling the large morphological and physiological changes between wild and domesticated forms: for example, genes related to the increase of fruit size [25] and seed mass [26] in tomato, or to the degree of shoot branching [27] as well as the regulation of flowering time [28] in maize.

Quantitative trait locus (QTL) mapping, however, has limitations. The power of any study is limited by the genotyping depth (although this is of lesser concern today, with the development of next-generation sequencing) and by the extent of the genetic variation captured in the population of wild and the domesticated individuals that may not be entirely representative of the domestication history of the species. Given that the QTL approach is also often restricted to the detection of polymorphisms associated with large phenotypic effects, these studies may give the impression that domestication was a relatively simple genetic process, controlled by few major loci. To overcome the limitations of QTL mapping, a second approach, that of linkage disequilibrium (LD) mapping, was introduced. Essentially, instead of studying the progeny of a controlled cross, LD mapping takes advantage of historical recombination which uses large panels of accessions, thus increasing the mapping resolution by checking for associations between phenotypes and genetic polymorphisms at the population level. Genome-wide associations of large sets of accessions have generally uncovered a more complex architecture of domestication, with the identification of numerous large and small effect loci underlying the differences between domesticated forms and their wild progenitors [29].

The second class of approaches for studying the genetic basis of domestication involve population-level analysis of genetic diversity. This approach is not biased towards a specific

phenotype; instead, the objective is to detect signatures of selection from DNA polymorphism data in the form of selective sweeps or population subdivision [30,31]. Genomic scans for selection in populations of wild and domesticated accessions have revealed a large number of putative regions under selection, thus uncovering many putative genes, not only those related to morphological changes, which could have been targeted by domestication [32]. However, it is important to note that, given the correct choice of parents in the development of the biparental crosses that underlie QTL mapping populations, insight into domestication could obviously be achieved. As such, the approaches should probably be regarded as complementary.

Although the studies mentioned so far have provided a large catalog of mutations – mostly SNPs – that are causal for traits affected by domestication and diversification, the impact of larger structural variants as well as genomic rearrangements and their relationships to the evolutionary processes occurring during domestication has received considerably less attention. The majority of causal polymorphisms identified so far, for example, have been assigned to SNPs, but more recent studies have started to uncover the contribution of larger genetic polymorphisms (e.g., copy-number variations and differential insertions of transposable elements, TEs) that affect domestication traits [19,33,34].

TEs, in particular, are ubiquitous genetic elements which possess the ability to integrate into different locations in the genome. TEs represent the largest majority of structural variants (SVs) and are present in nearly all genomes sequenced to date. In the case of *Arabidopsis thaliana* (arabidopsis), as well as other plants with small genome size, TEs may constitute up to 30% of the genome, but can reach proportions >80% in case of large genomes such as those of hexaploid wheat and maize [35]. Considering their large abundance, widespread distribution, and especially the mutagenic and regulatory roles TEs might have in plant genomes, it was long questioned whether these mobile elements could have contributed to the emergence of adaptive traits during plant evolution [36]. TE mobilization, in fact, in addition to mutations generated by the transposition itself (e.g., upon integration of the mobile element into a new genomic location), may also determine a wide range of effects on the function, processing, and expression of the genes in the vicinity of the integration site [37]. The largest majority of TE integration events were probably deleterious, such that these variants were promptly eliminated by natural selection. Although it is thus difficult to estimate the consequences of these TE-induced loss-of-function mutations over long evolutionary timescales, the process of plant domestication, which started at around 12 000 years ago [38], could offer instead the possibility to see the recent consequences of TE activity on the evolution of plant metabolic phenotypes under the influence of human selection. Indeed, some novel metabolic traits, apparently selected during domestication, have been documented to originate from TE insertions and subsequent rearrangements (Box 1), as well as larger insertions and deletions (e.g., [39]).

In addition, several typical phenotypes of domesticated plants, such as for example the gigantism of leaf and other organs, are reminiscent of the effects of polyploidy on organ size [40,41]. A recent study dated the occurrence and the frequency of polyploidization events in a large group of wild and domesticated plants, analyzing the data in a probabilistic framework across a large phylogenetic scale. In contrast to wild plants, polyploidization is a frequent feature in the genomes of domesticates, and its occurrence pre-dated, in most cases, the onset of domestication [42]. The reason for this higher preponderance of polyploids among domesticated forms is probably related to the higher adaptive value conferred by the presence of multiple genomes. Polyploidization increases allelic diversity, fueling evolutionary novelties upon which natural (and artificial) selection can act, and provides genetic buffering during the reduction of genetic diversity caused by domestication bottlenecks [42]. Thus, for most plant species analyzed, it seems probable that incipient domesticates were already polyploids [5].

### Box 1. Examples of TE-Induced Metabolic Phenotypes of Domestication

Arguably the best known example of a domestication phenotype induced by TE transposition into a metabolic gene is that in the gene encoding granule bound starch synthase (*GBSS1*) in foxtail millet (*Setaria italica*). Wild accessions of this species have a non-waxy endosperm, containing little or no amylose. The waxy endosperm in domesticated *S. italica* emerged from multiple integrations of various TEs into *GBSS1*, which is responsible for the synthesis of amylose. Although all wild accessions were non-waxy, and expressed a fully functional *GBSS1* gene, all Asian landraces instead carried alleles in which retrotransposons (RTs) and other class II TEs were inserted, causing loss-of-function mutations and resulting in the endosperm waxy phenotype [112,113]. Other cereals, such as barley and rice, also show convergence for the waxy phenotype upon domestication; in these cases, however, different mutations – not involving insertion of TEs – led to the loss of function of *GBSS1* and to the appearance of the waxy phenotype [114,115].

