Perspective – *Formatted version*

Emerging mechanistic insights into the regulation of specialized metabolism in plants

Louis-Valentin Méteignier[1], Hans-Wilhelm Nützmann[2], Nicolas Papon[3], Anne Osbourn[4,]✉ and Vincent Courdavault[1]✉

[1] Université de Tours, EA2106 Biomolécules et Biotechnologies Végétales, Tours, France.

[2] The Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AY, United Kingdom.

[3] Univ Angers, Univ Brest, IRF, SFR ICAT, F-49000 Angers, France.

[4] Department of Biochemistry and Metabolism, John Innes Centre, Norwich NR4 7UH, United Kingdom.

✉ to whom correspondence may be addressed: anne.osbourn@jic.ac.uk; vincent.courdavault@univ-tours.fr

ORCID: 0000-0001-9639-5737 (LVM); 0000-0002-6855-7556 (HWN); 0000-0001-6265-7321 (NP); 0000-0003-2195-5810 (AO); 0000-0001-8902-4532 (VC)

Abstract

Plants biosynthesize a broad range of natural products through specialized and species-specific metabolic pathways that are fueled by core metabolism, together forming a metabolic network. Specialized metabolites have important roles in development, adaptation to external cues, and also have invaluable pharmacological properties. A growing body of evidence has highlighted the impact of translational, transcriptional, epigenetic, and chromatin-based regulation and evolution of specialized metabolism genes and metabolic networks. Here, we review the forefront of this research field and extrapolate to medicinal plants that synthetize rare molecules. We also discuss how this new knowledge could help in improving strategies to produce useful plant-derived pharmaceuticals.

Introduction

In microbes and plants, core metabolism produces the precursor compounds that fuel specialized metabolism. Core and specialized metabolisms are thus evolutionarily and molecularly intertwined to form gene and metabolic networks. Indeed, enzymes involved in specialized metabolism have evolved from core metabolic enzymes[1], and core metabolism provides the precursors, scaffolding molecules and co-factors that enable specialized metabolite biosynthesis[2].

Plant specialized metabolic pathways produce structurally complex and functionally diverse molecules such as glucosinolates, terpenoids, cannabinoids, and alkaloids. In general, these molecules have roles in response to stress and in the regulation of growth and development[3]. A plethora of plant specialized metabolites also possess important pharmacological bioactivities[4]. Unfortunately, most plant-derived biopharmaceuticals that are beneficial for human health accumulate naturally in only very small quantities in plants, and their complex structures usually prevent industrial production by total chemical synthesis. Therefore, it is often for biological, chemical, and economic reasons that plant-derived drugs of interest for human health are not widely available on the market. In this respect, from the perspective of developing biotechnological alternatives to produce these precious compounds, some medicinal plants have progressively reached the status of model species for the elucidation of specialized metabolism and pathway architecture. Prominent examples include the anticancer and antihypertensive monoterpene indole alkaloids (MIAs) produced by the Madagascar periwinkle (*Catharanthus roseus*) and the painkiller benzylisoquinoline alkaloids (BIAs) from opium poppy (*Papaver somniferum*). More than 40 years of research have shed light on the highly organized biosynthetic pathways for these compounds at the subcellular, cellular and tissular levels[5,6]. It is now well documented in the literature that distinct parts of both the MIA and BIA biosynthetic routes take place in several cell types. Accordingly, biosynthetic gene expression is localized to specific cell types and organs, and correlates with the presence of the various metabolic pathway intermediates. However, the mechanisms that control the restriction of MIA and BIA biosynthesis to specific tissues and cell types remain unresolved.

Given their importance in plant growth and stress responses, specialized metabolic pathways are regulated at multiple levels, including posttranslational modification, allosteric feedback of enzymes, and regulation by transcription factors[7,8]. In recent years, experimental system biology approaches, mostly performed in the model plant species *Arabidopsis thaliana*, have further expanded the knowledge of how specialized metabolism is regulated and has evolved, notably by mechanisms involving non-coding RNAs, transposable elements (TEs), and epigenetic- and epigenomic-based mechanisms. In parallel, the modest but significant expansion of chromosome-scale genome sequence assemblies of medicinal plants now opens up new opportunities to decipher the genetic programs controlling the biosynthesis of plant natural products of pharmacological importance. By focusing on a selected set of recent studies, we cover here classical and new concepts regarding the regulation and evolution of plant specialized metabolism that could feed future metabolic engineering strategies.

