



Plant Metabolic Gene Clusters: Evolution, Organization, and Their Applications in Synthetic Biology

Revuru Bharadwaj¹, Sarma R. Kumar¹, Ashutosh Sharma^{2*} and Ramalingam Sathishkumar^{1*}

¹ Plant Genetic Engineering Laboratory, Department of Biotechnology, Bharathiar University, Coimbatore, India, ² Tecnológico de Monterrey, Centre of Bioengineering, Querétaro, Mexico

OPEN ACCESS

Edited by:

Jakob Franke,
Leibniz University Hannover, Germany

Reviewed by:

Raimund Nagel,
University of Leipzig, Germany
Mariam Gaid,
Independent Researcher,
Braunschweig, Germany

*Correspondence:

Ramalingam Sathishkumar
rsathish@buc.edu.in
Ashutosh Sharma
asharma@tec.mx

Specialty section:

This article was submitted to
Plant Metabolism and Chemodiversity,
a section of the journal
Frontiers in Plant Science

Received: 19 April 2021

Accepted: 05 July 2021

Published: 13 August 2021

Citation:

Bharadwaj R, Kumar SR, Sharma A
and Sathishkumar R (2021) Plant
Metabolic Gene Clusters: Evolution,
Organization, and Their Applications in
Synthetic Biology.
Front. Plant Sci. 12:697318.
doi: 10.3389/fpls.2021.697318

Plants are a remarkable source of high-value specialized metabolites having significant physiological and ecological functions. Genes responsible for synthesizing specialized metabolites are often clustered together for a coordinated expression, which is commonly observed in bacteria and filamentous fungi. Similar to prokaryotic gene clustering, plants do have gene clusters encoding enzymes involved in the biosynthesis of specialized metabolites. More than 20 gene clusters involved in the biosynthesis of diverse metabolites have been identified across the plant kingdom. Recent studies demonstrate that gene clusters are evolved through gene duplications and neofunctionalization of primary metabolic pathway genes. Often, these clusters are tightly regulated at nucleosome level. The prevalence of gene clusters related to specialized metabolites offers an attractive possibility of an untapped source of highly useful biomolecules. Accordingly, the identification and functional characterization of novel biosynthetic pathways in plants need to be worked out. In this review, we summarize insights into the evolution of gene clusters and discuss the organization and importance of specific gene clusters in the biosynthesis of specialized metabolites. Regulatory mechanisms which operate in some of the important gene clusters have also been briefly described. Finally, we highlight the importance of gene clusters to develop future metabolic engineering or synthetic biology strategies for the heterologous production of novel metabolites.

Keywords: plant gene clusters, specialized metabolites, defensive functions, gene duplications, metabolic engineering, synthetic biology

INTRODUCTION

Plants produce an array of specialized metabolites to evade biotic and abiotic stressors. Therefore, the production of specialized metabolites is influenced by various environmental cues. These metabolites have been extensively employed in preparing herbal formulations for human health care. For instance, specialized metabolites, such as vincristine, vinblastine, paclitaxel, and curcumin, are recognized as effective inhibitors of cell proliferation and being used in cancer therapeutics (Seca and Pinto, 2018). The significance of plant secondary metabolites in human medicine led researchers to explore the plant kingdom for understanding the biosynthetic machinery of novel metabolites. Plant specialized metabolites are classified according to their chemical backbone and functional groups. Biosynthetic pathways of several specialized metabolites have been elucidated by characterizing the pathway genes, regulators, and gene products involved

in their biosynthesis. Transcriptomic, functional genomics studies combined with metabolomic approaches revealed insights into the operational features of novel metabolite pathways in different medicinal plants (Verma et al., 2014; Meena et al., 2017; Anand et al., 2019; Nagegowda and Gupta, 2020). It is quite challenging to understand the evolutionary aspects of the plant metabolic diversity at the molecular level as several metabolite-encoding gene cascades might be present in the plant genome as well, which are yet to be deciphered (Nützmann et al., 2016).

The clustering of nonhomologous genes of catabolic enzymes and the genes involved in the biosynthesis of specialized metabolites is common in prokaryotes with Lac-operon being the best example (Jacob et al., 1960). Further, in *Streptomyces* sp., genes encoding enzymes involved in the biosynthesis of antibiotics such as granaticin, actinorhodin, are reported to be clustered (Caballero et al., 1991; Ichinose et al., 1998). A few classes of filamentous fungi are known to possess clusters of both primary and secondary metabolic pathway genes and are coordinately expressed (Nützmann et al., 2018). However, the functional genes in animals and plants are scattered throughout the genome, except in a few cases of gene complexes, such as homeobox (*Hox*) and major histocompatibility complexes (*MHCs*), which exist as clusters in animals and are expressed in a synchronized manner (Horton et al., 2004; Holland, 2013; Nützmann et al., 2018). In plants, until the discovery of a gene cluster in *Zea mays* (maize) involved in the biosynthesis of hydroxamic acid 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA), it was assumed that secondary metabolite-producing genes occur randomly in the plant genome (Frey et al., 1997). The genes encoding the enzymes for DIMBOA biosynthesis have been reported to be clustered together on the chromosome 4 in maize, and this cluster is found to be widely distributed among the monocots (Frey et al., 2009). In general, a biosynthetic gene cluster is defined as the occurrence of two or more non-homologous genes, in the vicinity on a particular chromosome, which are involved in a common biosynthetic pathway to produce a specialized metabolite or group of similar metabolites (Medema et al., 2015). In particular, plant gene cluster size ranges from 35 kb to several hundred kb, and gene clusters comprises primarily the genes responsible for determining class of metabolites and secondarily one or more genes whose role is to modulate metabolite scaffold to create metabolic diversity (Schneider et al., 2017; Nützmann et al., 2020).

Fungal and plant gene clusters generally share several similarities in the cluster architecture and evolutionary aspects except in the case of the concept of horizontal gene transfer (HGT) of clusters in fungi, which is absent in plants (Slot and Rokas, 2010; Nützmann et al., 2018). Filamentous fungi possess both primary and secondary metabolite producing gene clusters, however specialized metabolite producing gene clusters have been predominantly characterized in plants (Nützmann et al., 2018; Rokas et al., 2018). While fungal gene clusters are equipped with pathway specific transcriptional regulators and transporters, this salient feature is not common in plant gene clusters (Rokas et al., 2020). However, a few plant gene clusters

have been reported to possess transporters and coordinately activated regulatory genes (Darbani et al., 2016; Hen-Avivi et al., 2016; Shen et al., 2021). In addition, multigene clusters for a single biosynthetic pathway, the co-occurrence of two clusters of genes on a single chromosome, and intertwined clusters for producing different metabolites are common in filamentous fungi (Yu et al., 2000; Bradshaw et al., 2013; Wiemann et al., 2013). These distinct features of gene cluster organization is not common in plants, except for a few cases such as the co-occurrence of steroidal glycoalkaloid- (SGA) and acyl sugar producing gene clusters on a single chromosome in the members of the *Solanaceae* and the multi-functionality of CYP76M8 in the phytoalexin production in rice (Fan et al., 2020; Kitaoka et al., 2021).

In the past two decades, more than 20 specialized metabolite-producing gene clusters involved in the biosynthesis of different classes of compounds have been identified in various plants (Nützmann et al., 2016). Recent discoveries of high-value metabolite-producing noscapine and thebaine gene clusters in *Papaver somniferum* (poppy) contributed significantly to the understanding of cluster organization, and these findings could aid in developing metabolic engineering/synthetic biology strategies for overproducing these compounds in heterologous production platforms (Guo et al., 2018). Plant metabolic gene clusters typically contain primarily the genes of committed or rate limiting enzymes of various pathways and secondarily the other enzymes required for the modification of backbone to form end-products. It has been shown that, in most cases genes are recruited from the primary metabolism through gene duplications followed by neofunctionalization (Qi et al., 2006). Most of the gene clusters are inferred to have evolved as adaptive strategies to defend against pathogen attack by producing defense metabolites, and a few other clusters are found to have a role in plant development (Qi et al., 2006; Field and Osbourn, 2008; Krokida et al., 2013; **Table 1**). It could be possible that plant systems favor the clustering of genes for controlled and regulated accumulation of metabolites, thus avoiding the formation of toxic intermediates and co-inheritance to progeny (Nützmann et al., 2016). Moreover, the metabolites produced by the gene clusters have a significant agronomic and human therapeutic importance, and in-depth studies are needed to develop the strategies for improving the bioproduction of target metabolites in heterologous hosts. Furthermore, plant genome mining through the genomic resources followed by functional genomic approaches together with metabolomics could reveal the existence of novel gene clusters involved in the production of high-value metabolites.

Herein, we discuss the evolutionary aspects of plant gene clusters along with their molecular features related to the organization of the clusters and regulatory mechanisms governing the organization. In addition, we elaborate the physiological role of cluster-derived metabolites in plant defense and other metabolic functions. Finally, we propose guiding principles toward the development of novel strategies related to metabolic engineering and synthetic biology by utilizing the repository of studies on plant gene clusters.

TABLE 1 | Details of the gene clusters present in the different plant species and diversity of metabolites produced along with their physiological function.

S.no	Species	Metabolite	Compound class	Chromosome (core genes)	Function	Tissue of expression	References
1	<i>Arabidopsis thaliana</i>	Thalianol	Triterpenes	5	Unknown physiological function, unregulated expression of cluster genes leads to dwarfing of plant. Modulate root microbiome content	Roots	Field and Osbourn, 2008; Field et al., 2011; Chen et al., 2019a
2	<i>A. thaliana</i>	Marneral	Triterpenes	5	Unknown physiological function, unregulated expression of cluster genes leads to dwarfing of plant	Roots	Field and Osbourn, 2008; Field et al., 2011
3	<i>Avena strigosa</i>	Avenacins	Triterpenes	1	Defense against pathogens	Roots	Li et al., 2021a
4	<i>Cucumis sativus</i>	Cucurbitacins	Triterpenes	6	Insect deterrent properties and possess medicinal value	Leaves and fruits	Shang et al., 2014
5	<i>Ricinus communis</i> and <i>Jatropha curcas</i>	Casbenes	Diterpenes	1	Possess medicinal value and used in treating cancers and HIV infection	Constitutive expression in leaves, roots and stems	King et al., 2014, 2016
6	<i>Zea mays</i>	2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA)	Hydroxamic acids	4	Defense related activities	Mainly expressed during seedling stages and in roots.	Frey et al., 1997
7	<i>Oryza sativa</i>	Momilactones	Diterpenes	4	Insect deterring properties and anti-fungal properties	Induced expression during pathogen attack	Shimura et al., 2007; Wang et al., 2011
8	<i>O. sativa</i>	Phytocassanes Oryzalides	Diterpenes	2	Defense related activities	Induced expression during pathogen attack	Swaminathan et al., 2009; Wu et al., 2011
9	<i>Lotus japonicus</i>	Linamarin Lotustralin	Cyanogenic glucosides	3	Herbivore deterrent activities	Above ground plant parts	Takos et al., 2011
10	<i>Sorghum bicolor</i>	Dhurrin	Cyanogenic glucosides	1	Herbivore deterrent activities	Above ground plant parts	Takos et al., 2011
11	<i>L. japonicus</i>	20-hydroxy-betulinic acid	Triterpene	3	Possible role in plant development and nodule formation	Elevated expression in roots and nodules	Krokida et al., 2013
12	<i>Solanum lycopersicum</i>	α -tomatine	Steroidal glycoalkaloid	7	Anti-pathogenic activity and toxic to humans	Elevated expression during pathogen attack	Itkin et al., 2013
13	<i>Solanum tuberosum</i>	α -solanine	Steroidal alkaloid glycoalkaloid	7	Anti-pathogenic activity and toxic to humans	Elevated during stress conditions	Itkin et al., 2013
14	<i>S. lycopersicum</i>	Mono & di terpenes	β -phellandrene Lycosantalanol	8	Involved in attracting pollinators and possible role in herbivore and fungal defenses	Localized expression in young leaves and trichomes	Matsuba et al., 2013, 2015; Zhou and Pichersky, 2020
15	<i>Papaver somniferum</i>	Noscapine	Benzylisoquinoline alkaloid	11	High value secondary metabolites	Stem and capsule	Guo et al., 2018
16	<i>P. somniferum</i>	Thebaine	Benzylisoquinoline alkaloid	11	High value secondary metabolites	Stem and capsule	Guo et al., 2018
17	<i>S. lycopersicum</i>	Falcarindiol	Fatty acid cluster	12	Elicited by pathogen attack	Above ground parts	Jeon et al., 2020
18	<i>Capsicum annuum</i>	Capsidiol	Sesquiterpene	2, 12	Pathogen induced production	Above ground parts	Lee et al., 2017

EVOLUTIONARY DYNAMICS OF PLANT GENE CLUSTERS

Genes that exist in close proximity on chromosomes, are often co-expressed (Elizondo et al., 2009). In prokaryotes, a set of non-homologous genes form clusters, which are generally referred as operons, and the genes in the operons are transcribed together to form a polycistronic messenger RNA (mRNA) to encode the proteins involved in a specific metabolic function (Jacob et al., 1960). The “Selfish operon model” describes that cluster arrangement of genes can improve HGT to other species, thereby increasing the chances of cluster survival (Ballouz et al., 2010). In fungi, the genes encoding enzymes involved in formation of β -lactam antibiotics and the biosynthesis of nitrate, proline, and galactose (GAL) occur as clusters with a similar pattern of co-expression (Nützmann et al., 2018). β -lactam antibiotic clusters might be of bacterial origin and could have been transferred to fungi through HGT (Liras and Martín, 2006; Slot, 2017). Similarly, GAL clusters originated differently in three species of yeast, whereas in *Saccharomyces cerevisiae* and *Candida sp.* GAL clusters are known to be originated independently through gene relocation. Comparably, in the case of *Schizosaccharomyces*, GAL cluster is found to have been acquired from *Candida* species through HGT (Slot and Rokas, 2010).

It is inferred that plant gene clusters are not evolved through HGT. Further, it has been shown that the cluster development in plants could have occurred through gene duplications, relocalization, neofunctionalization, or an independent evolution of genes toward the acquisition of specialized metabolism (Nützmann et al., 2016). To understand the evolution of plant gene clusters, Liu et al. (2020a) explored the genomes of the members of *Brassicaceae* for genome neighborhood (GNs) regions spanning around oxidosqualene cyclases (OSCs are key enzymes involved in sterol biosynthesis in plants) and identified that clade II OSCs were surrounded by cytochrome P450s (CYP450s) and acyltransferase genes possibly indicating the cluster organization. These GNs were found to be in highly dynamic chromosomal regions and lacked synteny toward the regions originated from whole genome duplication (WGD) in *Brassicaceae*, depicting an independent mode of evolution. Interestingly, functional characterization of GNs together with OSC in different species revealed that these putative clusters are equipped with a similar set of genes, even though they exhibited a diverse spatial and temporal expression (Liu et al., 2020a). Thalianol-producing gene cluster is known to be identical in *Arabidopsis thaliana* and *Arabidopsis lyrata*. Genome analysis of various *Arabidopsis* species revealed the evolution of thalianol cluster occurred before the divergence of *A. thaliana* and *A. lyrata*. In both the species, the cluster is organized with four core genes, whereas in *A. thaliana* three additional genes (*THAA2*, *THAR1* and *THAR2*) are required for producing thalianin, which might have occurred through a chromosomal inversion event, but in *A. lyrata* five genes [four core genes; *THAS*, thalianol hydroxylase (*THAH*), *THAO*, *THAA1* and one linked gene; *THAA2*] are responsible for epithalianin production (Liu et al., 2020b). In addition, similar clusters from *Capsella rubella* and *Brassica rapa* are

known to produce tirucalol derivatives (produced in buds) and euphol, respectively (Liu et al., 2020a). Boutanaev and Osbourn (2018) reported that transposable elements, such as miniature inverted-repeat transposable elements (MITEs), are present within the gene clusters of eudicots and they are known to play a predominant role in cluster formation by chromosomal rearrangements such as deletions, translocations, and inversions. Accordingly, triterpene clusters of *A. thaliana* are rich in transposable elements, which might have contributed to cluster formation through the segmental duplication of a committed step followed by an independent recruitment of tailoring enzymes (Field et al., 2011). In addition, transposable element-mediated genetic recombination and duplications led to the formation of sesterterpene-producing gene clusters in *A. thaliana* (Chen et al., 2019a). In addition, *Brassicaceae* members also possess pairs of terpene synthase and prenyl transferase on the genome that produces different sesterterpenes (Huang et al., 2017).

Local gene duplication could also contribute to the formation of gene clusters which are explicitly lineage specific (Schlöpfer et al., 2017). For example, SGA gene clusters emerged in *Solanum lycopersicum* (tomato) and *Solanum tuberosum* (potato) through a duplication event from the common ancestor. Similar orthologous clusters have been identified in eggplant and pepper. In pepper, few genes for SGA biosynthesis underwent deletions during the course of evolution, and as a result pepper plant produces steroidal saponins instead of SGAs (Itkin et al., 2013; Barchi et al., 2019; **Table 1**). In addition, terpene synthase gene cluster in *Solanum* species might have evolved through several segment duplications of terpene synthase and *cis*-prenyltransferase genes, though few genes in the cluster are found to be non-functional (Matsuba et al., 2013). In *Solanaceae*, three genes *AsAT1*, *AACSI*, and *AECH1* are characterized to be (occurring in multi-chromosome synteny regions) responsible for the production of medium-chain acyl sugars. These genes might have evolved through the insertion and segmental duplication events that occurred in a common ancestor before divergence (Fan et al., 2020).