In *Vitis* spp., white grape varieties were found to be homozygous for the insertion of *Gret1* [a long terminal repeat (LTR) RT of the Ty3/gypsy family] immediately upstream of the coding sequence of *MybA1*, a transcription factor involved in the positive regulation of anthocyanin biosynthesis. The insertion of *Gret1* abolished or strongly reduced the expression of the *MybA1* gene; today, almost all white grape varieties are homozygous for this insertion. Because the two LTRs of *Gret1* have an almost identical DNA sequence, the insertion of this RT probably occurred recently in an ancestral black grape and was then brought to a homozygous state through crossing [116].

Plants of the *Citrus* genus (which includes sweet orange, mandarin, citron, and pummelo) offer another example of the impact of TEs in the diversification of a metabolic phenotype. Within the genus, the accumulation of anthocyanins in fruits, flowers, and leaves varies significantly among the various species and hybrids: the presence of these pigments has been linked to deletions, RT insertions, and nonfunctional alleles of the *Ruby* locus. The 'blood orange' phenotype, for example, was due to the insertion of a LTR-RT into the promoter of *Ruby*, causing its transcriptional activation [117]. Similarly, inactivating insertions of LTR-RTs have also been detected in *Noemi* [118,119]. Thus, during the domestication and subsequent diversification of *Citrus*, various selective pressures, both conscious and unconscious, led to the fixation of various combinations of functional and nonfunctional alleles of *Ruby* and *Noemi*. This allelic diversity probably explain the varying patterns of anthocyanin pigmentation observed in *Citrus* species today.

Other than polyploidization, few large-scale genomic surveys have addressed the impact of other chromosome-level mutations on domestication processes. It was long known, for example, that chromosome inversions may contribute to local adaptation, life history, and morphological traits, and thus to the spread of reproductive isolation barriers between populations [43,44]. Six large inversions detected in the genome of cucumber, *Cucumis sativus*, for example, seem to differentiate the wild (*C. sativus* var. *hardwickii*) from the cultivated forms (*C. sativus* var. *sativus*), suggesting a role for these inversions in the domestication of cucumber [45]. In general, however, assessing the impact of these chromosome rearrangements would require precise dating of when the inversion arose during species evolution. Most of the inversions characterized to date in the genomes of crops seem in fact to have been fixed well before the onset of domestication [46].

### Transcriptomics-Based Assessment of the Effects of Domestication

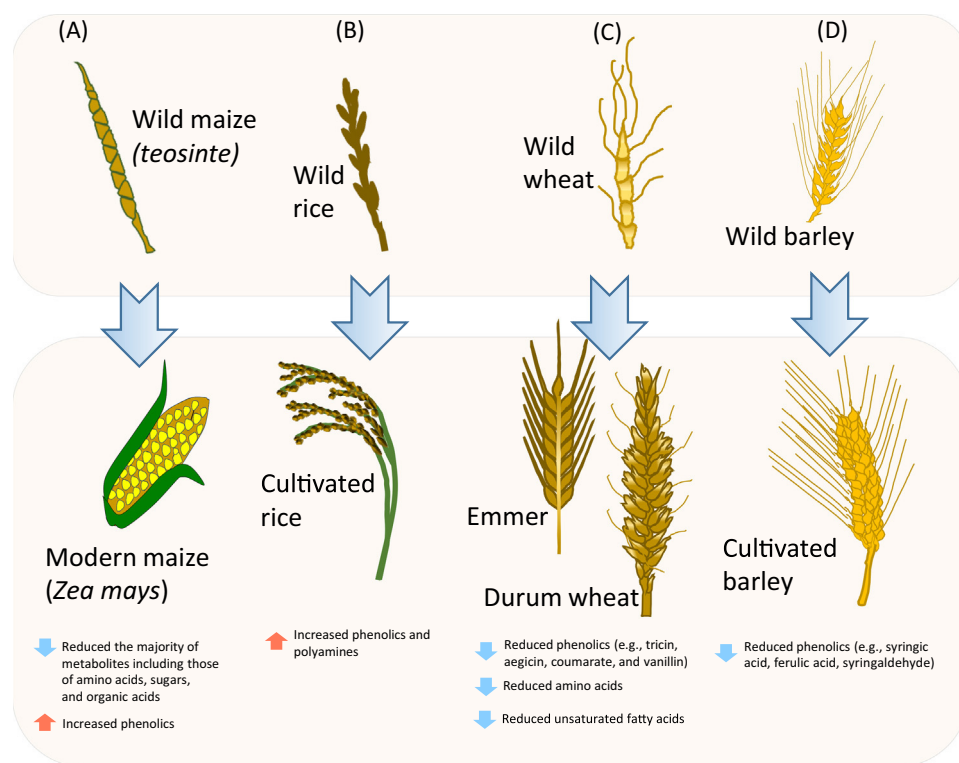
In recent years transcriptomics has been used to study the effect of domestication on a wide range of species including the major cereals rice [47,48], maize [49,50], wheat (*Triticum aestivum*) [51], and barley [52], as well as cotton (*Gossypium hirsutum*) [53], tomato [54,55], eggplant (*Solanum melongena*) [56,57], peppers [58], sunflower (*Helianthus annuus*) [59], common bean (*Phaseolus vulgaris*) [60], curcubita [61], and a range of more other crops including chickpea (*Cicer arietinum*) [62], lupin (*Lupinus albus*) [63], pistachio (*Pistacia vera*) [64], spinach (*Spinacia oleracea*) [65], lettuce (*Lactuca sativa*) [66], and jujube (*Ziziphus jujube*) [67]. A common theme that emerges from these studies is that domestication is associated with a large reduction in allelic diversity, with estimates suggesting that our current crops contain on average only 6% of the allelic diversity present in wild species gene pools [68]. Indeed, domestication results in massively decreased nucleotide diversity in common bean [60] and somewhat lower diversity in maize [50,69]. This reduced diversity is often associated with higher overall levels of expression, more uniform gene expression, and modified coexpression patterns. Because transcriptional changes that occurred during domestication and crop improvement have been recently reviewed elsewhere [70], we will not detail these studies extensively but instead focus on those changes that