The coordination of core and specialized metabolism by transcriptional and post-transcriptional mechanisms

Numerous examples describe how individual transcription factors (TFs) regulate one or a few key enzyme-encoding metabolic genes[8,9]. In addition, the formation of metabolons that are mediated by protein-protein interactions of successive enzymes for substrate channeling is a well-documented mechanism of regulation of plant metabolism[10–12]. However, metabolic pathways are also highly interconnected and thus require systems-level approaches for an integrative understanding. For instance, nitrogen metabolism is connected to specialized metabolic pathways such as the MIA pathway, which inherently requires aromatic amino acids, while nitrate-related transporters are involved in intra-cellular MIA trafficking[5]. The transcriptional regulation of nitrogen metabolism in Arabidopsis has been comprehensively characterized at the genome-wide level, leading to the discovery of an interactome of 1,660 unique interactions between transcription factors and promoters of nitrogen metabolism genes[13]. A similar but broader strategy has also been deployed for 224 gene promoters from core and specialized metabolism[14]. The screening of a library including 85% of Arabidopsis transcription factors (TFs) resulted in the identification of 27,485 unique interactions between 1,930 TFs and 220 promoters, representing 11 core and specialized metabolic pathways, and uncovered regulators of the tricarboxylic acid cycle. Interestingly, this revealed that TFs involved in the regulation of specific specialized pathways also directly bind to promoters of genes involved in the biosynthesis of the corresponding core precursors as observed for the glucosinolate pathway, which is regulated by TFs that also regulate central carbon metabolism. Importantly, and in contrast to unicellular organisms where distinct transcriptional regulatory networks correspond to specific pathways[15], the regulation of core and specialized pathways in plants thus appear to be controlled by common and multiple TFs in a coordinated manner[14]. However, the question of how and when the balance between core and specialized metabolism is altered, for instance during immune responses, remains unresolved. Several lines of evidence suggest that the conserved Target of Rapamycin (TOR) kinase may play a role in metabolic orchestration in plants. The TOR pathway responds to glucose[16], sulfur[17] and the polyamine spermidine[18]. Importantly, activation of the TOR kinase by glucose leads to repression of core metabolism genes and concurrent activation of specialized metabolic pathway genes[16], and TOR knock-down or overexpression respectively inhibits or promotes plant immunity[19,20]. TOR is a master regulator of translation in eukaryotes, and so may participate in the translationally dependent metabolic shift from core to specialized metabolism required for plant immunity. Interestingly, the translation of enzymes involved in spermidine biosynthesis is promoted by spermidine-mediated TOR activation, thus, establishing an example of metabolite-dependent positive feedback regulatory loop in plants[18]. In the future, the application of systems-level approaches, including strategies that shed light on transcriptional and translational regulation (e.g. ribosome profiling), will provide a fuller picture of the nature and mechanistic basis of the factors governing the dynamic interconnections between core and specialized metabolisms, and of the impact of the environmental factors on this balance.

Functional consequences of the interplay between metabolism and chromatin remodeling

In addition to TF-mediated regulation of plant metabolic pathways, it is clear that core metabolites influence gene expression and metabolic control[21–23]. Notably, S-Adenosylmethionine (SAM)