In monocots, avenacin cluster and metabolite biosynthesis is highly specific to *Avena strigosa* (diploid oat), and this cluster has been reported to have evolved independently (Qi et al., 2004). Genome analysis of *A. strigosa* revealed the presence of avenacin cluster in sub-telomeric regions, with the presence of the early pathway genes near to the telomere, and the existence of terminal pathway genes away from the telomere. These typical positioning co-linearity of genes in the cluster is considered to avoid the deletions of terminal pathway genes, thus preventing the accumulation of toxic intermediates (Li et al., 2021a). On the other hand, maize DIMBOA gene cluster is also located at the tip (telomeric region) of the chromosome for facilitating an adaptive evolution and a coordinated regulation (Dutartre et al., 2012). Hydroxamic acids of maize (DIMBOA) are also produced in rye and wheat, and DIMBOA-producing genes are present on two different chromosomes in respective species without disrupting metabolite biosynthesis (Frey et al., 1997, 2009). Barnyard grass, a noxious weed of rice fields, has acquired orthologous gene clusters involved in the biosynthesis of momilactone and DIMBOA, respectively. In addition, barnyard grass is known

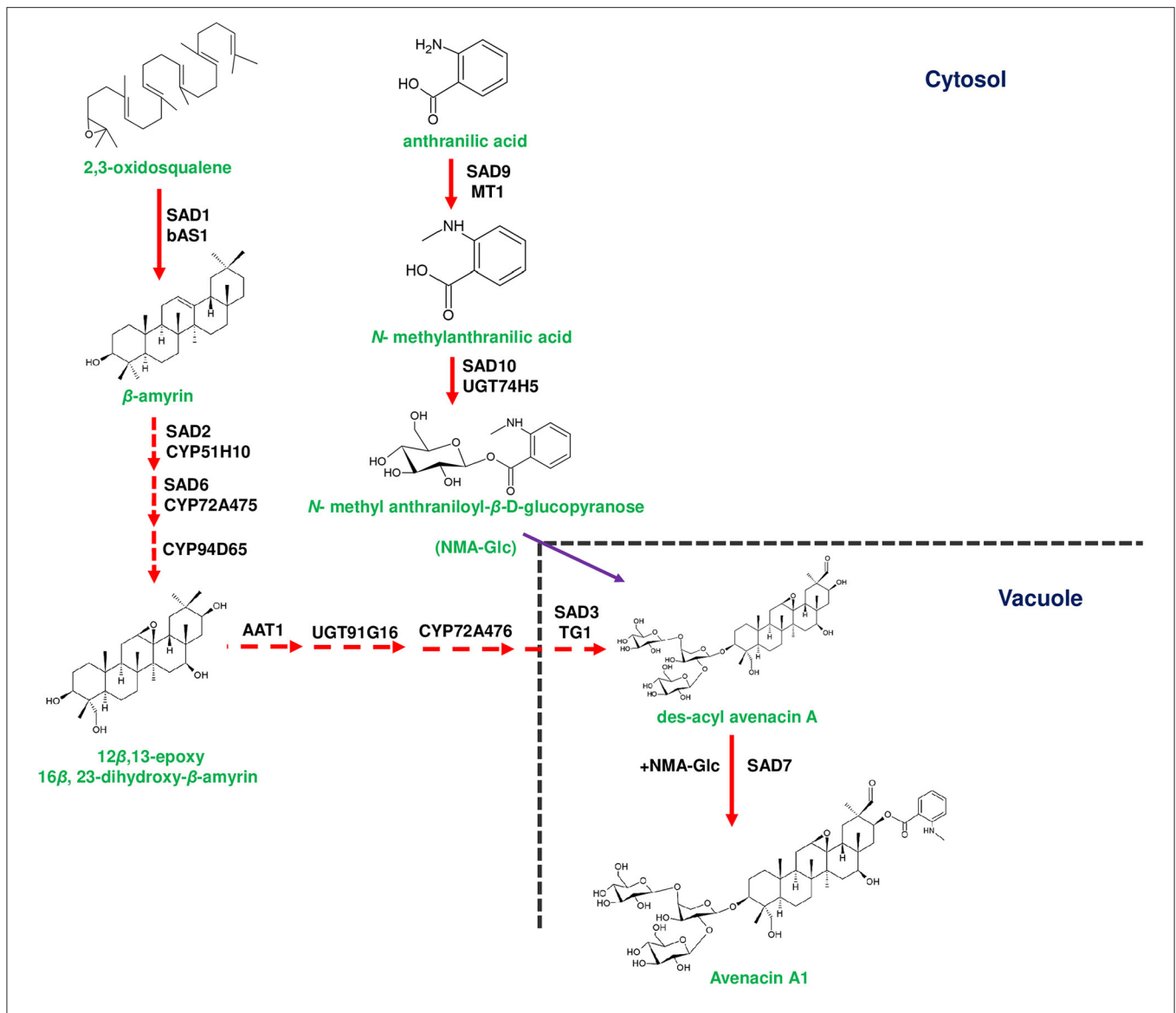


FIGURE 1 | Biosynthetic pathway of avenacin in the roots of *Avena strigosa*. Avenacin biosynthesis starts with the conversion of 2,3-oxidosqualene to β -amyrin by Sad1 (Saponin deficient 1). Conversion of β -amyrin to 12 β ,13-epoxy 16 β , 23-dihydroxy- β -amyrin that occurs in the cytosol is mediated by Sad2, Sad6 and CYP94D65, respectively. 12 β ,13-epoxy 16 β , 23-dihydroxy- β -amyrin is subsequently conjugated with one arabinose and two glucose moieties for the formation of des-acyl avenacin. Two glycosylation steps occur in the cytosol catalyzed by *A. strigosa* arabinosyltransferase (AsAAT1) and AsUGT91G16 (UDP-glucosyl transferase). The final glucose moiety is added by a unique vacuolar glycosyltransferase *A. strigosa* transglucosidase 1 (AsTG1). Des-acyl avenacin is acylated by Sad7 aided by acyl donors such as *N*-methyl anthranilate glucopyranose (NMA-Glc) to form avenacin A1, in the vacuole. NMA-Glc is formed from anthranilic acid through a two-step reaction catalyzed by Sad9 and Sad10, respectively. Genes of the plant metabolic gene cluster encoding enzymatic reactions are indicated in red arrows. Dotted arrows represents multi-step pathway.

to possess a quercetin producing gene cluster, which is highly upregulated during the interaction of the weed with the host (Sultana et al., 2019). Gene clusters of rice (phytocassanes and momilactones) might have evolved as an adaptive strategy to counter pathogenic invasions (Swaminathan et al., 2009). Miyamoto et al. (2016) suggested that the evolution of these two gene clusters could have occurred in the common ancestor of *Oryza* species before domestication. In addition, momilactone cluster might have evolved through an assembly of individual

genes into a physical proximity through duplications, whereas the phytocassane cluster might have already existed in the common ancestor of *Oryza* species but got lost in different lineages during the course of evolution (Miyamoto et al., 2016). Fascinating occurrence of the momilactone biosynthetic cluster in *Calohyphnum plumiforme*, a bryophyte species has opened up the speculations of convergent evolution of gene clusters for species survival in challenging environments (Mao et al., 2020; Zhang and Peters, 2020). Casbene diterpenoid producing gene

cluster in rice is specific to *Oryza* genus and have evolved through gene duplications (Zhan et al., 2020). Furthermore, an intact casbene production is observed in japonica cultivars (Medema et al., 2015) compared to indica cultivars of rice. This has been attributed to the natural selection of cluster in japonica varieties during the process of domestication to impart innate resistance against blight disease (Zhan et al., 2020). In addition, recent documentation of hydroxycinnamoyl tyramine producing gene cluster being specific to rice lineage and a high level of induced expression of the clustered genes during pathogen attack sheds light on the adaptive evolution of plant gene clusters in response to environmental cues (Shen et al., 2021).

In the case of cyanogenic glucoside clusters, parallel and independent evolution might have occurred in higher plants possessing the same scaffold of genes. *Lotus japonicus* and *Sorghum bicolor* are known to possess similar classes of genes in the cluster responsible for the production of plant-specific glucosides (Takos et al., 2011; **Table 1**). Intriguingly, an orthologous gene cluster of cyanogenic glucosides was also observed in white clover, a distant relative of *L. japonicus* (both belong to the subfamily *Papilionoideae*). They are considered to share a common ancestry, even though other members of *Papilionoideae* lost this cluster during the course of evolution as an adaptive strategy toward specific environmental niche (Olsen and Small, 2018). An *in silico* analysis using Plantismash revealed the presence of 10 orthologous gene clusters in the genomes of *Amaranthus cruentus* and *Amaranthus hypochondriacus* related to secondary metabolism. However, their *in planta* functional role is not conclusive (Ma et al., 2021). Comparative genome analysis of four genera of *Amaranthaceae* (*Amaranthus*, *Beta*, *Chenopidium*, and *Spinacia*) provided conclusive evidence about the co-occurrence of betalain pigment-producing genes on a specific chromosome, but in spinach (*Spinacia oleracea*) and quinoa (*Chenopodium quinoa*) additional copies of the genes were found to exist. This could be due to tandem gene duplication mechanisms (Ma et al., 2021).

Available reports on gene cluster evolution indicate that plants developed gene clusters to reprogram their functional attributes toward an adaptation to different ecological niches by recruiting genes through the duplication of primary metabolism genes and acquiring new functions to them. For instance, *Sad1* (saponin deficient), *Sad2* genes in oat avenacin cluster are recruited from sterol metabolism, and novel functions were acquired subsequently (Qi et al., 2004, 2006; **Figure 1; Table 2**). Sonawane et al. (2016) also reported the duplication events followed by the neofunctionalization of primary metabolite biosynthesis led to the evolution of cholesterol biosynthetic genes in *Solanum* species. Moreover, the formation of these gene clusters is facilitated by various chromosomal recombination events, the presence of transposable elements, and sub-telomeric positions of clusters, and these events further support the notion that gene clusters are dynamic and evolving rapidly. Further, these gene clusters are co-inherited to progeny as an environmental adaptation for functions such as development and defense responses. In addition, negative selection pressure against the accumulation of toxic intermediates could lead to the formation of gene clusters in plants (Li et al., 2021a). Based on the

observations related to the dynamic evolution of gene clusters, it is conclusive that plant genome is highly flexible and capable of associating non-homologous genes into a single coordinated cluster. Nevertheless, the lack of genome information of several plant species is a major constraint to reach a plausible conclusion on the evolution of gene clusters. It will be interesting to study the precise molecular mechanisms of evolutionary pressure that prompted plant genome plasticity. Genome mining of several plants through bioinformatics tools followed by analyzing the emergence of gene clusters in closely related and as well as distantly related species through phylogenomics, and comparative genomics approaches could help in understanding the evolutionary dynamics. Finally functional characterization of these clusters help to identify their *in planta* role as part of a major adaptive evolution.

ORGANIZATION OF PLANT GENE CLUSTERS

General Rules of Plant Gene Clusters

Plant gene clusters have been characterized in both monocots (phytoalexin clusters in rice, DIMBOA cluster in maize, and avenacin cluster in oat) and dicots (thalianol, marneral clusters of *Arabidopsis*, noscapine, and thebaine clusters of poppy; **Table 1**). The common principle of gene clusters is the occurrence of set of non-homologous genes producing a specific metabolite in physical proximity on a chromosome (Medema et al., 2015; **Table 1**). Recent chromosomal analysis by Nützmann et al. (2020) in *Arabidopsis* revealed that the active cluster regions occur in special local hotspot regions away from the heterochromatin region and nuclear periphery. Furthermore, a similar kind of organization has been reported in rice, tomato, and maize (Nützmann et al., 2020). Genes of the first committed pathway step followed by downstream tailoring enzymes are the typical components of plant gene clusters. The number of downstream tailoring enzymes in a cluster ranges from 3 to 12 depending on the complexity of different metabolic pathways (**Table 2**). A signature enzyme catalyzes the first step of the pathway, which outlines the class of metabolite to be produced, and this signature enzyme is assigned to draw the primary metabolite flux toward a more specialized metabolism (Nützmann et al., 2016; **Figures 1–6**). However, cyanogenic glucoside, SGA pathways of *L. japonicus*, and several other members of the *Solanaceae* deviate from the above rule by catalyzing the first step of the pathway through CYP450s (Takos et al., 2011; Itkin et al., 2013; **Figure 3**). The arrangement of genes in plant gene clusters differs significantly, for instance, noscapine and faltarindiol clusters possess non-functional genes whose role in the metabolism is yet to be deciphered while thebaine cluster is tightly packed and contains the genes responsible for the biosynthesis of thebaine from *R*-reticuline (Guo et al., 2018; Jeon et al., 2020; Li et al., 2020; **Table 1; Figure 5**). In addition, genome analysis of poppy revealed that noscapine and thebaine clusters occur separately on the same chromosome while *R*-reticuline-synthesizing genes were found to be loosely clustered (Guo et al., 2018; Li et al., 2020). A similar pattern of compact clustering is

TABLE 2 | Different classes of tailoring enzymes occurring in the plant gene clusters, including CYP450s, acyl transferases, glycosyl transferases, and alcohol dehydrogenases.

S. no	Cluster	CYP450s	Acyl transferases	Glycosyl transferases	Alcohol dehydrogenases	Other classes of enzymes	References
1	Thalianol	THAH-CYP708A2 THAD-CYP705A5, THAR1, THAR2	BAHD acyltransferases THAA1, THAA2, THAA3				Field and Osbourn, 2008; Liu et al., 2020a,b; Bai et al., 2021
2	Marneral	MRO-CYP71A16 CYP705A12					Field and Osbourn, 2008; Field et al., 2011
3	Avenacin	Sad2-CYP51H10 Sad6-CYP72A475, CYP72A476 CYP94D65	Sad7 acyl transferase (Serine carboxypeptidase like protein)	Sad3-glucosyl hydrolase glycosyl transferases- Sad10, AsAAT1 and UGT91G16		Sad9 Methyl transferase	Qi et al., 2006 and references related to avenacin biosynthesis mentioned in main text
4	Cucurbitacins	CYP81Q58, CYP89A140 CYP87D19 (within cluster) CYP87D20 (Chromosome1) CYP712D8 CYP88L2 CYP88L3 (Chromosome 3)	Acyl transferase (Csa6G088700)				Shang et al., 2014; Zhou et al., 2016
5	Casbenes	CYP76A14 CYP76A15 CYP76A16 CYP76A19 (<i>R. communis</i>) CYP71D495 CYP726A35 (<i>Jatropha carcus</i>)					King et al., 2014, 2016
6	Momilactones	CYP99A2 CYP99A3 CYP701A8 CYP76M14			OsMAS Short chain alcohol dehydrogenase		Shimura et al., 2007; Wang et al., 2011; De La Peña and Sattely, 2021
7	<i>Oryzalides</i> <i>Phytocassanes</i>	CYP71Z6 CYP71Z7 CYP76M7 CYP76M8					Swaminathan et al., 2009; Wu et al., 2011
8	Cyanogenic glucosides (<i>L.japonicus</i>)	CYP736A2 CYP79D3/D4		UGT85K2 (UDP-glucosyltransferase)			Takos et al., 2011
9	Cyanogenic glucosides (<i>S.bicolor</i>)	CYP71E1 CYP79A1		UGT85B1 (UDP-glucosyltransferase)			Takos et al., 2011
10	20-hydroxy-betulinic acid	CYP71D353					Krokida et al., 2013
11	Steroidal glucoside alkaloid cluster	GAME6 GAME7 GAME4 GAME8 (CYP72)		GAME1, GAME17, GAME18, GAME2 (UDP-glycosyl transferases)	GAME25 (Short chain dehydrogenase)	GAME11 (Putative dioxygenase) GAME12 (Transaminase)	Itkin et al., 2013; Sonawane et al., 2018

(Continued)

TABLE 2 | Continued

S. no	Cluster	CYP450s	Acyl transferases	Glycosyl transferases	Alcohol dehydrogenases	Other classes of enzymes	Reference
12	Falcarindiol cluster					Acetylase Fatty acid desaturase Decarboxylase	Jeon et al., 2020
13	Capsidiol					5-epi-aristolochene synthase (EAS) 5-epi-Aristolochene 3-hydroxylase, (EAH)	Lee et al., 2017

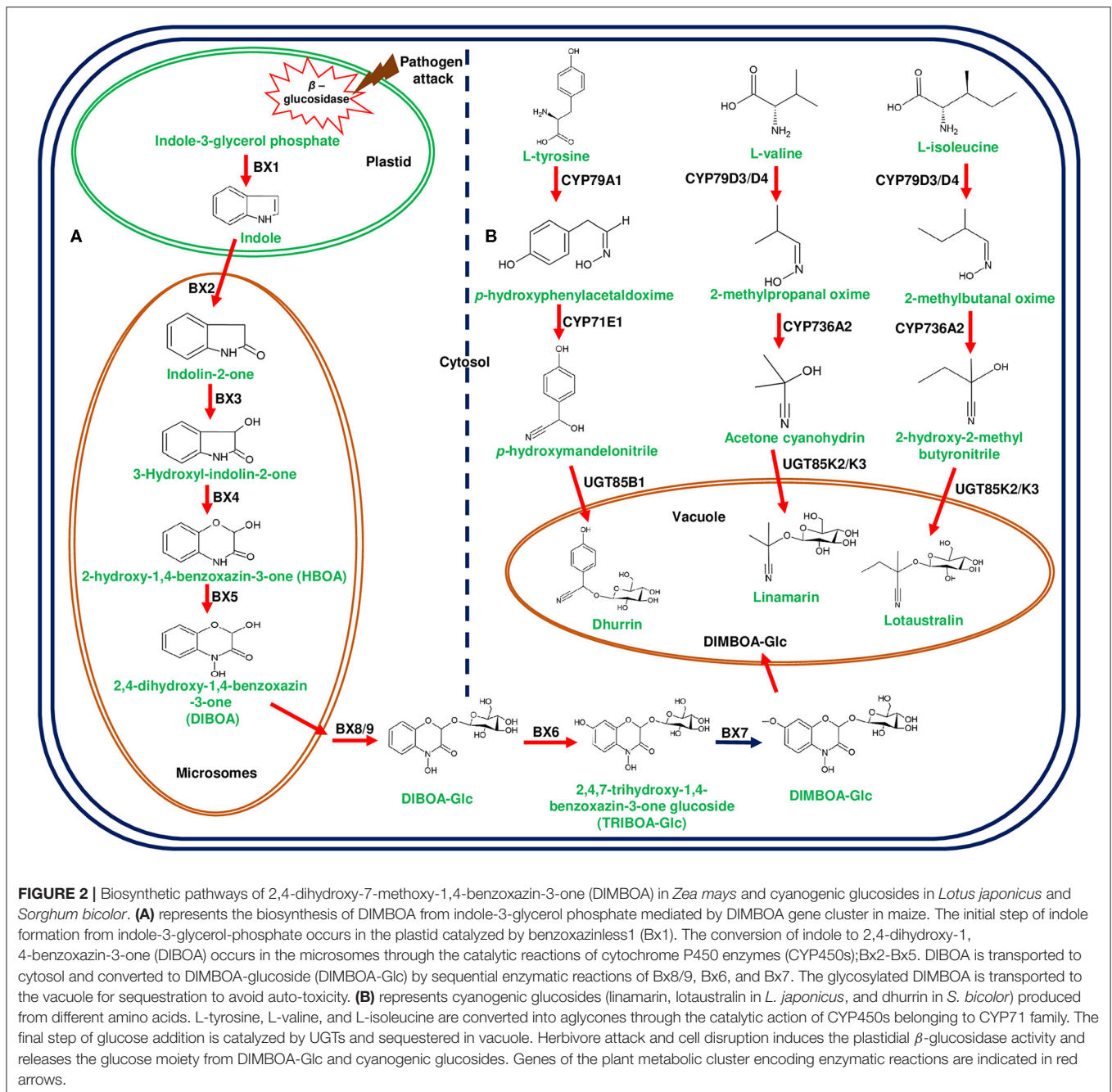
also observed in oat avenacin cluster (Qi et al., 2004; Li et al., 2021a). Recently, elucidation of complete cluster organization of avenacins revealed that all genes are organized in a co-linear manner in relation to the biosynthetic steps of the pathway, and ten genes are arranged on the end of long arm of the chromosome 1 while the other two glycosyl transferases (GTs) are present in proximal scaffold (Li et al., 2021a).