are known, or could be anticipated, to affect metabolism. Given that increased overall gene expression is a relatively highly conserved domestication trait, the fact that so many of the domestication genes cloned to date are transcription factors [3,10,19] is perhaps not surprising. Although many of these genes, such as *Rht1* in wheat [71], *VRN1* and 2 in barley and wheat [72,73], *Sh1* in sorghum, rice, and maize [74], and *GW8* in rice [75], control reproductive and morphological traits, we contend that transcription factor-mediated alterations in metabolism are also prominent manifestations of the domestication and improvement processes. Table 1 lists examples of domestication genes directly related to metabolism that have been identified. The majority of these transcription factors, including those underlying pigmentation, sweetness, or bitterness, correspond to traits such as color or taste that are sensed by humans. Moreover, almost half of the genes in Table 1, including some of the genes invoking the above-mentioned traits, are the result of transcription factor-mediated changes in metabolism. Interestingly, the classical domestication gene *Tb1* is on this list because studies in maize identified that it not only affects hormone-mediated branching [76,77] but also displays highly altered energy metabolism [76]. Given these facts, alongside the likelihood that mutation of many other of the domestication genes has consequences for metabolism, these combined data suggest that metabolic consequences of domestication and improvement are considerably broader than the morphological traits classically associated with the domestication syndrome.

### Metabolomics-Based Assessment of the Effects of Domestication

Variation in crop species has been studied for decades for specific metabolic traits linked to agronomic traits; for example, the Illinois maize continuous selection experiment now running for well over 100 years has followed kernel protein and oil content [78]. Other examples of tracking metabolic variation include longstanding breeding for (i) reduced bitterness in potato (*S. tuberosum*), tomato, and cucumber [79], (ii) modified acidity in sweet melon (*Cucumis melo* var. *cantalupo*) [80], apple (*Malus domestica*) [81], (iii) modified sweetness in watermelon (*Citrullus lanatus*) [82], (iv) attractive color in citrus [83], melon [77], tomato [54], maize, rice, barley, soybean (*Glycine max*), grape (*Vitis vinifera*), apple, and common bean [84], (v) starchiness in rice [47,85], and (vi) aroma in a wide range of crops including tomato, pepper, cucumber, Brassicaceae, and onion (*Allium cepa*) [86]. Although the previously-mentioned studies were invaluable in identifying the genomic loci or even the genes underlying metabolic changes occurring on domestication, the combination of next-generation sequencing and metabolomics has greatly accelerated advances in our understanding of the metabolic changes accompanying domestication.

Early studies focusing on the levels of phenolics across an eggplant panel set up to establish the effects of domestication revealed that agronomic features aside from nutrition were prominently selected for [87], an observation that now appears to be a feature that is conserved throughout crop breeding [88]. Many of the early studies that adopted metabolomics to evaluate the domestication and improvement processes focused on cereals [24,89–93]. Indeed, the first such study was that of Beleggia *et al.* [24] who performed a relatively simple, but ground-breaking, evaluation of metabolic changes occurring during wheat domestication. They investigated the levels of 51 primary metabolites in the kernels of three *Triticum turgidum* L. subspecies (wild emmer, emmer, and durum wheat), finding that primary domestication was associated with a reduction in unsaturated fatty acids whereas a decrease in amino acid levels characterized secondary domestication (Figure 1). Importantly, these effects were partially independent of the associations of any of these metabolites with other domestication-related kernel traits. Moreover, the changes in metabolite content were coupled to alterations in metabolite correlation networks, suggesting that deep metabolic restructuring may have taken place during domestication and that this often resulted in a loss of nutritional quality [94]. In particular, the massive metabolic changes occurring specifically during the diversification phase of crops have been mostly driven by the strong



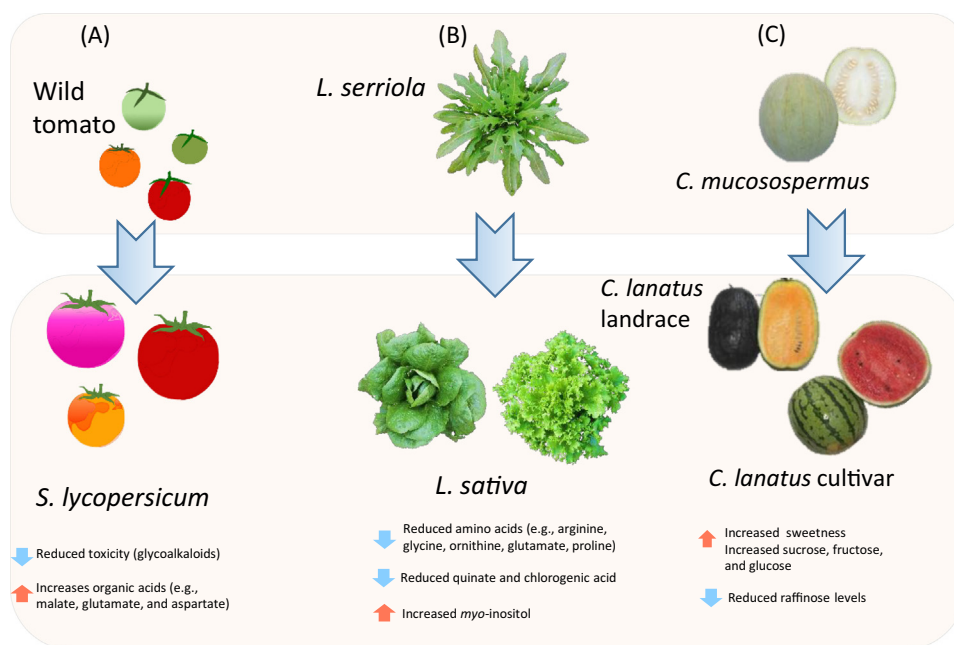
## Trends In Plant Science

**Figure 1. Metabolic Changes Concurrent with the Domestication and Improvement of Major Cereals.** The top part of the lower panel shows morphological changes in the crops (A) maize, (B) rice, (C) wheat, and (D) barley. The lower part of the lower panel shows the metabolic changes associated with primary and secondary domestication in wheat alongside domestication in maize, rice, and barley. Data from [24,49,90,92].