metabolism broadly modulates plant immunity by regulating DNA methylation and biosynthesis of the ethylene defense phytohormone[24,25]. Recent work also show that metabolic enzymes of primary metabolism can regulate transcription *in situ* by physically interacting with chromatin-remodeling enzymes in plants. Indeed, physical interaction between a SAM synthase and a DNA topoisomerase has been shown to specifically regulate genes interspersed within heterochromatin in Arabidopsis[26]. Likewise, the Arabidopsis myo-inositol phosphate synthase 1 (MIPS1) interacts with the histone methyltransferases ATXR5/6 to regulate its own gene expression. Moreover, the MIPS1-ATXR5/6 module is regulated by the MAPK4 kinase upon defense response activation[27]. At the genome-wide level, defects in SAM profoundly alter DNA methylation and H3K9me2 patterns[28], while the effects of myo-inositol are not yet characterized. In addition, similar to core metabolites, specialized metabolites have also been shown to be directly involved in the regulation of chromatin through the allosteric inhibition of histone deacetylases (HDACs)[29]. The benzoxazinoids, a type of allelochemical alkaloids produced in grasses, directly bind HDACs. Docking simulations and *in vitro* experiment showed that benzoxazinoids inhibit HDACs in a similar way to well-characterized allosteric inhibitors of HDACs. Treatment of plants with benzoxazinoids thus led to marked changes in histone acetylation levels, gene expression and plant growth. Histone modifications also impact synthesis of the potent defense alkaloid camalexin, which is rapidly induced in response to pathogen attack. In this context, the regulation of camalexin biosynthetic genes appears to be in part mediated by two histone modifications with opposing function. The bivalent modification of chromatin with histone H3 lysine 27 trimethylation H3K27me3 (repression) and histone H3 lysine 18 acetylation H3K18ac (activation) was shown to be required for the precise timing of the expression of camalexin biosynthetic genes upon flagellin sensing[30]. Indeed, defense response activation led to a marked change in the ratio of H3K27me3/H3K18ac in favor of H3K18ac that likely mediated camalexin gene activation[30] (Fig. 1a).

A study of the specialized metabolome of epigenetic recombinant inbred lines[31] (epiRILs *i.e.* a set of fixed homozygous lines with different DNA methylation profiles obtained from near-isogenic parents) of Arabidopsis revealed substantial qualitative and quantitative differences between these lines when compared to the wild type and mutant parents[32]. Epigenetic QTL analysis of the epiRIL population identified differentially methylated regions (DMRs) in *cis* and *trans* of specialized metabolic genes. Interestingly, *cis* and *trans* DMRs were associated with changes in expression of metabolic genes, and *trans* DMRs corresponded to sRNAs homologous to metabolic genes, hence providing mechanistic insights into the regulation of genes for specialized metabolism and/or epiallele inheritance[32].

Another independent study correlated the genome-wide DNA methylation patterns and gene expression of 620 geographically diverse Arabidopsis accessions in order to obtain a better understanding of DNA methylation-driven evolution. Strikingly, functional categories having higher expressional variation between accessions corresponded to specialized metabolism genes, notably for terpenoid biosynthetic genes, with differential DNA methylation patterns. DNA methylation can occur in the CG, CHG or CHH contexts (where H is A, C, or T). Importantly, CG methylation in promoter regions of terpenoid biosynthetic genes was positively selected among the 620 Arabidopsis accessions, thereby demonstrating the importance of epigenetic regulation of terpenoid genes in the adaptation of plants to their ecological niches[33]. Finally, the tissue-specific methylome of the MIA-accumulating plant *C. roseus* has revealed contrasted methylation status for most of the TFs involved in MIA metabolism and for some MIA metabolic genes[34].

Thus, metabolites and metabolic enzymes can have unanticipated effects on regulation of gene expression. Future investigations with natural plant populations may reveal insights into the evolutionary forces and mechanisms that govern the regulation of gene expression at the level of chromatin, and into the ways in which plants adapt to their native environments. In addition, the analysis of the natural variability by metabolomics and transcriptomics may result in the discovery of missing metabolic factors through gene regulatory network reconstruction and genome-wide association studies as performed in crops[35–40]. For a biotechnological perspective and upon characterization of the chromatin landscape in medicinal plants, epigenetic modifications of histones and DNA from specialized metabolism genes could be selectively conducted through the use of CRISPR-Cas9 fused to histone/DNA modifying enzymes to engineer natural product biosynthesis. Of note, CRISPR-Cas9-mediated mutagenesis and epi-mutagenesis has been successfully applied in *C. roseus*[41] and crops/Arabidopsis[42], respectively.

Novel mechanisms of evolution of specialized metabolic genes

The evolution of novel plant specialized metabolic pathways is promoted by gene and whole genome duplication, gene neo-functionalization, as well as chromosomal rearrangements, horizontal gene transfer (HGT) and chromatin remodeling[35,43,44]. In this respect, multiple studies have revealed an important role of transposable elements in the evolution of specialized metabolic pathways. For example, in hemp (*Cannabis sativa*), the tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA) synthases steps are bottlenecks for cannabinoid accumulation. Interestingly, genes encoding these enzymes are located in transposable element (TE)-rich regions that are nonhomologous between hemp and other drug-accumulating chemotypes, suggesting that TE-mediated chromosomal rearrangements have likely participated in the emergence of the cannabinoid pathway[45].