Few metabolite pathway genes are known to exist adjacently in a cluster, and the three genes, viz., *Sad7*, *Sad9*, and *Sad10* are found to form an acylation module, which is responsible for acylating avenacins (Mugford et al., 2013; Nützmänn et al., 2016). In contrast to the gene arrangement of avenacin cluster, *Glycoalkaloid Metabolism 7 (GAME7)* and *GAME8* genes of SGA biosynthesis are found to be located away from the cluster in both tomato and potato (Itkin et al., 2013). *Berberine bridge enzyme (BBE)* and *tetrahydroprotoberberine N-methyltransferase (TNMT)* genes of noscapine biosynthesis are known to lack a clustered arrangement due to their additional functions in sanguinarine biosynthesis (Guo et al., 2018). In addition, two CYP450s (*THAR1* and *THAR2*) involved in the production of thalianin (the final product derived from thalianol) are located away from the thalianol cluster (Liu et al., 2020a,b). In few cases, an identical set of homologous genes are known to produce slightly different compounds. For instance, gene clusters of tomato and potato produce two different SGAs, viz., α -tomatine and α -solanine, respectively (Itkin et al., 2013; **Figure 3B**). Conversely, the orthologous genes of *L. japonicus* and *S. bicolor* encode a similar class of enzymes *albeit* producing different cyanogenic glucosides (Takos et al., 2011; **Figure 2B**). In addition, a few plants are known to possess multifunctional gene clusters equipped with the genes responsible for producing different compounds, for example, rice phytoalexin gene cluster contain the genes involved in the bioproduction of both phytocassanes and oryzalides (Swaminathan et al., 2009; **Figure 4**). Similarly, tomato terpene gene cluster is known to contain the genes producing mono- and di-terpenes such as lycosantalol and β -phellandrene (Matsuba et al., 2015). In yet another situation, a few genes of monoterpene indole alkaloid (MIA) pathway are arranged together in different scaffolds in *Catharanthus roseus*. Interestingly, similar small clusters have also been observed in *Gelsemium sempervirens*, which produces the similar MIAs depicting the conserved nature of genes in different plant families (Kellner et al., 2015; Franke et al., 2019). A gene cluster

producing medium-chain acyl sugars in the *Solanaceae* occurs adjacent to the SGA gene cluster on chromosome 7. However, the evolutionary aspects regarding the co-localization of both clusters are yet to be deciphered (Fan et al., 2020). Despite the occurrence of several complicated organizational patterns of gene clusters in different plant species, the genes responsible for the bioproduction of specialized metabolites are found to be expressed in a coordinated manner independent of their occurrence within or outside the cluster. Dhurrin producing gene cluster of *S. bicolor* is found to be stringently regulated and forms metabolon for channeling intermediates, but the enzymes of avenacin and DIMBOA biosynthesis occur in different cellular compartments highlighting the complexity of cluster expression, channeling of the metabolic intermediates (Frey et al., 2009; Takos et al., 2011; Li et al., 2021a). Along with the genes responsible for metabolite bioproduction, a few clusters are also equipped with transporters and other regulatory elements. For example, the dhurrin cluster contains a co-expressed multi-antimicrobial extrusion protein (MATE) transporter gene (multidrug and toxic compound extrusion), which could bind and transport dhurrin (Darbani et al., 2016). Additionally, a cofactor synthase *OsPDX3* of rice, which is involved in the biosynthesis of hydroxycinnamoyl tyramine, occurs within the cluster (Shen et al., 2021). Hen-Avivi et al. (2016) reported the presence of a regulatory gene within the vicinity of the β -diketone cluster in wheat. Considering all the above mentioned examples, at this point of time, it is indeed difficult to draw firm conclusions related to the general organization and behavior of plant gene clusters. An in-depth analysis could reveal not only their functional role but also indicate novel evolutionary strategies of these gene clusters in conferring adaptability to the ever-changing environmental conditions.

Significance of First Pathway Step in Metabolite Production

In most of the gene clusters identified so far, the first step of the pathway is catalyzed by a specific class of enzymes (e.g., terpene cyclases and terpene synthases) that divert the flux of primary metabolites to synthesize the cluster-specific specialized metabolites. For instance, OSC of thalianol (*THAS*; *thalianol synthase*) and marneral (*MRNI*; *marneral synthase*) clusters initiate metabolite bioproduction in *Arabidopsis* by converting 2,3-oxidosqualene, a branch point intermediate to a different set



of metabolites. *THAS* and *MRN1* belong to the clade II OSC of *Arabidopsis*, and other enzymes belonging to this clade are also found to be flanked and are co-expressed (Field and Osbourn, 2008; Field et al., 2011; Liu et al., 2020a,b). Interestingly, oat-specific *A. strigosa* β -amyryn synthase (*AsbAS1*) gene is similar to cycloartenol synthase (has a role in sterol precursor production) and mediate the conversion of 2,3-oxidosqualene into β -amyryn (Haralampidis et al., 2001; Qi et al., 2004; **Figure 1**). OSC encoded by *bitterness* (*Bi*) gene of the cucurbitacin cluster catalyzes the first step to produce cucurbitadienol, and *Bi* is known to be highly conserved across cucurbits (Shang et al., 2014; **Figure 3A**).

Additionally, *AMY2* (catalyzes the first step of 20-hydroxybetulinic acid synthesis) in *L. japonicus*, encodes an unusual OSC to produce dihydro-lupeol along with a relatively lower amounts of β -amyryn (Krokida et al., 2013; **Figure 6A**).

Diterpene synthases of gene cluster of *Ricinus communis* are involved in the diversion of geranylgeranyl pyrophosphate (GGPP) pool from the primary metabolism to the synthesis of casbene oxidases (King et al., 2014). Similarly, phytoalexins produced by diterpene synthases of *O. sativa* (rice) share an analogy with the enzymes of gibberellin biosynthesis (Wilderman et al., 2004; Zhang and Peters, 2020). Interestingly,

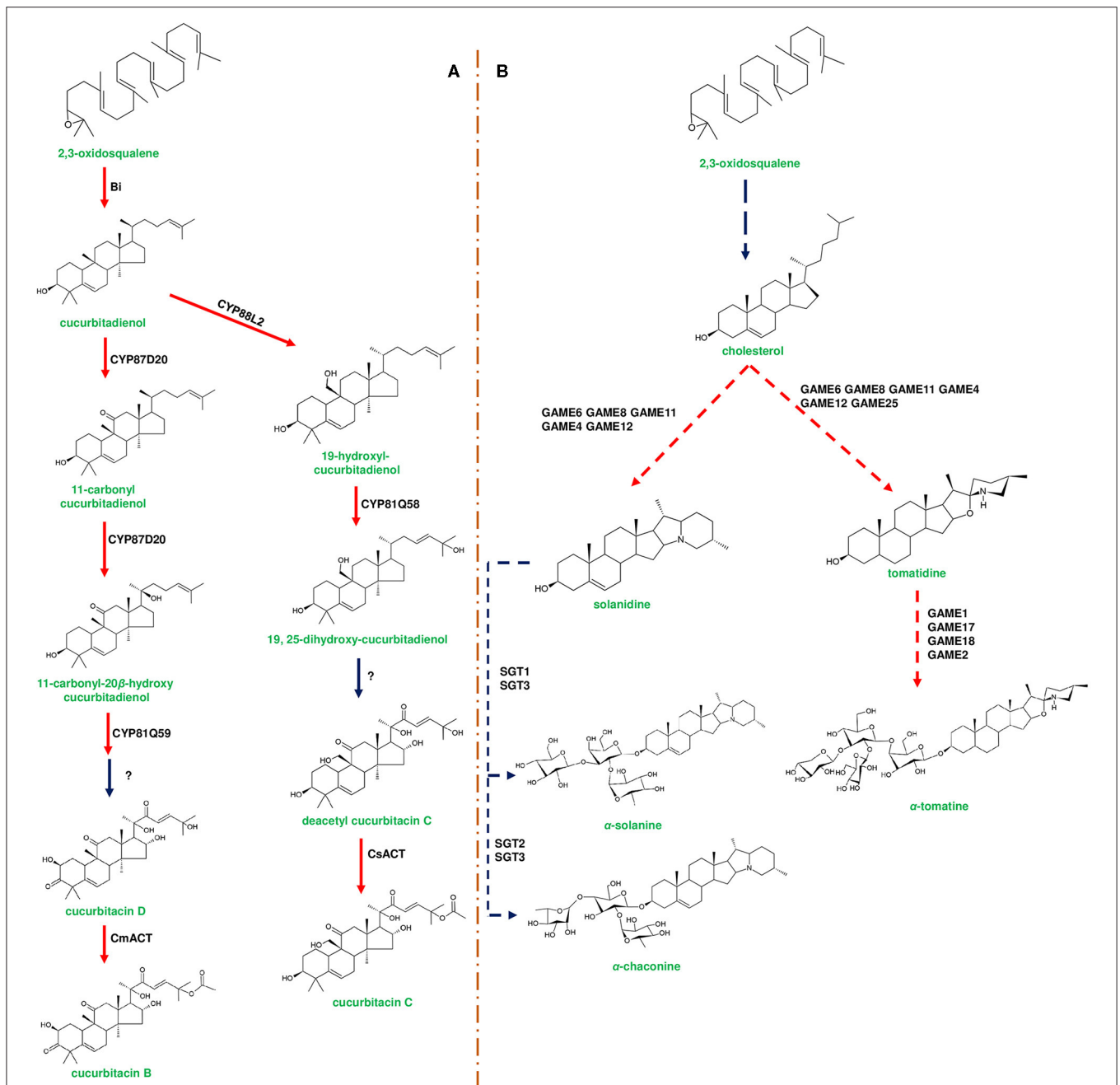
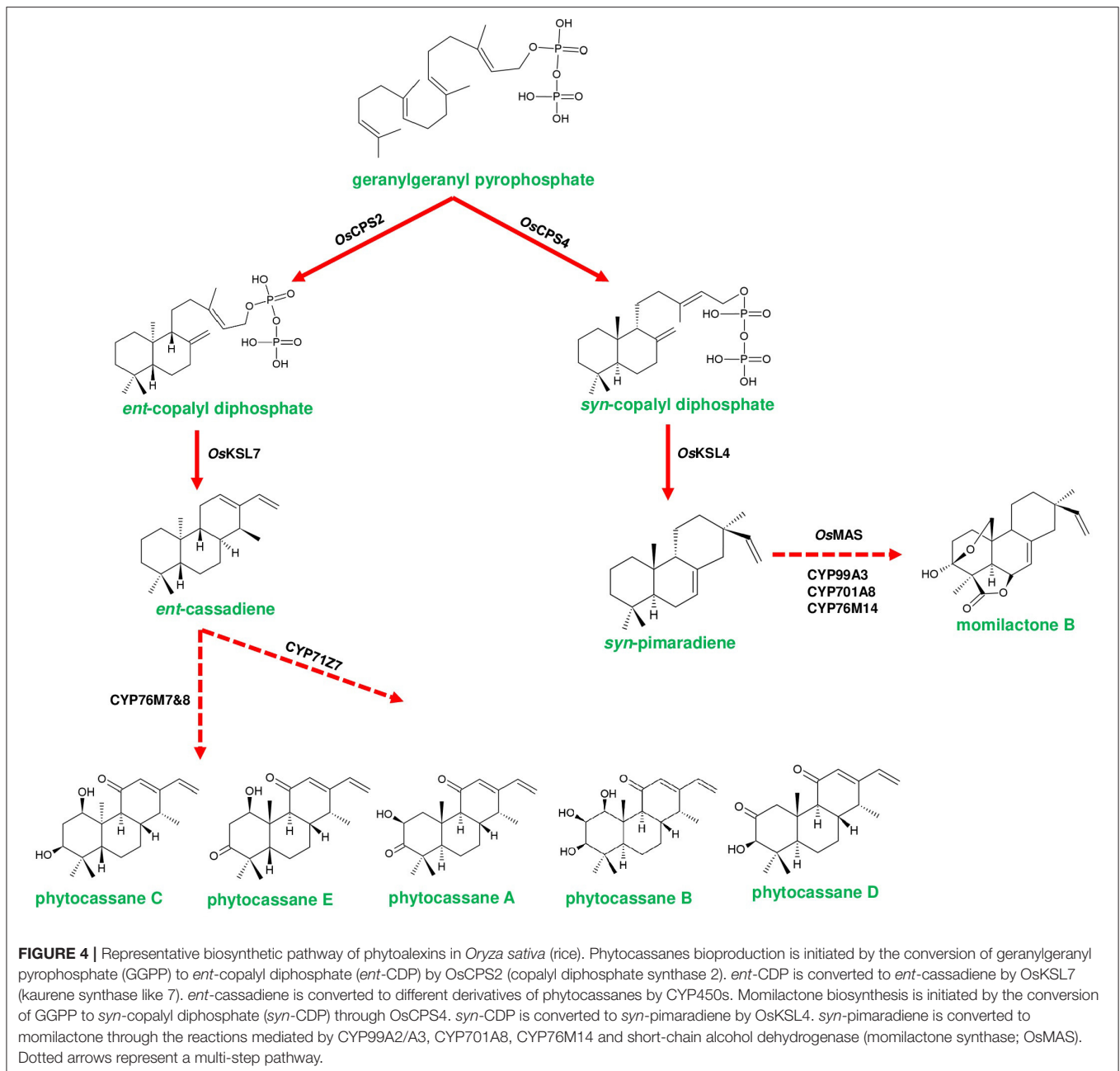


FIGURE 3 | Representative biosynthetic pathway of cucurbitacins in *Cucurbitaceae* and steroidal glycoalkaloids (SGAs) in *Solanaceae*. **(A)** represents the biosynthesis of cucurbitacins through enzymes encoded by the cucurbitacin gene cluster. The first step in cucurbitacin biosynthesis occurs by the enzymatic conversion of 2,3-oxidosqualene to cucurbitadienol mediated by enzyme Bitterness (Bi) belonging to oxidosqualene cyclase family. Cucurbitadienol is converted to cucurbitacin C (CuC) in *Cucumis sativus* mediated by CYP450s and acyltransferase [*C. sativus* acyltransferase (CsACT)]. In *Cucumis melo*, cucurbitadienol is converted to cucurbitacin B (CuB) by CYP450s and *C. melo* acyl transferase (CmACT). Modification of the backbone in different *Cucumis* species is attributed to the activity of CYP450s. **(B)** represents the biosynthesis of SGAs from cholesterol using the enzymes encoded by the SGA gene cluster. α -solanine and α -chaconine are the signature metabolites of potato (*Solanum tuberosum*). In potato, cholesterol is converted to solanidine by catalytic reactions of GAME8 (glycoalkaloid metabolism), GAME6, GAME11, and GAME12. Further, solanidine is converted to α -solanine and α -chaconine by Sterol alkaloid glycosyl transferases (SGT1, SGT2, SGT3). α -tomatine is produced in *Solanum lycopersicum* (tomato) through a similar set of genes present in the SGA gene cluster. Tomatidine is formed by the modification of cholesterol mediated by GAME6, GAME8, GAME11, GAME4, GAME12, and GAME25, respectively. Tomatidine is conjugated with four sugar moieties [one moiety each of galactose and xylose and two moieties of glucose] by UDP-glycosyl transferases. GAME1, GAME17, GAME18, and GAME2, respectively. Genes of the plant metabolic cluster encoding enzymatic reactions are indicated in red arrows. Dotted arrows represent a multi-step pathway.