emphasis that breeders have put on yield-related traits, with consequent erosion of the genetic diversity that is important for quality traits (color, flavor, taste, and nutritional potential [95,96]). In another study, the metabolic divergence between maize and its wild ancestor teosinte was analyzed [91]. This study revealed that evolutionary transitions in the levels of alkaloids, terpenoids, and lipids were targeted at the divergence between teosinte and tropical maize, whereas benzoxazinoids were targeted at the divergence between tropical and temperate maize. This study and another like it additionally examined either maize–teosinte cross populations or maize and teosinte populations separately, and identified candidate genes underlying differences in metabolite composition [90,91] or sulfur-rich storage proteins [97], respectively. The changes associated with domestication in rice, as compared to maize, were recently addressed in a metabolomic study [89], revealing that these species displayed different metabolomic shifts during their evolution, and moreover suggests that those shifts were different from those reported in wheat. Two recent studies assessed changes in the metabolome during the domestication of barley – again showing diversity in their domestication-driven changes. In the first of these studies, the levels of secondary metabolites, particularly those known to confer UV-tolerance, were found to be considerably higher in quinke, a hull-less naked grained barley grown on the Tibetan plains [93]. The second study revealed differences in the allelopathic alkaloids gramine and hordenine in cultivars that likely had different domestication histories [92].

Although many studies have been carried out in cereals, a wide range of non-grass studies have been performed, providing insight into the metabolic changes that occurred during domestication

of a range of other species including tomato [54,98], lettuce [99], tea [100], watermelon [101], soybean [102], strawberry (*Fragaria × ananassa*) [103], and citrus [104]. The evaluation of the tomato metabolome was carried out at a previously unprecedented level in a large-scale multi-omic study including 610 tomato accessions, extending work on a previous population assembled to gain insight into the processes of tomato domestication and improvement [105]. This study demonstrated that selection of alleles of genes associated with larger fruits altered the metabolome as a consequence of linkage drag, whereas selection at five major loci reduced the accumulation of the anti-nutritional steroidal glycoalkaloids (SGAs) in ripe fruits. This study concurrently identified a second gene cluster for SGAs in addition to that previously reported by Itkin *et al.* [106]. Moreover, the study by Zhu *et al.* revealed that breeding for pink fruits, favored by Asian consumers, also modified the content of >100 metabolites, whereas the introgression of resistance genes from wild relatives also caused unexpected disturbance to the metabolome (Figure 2). Study of the primary metabolite complement of the lettuce population described previously revealed that metabolites in the wild species *L. serriola* differed from those of all evaluated cultivated lettuces, supporting a single domestication event for this species [99]. Moreover, galactinol, raffinose, malate, quinate, and threonate were affected by the domestication and cultivar differentiation processes, with the first two metabolites likely being selected during stem lettuce cultivation as an adaptation to the environment in China. Furthermore, quinate and chlorogenic acid levels were strongly reduced upon domestication of lettuce, probably as a consequence of the desire to reduce bitterness. Similarly, melon domestication was characterized by negative selection of cucurbitacins and flavonoids but also by elevated levels of sugars, carotenoids, and citrate:malate ratios [107], soybean domestication by negative selection of isoflavones [102], strawberry by shifts in ripening regulated metabolites [103], and citrus by negative selection on terpenoids [104]. By contrast, a pedigree



Trends in Plant Science

**Figure 2. Metabolic Changes Concurrent with Domestication and Improvement of Major Fruit and Vegetables.** The top part of the lower panel shows morphological changes in the crops (A) tomato, (B) lettuce, and (C) watermelon. The lower part of the lower panel shows the metabolic changes associated with increase in size, selection for color, and introgression of resistance genes in tomato, domestication in lettuce, and the stepwise domestication of watermelon. Data from [54,77,82,84,90].



and metabolite analysis in tea, which has a more complex domestication history, revealed virtually no long-term artificial directional selection for flavor-related metabolites [100].