a tryptophan synthase homolog [*benzoxazinless1 (Bx1)*] of maize diverts the flux of indole-3-glycerol phosphate from tryptophan synthesis to indole for the biosynthesis of benzoxazinoids, and *Bx1* is considered to have evolved independently in monocots through duplication and neofunctionalization (Frey et al., 1997; **Figure 2A**). In addition, the identification of orthologous *Bx1* gene in dicots implicates the convergent evolution of benzoxazinoid production across the plant kingdom (Schullehner et al., 2008; Dick et al., 2012). Tomato faltarindiol cluster (a modified fatty acid) involved in the biosynthesis of specialized metabolite was shown to be initiated by acetaldehyde enzyme catalyzing the conversion of the primary metabolite

linoleic acid to crepenylic acid (Jeon et al., 2020; **Figure 6B**). There are a few other gene clusters in which the first step of the metabolic pathway is not catalyzed by any signature enzyme. An example of this kind of gene cluster function is found in *L. japonicus*, where cyanogenic glucoside bioproduction is initiated by CYP450 (CYP79D3/D4), which catalyzes the conversion of amino acids (L-valine, and L-isoleucine) into respective oximes (2-methylpropanal oxime, 2-methylbutanal oxime). These oximes are subsequently converted to linamarin and lotaustralin (Takos et al., 2011; **Figure 2B**). CYP79 class is found to be highly conserved among higher plants and is involved in the conversion of amino acids to oximes

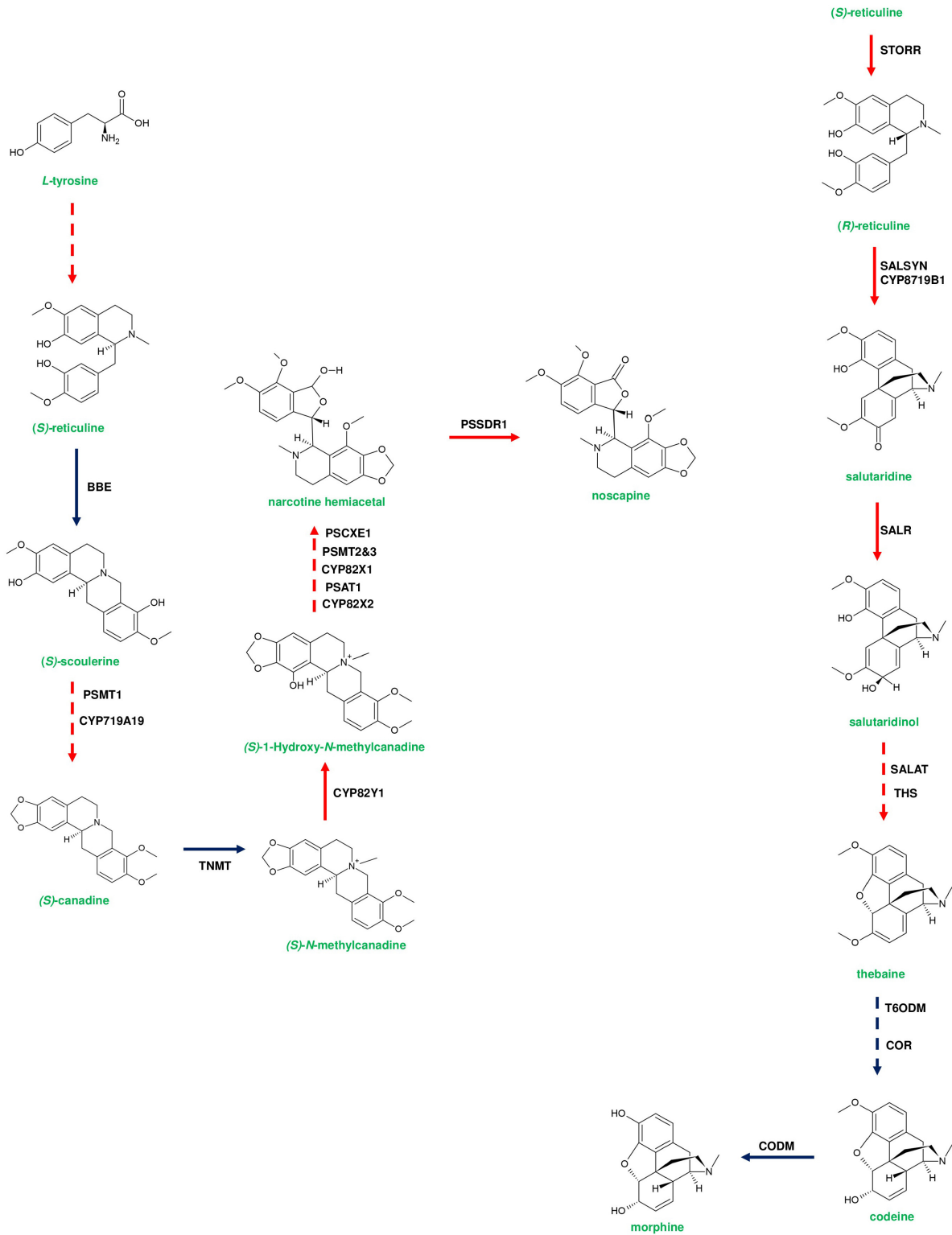


FIGURE 5 | Representative pathway showing the biosynthetic events of benzylisoquinoline alkaloids (BIAs) in *Papaver somniferum*. Genes responsible for (S)-reticuline biosynthesis from L-tyrosine are loosely clustered in the poppy genome. (S)-reticuline is converted to noscapine by a set of enzymes; berberine bridge enzyme (Continued)

FIGURE 5 | (BBE), *P. somniferum* methyltransferase 1 (PSMT1), Tetrahydroprotoberberine N-methyltransferase (TNMT), *P. somniferum* acyltransferase 1 (PSAT1), PSMT2&3, *P. somniferum* carboxylesterase 3 (PSCXE3), and *P. somniferum* shortchain dehydrogenase 1 (PSSDR1). BBE and TNMT genes are not present in the noscapine gene cluster due to their additional function in sanguinarine biosynthesis. Thebaine-producing genes are clustered and occur adjacent to the noscapine gene cluster. Thebaine biosynthesis is initiated from (*R*)-reticuline. (*S*)-reticuline is converted to (*R*)-reticuline by (*S*)- to (*R*)-reticuline (STORR) enzyme. The formation of thebaine occurs through the activity of salutaridine synthase (SALSYN), salutaridine reductase (SALR), salutaridinol-7-*O*-acetyl transferase (SALAT), and thebaine synthase (THS). Genes involved in morphine biosynthesis are not clustered. Thebaine is converted to codeine by two subsequent steps catalyzed by thebaine 6-*O*-demethylase (T6ODM), and codeinone reductase (COR). Finally, codeine is converted to morphine by Codeine3-*O*-demethylase (CODM). Genes of plant metabolic gene cluster encoding enzymatic reactions are indicated in red arrows. Dotted arrows represent a multi-step pathway.

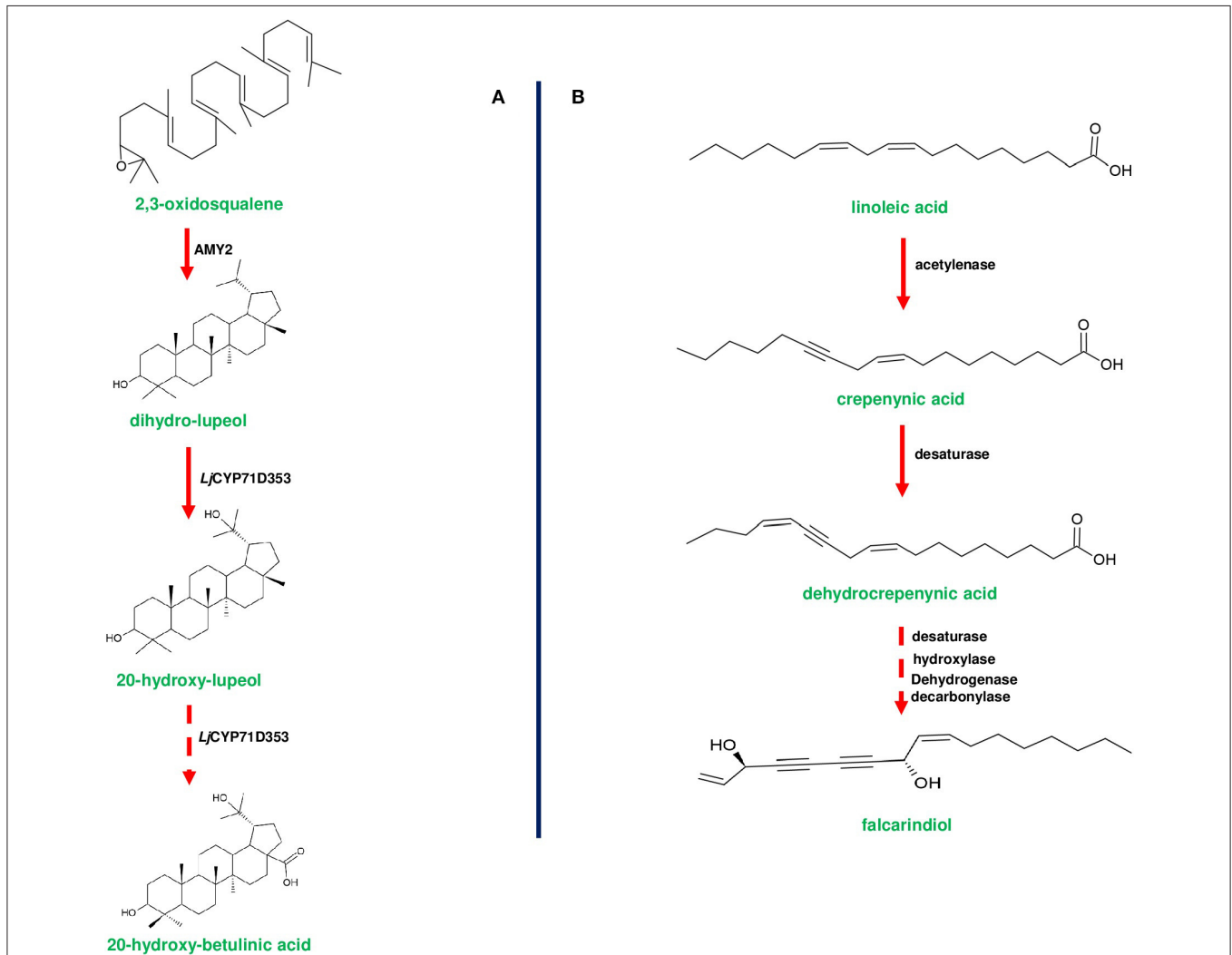


FIGURE 6 | Biosynthetic pathways of 20-hydroxy-betulinic acid in *Lotus japonicus* and anti-pathogenic fatty acid falcarindiol in tomato. **(A)** shows the biosynthesis of 20-hydroxybetulinic acid from 2,3-oxidosqualene by the enzymes encoded by a triterpene producing gene cluster in *Lotus japonicus*. AMY2 (β -amyrin synthase) converts 2,3-oxidosqualene into a unique metabolite dihydro-lupeol. Subsequently dihydro-lupeol is converted to 20-hydroxy-betulinic acid by two consecutive reactions catalyzed by *L. japonicus* CYP71D353. **(B)** represents the bioproduction of falcarindiol from the primary metabolite linoleic acid by the enzymes of anti-pathogenic fatty acid gene cluster in *S. lycopersicum* (tomato). Linoleic acid is converted to crepenynic acid by acetylenase enzyme encoded by *Solyc12g100240* gene. Crepenynic acid is converted into dehydrocrepenynic acid by a desaturase (*Solyc12g100250*). Subsequently, dehydrocrepenynic acid is converted to falcarindiol by a set of desaturase, hydroxylase, and decarboxylase. Dotted arrows represent a multi-step pathway.

for producing defense compounds such as glucosinolates and camalexin (Halkier et al., 2002; Takos et al., 2011). Similarly, *GAME7* encoding CYP72 class CYP450 modifies cholesterol to 22-hydroxycholesterol for SGA production in *Solanum* species (**Figure 3B**). Interestingly, *GAME7* was reported to be 7,880 kb

away from the SGA cluster on chromosome 7 (Itkin et al., 2013; Cárdenas et al., 2016; Nützmann et al., 2016). Yet another interesting inference has been made in *Solanum* species where the biosynthetic machinery of cholesterol, the precursor for SGAs is evolved from phytosterol pathway through gene duplications

and also *GAME12* of SGA pathway is likely to be evolved from gamma amino butyric acid transaminases (*GABA2*) to impart nitrogen into the steroidal backbone (Sonawane et al., 2016, 2020). In poppy, a loosely arranged gene cluster (seven genes, spanning a distance of 5 Mb across the chromosome) diverts the flux of tyrosine toward (S) - reticuline biosynthesis, which is the precursor for the biosynthesis of benzylisoquinoline alkaloids (BIA). On the other hand, *P. somniferum methyltransferase 1 (PSMT1)* and (S)- to (R)-reticuline (*STORR*) mediate the committed step in noscapine and thebaine biosynthesis, and these two genes are found to be arranged in proximity within their respective clusters (Winzer et al., 2015; Guo et al., 2018; Li et al., 2020; **Figure 5**).

Tailoring Enzymes of Plant Gene Clusters: Importance of Cytochrome P450s

Plant gene clusters possess various tailoring enzymes for modifying the backbone of signature metabolite produced through the first catalytic reaction. These enzymes include CYP450s, acyl transferases, UDP-glycosyl transferases (UGTs), short-chain alcohol dehydrogenases, transaminases, and decarboxylases (**Table 2**). CYP450s represent the largest enzyme family playing a significant role in the structural diversification of terpene scaffolds (Ghosh, 2017). Certain classes of terpene synthases and CYP450s (TPS/CYP450) form non-random pairing and are distributed among different gene clusters in both eudicots and monocots (Boutanaev et al., 2015). Boutanaev et al. (2015) reported that TPS/CYP450s pairs might have an independent evolutionary origin among different families of angiosperms. In *Arabidopsis*, triterpene cyclases (TTC) and CYP71 clan (CYP705 family) exist as pairs in both thalianol and marneral clusters, and it has been observed that TTC/CYP71 pairs are duplicated from a single founder cluster and then both clusters independently recruited another CYP71 family genes, which include *THAH* of thalianol, marneral oxidase (*MRO*) of marneral clusters to complete the cluster organization (Field et al., 2011; **Table 2**). Two CYP450s of thalianol cluster are involved in the conversion of thalianol to (-)-16-keto-3 β ,7 β ,15-thaliantriol, and this metabolic intermediate is converted to thalianin upon subsequent reactions catalyzed by four gene products (two genes encoding BAHD acyl transferases: *THAA1*, *THAA2*, and other two oxido-reductases: *THAR1*, *THAR2*), a key metabolite involved in modulating microbiome of the *Arabidopsis* roots (Huang et al., 2019). Further two oxido-reductases, *THAR1*, *THAR2* are unlinked to the thalianol cluster and can also act on other triterpenes of *Arabidopsis* such as arabidiol and tirucalladienol resulting in their respective oxidative products (Huang et al., 2019; Liu et al., 2020a,b).

CYP71D353 belonging to the triterpene cluster of *L. japonicus* catalyzes the final steps of converting dihydro lupeol to 20-hydroxy-betulinic acid, and this enzyme is found to be phylogenetically similar to *MRO* (CYP71A16; marneral oxidase) of marneral cluster belonging to CYP71 clan CYP450s (Krokida et al., 2013; **Figure 6A**). Additionally, it is reported that CYP71 clan enzymes can oxidize various triterpenes,

such as, lupeol, β -amyrin, and α -amyrin, leading to more specialized metabolites (Yasumoto et al., 2016). TTC/CYP81Q58 pair along with CYP88L2 and *Cs sativus* acyl transferase (*csACT*) accounts for the bioproduction of cucurbitadienol and its further conversion to cucurbitacin C (CuC) in *Cucumis sativus* (Shang et al., 2014). Interestingly, the other two *Cucumis* species (musk melon and water melon) do not produce CuC due to non-functional CYP88L2, instead they produce cucurbitacin B (CuB) and cucurbitacin E (CuE) through the catalytic activity of CYP87D20 and CYP81Q59, respectively (Zhou et al., 2016; **Figure 3A**). Moreover, CYP450s and tissue specific regulators involved in the biosynthesis of cucurbitacins were differentially expressed in both wild and domesticated species of *Cucumis*, indicating the operation of environmental selection force on cluster organization (Zhou et al., 2016).

Rice phytoalexin gene cluster (phytocassanes and oryzalides) has a combination of diterpene synthase genes with CYP450s of CPY71 clan (TPS/CYP71). GGPP is converted into *ent*-copalyl diphosphate (*ent*-CDP) by CPS2, then two kauerene synthase like genes (*KSL6* and *KSL7*) produce *ent*-isokaurene and *ent*-cassadiene from *ent*-CDP, which are the immediate precursors for oryzalides and phytocassanes, respectively (Wu et al., 2011; **Figure 4**). CYP71Z6 mediates primarily the biosynthesis of oryzalides, and on the other hand, CYP76M7 & CYP76M8 and CYP71Z7 catalyze subsequent steps of phytocassane biosynthesis (Swaminathan et al., 2009; Wu et al., 2011; **Figure 4**). Similarly, rice momilactone cluster contains CDP synthase (*OsCPS4*), kaurene synthase like (*KSL4*), CYP99 family (*CYP99A2/CYP99A3*) genes, and dehydrogenase (*OsMAS*), which are collectively involved in the production of momilactone (Shimura et al., 2007; Wang et al., 2011; **Figure 4**). The elucidation of the momilactone pathway by De La Peña and Sattely (2021) led to the identification of two other CYP450s (CYP701A8 and CYP76M14) that are responsible for the final conversion of momilactone A to momilactone B (De La Peña and Sattely, 2021). Furthermore, Kitaoka et al. (2021) proposed a novel biosynthetic route of momilactone biosynthesis through deciphering the role of CYP76M8, which converts *syn*-pimaradien-19-al to 6 β -hydroxy-*syn*-pimaradien-19-al, a key intermediate in momilactone production. Afterward, this metabolite gets converted to momilactone through subsequent steps catalyzed by *OsMAS1* or *OsMAS2* and CYP701A8, respectively (Kitaoka et al., 2021). CYP76M8 is located in the phytocassane cluster, which is a close relative of CYP76M7. However, the functioning of CYP76M8 together with momilactone cluster genes for metabolite biosynthesis indicates the interdependent evolution of gene clusters as a selective advantage (Kitaoka et al., 2021). The promiscuous activity of CYP76M8 is further supported by its ability to hydroxylate several diterpene metabolites of rice (Wang et al., 2012). Moreover, the bryophyte *C. plumiforme* possess four genes *CpDTC1/HpDTC1*, *CpMAS*, *CpCYP970A14*, and *CpCYP964A1* related to momilactone production, and the complete biosynthetic pathway is yet to be elucidated (Mao et al., 2020; Zhang and Peters, 2020).

CYP726A14 and CYP726A15 of casbene cluster in *R. communis* perform the conserved step reaction by oxidizing casbene and neocembrene at the fifth position, and the oxidation of the fifth carbon atom is conserved among the casbene-derived metabolites distributed across *Euphorbiaceae* (King et al., 2014). In *Jatropha curcas*, CYP726A35, orthologous to the CYP726A18 oxidizes the fifth position of casbene metabolites (King et al., 2016). Few other CYP450s (CYP726A19, CYP71D495) in the clusters of *R. communis* and *J. curcas* catalyzes the oxidation of casbene and its derivatives for further metabolic diversification (King et al., 2014, 2016). Recently, Zhan et al. (2020) identified a similar 5,10-diketo-casbene producing gene cluster (*OsTPS28*, *OsCYP71Z2*, and *OsCYP71Z21*) on the seventh chromosome of rice while *OsTPS28* converts GGPP to casbene and the other two CYP450s converts casbene to 5,10-diketo-casbene (Zhan et al., 2020).

Oat avenacin cluster possesses an unique combination of TTC and CYP51 (*AsCyp51H10*) encoding enzymes of the first two steps of avenacin biosynthesis (Figure 1). CYP51 family has been shown to play a specific role in sterol biosynthesis, but it has been demonstrated that *AsCyp51H10* was recruited from ancient CYP51 family and acquired new functions in avenacin metabolism (Haralampidis et al., 2001; Qi et al., 2006; Figure 1). On the other hand, CYP51 gene of sterol biosynthesis in *Solanum* species performs additional function in cholesterol biosynthesis, which is a precursor of SGAs (Sonawane et al., 2016). Moreover, oat avenacin cluster has been proposed to have evolved independently and orthologous genes and metabolite production were not found in other monocots (Qi et al., 2004). Meanwhile, similar pairing of TTC/CYP51 is observed in a cereal member, viz., *Brachypodium distachyon*, its functional similarity with the avenacin cluster is still found to be inconclusive (Boutanaev et al., 2015). Along with *CYP51H10*, three more CYP450s have been shown to be associated with avenacin biosynthesis, out of which two CYP450s belong to the CYP72 clan. CYP72A475 and CYP72A476 participate in C-21- β -hydroxylation and C-30 aldehyde group addition to β -amyrin scaffold, respectively (Leveau et al., 2019; Li et al., 2021a). CYP94D65 is responsible for adding a hydroxyl group at C-23 position of β -amyrin scaffold (Li et al., 2021a). In a comparable way, CYP71C family genes (*Bx2* to *Bx5*) of maize involved in DIMBOA biosynthesis are of monophyletic origin and homologous genes were not observed in other monocots other than *Bx2* (Frey et al., 2009; Figure 2A). These set of genes might have evolved before divergence of *Triticeae* and *Panicoideae* (Frey et al., 2009).

The non-random pairing of TPS and CYP450s were also present in the genomes of the members of *Lamiaceae* (Lichman et al., 2020). For instance, the genome assembly of *Tectona grandis* revealed the presence of 41 terpene synthase genes in 14 tandem clusters, and 20 TPS together with 31 CYP450s were physically clustered in the genome. *TPS-c/CYP76AH31* combination of *T. grandis* is involved in diterpene biosynthesis displaying a high percentage of homology toward *Salvia miltiorrhiza* *SmCPS1/CYP76AH12* responsible for the production of tanshinones (Zhao et al., 2019). In the case of *S. miltiorrhiza* *SmCPS1* and *SmCPS2*, genes were

flanked with CYP76AH subfamily genes involved in tanshinone biosynthesis, whereas these clusters might have formed from the ancestral duplication event of *CPS/CYP76AH* pair (Xu et al., 2016). Similarly, in *Salvia splendens* genome, eight clusters of TPS/CYP450 combinations were found, even though the functional role of these combinations are yet to be deciphered (Dong et al., 2018). The occurrence of gene pairs of terpenoid biosynthesis is not restricted to TPS and CYP450 combinations alone in *Brassicaceae*. The distinct pairs of prenyl transferase and terpene synthases were spread across the genome of several members, which are responsible for the biosynthesis of sesterterpenoids and these pairs might have evolved from a common ancestral pair via duplication and functional divergence (Huang et al., 2017). It has been reported that the construction of Lavender (*Lavandula angustifolia*) genome assembly led to the identification of TPS-TPS, TPS-BAHD acyltransferases, and TPS-CYP450 gene pairs, which are found to be induced upon stress conditions (Li et al., 2021b). It has been inferred that tandem duplications might have been the driving force for the emergence of these gene combinations (Li et al., 2021b). In addition, it has been reported that the coupling of terpene synthases and CYP450s P450s in gene clusters are conserved in eudicots, but not in monocots. The occurrence of transposable elements (in eudicots) and the sub-telomeric position of clusters (in monocots) indicate the possibilities of the occurrence of recombination events leading to a novel cluster formation resulting in metabolic diversity (Qi et al., 2006; Field et al., 2011; Boutanaev et al., 2015).

Acyl Transferases and Glycosyl Transferases: Avoiding the Accumulation of Toxic Intermediates

Plants have evolved in such a way to physically link genes as clusters to avoid the accumulation of toxic intermediates (Nützmann et al., 2016). The addition of sugar moieties and acyl groups to metabolic pathway intermediates, respectively, through glycosyl transferases and acyl transferases results in reducing the cytotoxicity (Itkin et al., 2011). It is known that the biological activity of triterpenes requires acylation at C-21 position (Podolak et al., 2010). Accordingly, des-acyl avenacins are acylated at C-21 position by serine carboxypeptidase like (SCPL) acyltransferase (*Sad7*), which incorporates *N*-methyl anthraniloyl group to produce avenacin A1 (Mugford et al., 2009; Table 2; Figure 1). In general, SCPL acyltransferases utilize β -acetyl glucose esters as acyl donors and, in oat *Sad7* utilizes *N*-methyl anthraniloyl-*O*-glucopyranose (NMA-Glc) as the acyl donor (Mugford and Osbourn, 2010; Ciarkowska et al., 2019). NMA-Glc is synthesized from anthranilic acid through two sequential steps catalyzed by methyltransferase (*Sad9*) and UDP-glucosyltransferase (*Sad10*), respectively (Owatworakit et al., 2012; Mugford et al., 2013; Figure 1). It is interesting to note that *Sad7*, *Sad9*, and *Sad10* occur adjacently in the avenacin gene cluster and regarded as acylation module (Mugford et al., 2013). In addition, *Sad7* can acylate des-acyl avenacins with benzoyl group donated by benzoyl- β -glucopyranose to form avenacin A2 (Mugford et al., 2009; Owatworakit et al., 2012). Hence,

the scaffold diversity of avenacins is attributed to the SCPL acyltransferase (Sad7) and availability of acyl donors (Figure 1).

Besides acylation, triterpene backbone of avenacin is glycosylated with branched trisaccharide moiety (one arabinose and two glucose molecules), which is essential for its antimicrobial activity (Louveau et al., 2018). In the first instance, the addition of arabinose to β -amyrin scaffold is mediated by an arabinosyltransferase (AsAAT1) and the mutants of *asaat1* exhibited susceptibility to pathogenic *Gaeumannomyces* sp. (Louveau et al., 2018). Subsequently, two glucose molecules are attached to L-arabinose through glucosyltransferase and glycosyl hydrolase encoded by *AsUGT91G16* and *A. strigosa* transglucosidase 1 (*AsTG1*), respectively (Figure 1). *AsUGT91G16* adds 1,2 linked glucose to arabinose, and *AsTG1* performs the final addition of 1,4 linked glucose (Orme et al., 2019). Remarkably, *AsTG1* (synonymous to *Sad3*; a core gene of avenacin cluster) encodes a unique vacuolar glycosyl hydrolase 1 (GH1), and the final addition of glucose moiety is known to occur in the vacuole, although an activity of glycosyl transferases is predominant in the cytosol as well (Hansen et al., 2013; Orme et al., 2019; Figure 1). Furthermore, *sad3* mutants accumulated non-glycosylated avenacins, which resulted in stunted root growth and deformed root hairs in oat, implicating the importance of glycosylation for normal avenacin bioactivity (Mylona et al., 2008). While the core cluster of avenacin exists in a sub-telomeric region of chromosome 1, *AsUGT91G16* and *AsTG1* genes exist in a location away from the core cluster in a proximal scaffold to avoid telomere gene deletions, which in turn help to avoid the accumulation of toxic intermediates (Li et al., 2021a). In *Arabidopsis*, three BAHD acyltransferases, *THAA1*, *THAA2*, and *THAA3*, are involved in triterpene metabolism, and acetylates signature metabolites, such as thalianol, arabidiol, and deletions of the *THAA2* locus, could lead to the development of short roots as compared to the wild type although an internal molecular cue is yet to be deciphered (Huang et al., 2019; Liu et al., 2020a,b; Bai et al., 2021).

Plants producing benzoxazinoids (DIMBOA) glycosylate them to avoid auto toxicity and store these biomolecules in the vacuole, while during defense response β -glucosidase cleaves sugar molecule and activates the function (Von Rad et al., 2001; Dick et al., 2012; Figure 2A). *Solanum* species also glycosylate SGAs to avoid auto toxic effects, and four UGTs (*GAME1*, *GAME17*, *GAME18*, and *GAME2*) of α -tomatine gene cluster adds tetrasaccharide moieties (one galactose, one xylose, and two glucose molecules) to tomatidine for the formation of α -tomatine (Itkin et al., 2013; Figure 3B). Similarly, α -solanine and α -chaconine are glycosylated with glucose:galactose:rhamnose and rhamnose:galactose:rhamnose, respectively (Ohyama et al., 2013). Alterations in *GAME1* activity have been shown to result in the accumulation of high levels of tomatidine and as a consequence plants exhibited defective fruit development and increased susceptibility to pathogen attack. It has been reported that the accumulation of α -tomatine and its subsequent conversion to esculeoside A is important for the normal development of tomato fruits (Itkin et al., 2011). Additionally, *GAME25* encodes a short-chain dehydrogenase, which catalyzes the reduction of a double bond at the C-5,C-6 position and is

considered as a key step in determining the diversity of SGAs in *Solanum* species. In addition, this modification has been shown to result in the reduction of the toxic effects of SGAs (Sonawane et al., 2018).

In a similar way, cyanogenic glucosides are glycosylated at the final step and transported to the vacuole (Gleadow and Møller, 2014). Cyanogenic glucoside producing the gene clusters of *L. japonicus* and *S. bicolor* are equipped with UGTs so as to catalyze the final glycosylation step for producing linamarin, lotaustralin, and dhurrin (Tako et al., 2011; Figure 2B). Independent recruitment of genes and their coordinated expression for metabolic channeling are the characteristic features of plant gene clusters to avoid the accumulation of toxic metabolic intermediates. Extensive mining of plant genomes can reveal the novel gene clusters, and it would be interesting to study their architecture related to the arrangement of the constituent genes of the cluster and their functional attributes in metabolite production.

REGULATION OF PLANT GENE CLUSTERS

Plant Gene Clusters Are Stress-Induced and Exhibit Spatiotemporal Expression

In general, plant gene clusters are spatiotemporally regulated (Qi et al., 2006; Shimura et al., 2007; Table 1). The expression of *AsbAs1* of avenacin cluster and other genes are found to be localized specifically in root tips, depicting that the biosynthesis and storage of metabolites are tissue specific (Haralampidis et al., 2001; Qi et al., 2004, 2006). In maize, the biosynthesis of DIMBOA is reported to be developmentally regulated with a localized expression restricted to leaves and roots (Frey et al., 1997). The expression of thalianol and marneral gene clusters of *Arabidopsis* are localized in root epidermis. However, the physiological function of the clusters is yet to be deciphered. The overexpression of these clusters resulted in dwarfing and negatively affected the overall plant development in *Arabidopsis* (Field and Osbourn, 2008; Field et al., 2011). In addition, triterpenes produced by *Arabidopsis* roots are reported to modulate the microbial diversity in the rhizosphere by promoting the enrichment of species-specific bacteria on root surfaces (Huang et al., 2019) though the role of microbiota in plant development is not conclusive. Mutant *Arabidopsis* lines of triterpene biosynthetic genes significantly reduced microbial operational taxonomical units (OTUs) compared to wild type (Huang et al., 2019). In continuation with this, Chen et al. (2019a) identified a group of sesterterpene cluster-derived metabolites in *A. thaliana*, which were shown to modulate the density of root microbiota. This observation was further confirmed by analyzing OTUs in wild-type and mutant lines (Chen et al., 2019a).

Diterpene clusters in rice, such as phytocassanes, momilactones, and oryzalides, are activated by chitin oligosaccharide-based elicitation and UV irradiations, and their biosynthesis is localized in leaves and roots (Shimura et al., 2007; Swaminathan et al., 2009). Pathogen-induced capsidiol gene cluster has been identified in plants such as capsicum and tobacco, and the cluster is highly conserved in both the species

(Lee et al., 2017; Chen et al., 2019b). CYP450s catalyzing the biosynthesis of linamarin and lotaustralin was found to be highly expressed in younger apical leaves and the expression gradually diminished with plant age (Takos et al., 2011). In a similar way, the expression of lycosantalonal-producing genes leaf petiole, and β -phellandrene gene expression is confined to leaves (Matsuba et al., 2015; Zhou and Pichersky, 2020). Interestingly, 20-hydroxy-betulinic acid-producing gene cluster in *L. japonicus* is highly expressed during nodule formation, thus portraying its role in plant development (Krokida et al., 2013). Therefore, it is evident that gene clusters evolved as adaptive strategies to produce defense-specific metabolites in a rapid and coordinated fashion to evade both biotic and abiotic stress conditions.

Role of Transcription Factors and Chromatin Remodeling

Three basic helix–loop–helix (bHLH) transcription factors (TFs) related to cucurbitacin biosynthesis have been identified in cucurbits, viz., *bitter fruit* (*Bt*), *bitter leaf* (*Bl*), and *bitter root* (*Br*), and the expression of these TFs is specific to fruits, leaves, and roots, respectively. These regulators can strongly bind to the promoter regions of nine cucurbitacin biosynthesis genes and influence the biosynthesis of CuC, CuB, and CuE (Shang et al., 2014; Zhou et al., 2016). In addition, it has been shown that domesticated cucurbit fruits (cucumber, melon, and watermelon) lost their bitterness due to a mutation in the *Bt* gene affecting its binding properties to promoter regions, although the wild species retained *Bt* activity (Zhou et al., 2016; Chomiccki et al., 2020). In tomato and potato, *GAME9* TF [belonging to a class of Ethylene Responsive Factor (ERF)] is found to regulate SGA bioproduction. While the overexpression of *GAME9* leads to increased levels of α -tomatine, α -chaconine, the knockout of *GAME9* activity reversed the expression of SGA biosynthetic genes (Cárdenas et al., 2016). Recently, Yu et al. (2020) identified two allelic variants (*GAME9*^{135A} and *GAME9*^{135V}) of *GAME9* in wild and domesticated *Solanum* species. In the wild species, *GAME9*^{135A} exhibited a strong binding affinity to *GAME7* and *GAME17* gene promoters (Yu et al., 2020). However, *GAME9*^{135V} did not display strong interactions with the promoter regions and is coordinated with another TF MYC2 to regulate SGA biosynthesis (Yu et al., 2020). In addition, JRE4 TF of tomato increased SGA accumulation by binding to the promoter region of *GAME4* gene. In addition, overexpression and knockout experiments established the role of JRE4 in SGA biosynthesis (Thagun et al., 2016; Nakayasu et al., 2017).