### Conclusion and Future Prospects

Although by no means comprehensive in terms of the crop species studied, the massive qualitative and quantitative changes that have been documented to occur during domestication are consistent. Intriguingly, unlike the changes that occur in the genome and transcriptome during domestication, there is little commonality within the changes of the metabolomes across species. Importantly, the study on changes in the tomato metabolome, which remains the largest in scale to date, suggests that many more changes occurred during the crop improvement process than during domestication *per se*. This study further indicated that many of the changes in metabolism were unintended, suggesting that metabolomic studies will be highly instructive in assessing the nutritional composition of our crops and their progenitors. The large-scale metabolic reprogramming underlying both domestication and improvement must also be taken into account when considering both the use of *de novo* domestication of wild crop relatives and approaches concerned with engineering metabolically valuable crops [3]. In this regard, the Genotype–Phenotype Database ([www.gephebase.org](http://www.gephebase.org)) aims to collect data from eukaryotic organisms regarding the 'quantitative trait nucleotides' (QTNs) that are directly related to evolutionary changes. The database could thus provide an inventory of suitable genes, for which a clear genotype–phenotype link has been already established, to be engineered or introgressed into crops for the transfer of useful traits. Analysis of the Gephebase database also confirms that the largest majority of the mutations selected during the breeding process, both in animal and plants, are mostly loss-of-function mutations (nonsense, frameshift, or regulatory mutations abolishing gene expression). As such, the global view of evolution under crop diversification is consistent with fixation of large-effect loss-of-function mutations that generate evolutionary novelties in the human-controlled setting, but are then maladaptive under natural or highly variable environments [108]. This point aside, the database additionally contains several other metabolism-associated domestication genes of interest, including the *BADH2* gene in rice, inactivation of which has been demonstrated to enrich aromatic properties [109]. Taken together, from the multiple examples described previously it is safe to say that domestication and subsequent crop improvement have been characterized not only by substantial changes in transcription and visible phenotype but also by considerable changes in metabolism. In terms of utilizing characterized domestication genes for *de novo* domestications of new species (e.g., [110,111]), it is important to note that, aside from *tb1*, the metabolic consequences of key developmental genes associated with domestication have rarely been studied. Such research is a clear priority for better integration of metabolic data to understand and influence yield and quality traits of current and novel crops. With respect to understanding metabolic changes on domestication *per se* as we list in the Outstanding Questions, a crucial point will be to include more landraces in future studies to allow us to determine whether changes occurred during the domestication event or instead in the subsequent process of crop improvement.

### Acknowledgments

A.R.F. and S.A. are supported by the EU Horizon 2020 Research and Innovation Programme, project PlantaSYST (SGA-CSA No. 739582 under FPA 664620). R.P. is grateful to the European Research Area Network for Coordinating Action in Plant Sciences (ERA-CAPS) project BEAN\_ADAPT and R.P. to the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) 2017 project PARDOM.

### Declaration of Interests

The authors declare no conflicts of interest.

### Outstanding Questions

Although the metabolic changes on domestication and improvement processes combined are divergent in diverse species, it will be interesting to dissect this at higher resolution by the inclusion of more landraces to see if the domestication event *per se* triggered similar changes in the metabolome.

Similarly, higher-level resolution of the type performed in tomato to establish the metabolomic consequences of wild species introgressions in a wider range of species will likely be highly informative.

Finally, more extensive studies concerning the role of adaptive changes when a crop is introduced into a novel habitat will be highly useful in understanding the historical expansion of particular crops as well as for planning future agricultural strategies.

**References**

1. Purugganan, M.D. and Fuller, D.Q. (2009) The nature of selection during plant domestication. *Nature* 457, 843–848
2. Meyer, R.S. and Purugganan, M.D. (2013) Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852
3. Fernie, A.R. and Yan, J. (2019) De novo domestication: an alternative route toward new crops for the future. *Mol. Plant* 12, 615–631
4. Purugganan, M.D. (2019) Evolutionary insights into the nature of plant domestication. *Curr. Biol.* 29, R705–R714
5. Salman-Minkov, A. *et al.* (2016) Whole-genome duplication as a key factor in crop domestication. *Nat. Plants* 2, 16115
6. Bitocchi, E. *et al.* (2012) Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc. Natl Acad. Sci. U. S. A.* 109, E788–E796
7. Liu, H. and Yan, J. (2017) Rice domestication: an imperfect African solution. *Nat. Plants* 3, 17083
8. Gross, B.L. and Olsen, K.M. (2010) Genetic perspectives on crop domestication. *Trends Plant Sci.* 15, 529–537
9. Gaut, B.S. *et al.* (2018) Demography and its effects on genomic variation in crop domestication. *Nat. Plants* 4, 512–520
10. Lenser, T. and Theißen, G. (2013) Molecular mechanisms involved in convergent crop domestication. *Trends Plant Sci.* 18, 704–714
11. Wang, M. *et al.* (2018) Parallel selection on a dormancy gene during domestication of crops from multiple families. *Nat. Genet.* 50, 1435–1441
12. Paran, I. and van der Knaap, E. (2007) Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. *J. Exp. Bot.* 58, 3841–3852
13. Paterson, A.H. *et al.* (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic Loci. *Science* 269, 1714–1718
14. Li, W. and Gill, B.S. (2006) Multiple genetic pathways for seed shattering in the grasses. *Funct. Integr. Genomics* 6, 300–309
15. Doebley, J. (2004) The genetics of maize evolution. *Annu. Rev. Genet.* 38, 37–59
16. Doust, A.N. *et al.* (2004) Genetic control of branching in foxtail millet. *Proc. Natl. Acad. Sci. U. S. A.* 101, 9045–9050
17. Pourkheirandish, M. *et al.* (2015) Evolution of the grain dispersal system in barley. *Cell* 162, 527–539
18. Denham, T. *et al.* (2020) The domestication syndrome in vegetatively propagated field crops. *Ann. Bot.* 125, 581–597
19. Lye, Z.N. and Purugganan, M.D. (2019) Copy number variation in domestication. *Trends Plant Sci.* 24, 352–365
20. Martínez-Ainsworth, N.E. and Tenailon, M.I. (2016) Superheroes and masterminds of plant domestication. *C. R. Biol.* 339, 268–273
21. Meyer, R.S. *et al.* (2012) Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48
22. Spengler 3rd, R.N. (2020) Anthropogenic seed dispersal: rethinking the origins of plant domestication. *Trends Plant Sci.* 25, 340–348
23. Shang, Y. *et al.* (2014) Plant science. Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* 346, 1084–1088
24. Beleggia, R. *et al.* (2016) Evolutionary metabolomics reveals domestication-associated changes in tetraploid wheat kernels. *Mol. Biol. Evol.* 33, 1740–1753
25. Frary, A. *et al.* (2000) *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 289, 85–88
26. Orsi, C.H. and Tanksley, S.D. (2009) Natural variation in an ABC transporter gene associated with seed size evolution in tomato species. *PLoS Genet.* 5, e1000347
27. Doebley, J. *et al.* (1990) Genetic and morphological analysis of a maize-teosinte F2 population: implications for the origin of maize. *Proc. Natl. Acad. Sci. U. S. A.* 87, 9888–9892
28. Jiménez-Gómez, J.M. *et al.* (2007) Quantitative genetic analysis of flowering time in tomato. *Genome* 50, 303–315
29. Wang, H. *et al.* (2016) The power of inbreeding: NGS-based GWAS of rice reveals convergent evolution during rice domestication. *Mol. Plant* 9, 975–985
30. Booker, T.R. *et al.* (2017) Detecting positive selection in the genome. *BMC Biol.* 15, 98
31. Scoossa, F. and Fernie, A.R. (2020) The evolution of metabolism: how to test evolutionary hypotheses at the genomic level. *Comput. Struct. Biotechnol. J.* 18, 482–500
32. Kantar, M.B. *et al.* (2017) The genetics and genomics of plant domestication. *BioScience* 67, 971–982
33. Alonge, M. *et al.* (2020) Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* 182, 145–161
34. Domínguez, M. *et al.* (2020) The impact of transposable elements on tomato diversity. *Nat. Commun.* 11, 4058
35. Akakpo, R. *et al.* (2020) The impact of transposable elements on the structure, evolution and function of the rice genome. *New Phytol.* 226, 44–49
36. Lisch, D. (2013) How important are transposons for plant evolution? *Nat. Rev. Genet.* 14, 49–61
37. Dubin, M.J. *et al.* (2018) Transposons: a blessing curse. *Curr. Opin. Plant Biol.* 42, 23–29
38. Stetter, M.G. *et al.* (2017) How to make a domesticate. *Curr. Biol.* 27, R896–r900
39. Gao, L. *et al.* (2019) The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* 51, 1044–1051
40. Doyle, J.J. and Coate, J.E. (2019) Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. *Int. J. Plant Sci.* 180, 1–52
41. Robinson, D.O. *et al.* (2018) Ploidy and size at multiple scales in the *Arabidopsis* sepal. *Plant Cell* 30, 2308–2329
42. Nieto Feliner, G. *et al.* (2020) Genomics of evolutionary novelty in hybrids and polyploids. *Front. Genet.* 11, 792
43. Lowry, D.B. and Willis, J.H. (2010) A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* 8, e1000500
44. Huang, K. and Rieseberg, L.H. (2020) Frequency, origins, and evolutionary role of chromosomal inversions in plants. *Front. Plant Sci.* 11, 296
45. Yang, L. *et al.* (2012) Chromosome rearrangements during domestication of cucumber as revealed by high-density genetic mapping and draft genome assembly. *Plant J.* 71, 895–906
46. Wellenreuther, M. and Bernatchez, L. (2018) Eco-evolutionary genomics of chromosomal inversions. *Trends Ecol. Evol.* 33, 427–440
47. Zheng, X.M. *et al.* (2019) Genome-wide analyses reveal the role of noncoding variation in complex traits during rice domestication. *Sci. Adv.* 5, eaax3619
48. Zheng, J. *et al.* (2019) Determining factors, regulation system, and domestication of anthocyanin biosynthesis in rice leaves. *New Phytol.* 223, 705–721
49. Yan, L. *et al.* (2019) Parallels between natural selection in the cold-adapted crop-wild relative *Tripsacum dactyloides* and artificial selection in temperate adapted maize. *Plant J.* 99, 965–977
50. Swanson-Wagner, R. *et al.* (2012) Reshaping of the maize transcriptome by domestication. *Proc. Natl. Acad. Sci. U. S. A.* 109, 11878–11883
51. Xiang, D. *et al.* (2019) The transcriptional landscape of polyploid wheats and their diploid ancestors during embryogenesis and grain development. *Plant Cell* 31, 2888–2911
52. Dai, F. *et al.* (2014) Transcriptome profiling reveals mosaic genomic origins of modern cultivated barley. *Proc. Natl. Acad. Sci. U. S. A.* 111, 13403–13408
53. Bao, Y. *et al.* (2019) Unraveling cis and trans regulatory evolution during cotton domestication. *Nat. Commun.* 10, 5399
54. Zhu, G. *et al.* (2018) Rewiring of the fruit metabolome in tomato breeding. *Cell* 172, 249–261 e12
55. Koenig, D. *et al.* (2013) Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci. U. S. A.* 110, E2655–E2662
56. Wei, Q. *et al.* (2019) Comparative transcriptome analysis in eggplant reveals selection trends during eggplant domestication. *Int. J. Genomics* 2019