In rice, elicitor-induced bZIP TF *OsTGAP1* has been identified along with its possible role in momilactone biosynthesis, and it has been shown that the overexpression of *OsTGAP1* significantly enhanced metabolite accumulation (Okada et al., 2009). Intriguingly, *OsTGAP1* does not directly regulate the transcription of the momilactone cluster, instead it binds to the intergenic regions adjacent to the momilactone cluster (Miyamoto et al., 2014). However, *OsTGAP1* binds strongly to the promoter region of *OsDXS* gene and upregulates its transcription, probably to increase the precursor pool (Miyamoto

et al., 2014). Interestingly, the overexpression of TGA factor, *OsZIP79* suppresses momilactone biosynthesis, and *OsZIP79* and *OsTGAP1* interact with each other to form a heterodimer, which might have a role in phytoalexin biosynthesis in rice (Miyamoto et al., 2015). In addition, TFs regulating the synthesis of nicotine and terpene indole alkaloids are clustered in the genome for facilitating a coordinated activity in *Nicotiana tabacum* and *C. roseus*, respectively (Shoji et al., 2010; Singh et al., 2020).

In *Arabidopsis*, metabolic clusters possess histone 3 lysine trimethylation (H3K27me3) and histone 2 variant H2A.Z chromatin signatures regulated by sick with RSC/Rat1 (SWR) complex-mediated chromatin remodeling (Yu et al., 2016). Actin-related protein (Arp6), a subunit of SWR complex, incorporates a histone 2 variant (H2A.Z) into nucleosomes of cluster genes and facilitate metabolite bioproduction. The downregulation of the cluster gene expression in the mutants of *arp6* and *h2a.z* further confirmed their regulatory role in the cluster (Nützmann and Osbourn, 2015). Reimegård et al. (2017) identified the similar modifications performed by histone methylases and histone deacetylases to facilitate the expression of development-specific gene clusters of *Arabidopsis*. Zhan et al. (2020) reported a histone demethylase JM1705 activated upon methyl jasmonate treatment, which acted antagonistically on H3K27me3 of a chromatin region and upregulated the expression of rice casbene gene cluster. Recently, Nützmann et al. (2020) reported that metabolic gene clusters occurred in the local interactive domains of plant genome, and these domains surrounding the clusters facilitated the tightly coordinated expression of genes in the cluster. Further, the localization of the silenced clusters occurred in the periphery of the chromosomes while the expressing clusters are positioned in a location away from the periphery (Nützmann et al., 2020). In addition, chromatin-level remodeling is evident in the regulation of specialized metabolite gene clusters in filamentous fungi (Bok et al., 2009). Wegel et al. (2009) demonstrated that chromatin decondensation at *Sad1*, *Sad2* genes occurs only in root tip epidermis to initiate avenacin synthesis. Similar decondensation process was minimal in other tissues, although internal molecular cascades of cell-specific chromatin decondensation is yet to be understood.

PHYSIOLOGICAL ACTIVITIES OF PLANT GENE CLUSTER-DERIVED METABOLITES

Metabolites produced by gene clusters exhibit a specific role in defense responses and few other metabolites are involved in regulating the cell physiology during plant development (Krokida et al., 2013; Louveau et al., 2018; **Table 1**). Avenacins synthesized by oat exhibit antifungal properties and are known to inhibit fungal infection, by acting against *Blumeria graminis*, *Bipolaris oryzae*, and *Magnaporthe oryzae* (Inagaki et al., 2013). Non-glycosylated avenacins (*Sad3* mutants) severely affected the formation of root epidermis, implicating the biosynthesis and accumulation of specific intermediates of specialized metabolites is cytotoxic to the plant (Mylona

et al., 2008). On the other hand, rice momilactones are induced by UV irradiation and are active against a wide range of fungal pathogens such as *Magnaporthe grisea*, *Botrytis cinerea*, *Fusarium solani*, *Colletotrichum gloeosporioides*, and *Phytophthora infestans* (Zhao et al., 2018). Similarly, DIMBOA and its derivatives are found to be effective against *Ralstonia solanacearum* and *Rhizoctonia solani*, which are the causative agents of bacterial wilt and sheath blight disease, respectively (Song et al., 2011; Guo et al., 2016). Besides, cyanogenic glucosides, such as, linamarin, lotaustralin, and dhurrin, have been shown to exhibit anti-herbivore and anti-insect properties (Gleadow and Møller, 2014). The removal of the saccharide moiety by glucosidase from cyanogenic glucosides and DIMBOA is essential for its activity. Plants sequester specific glucosidases in plastids to avoid auto toxicity. When cells get ruptured due to mechanical injury, these enzymes get released and cleave glucose moiety to induce defense responses (Gleadow and Møller, 2014). Few insects such as burnet moth can ingest cyanogenic glucosides and use it for its own defense, and also this moth has been shown to possess cyanogenic glucoside-producing genes and capable of synthesizing them on its own (Jensen et al., 2011). Cyanogenic glucoside gene machinery might have co-evolved with insects to develop an immunity against plant defense.

Metabolites, such as α -tomatine and tomatidine, display neurotoxic effects in humans by producing truncated proteins and deactivating proteasomes (da Silva et al., 2017). Unripe fruits are known to produce relatively high amounts of α -tomatine and during the process of ripening it gets gradually converted into esculoside A, while the disruption of α -tomatine biosynthesis leads to the development of deformed fruits (Itkin et al., 2011). SGAs also possess anticancer, anti-inflammatory, antioxidant, and cardiovascular curative properties (Friedman, 2013; Al Sinani and Eltayeb, 2017). α -tomatine suppressed the metastasis of human lung cells by downregulating focal adhesion kinase (FAK), phosphatidylinositol 3-kinase (PI3K), and nuclear factor kappa B (NF- κ B), which are potentially involved in cancer cell migration (Shieh et al., 2011). Interestingly, *Fusarium oxysporum*-derived tomatinases detoxify the effects of α -tomatine and suppresses host defense mechanisms (Ito et al., 2004). It can be postulated that pathogenic fungi might produce these enzymes as adaptive mechanisms to counteract host immune responses.

Casbene-derived metabolites, such as prostratin and ingenol-3,20-dibenzoate (IDB), are known to activate protein kinase C-(PKC-) mediated signaling, which is involved in repolarizing cardiac muscle cells, and could be used for treating cardiovascular disorders (Jiang et al., 2018). Prostratin and Ingenol derivatives are known to be potential latency reversal agents (LRA) and are found to clear the latently infected cells of HIV (Sloane et al., 2020). In addition, casbene-derived ingenol mebutate upsurges the neutrophil-mediated tumor cell degradation of subcutaneous melanoma (Braun et al., 2018). Noscapine, a notable BIA, is analgesic and used in the preparation of cough syrups due to its antitussive properties (Winzer et al., 2015). Thebaine is widely used in industries to develop semisynthetic drugs, such as oxycodone, oxymorphone, etorphine, nalbuphine,

and naloxone, which are used as analgesics and also in the treatment of opioid poisoning (Hagel and Facchini, 2013; Singh et al., 2019). Metabolic engineering approaches for the heterologous expression of these clusters could be developed for the heterologous production of these metabolites for agriculture and pharmaceutical applications.

PLANT GENE CLUSTERS FOR METABOLIC ENGINEERING

The biosynthesis of specialized metabolites in plants underpins many traits of ecological, pharmaceutical, and agronomic importance. However, plant-derived products have not gained importance till the recent times due to the domination by synthetic chemical analogs. Recent era has evidenced the emergence of plant-based natural compounds as the potential alternatives against synthetic counterparts in both agricultural and pharmaceutical sectors. Furthermore, advances in the omics approach led to the identification of plant metabolic gene clusters involved in the biosynthesis of terpenoids, alkaloids, benzoxazinoids, and cyanogenic glycosides, that have several applications. Novel findings from several studies shed light on developing sustainable metabolic engineering strategies for the overproduction of some of these specialized metabolites to meet market demands. In this regard, *in silico* approaches, including cluster predicting tools, such as PhytoClust, plantiSMASH, ClusterFinder, and ClustScan to name a few, have accelerated the discovery of diverse metabolic pathway-related gene clusters in plants (Chavali and Rhee, 2018). Technological advances in genome sequencing and availability of high-throughput functional genomics tools resulted in a shift from a single-step characterization to the validation of entire metabolic pathways. Increased knowledge of metabolic gene clusters accelerated genome engineering strategies for the biosynthesis of alkaloids and terpenoids in heterologous host systems. Nevertheless, low precursor availability, the accumulation of unwanted intermediates, and hindrance due to the lack of information regarding genes/regulatory steps are some of the bottlenecks to overproduce these compounds through synthetic biology approaches.

Primarily, the biosynthetic pathways of plant specialized metabolism are complex with several gene cascades, enzymatic reactions, and compartmentalized intermediates and or end products. Hence, a thorough understanding of the pathway dynamics is necessary to develop a strategy for heterologous metabolite production. For instance, the introduction of 12 artemisinin biosynthetic genes into tobacco chloroplasts severely affected the growth of the plant (Saxena et al., 2014). However, dividing the pathway and expressing the genes into two compartments (plastids and cytosol) successfully yielded a reasonable titer of artemisinin in tobacco that could be further used for commercial application (Malhotra et al., 2016). De La Peña and Sattely (2021) increased the biosynthesis of momilactone by expressing plastid localized geranylgeranyl diphosphate synthases (GGPPS) and CPS in cytosol through

plastid tag truncation (De La Peña and Sattely, 2021). Another strategy is to modulate the precursor flux by engineering additional copies of rate limiting step genes or by silencing the branch point genes so that the flux can be diverted to the synthesis of the desired metabolite. This approach was attempted in *Artemisia annua* by downregulating the expression of β -caryophyllene synthase gene to increase the bioproduction of artemisinin (Lv et al., 2016). A similar strategy of pathway shunting can be achieved by gene editing tools, such as CRISPR-Cas9, which is currently being used widely in metabolic engineering (Sabzehzari et al., 2020). Furthermore, the identification of TFs that regulate multiple steps in a biosynthetic pathway is yet another approach for overproducing specific metabolites. GAME9 TF of the SGA cluster regulates both the MEP pathway and SGA pathway (Cárdenas et al., 2016). Finally, improving the cell numbers in the plant, such as increasing the trichome on the leaves, can also improve the metabolite bioproduction specific to a particular tissue (Fu et al., 2018).

The use of robust microbial hosts, such as *Escherichia coli* and yeast, is another promising approach for overproducing target metabolites by synthetic biology (Pyne et al., 2019). Celedon et al. (2016) engineered a multistep pathway in yeast cells for producing sandalwood oil similar to the one extracted from the heartwood of *Santalum album*. In addition, opioids were successfully produced in yeast through

the introduction of morphine biosynthesis genes (Galanie et al., 2015). Nevertheless, some of the challenges which the researchers should anticipate and foresee while using a microbial host include: (a) ways to avoid the accumulation of toxic metabolic intermediates; (b) a fine control of gene expression; and (c) precursor limitation. Taken together, the combination of *in silico* analysis, the availability of robust functional genomic tools, and the knowledge about metabolic gene clusters could shed light into new directions in synthetic biology research at an accelerated level. The present scenario in this area of research holds a great promise to translate the basic knowledge of plant metabolism into tangible benefits for agricultural and pharmaceutical applications.

AUTHOR CONTRIBUTIONS

RB, SRK, and AS wrote the manuscript. RS revised it. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors thank Dr. V. S. Pragadheesh, Scientist, CSIR-Central Institute of Medicinal and Aromatic Plants, Bengaluru, India for the help rendered with chemical structures.

REFERENCES

- Al Sinani, S. S. S., and Eltayeb, C. A. (2017). The steroidal glycoalkaloids solamargine and solasonine in solanum plants. *S. Afr. J. Bot.* 112, 253–269. doi: 10.1016/j.sajb.2017.06.002
- Anand, A., Jayaramaiah, R. H., Beedkar, S. D., Dholakia, B. B., Lavhale, S. G., Puneekar, S. A., et al. (2019). Terpene profiling, transcriptome analysis and characterization of cis- β -terpineol synthase from ocimum. *Physiol. Mol. Biol. Plants*. 25, 47–57. doi: 10.1007/s12298-018-0612-6
- Bai, Y., Fernández-Calvo, P., Ritter, A., Huang, A. C., Morales-Herrera, S., Bicalho, K. U., et al. (2021). Modulation of *Arabidopsis* root growth by specialized triterpenes. *New Phytol.* 230, 228–243. doi: 10.1111/nph.17144
- Ballouz, S., Francis, A. R., Lan, R., and Tanaka, M. M. (2010). Conditions for the evolution of gene clusters in bacterial genomes. *PLoS Comput. Biol.* 6:e1000672. doi: 10.1371/journal.pcbi.1000672
- Barchi, L., Pietrella, M., Venturini, L., Minio, A., Toppino, L., Acquadro, A., et al. (2019). A chromosome-anchored eggplant genome sequence reveals key events in solanaceae evolution. *Sci. Rep.* 9:11769. doi: 10.1038/s41598-019-47985-w
- Bok, J. W., Chiang, Y. M., Szewczyk, E., Reyes-Dominguez, Y., Davidson, A. D., Sanchez, J. F., et al. (2009). Chromatin-level regulation of biosynthetic gene clusters. *Nat. Chem. Biol.* 5, 462–464. doi: 10.1038/nchembio.177
- Boutanaev, A. M., Moses, T., Zi, J., Nelson, D. R., Mugford, S. T., Peters, R. J., et al. (2015). Investigation of terpene diversification across multiple sequenced plant genomes. *Proc. Natl. Acad. Sci. U.S.A.* 112, E81–E88. doi: 10.1073/pnas.1419547112
- Boutanaev, A. M., and Osbourn, A. E. (2018). Multigenome analysis implicates miniature inverted-repeat transposable elements (MITEs) in metabolic diversification in eudicots. *Proc. Natl. Acad. Sci. U.S.A.* 115, E6650–E6658. doi: 10.1073/pnas.1721318115
- Bradshaw, R. E., Slot, J. C., Moore, G. G., Chettri, P., de Wit, P. J. G. M., Ehrlich, K. C., et al. (2013). Fragmentation of an aflatoxin-like gene cluster in a forest pathogen. *New Phytol.* 198, 525–535. doi: 10.1111/nph.12161
- Braun, S. A., Baran, J., Schrupf, H., Buhren, B. A., Bölke, E., Homey, B., et al. (2018). Ingenol mebutate induces a tumor cell-directed inflammatory response and antimicrobial peptides thereby promoting rapid tumor destruction and wound healing. *Eur. J. Med. Res.* 23:45. doi: 10.1186/s40001-018-0343-8
- Caballero, J. L., Martinez, E., Malpartida, F., and Hopwood, D. A. (1991). Organisation and functions of the actVA region of the actinorhodin biosynthetic gene cluster of *Streptomyces coelicolor*. *Mol. Genet. Genom.* 230, 401–412. doi: 10.1007/BF00280297
- Cárdenas, 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
- Celedon, J. M., Chiang, A., Yuen, M. M., Diaz-Chavez, M. L., Madilao, L. L., Finnegan, P. M., et al. (2016). Heartwood-specific transcriptome and metabolite signatures of tropical sandalwood (*Santalum album*) reveal the final step of (Z)-santalol fragrance biosynthesis. *Plant J.* 86, 289–299. doi: 10.1111/tpj.13162
- Chavali, A. K., and Rhee, S. Y. (2018). Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. *Brief Bioinform.* 19, 1022–1034. doi: 10.1093/bib/bbx020
- Chen, Q., Jiang, T., Liu, Y. X., Liu, H., Zhao, T., Liu, Z., et al. (2019a). Recently duplicated sesterterpene (C25) gene clusters in *Arabidopsis thaliana* modulate root microbiota. *Sci. China Life Sci.* 62, 947–958. doi: 10.1007/s11427-019-9521-2
- Chen, X., Liu, F., Liu, L., Fang, D., Wang, W., Zhang, X., et al. (2019b). Characterization and evolution of gene clusters for terpenoid phytoalexin biosynthesis in tobacco. *Planta* 250, 1687–1702. doi: 10.1007/s00425-019-03255-7
- Chomicki, G., Schaefer, H., and Renner, S. S. (2020). Origin and domestication of *Cucurbitaceae* crops: insights from phylogenies, genomics and archaeology. *New Phytol.* 226, 1240–1255. doi: 10.1111/nph.16015
- Ciarkowska, A., Ostrowski, M., Starzyńska, E., and Jakubowska, A. (2019). Plant SCPL acyltransferases: multiplicity of enzymes with various functions in secondary metabolism. *Phytochem. Rev.* 18, 303–316. doi: 10.1007/s11101-018-9593-x
- da Silva, D. C., Andrade, P. B., Valentão, P., and Pereria, D. M. (2017). Neurotoxicity of the steroidal alkaloids tomatine and tomatidine is RIP1