57. Page, A. *et al.* (2019) Eggplant domestication: pervasive gene flow, feralization, and transcriptomic divergence. *Mol. Biol. Evol.* 36, 1359–1372
58. Díaz-Valenzuela, E. *et al.* (2020) Cis- and trans-regulatory variations in the domestication of the chili pepper fruit. *Mol. Biol. Evol.* 37, 1593–1603
59. Badouin, H. *et al.* (2017) The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546, 148–152
60. Bellucci, E. *et al.* (2014) Decreased nucleotide and expression diversity and modified coexpression patterns characterize domestication in the common bean. *Plant Cell* 26, 1901–1912
61. Montero-Pau, J. *et al.* (2018) De novo assembly of the zucchini genome reveals a whole-genome duplication associated with the origin of the *Cucurbita* genus. *Plant Biotechnol. J.* 16, 1161–1171
62. Moenga, S.M. *et al.* (2020) Gene co-expression analysis reveals transcriptome divergence between wild and cultivated chickpea under drought stress. *Plant J.* 104, 1195–1214
63. Hufnagel, B. *et al.* (2020) High-quality genome sequence of white lupin provides insight into soil exploration and seed quality. *Nat. Commun.* 11, 492
64. Zeng, L. *et al.* (2019) Whole genomes and transcriptomes reveal adaptation and domestication of pistachio. *Genome Biol.* 20, 79
65. Xu, C. *et al.* (2017) Draft genome of spinach and transcriptome diversity of 120 *Spinacia* accessions. *Nat. Commun.* 8, 15275
66. Zhang, L. *et al.* (2017) RNA sequencing provides insights into the evolution of lettuce and the regulation of flavonoid biosynthesis. *Nat. Commun.* 8, 2264
67. Guo, M. *et al.* (2020) Genomic analyses of diverse wild and cultivated accessions provide insights into the evolutionary history of jujube. *Plant Biotechnol. J.* Published online September 18, 2020. <https://doi.org/10.1111/pbi.13480>
68. Fernie, A.R. *et al.* (2006) Natural genetic variation for improving crop quality. *Curr. Opin. Plant Biol.* 9, 196–202
69. Hufford, M.B. *et al.* (2012) Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44, 808–811
70. Schreiber, M. *et al.* (2018) Genomic approaches for studying crop evolution. *Genome Biol.* 19, 140
71. Doebley, J.F. *et al.* (2006) The molecular genetics of crop domestication. *Cell* 127, 1309–1321
72. Asp, T. *et al.* (2011) Comparative sequence analysis of VRN1 alleles of *Lolium perenne* with the co-linear regions in barley, wheat, and rice. *Mol. Genet. Genomics* 286, 433–447
73. Alonso-Blanco, C. *et al.* (2009) What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell* 21, 1877–1896
74. Lin, Z. *et al.* (2012) Parallel domestication of the Shattering1 genes in cereals. *Nat. Genet.* 44, 720–724
75. Martin, A. and Orgogozo, V. (2013) The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* 67, 1235–1250
76. Dong, Z. *et al.* (2019) The regulatory landscape of a core maize domestication module controlling bud dormancy and growth repression. *Nat. Commun.* 10, 3810
77. Zhao, G. *et al.* (2019) A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. *Nat. Genet.* 51, 1607–1615
78. Moose, S.P. *et al.* (2004) Maize selection passes the century mark: a unique resource for 21st century genomics. *Trends Plant Sci.* 9, 358–364
79. Zhou, Y. *et al.* (2016) Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. *Nat. Plants* 2, 16183
80. Cohen, S. *et al.* (2014) The PH gene determines fruit acidity and contributes to the evolution of sweet melons. *Nat. Commun.* 5, 4026
81. Ma, B. *et al.* (2018) Determination of predominant organic acid components in malus species: correlation with apple domestication. *Metabolites* 8, 74
82. Guo, S. *et al.* (2019) Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. *Nat. Genet.* 51, 1616–1623
83. Butelli, E. *et al.* (2017) Changes in anthocyanin production during domestication of *Citrus*. *Plant Physiol.* 173, 2225–2242
84. Paauw, M. *et al.* (2019) Alteration of flavonoid pigmentation patterns during domestication of food crops. *J. Exp. Bot.* 70, 3719–3735
85. Parween, S. *et al.* (2020) Balancing the double-edged sword effect of increased resistant starch content and its impact on rice texture: its genetics and molecular physiological mechanisms. *Plant Biotechnol. J.* 18, 1763–1777
86. Zhu, G. *et al.* (2019) Next-gen approaches to flavor-related metabolism. *Annu. Rev. Plant Biol.* 70, 187–212
87. Meyer, R.S. *et al.* (2015) Parallel reductions in phenolic constituents resulting from the domestication of eggplant. *Phytochemistry* 115, 194–206
88. Hufford, M.B. *et al.* (2019) Crop biodiversity: an unfinished magnum opus of nature. *Annu. Rev. Plant Biol.* 70, 727–751
89. Deng, M. *et al.* (2020) Metabolomic analysis reveals differences in evolution between maize and rice. *Plant J.*
90. Li, K. *et al.* (2019) Large-scale metabolite quantitative trait locus analysis provides new insights for high-quality maize improvement. *Plant J.* 99, 216–230
91. Xu, G. *et al.* (2019) Evolutionary metabolomics identifies substantial metabolic divergence between maize and its wild ancestor, teosinte. *Plant Cell* 31, 1990–2009
92. Mavor, M. *et al.* (2020) New insights in the allelopathic traits of different barley genotypes: Middle Eastern and Tibetan wild-relative accessions vs. cultivated modern barley. *PLoS One* 15, e0231976
93. Zeng, X. *et al.* (2020) Genome-wide dissection of co-selected UV-B responsive pathways in the UV-B adaptation of qingke. *Mol. Plant* 13, 112–127
94. Hebelstrup, K.H. (2017) Differences in nutritional quality between wild and domesticated forms of barley and emmer wheat. *Plant Sci.* 256, 1–4
95. Morris, C.E. and Sands, D.C. (2006) The breeder's dilemma – yield or nutrition? *Nat. Biotechnol.* 24, 1078–1080
96. Folta, K.M. and Klee, H.J. (2016) Sensory sacrifices when we mass-produce mass produce. *Hortic. Res.* 3, 16032
97. Li, X. *et al.* (2018) Genetic diversity and evolution of reduced sulfur storage during domestication of maize. *Plant J.* 94, 943–955
98. Schauer, N. *et al.* (2005) Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex. *J. Exp. Bot.* 56, 297–307
99. Zhang, W. *et al.* (2020) Dissection of the domestication-shaped genetic architecture of lettuce primary metabolism. *Plant J.* 104, 613–630
100. Zhang, W. *et al.* (2020) Genome assembly of wild tea tree DASZ reveals pedigree and selection history of tea varieties. *Nat. Commun.* 11, 3719
101. Ren, Y. *et al.* (2014) An integrated genetic map based on four mapping populations and quantitative trait loci associated with economically important traits in watermelon (*Citrullus lanatus*). *BMC Plant Biol.* 14, 33
102. Yun, D.Y. *et al.* (2020) Metabotyping of different soybean genotypes and distinct metabolism in their seeds and leaves. *Food Chem.* 330, 127198
103. Vallarino, J.G. *et al.* (2018) Genetic diversity of strawberry germplasm using metabolomic biomarkers. *Sci. Rep.* 8, 14386
104. Zhang, H. *et al.* (2020) Transcriptomic and metabolomic analyses provide insight into the volatile compounds of citrus leaves and flowers. *BMC Plant Biol.* 20, 7
105. Lin, T. *et al.* (2014) Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* 46, 1220–1226
106. Itkin, M. *et al.* (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science* 341, 175–179
107. Yuan, P. *et al.* Watermelon domestication was shaped by stepwise selection and regulation of the metabolome. *Nat. Commun.* (in press)
108. Courtier-Orgogozo, V. and Martin, A. (2020) The coding loci of evolution and domestication: current knowledge and implications for bio-inspired genome editing. *J. Exp. Biol.* 223, jeb208934

109. Shan, Q. *et al.* (2015) Creation of fragrant rice by targeted knockout of the OsBADH2 gene using TALEN technology. *Plant Biotechnol. J.* 13, 791–800
110. Zsögön, A. *et al.* (2018) De novo domestication of wild tomato using genome editing. *Nat. Biotechnol.* 36, 1211–1216
111. Zsögön, A. *et al.* (2017) Genome editing as a tool to achieve the crop ideotype and de novo domestication of wild relatives: case study in tomato. *Plant Sci.* 256, 120–130
112. Kawase, M. *et al.* (2005) Diverse origins of waxy foxtail millet crops in East and Southeast Asia mediated by multiple transposable element insertions. *Mol. Gen. Genomics.* 274, 131–140
113. Kuo, S.M. *et al.* (2018) Waxy allele diversification in foxtail millet (*Setaria italica*) landraces of Taiwan. *PLoS One* 13, e0210025
114. Olsen, K.M. and Purugganan, M.D. (2002) Molecular evidence on the origin and evolution of glutinous rice. *Genetics* 162, 941–950
115. Dornon, E. *et al.* (2002) The insertion/deletion polymorphisms in the waxy gene of barley genetic resources from East Asia. *Theor. Appl. Genet.* 104, 132–138
116. Kobayashi, S. *et al.* (2004) Retrotransposon-induced mutations in grape skin color. *Science* 304, 982
117. Butelli, E. *et al.* (2012) Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* 24, 1242–1255
118. Butelli, E. *et al.* (2019) Noemi controls production of flavonoid pigments and fruit acidity and illustrates the domestication routes of modern *Citrus* varieties. *Curr. Biol.* 29, 158–164
119. Zhu, F. *et al.* (2019) Finding Noemi: the transcription factor mutations underlying trait differentiation amongst *Citrus*. *Trends Plant Sci.* 24, 384–386
120. Bai, Y. *et al.* (2012) A natural mutation-led truncation in one of the two aluminum-activated malate transporter-like genes at the Ma locus is associated with low fruit acidity in apple. *Mol. Gen. Genomics.* 287, 663–678
121. Sun, X. *et al.* (2020) Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat. Genet.* 52, 1423–1432
122. Nimmakayala, P. *et al.* (2014) Single nucleotide polymorphisms generated by genotyping by sequencing to characterize genome-wide diversity, linkage disequilibrium, and selective sweeps in cultivated watermelon. *BMC Genomics* 15, 767
123. Sánchez-Pérez, R. *et al.* (2019) Mutation of a bHLH transcription factor allowed almond domestication. *Science* 364, 1095–1098
124. This, P. *et al.* (2007) Wine grape (*Vitis vinifera* L.) color associates with allelic variation in the domestication gene VvmybA1. *Theor. Appl. Genet.* 114, 723–730
125. Walker, A.R. *et al.* (2007) White grapes arose through the mutation of two similar and adjacent regulatory. *genes* 49, 772–785
126. Tieman, D. *et al.* (2017) A chemical genetic roadmap to improved tomato flavor. *Science* 355, 391–394
127. Garbowicz, K. *et al.* (2018) Quantitative trait loci analysis identifies a prominent gene involved in the production of fatty acid-derived flavor volatiles in tomato. *Mol. Plant* 11, 1147–1165
128. Gillman, J.D. *et al.* (2011) Loss-of-function mutations affecting a specific *Glycine max* R2R3 MYB transcription factor result in brown hilum and brown seed coats. *BMC Plant Biol.* 11, 155
129. Zhu, B.-F. *et al.* (2011) Genetic control of a transition from black to straw-white seed hull in rice domestication. *Plant Physiol.* 155, 1301
130. Jarvis, D.E. *et al.* (2017) The genome of *Chenopodium quinoa*. *Nature* 542, 307–312
131. Furukawa, T. *et al.* (2007) The Rc and Rd genes are involved in proanthocyanidin synthesis in rice pericarp. *Plant J.* 49, 91–102
132. Zhang, Y. *et al.* (2009) The potato R locus codes for dihydroflavonol 4-reductase. TAG. Theoretical and applied genetics. *Theor. Angew. Genet.* 119, 931–937
133. Dias, A.P. *et al.* (2003) Recently duplicated maize R2R3 Myb genes provide evidence for distinct mechanisms of evolutionary divergence after duplication. *Plant Physiol.* 131, 610–620
134. Jia, Q. *et al.* (2009) GA-20 oxidase as a candidate for the semidwarf gene sdw1/denso in barley. *Funct. Integr. Genomics* 9, 255–262