- kinase- and caspase-independent and involves the eIF2 α branch of the endoplasmic reticulum. *J. Steroid Biochem. Mol. Biol.* 171, 178–186. doi: 10.1016/j.jsbmb.2017.03.009
- Darbani, B., Motawia, M. S., Olsen, C. E., Nour-Eldin, H. H., Møller, B. L., and Rook, F. (2016). The biosynthetic gene cluster for the cyanogenic glucoside dhurrin in *Sorghum bicolor* contains its co-expressed vacuolar MATE transporter. *Sci. Rep.* 6:37079. doi: 10.1038/srep37079
- De La Peña, R., and Sattely, E. S. (2021). Rerouting plant terpene biosynthesis enables momilactone pathway elucidation. *Nat. Chem. Biol.* 17, 205–212. doi: 10.1038/s41589-020-00669-3
- Dick, R., Rattei, T., Haslbeck, M., Schwab, W., Gierl, A., and Frey, M. (2012). Comparative analysis of benzoxazinoid biosynthesis in monocots and dicots: independent recruitment of stabilization and activation functions. *Plant Cell.* 24, 915–928. doi: 10.1105/tpc.112.096461
- Dong, A. X., Xin, H. B., Li, Z. J., Liu, H., Sun, Y. Q., Nie, S., et al. (2018). High-quality assembly of the reference genome for scarlet sage, *Salvia splendens*, an economically important ornamental plant. *Gigascience* 7:giy068. doi: 10.1093/gigascience/giy068
- Dutartre, L., Hilliou, F., and Feyerisen, R. (2012). Phylogenomics of the benzoxazinoid biosynthetic pathway of *Poaceae*: gene duplications and origin of the Bx cluster. *BMC Evol. Biol.* 12:64. doi: 10.1186/1471-2148-12-64
- Elizondo, L. I., Jafar-Nejad, P., Clewing, J. M., and Boerkoel, C. F. (2009). Gene clusters, molecular evolution and disease: a speculation. *Curr. Genomics* 10, 64–75. doi: 10.2174/138920209787581271
- Fan, P., Wang, P., Lou, Y. R., Leong, B. J., Moore, B. M., and Schenck, C. A. (2020). Evolution of a plant gene cluster in Solanaceae and emergence of metabolic diversity. *Elife* 9:e56717. doi: 10.7554/eLife.56717
- Field, B., Fiston-Lavier, A. S., Kemen, A., Geisler, K., Quesneville, H., and Osbourn, A. E. (2011). Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16116–16121. doi: 10.1073/pnas.1109273108
- Field, B., and Osbourn, A. E. (2008). Metabolic diversification-independent assembly of operon-like gene clusters in different plants. *Science* 320, 543–547. doi: 10.1126/science.1154990
- Franke, J., Kim, J., Hamilton, J. P., Zhao, D., Pham, G. M., Wiegert-Rininger, K., et al. (2019). Gene discovery in gelsemium highlights conserved gene clusters in monoterpene indole alkaloid biosynthesis. *Chembiochem* 20, 83–87. doi: 10.1002/cbic.201800592
- Frey, M., Chomet, P., Glawischmig, E., Stettner, C., Grün, S., Winklmaier, A., et al. (1997). Analysis of a chemical plant defense mechanism in grasses. *Science* 277, 696–699. doi: 10.1126/science.277.5326.696
- Frey, M., Schullehner, K., Dick, R., Fiesselmann, A., and Gierl, A. (2009). Benzoxazinoid biosynthesis, a model for evolution of secondary metabolic pathways in plants. *Phytochemistry* 70, 1645–1651. doi: 10.1016/j.phytochem.2009.05.012
- Friedman, M. (2013). Anticarcinogenic, cardioprotective, and other health benefits of tomato compounds lycopene, α -tomatine, and tomatidine in pure form and in fresh and processed tomatoes. *J. Agric. Food Chem.* 61, 9534–9550. doi: 10.1021/jf402654e
- Fu, R., Martin, C., and Zhang, Y. (2018). Next-Generation plant metabolic engineering, inspired by an ancient chinese irrigation system. *Mol. Plant* 11, 47–57. doi: 10.1016/j.molp.2017.09.002
- Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M., and Smolke, C. D. (2015). Complete biosynthesis of opioids in yeast. *Science* 349, 1095–1100. doi: 10.1126/science.aac9373
- Ghosh, S. (2017). Triterpene structural diversification by plant cytochrome P450 enzymes. *Front. Plant Sci.* 8:1886. doi: 10.3389/fpls.2017.01886
- Gleadow, R. M., and Møller, B. L. (2014). Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. *Annu. Rev. Plant Biol.* 65, 155–185. doi: 10.1146/annurev-arplant-050213-040027
- Guo, B., Zhang, Y., Li, S., Lai, T., Yang, L., Chen, J., et al. (2016). Extract from maize (*Zea mays* L.): antibacterial activity of DIMBOA and its derivatives against *Ralstonia solanacearum*. *Molecules* 21:1397. doi: 10.3390/molecules21101397
- Guo, L., Winzer, T., Yang, X., Li, Y., Ning, Z., He, Z., et al. (2018). The opium poppy genome and morphinan production. *Science* 362, 343–347. doi: 10.1126/science.aat4096
- Hagel, J. M., and Facchini, P. J. (2013). Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. *Plant Cell Physiol.* 54, 647–672. doi: 10.1093/pcp/pct020
- Halkier, B. A., Hansen, C. H., Mikkelsen, M. D., Mikkelsen, M. D., Naur, P., and Wittstock, U. (2002). “The role of cytochromes P450 in biosynthesis and evolution of glucosinolates,” in *Recent Advances in Phytochemistry*, Vol. 36, ed J. T. Romeo, and R. A. Dixon (Elsevier), 223–248. doi: 10.1016/S0079-9920(02)80029-6
- Hansen, S. F., Harholt, J., Oikawa, A., and Scheller, H. V. (2013). Plant glycosyltransferases beyond CAZy: a perspective on DUF families. *Front. Plant Sci.* 3:59. doi: 10.3389/fpls.2012.00059
- Haralampidis, K., Bryan, G., Qi, X., Papadopoulou, K., Bakht, S., Melton, R., et al. (2001). A new class of oxidosqualene cyclases directs synthesis of antimicrobial phytoprotectants in monocots. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13431–13436. doi: 10.1073/pnas.231324698
- Hen-Avivi, S., Savin, O., Racovita, R. C., Lee, W. S., Adamski, N. M., Malitsky, S., et al. (2016). A metabolic gene cluster in the wheat W1 and the barley Cer-cq loci determines β -diketone biosynthesis and glucousness. *Plant Cell* 28, 1440–1460. doi: 10.1105/tpc.16.00197
- Holland, P. W. (2013). Evolution of homeobox genes. *WIREs Dev. Biol.* 2, 31–45. doi: 10.1002/wdev.78
- Horton, R., Wilming, L., Rand, V., Lovering, R. C., Bruford, E. A., Khodiyar, V. K., et al. (2004). Gene map of the extended human MHC. *Nat. Rev. Genet.* 5, 889–899. doi: 10.1038/nrg1489
- Huang, A. C., Jiang, T., Liu, Y. X., Bai, Y. C., Reed, J., Qu, B., et al. (2019). A specialized metabolic network selectively modulates *Arabidopsis* root microbiota. *Science* 364:eaau6389. doi: 10.1126/science.aau6389
- Huang, A. C., Kautsar, S. A., Hong, Y. J., Medema, M. H., Bond, A. D., Tantillo, D. J., et al. (2017). Unearthing a sesiterpene biosynthetic repertoire in the Brassicaceae through genome mining reveals convergent evolution. *Proc. Natl. Acad. Sci. U.S.A.* 114, E6005–E6014. doi: 10.1073/pnas.1705567114
- Ichinose, K., Bedford, D. J., Tornus, D., Bechthold, A., Bibb, M. J., Revill, W. P., et al. (1998). The granatinic biosynthetic gene cluster of *Streptomyces violaceoruber* Tü22: sequence analysis and expression in a heterologous host. *Chem. Biol.* 5, 647–659. doi: 10.1016/S1074-5521(98)90292-7
- Inagaki, Y., Noutoshi, Y., Fujita, K., Imaoka, A., Arase, S., Toyoda, K., et al. (2013). Infection-inhibition activity of avenacin saponins against the fungal pathogens *Blumeria graminis* f. sp. hordei, *Bipolaris oryzae*, and *Magnaporthe oryzae*. *J. Gen. Plant Pathol.* 79, 69–73. doi: 10.1007/s10327-012-0412-8
- 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
- Itkin, M., Rogachev, I., Alkan, N., Rosenberg, T., Malitsky, S., Masini, L., et al. (2011). Glycoalkaloid Metabolism1 is required for steroidal alkaloid glycosylation and prevention of phytoxicity in tomato. *Plant Cell* 23, 4507–4525. doi: 10.1105/tpc.111.088732
- Ito, S., Eto, T., Tanaka, S., Yamauchi, N., Takahara, H., and Ikeda, T. (2004). Tomatidine and lycotetraose, hydrolysis products of α -tomatine by *Fusarium oxysporum* tomatinase, suppress induced defense responses in tomato cells. *FEBS Lett.* 571, 31–34. doi: 10.1016/j.febslet.2004.06.053
- Jacob, F., Perrin, D., Sánchez, C., Monod, J., and Edelman, S. (1960). The operon: a group of genes with expression coordinated by an operator. *C. R. Acad. Sci. Paris* 250, 1727–1729.
- Jensen, N. B., Zagrobelyny, M., Hjærn, Ø. K., Olsen, C. E., Houghton-Larsen, J., Borch, J., et al. (2011). Convergent evolution in biosynthesis of cyanogenic defence compounds in plants and insects. *Nat. Commun.* 3:820. doi: 10.1038/ncomms1865
- Jeon, J. E., Kim, J. G., Fischer, C. R., Mehta, N., Dufour-Schroif, C., Wemmer, K., et al. (2020). A pathogen-responsive gene cluster for highly modified fatty acids in tomato. *Cell* 180, 176–187.e19. doi: 10.1016/j.cell.2019.11.037
- Jiang, Q., Li, K., Lu, W. J., Li, S., Chen, X., Liu, X. J., et al. (2018). Identification of small-molecule ion channel modulators in *C. elegans* channelopathy models. *Nat Commun.* 9:3941. doi: 10.1038/s41467-018-06514-5
- Kellner, F., Kim, J., Clavijo, B. J., Hamilton, J. P., Childs, K. L., Vaillancourt, B., et al. (2015). Genome-guided investigation of plant natural product biosynthesis. *Plant J.* 82, 680–692. doi: 10.1111/tpj.12827

- King, A. J., Brown, G. D., Gilday, A. D., Forestier, E., Larson, T. R., and Graham, I. A. (2016). A cytochrome P450-mediated intramolecular carbon-carbon ring closure in the biosynthesis of multidrug-resistance-reversing lathyrane diterpenoids. *Chembiochem* 17, 1593–1597. doi: 10.1002/cbic.201600316
- King, A. J., Brown, G. D., Gilday, A. D., Larson, T. R., and Graham, I. A. (2014). Production of bioactive diterpenoids in the euphorbiaceae depends on evolutionarily conserved gene clusters. *Plant Cell* 26, 3286–3298. doi: 10.1105/tpc.114.129668
- Kitaoka, N., Zhang, J., Oyagbenro, R. K., Brown, B., Wu, Y., Yang, B., et al. (2021). Interdependent evolution of biosynthetic gene clusters for momilactone production in rice. *Plant Cell* 33, 290–305. doi: 10.1093/plcell/koaa023
- Krokida, A., Delis, C., Geisler, K., Garagounis, C., Tsikou, D., Peña-Rodríguez, L. M., et al. (2013). A metabolic gene cluster in *Lotus japonicus* discloses novel enzyme functions and products in triterpene biosynthesis. *New Phytol.* 200, 675–690. doi: 10.1111/nph.12414
- Lee, H. A., Kim, S., Kim, S., and Choi, D. (2017). Expansion of sesquiterpene biosynthetic gene clusters in pepper confers nonhost resistance to the irish potato famine pathogen. *New Phytol.* 215, 1132–1143. doi: 10.1111/nph.14637
- Leveau, A., Reed, J., Qiao, X., Stephenson, M. J., Mugford, S. T., Melton, R. E., et al. (2019). Towards take-all control: a C-21 β oxidase required for acylation of triterpene defence compounds in oat. *New Phytol.* 221, 1544–1555. doi: 10.1111/nph.15456
- Li, J., Wang, Y., Dong, Y., Zhang, W., Wang, D., Bai, H., et al. (2021b). Correction: The chromosome-based lavender genome provides new insights into Lamiaceae evolution and terpenoid biosynthesis. *Hortic Res.* 8:90. doi: 10.1038/s41438-021-00536-9
- Li, Q., Ramasamy, S., Singh, P., Hagel, J. M., Dunemann, S. M., Chen, X., et al. (2020). Gene clustering and copy number variation in alkaloid metabolic pathways of opium poppy. *Nat. Commun.* 11:2899. doi: 10.1038/s41467-020-16467-3
- Li, Y., Leveau, A., Zhao, Q., Feng, Q., Lu, H., Miao, J., et al. (2021a). Subtelomeric assembly of a multi-gene pathway for antimicrobial defense compounds in cereals. *Nat. Commun.* 12:2563. doi: 10.1038/s41467-021-22920-8
- Lichman, B. R., Godden, G. T., and Buell, C. R. (2020). Gene and genome duplications in the evolution of chemodiversity: perspectives from studies of *Lamiaceae*. *Curr. Opin. Plant Biol.* 55, 74–83. doi: 10.1016/j.pbi.2020.03.005
- Liras, P., and Martín, J. F. (2006). Gene clusters for beta-lactam antibiotics and control of their expression: why have clusters evolved, and from where did they originate? *Int. Microbiol.* 9, 9–19.
- Liu, Z., Cheema, J., Vigouroux, M., Hill, L., Reed, J., Paajanen, P., et al. (2020b). Formation and diversification of a paradigm biosynthetic gene cluster in plants. *Nat. Commun.* 11:5354. doi: 10.1038/s41467-020-19153-6
- Liu, Z., Suarez Duran, H. G., Harnvanichvech, Y., Stephenson, M. J., Schranz, M. E., Nelson, D., et al. (2020a). Drivers of metabolic diversification: how dynamic genomic neighbourhoods generate new biosynthetic pathways in the *Brassicaceae*. *New Phytol.* 227, 1109–1123. doi: 10.1111/nph.16338
- Louveau, T., Orme, A., Pfalzgraf, H., Stephenson, M. J., Melton, R., Saalbach, G., et al. (2018). Analysis of two new arabinosyltransferases belonging to the carbohydrate-active enzyme (CAZY) glycosyl transferase family1 provides insights into disease resistance and sugar donor specificity. *Plant Cell* 30, 3038–3057. doi: 10.1105/tpc.18.00641
- Lv, Z., Zhang, F., Pan, Q., Fu, X., Jiang, W., Shen, Q., et al. (2016). Branch pathway blocking in *artemisia annua* is a useful method for obtaining high yield artemisinin. *Plant Cell Physiol.* 57, 588–602. doi: 10.1093/pcp/pcw014
- Ma, X., Vaistij, F. E., Li, Y., Jansen van Rensburg, W. S., Harvey, S., and Bairu, M. W. (2021). A chromosome-level *Amaranthus cruentus* genome assembly highlights gene family evolution and biosynthetic gene clusters that may underpin the nutritional value of this traditional crop. *Plant J.* doi: 10.1111/tpj.15298
- Malhotra, K., Subramanian, M., Rawat, K., Kalamuddin, M., Qureshi, M. I., Malhotra, P., et al. (2016). Compartmentalized metabolic engineering for artemisinin biosynthesis and effective malaria treatment by oral delivery of plant cells. *Mol. Plant.* 9, 1464–1477. doi: 10.1016/j.molp.2016.09.013
- Mao, L., Kawaide, H., Higuchi, T., Chen, M., Miyamoto, K., Hirata, Y., et al. (2020). Genomic evidence for convergent evolution of gene clusters for momilactone biosynthesis in land plants. *Proc. Natl. Acad. Sci. U.S.A.* 117, 12472–12480. doi: 10.1073/pnas.1914373117
- Matsuba, Y., Nguyen, T. T., Wiegert, K., Falara, V., Gonzales-Vigil, E., Leong, B., et al. (2013). Evolution of a complex locus for terpene biosynthesis in *Solanum*. *Plant Cell.* 25, 2022–2036. doi: 10.1105/tpc.113.111013
- Matsuba, Y., Zi, J., Jones, A. D., Peters, R. J., and Pichersky, E. (2015). Biosynthesis of the diterpenoid lycosantalonal via nerylneryl diphosphate in *Solanum lycopersicum*. *PLoS ONE* 10:e0119302. doi: 10.1371/journal.pone.0119302
- Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., et al. (2015). Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* 11, 625–631. doi: 10.1038/nchembio.1890
- Meena, S., Rajeev Kumar, S., Dwivedi, V., Singh, A. K., Chanotiya, C. S., Akthar, M. Q., et al. (2017). Transcriptomic insight into terpenoid and carbazole alkaloid biosynthesis, and functional characterization of two terpene synthases in curry tree (*Murraya koenigii*). *Sci. Rep.* 7:44126. doi: 10.1038/srep44126
- Miyamoto, K., Fujita, M., Shenton, M. R., Akashi, S., Sugawara, C., Sakai, A., et al. (2016). Evolutionary trajectory of phytoalexin biosynthetic gene clusters in rice. *Plant J.* 87, 293–304. doi: 10.1111/tpj.13200
- Miyamoto, K., Matsumoto, T., Okada, A., Komiyama, K., Chujo, T., Yoshikawa, H., et al. (2014). Identification of target genes of the bZIP transcription factor OsTGAP1, whose overexpression causes elicitor-induced hyperaccumulation of diterpenoid phytoalexins in rice cells. *PLoS ONE* 9:e105823. doi: 10.1371/journal.pone.0105823
- Miyamoto, K., Nishizawa, Y., Minami, E., Nojiri, H., Yamane, H., and Okada, K. (2015). Overexpression of the bZIP transcription factor OsbZIP79 suppresses the production of diterpenoid phytoalexin in rice cells. *J. Plant Physiol.* 173, 19–27. doi: 10.1016/j.jplph.2014.09.001
- Mugford, S. T., Louveau, T., Melton, R., Qi, X., Bakht, S., Hill, L., et al. (2013). Modularity of plant metabolic gene clusters: a trio of linked genes that are collectively required for acylation of triterpenes in oat [published correction appears. *Plant Cell.* 25, 1078–1092. doi: 10.1105/tpc.113.110551
- Mugford, S. T., and Osbourn, A. (2010). Evolution of serine carboxypeptidase-like acyltransferases in the monocots. *Plant Signal Behav.* 5, 193–195. doi: 10.4161/psb.5.2.11093
- Mugford, S. T., Qi, X., Bakht, S., Hill, L., Wegel, E., Hughes, R. K., et al. (2009). A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. *Plant Cell.* 21, 2473–2484. doi: 10.1105/tpc.109.065870
- Mylona, P., Owatworakit, A., Papadopoulou, K., Jenner, H., Qin, B., Findlay, K., et al. (2008). Sad3 and sad4 are required for saponin biosynthesis and root development in oat. *Plant Cell* 20, 201–212. doi: 10.1105/tpc.107.056531
- Nagegowda, D. A., and Gupta, P. (2020). Advances in biosynthesis, regulation, and metabolic engineering of plant specialized terpenoids. *Plant Sci.* 294:110457. doi: 10.1016/j.plantsci.2020.110457
- Nakayasu, M., Umemoto, N., Ohyama, K., Ujimoto, Y., Lee, H. J., and Watanabe, B. (2017). A dioxygenase catalyzes steroid 16 α -hydroxylation in steroidal glycoalkaloid biosynthesis. *Plant Physiol.* 175, 120–133. doi: 10.1104/pp.17.00501
- Nützmann, H. W., Doerr, D., Ramírez-Colmenero, A., Sotelo-Fonseca, J. E., Wegel, E., Di Stefano, M., et al. (2020). Active and repressed biosynthetic gene clusters have spatially distinct chromosome states. *Proc. Natl. Acad. Sci. U.S.A.* 117, 13800–13809. doi: 10.1073/pnas.1920474117
- Nützmann, H. W., Huang, A., and Osbourn, A. (2016). Plant metabolic clusters - from genetics to genomics. *New Phytol.* 211, 771–789. doi: 10.1111/nph.13981
- Nützmann, H. W., and Osbourn, A. (2015). Regulation of metabolic gene clusters in *Arabidopsis thaliana*. *New Phytol.* 205, 503–510. doi: 10.1111/nph.13189
- Nützmann, H. W., Scazzocchio, C., and Osbourn, A. (2018). Metabolic gene clusters in eukaryotes. *Annu. Rev. Genet.* 52, 159–183. doi: 10.1146/annurev-genet-120417-031237
- Ohyama, K., Okawa, A., Moriuchi, Y., and Fujimoto, Y. (2013). Biosynthesis of steroidal alkaloids in Solanaceae plants: involvement of an aldehyde intermediate during C-26 amination. *Phytochemistry* 89, 26–31. doi: 10.1016/j.phytochem.2013.01.010
- Okada, A., Okada, K., Miyamoto, K., Koga, J., Shibuya, N., Nojiri, H., et al. (2009). OsTGAP1, a bZIP transcription factor, coordinately regulates the inductive production of diterpenoid phytoalexins in rice. *J. Biol. Chem.* 284, 26510–26518. doi: 10.1074/jbc.M109.036871
- Olsen, K. M., and Small, L. L. (2018). Micro- and macroevolutionary adaptation through repeated loss of a complete metabolic pathway. *New Phytol.* 219, 757–766. doi: 10.1111/nph.15184

- Orme, A., Louveau, T., Stephenson, M. J., Appelhagen, I., Melton, R., Cheema, J., et al. (2019). A noncanonical vacuolar sugar transferase required for biosynthesis of antimicrobial defense compounds in oat. *Proc. Natl. Acad. Sci. U.S.A.* 116, 27105–27114. doi: 10.1073/pnas.1914652116
- Owatworakit, A., Townsend, B., Louveau, T., Jenner, H., Rejzek, M., Hughes, R. K., et al. (2012). Glycosyltransferases from oat (*Avena*) implicated in the acylation of avenacins. *J. Biol. Chem.* 288, 3696–3704. doi: 10.1074/jbc.M112.426155
- Podolak, I., Galanty, A., and Sobolewska, D. (2010). Saponins as cytotoxic agents: a review. *Phytochem. Rev.* 9, 425–474. doi: 10.1007/s11101-010-9183-z
- Pyne, M. E., Narcross, L., and Martin, V. J. J. (2019). Engineering plant secondary metabolism in microbial systems. *Plant Physiol.* 179, 844–861. doi: 10.1104/pp.18.01291
- Qi, X., Bakht, S., Leggett, M., Maxwell, C., and Osbourn, A. (2004). A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. *Proc. Natl. Acad. Sci. U.S.A.* 101, 8233–8238. doi: 10.1073/pnas.0401301101
- Qi, X., Bakht, S., Qin, B., Leggett, M., Hemmings, A., Mellon, F., et al. (2006). A different function for a member of an ancient and highly conserved cytochrome P450 family: from essential sterols to plant defense. *Proc. Natl. Acad. Sci. U.S.A.* 103, 18848–18853. doi: 10.1073/pnas.0607849103
- Reimegård, J., Kundu, S., Pendle, A., Irish, V. F., Shaw, P., Nakayama, N., et al. (2017). Genome-wide identification of physically clustered genes suggests chromatin-level co-regulation in male reproductive development in *Arabidopsis thaliana*. *Nucleic Acids Res.* 45, 3253–3265. doi: 10.1093/nar/gkx087
- Rokas, A., Mead, M. E., Steenwyk, J. L., Raja, H. A., and Oberlies, N. H. (2020). Biosynthetic gene clusters and the evolution of fungal chemodiversity. *Nat. Prod. Rep.* 37, 868–878. doi: 10.1039/c9np00045c
- Rokas, A., Wisecaver, J. H., and Lind, A. L. (2018). The birth, evolution and death of metabolic gene clusters in fungi. *Nat. Rev. Microbiol.* 16, 731–744. doi: 10.1038/s41579-018-0075-3
- Sabzehzari, M., Zeinali, M., and Naghavi, M. R. (2020). CRISPR-based metabolic editing: Next-generation metabolic engineering in plants. *Gene* 759:144993. doi: 10.1016/j.gene.2020.144993
- Saxena, B., Subramaniam, M., Malhotra, K., Bhavesh, N. S., Potlakayala, S. D., and Kumar, S. (2014). Metabolic engineering of chloroplasts for artemisinin acid biosynthesis and impact on plant growth. *J. Biosci.* 39, 33–41. doi: 10.1007/s12038-013-9402-z
- Schläpfer, P., Zhang, P., Wang, C., Kim, T., Banf, M., Chae, L., et al. (2017). Genome-Wide prediction of metabolic enzymes, pathways, and gene clusters in plants. *Plant Physiol.* 173, 2041–2059. doi: 10.1104/pp.16.01942
- Schneider, L. M., Adamski, N. M., Christensen, C. E., Stuart, D. B., Vautrin, S., Hansson, M., et al. (2017). The Cer-cqu gene cluster determines three key players in a β -diketone synthase polyketide pathway synthesizing aliphatics in epicuticular waxes. *J. Exp. Bot.* 68, 2715–2730. doi: 10.1093/jxb/erw105
- Schullehner, K., Dick, R., Vitzthum, F., Schwab, W., Brandt, W., Frey, M., et al. (2008). Benzoxazinoid biosynthesis in dicot plants. *Phytochemistry* 69, 2668–2677. doi: 10.1016/j.phytochem.2008.08.023
- Seca, A. M. L., and Pinto, D. C. G. A. (2018). Plant secondary metabolites as anticancer agents: successes in clinical trials and therapeutic application. *Int. J. Mol. Sci.* 19:263. doi: 10.3390/ijms19010263
- Shang, Y., Ma, Y., Zhou, Y., Zhang, H., Duan, L., Chen, H., et al. (2014). Plant science. Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* 346, 1084–1088. doi: 10.1126/science.1259215
- Shen, S., Peng, M., Fang, H., Wang, Z., Zhou, S., Jing, X., et al. (2021). An Oryza-specific hydroxycinnamoyl tyramine gene cluster contributes to enhanced disease resistance. *Sci. Bull.* doi: 10.1016/j.scib.2021.03.015
- Shieh, J. M., Cheng, T. H., Shi, M. D., Wu, P. F., Chen, Y., Ko, S. C., et al. (2011). α -Tomatine suppresses invasion and migration of human non-small cell lung cancer NCI-H460 cells through inactivating FAK/PI3K/Akt signaling pathway and reducing binding activity of NF- κ B. *Cell Biochem. Biophys.* 60, 297–310. doi: 10.1007/s12013-011-9152-1
- Shimura, K., Okada, A., Okada, K., Jikumaru, Y., Ko, K. W., Toyomasu, T., et al. (2007). Identification of a biosynthetic gene cluster in rice for momilactones. *J. Biol. Chem.* 282, 34013–34018. doi: 10.1074/jbc.M703344200
- Shoji, T., Kajikawa, M., and Hashimoto, T. (2010). Clustered transcription factor genes regulate nicotine biosynthesis in tobacco. *Plant Cell.* 22, 3390–3409. doi: 10.1105/tpc.110.078543
- Singh, A., Menéndez-Perdomo, I. M., and Facchini, P. J. (2019). Benzylisoquinoline alkaloid biosynthesis in opium poppy: an update. *Phytochem. Rev.* 18, 1457–1482. doi: 10.1007/s11101-019-09644-w
- Singh, S. K., Patra, B., Paul, P., Liu, Y., Pattanaik, S., and Yuan, L. (2020). Revisiting the ORCA gene cluster that regulates terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant Sci.* 293:110408. doi: 10.1016/j.plantsci.2020.110408
- Sloane, J. L., Benner, N. L., Keenan, K. N., Zang, X., Soliman, M. S. A., Wu, X., et al. (2020). Prodrugs of PKC modulators show enhanced HIV latency reversal and an expanded therapeutic window. *Proc. Natl. Acad. Sci. U.S.A.* 117, 10688–10698. doi: 10.1073/pnas.1919408117
- Slot, J. C. (2017). Fungal gene cluster diversity and evolution. *Adv. Genet.* 100, 141–178. doi: 10.1016/bs.adgen.2017.09.005
- Slot, J. C., and Rokas, A. (2010). Multiple GAL pathway gene clusters evolved independently and by different mechanisms in fungi. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10136–10141. doi: 10.1073/pnas.0914418107
- Sonawane, P. D., Heinig, U., Panda, S., Gilboa, N. S., Yona, M., Kumar, S. P., et al. (2018). Short-chain dehydrogenase/reductase governs steroidal specialized metabolites structural diversity and toxicity in the genus *Solanum*. *Proc. Natl. Acad. Sci. U.S.A.* 115, E5419–E5428. doi: 10.1073/pnas.1804835115
- Sonawane, P. D., Jozwiak, A., Panda, S., and Aharoni, A. (2020). 'Hijacking' core metabolism: a new panache for the evolution of steroidal glycoalkaloids structural diversity. *Curr. Opin. Plant Biol.* 55, 118–128. doi: 10.1016/j.pbi.2020.03.008
- Sonawane, P. D., Pollier, J., Panda, S., Szymanski, J., Massalha, H., Yona, M., et al. (2016). Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. *Nat. Plants* 22:17101. doi: 10.1038/nplants.2017.101
- Song, Y. Y., Cao, M., Xie, L. J., Liang, X. T., Zeng, R. S., Su, Y. J., et al. (2011). Induction of DIMBOA accumulation and systemic defense responses as a mechanism of enhanced resistance of mycorrhizal corn (*Zea mays* L.) to sheath blight. *Mycorrhiza* 21, 721–731. doi: 10.1007/s00572-011-0380-4
- Sultana, M. H., Liu, F., Alamin, M., Mao, L., Jia, L., Chen, H., et al. (2019). ene modules co-regulated with biosynthetic gene clusters for allelopathy between rice and barnyardgrass. *Int. J. Mol. Sci.* 20:3846. doi: 10.3390/ijms20163846
- Swaminathan, S., Morrone, D., Wang, Q., Fulton, D. B., and Peters, R. J. (2009). CYP76M7 is an ent-cassadiene C11 α -hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *Plant Cell* 21, 3315–3325. doi: 10.1105/tpc.108.063677
- Takos, A. M., Knudsen, C., Lai, D., Kannangara, R., Mikkelsen, L., and Motawia, M. S. (2011). Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in *Lotus japonicus* and suggests the repeated evolution of this chemical defence pathway. *Plant J.* 68, 273–286. doi: 10.1111/j.1365-313X.2011.04685.x
- Thagun, C., Imanishi, S., Kudo, T., Nakabayashi, R., Ohshima, K., Mori, T., et al. (2016). Jasmonate-Responsive ERF transcription factors regulate steroidal glycoalkaloid biosynthesis in tomato. *Plant Cell Physiol.* 57, 961–975. doi: 10.1093/pcp/pcw067
- Verma, M., Ghangal, R., Sharma, R., Sinha, A. K., and Jain, M. (2014). Transcriptome analysis of *Catharanthus roseus* for gene discovery and expression profiling. *PLoS ONE* 9:e103583. doi: 10.1371/journal.pone.0103583
- Von Rad, U., Hüttel, R., Lottspeich, F., Gierl, A., and Frey, M. (2001). Two glucosyltransferases are involved in detoxification of benzoxazinoids in maize. *Plant J.* 28, 633–642. doi: 10.1046/j.1365-313x.2001.01161.x
- Wang, Q., Hillwig, M. L., Okada, K., Yamazaki, K., Wu, Y., Swaminathan, S., et al. (2012). Characterization of CYP76M5-8 indicates metabolic plasticity within a plant biosynthetic gene cluster. *J. Biol. Chem.* 287, 6159–6168. doi: 10.1074/jbc.M111.305599
- Wang, Q., Hillwig, M. L., and Peters, R. J. (2011). CYP99A3: functional identification of a diterpene oxidase from the momilactone biosynthetic gene cluster in rice. *Plant J.* 65, 87–95. doi: 10.1111/j.1365-313X.2010.04408.x
- Wegel, E., Koumproglou, R., Shaw, P., and Osbourn, A. (2009). Cell type-specific chromatin decondensation of a metabolic gene cluster in oat. *Plant Cell.* 21, 3926–3936. doi: 10.1105/tpc.109.072124
- Wiemann, P., Guo, C. J., Palmer, J. M., Sekonyela, R., Wang, C. C., and Keller, N. P. (2013). Prototype of an intertwined secondary-metabolite supercluster. *Proc. Natl. Acad. Sci. U.S.A.* 110, 17065–17070. doi: 10.1073/pnas.1313258110
- Wilderman, P. R., Xu, M., Jin, Y., Coates, R. M., and Peters, R. J. (2004). Identification of syn-pimara-7,15-diene synthase reveals functional clustering

- of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis. *Plant Physiol.* 135, 2098–2105. doi: 10.1104/pp.104.045971
- Winzer, T., Kern, M., King, A. J., Larson, T. R., Teodor, R. I., Donniger, S. L., et al. (2015). Plant science. Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. *Science* 349, 309–312. doi: 10.1126/science.aab1852
- Wu, Y., Hillwig, M. L., Wang, Q., and Peters, R. J. (2011). Parsing a multifunctional biosynthetic gene cluster from rice: Biochemical characterization of CYP71Z6 & 7. *FEBS Lett.* 585, 3446–3451. doi: 10.1016/j.febslet.2011.09.038
- Xu, H., Song, J., Luo, H., Zhang, Y., Li, Q., Zhu, Y., et al. (2016). Analysis of the Genome Sequence of the medicinal plant *Salvia miltiorrhiza*. *Mol. Plant* 9, 949–952. doi: 10.1016/j.molp.2016.03.010
- Yasumoto, S., Fukushima, E. O., Seki, H., and Muranaka, T. (2016). Novel triterpene oxidizing activity of *Arabidopsis thaliana* CYP716A subfamily enzymes. *FEBS Lett.* 590, 533–540. doi: 10.1002/1873-3468.12074
- Yu, G., Li, C., Zhang, L., Zhu, G., Munir, S., Shi, C., et al. (2020). An allelic variant of GAME9 determines its binding capacity with the GAME17 promoter in the regulation of steroidal glycoalkaloid biosynthesis in tomato. *J. Exp. Bot.* 71, 2527–2536. doi: 10.1093/jxb/era014
- Yu, J., Chang, P., Bhatnagar, D., and Cleveland, T. E. (2000). Cloning of a sugar utilization gene cluster in *Aspergillus parasiticus*. *Biochim. Biophys. Acta.* 1493, 211–214. doi: 10.1016/S0167-4781(00)00148-2
- Yu, N., Nützmann, H. W., MacDonald, J. T., Moore, B., Field, B., Berriri, S., et al. (2016). Delineation of metabolic gene clusters in plant genomes by chromatin signatures. *Nucleic Acids Res.* 44, 2255–2265. doi: 10.1093/nar/gkw100
- Zhan, C., Lei, L., Liu, Z., Zhou, S., Yang, C., Zhu, X., et al. (2020). Selection of a subspecies-specific diterpene gene cluster implicated in rice disease resistance. *Nat Plants.* 6, 1447–1454. doi: 10.1038/s41477-020-00816-7
- Zhang, J., and Peters, R. J. (2020). Why are momilactones always associated with biosynthetic gene clusters in plants? *Proc. Natl. Acad. Sci. U.S.A.* 117, 13867–13869. doi: 10.1073/pnas.2007934117
- Zhao, D., Hamilton, J. P., Bhat, W. W., Johnson, S. R., Godden, G. T., Kinsler, T. J., et al. (2019). A chromosomal-scale genome assembly of *Tectona grandis* reveals the importance of tandem gene duplication and enables discovery of genes in natural product biosynthetic pathways. *Gigascience* 8:giz005. doi: 10.1093/gigascience/giz005
- Zhao, M., Cheng, J., Guo, B., Duan, J., and Che, C. T. (2018). Momilactone and related diterpenoids as potential agricultural chemicals. *J. Agric. Food Chem.* 66, 7859–7872. doi: 10.1021/acs.jafc.8b02602
- Zhou, F., and Pichersky, E. (2020). The complete functional characterisation of the terpene synthase family in tomato. *New Phytol.* 226, 1341–1360. doi: 10.1111/nph.16431
- 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

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Bharadwaj, Kumar, Sharma and Sathishkumar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.