Another fascinating example of TE-mediated evolution of specialized metabolisms comes from the synthesis of geraniol in rose petals, which confers a subtle scent. The geraniol-synthesizing Nudix hydrolase 1-1a (NUDX1-1a) has arisen from more ancient and ubiquitous Nudix (NUDX1-1b) hydrolases required for nucleoside diphosphate hydrolysis through a complex series of TE-mediated *trans* and *cis* duplication events[46]. Indeed, NUDX1-1a first emerged from the more ancient NUDX1-1b gene by *trans* duplication. Then, NUDX1-1a *cis*-duplicated, leading to a gene dosage effect on the production of geraniol. After *cis/trans* duplication events that were potentially driven by MITE TEs, the promoter of NUDX1-1a evolved by the incorporation of a *Copia* TE, conferring rose petal-specific expression[46] (Fig. 1b). Although most plant specialized metabolic pathways have likely arisen by successive duplication and neofunctionalization of endogenous genes[35,43,44], at least one example suggests a transfer of metabolic pathway genes between species by HGT as observed in fungi[47,48]. The metabolic genes required for the synthesis of the benzoxazinoid allelochemicals indeed show signatures of HGT from Pooideae to Panicoideae, grass families that have diverged 50 million years ago[49].

Until recently, the mechanisms by which duplicated genes were wired *de novo* into pre-existing metabolic pathways to allow metabolic expansion remained unclear. A seminal study has shown that a duplicated P450 enzyme was neo-functionalized by a TE-mediated mechanism to participate in the biosynthesis of tryptophan-derived defense chemicals upon pathogen attack[50]. This specific LINE (Long Interspersed Nuclear Element) TE ancestor contained a WRKY TF-binding site and

retrotransposed into the P450 *cis*-regulatory region (Fig. 1b). However, and in contrast with the retrotransposed LINE, the LINE ancestor did not display histone modifications compatible with transcriptional activation. Indeed, the LINE ancestor was enriched with H3K27me3, a histone modification characteristic of facultative heterochromatin, that was lost in the *de novo* inserted LINE. In contrast, the retrotransposed LINE was enriched with H3K4me2 modification, a histone modification characteristic of euchromatin and transcriptional activation. Therefore, *de novo* TE insertion in combination with evolutionarily-acquired chromatin features conferred to this specific duplicated P450 gene responsiveness to pathogen attack, with associated triggering of the biosynthesis of tryptophan-derived defense compounds[50]. While *cis* (DNA sequence) and *trans* (DNA-binding protein) factors that regulate specialized metabolism genes represent powerful tools for the genetic engineering of plant natural product biosynthesis[51], reprogramming of cell fate and function through the use of bioinspired synthetic gene regulatory modules will open wider opportunities as recently reported for the reprogramming of Arabidopsis root architecture[52].

Beyond histone and DNA marking, higher-order chromatin regulation of biosynthetic gene clusters

In plants as in other living organisms, non-homologous specialized metabolism genes can be organized in operon-like structures called biosynthetic gene clusters (BGCs)[43,53] on a genomic linear view. To date, the largest known plant BGC occurs in opium poppy, which evolved a 0.5 Mb, is a 15 gene-containing cluster encoding BIA genes that are strongly co-expressed in stem tissues[54]. While the evolution of BGCs has been suggested to be under antagonistic selective pressures - purifying selection of individual biosynthesis genes between species and positive selection of the entire BGC within a species[55], higher-order chromatin regulation also impacts the expression of clustered genes. Indeed, histone and DNA modifications influence 3D chromatin architecture[56,57]. At the 3D level of chromatin organization, neighboring or distant genes can form functional units within the nuclear space called topologically-associated domains (TADs)[58]. The power of chromosome conformation capture assays (Hi-C, capture Hi-C) has recently revealed the TAD structure of BGCs in several plant species, which suggest that TAD formation may be a general principle in the control of BGC expression. With respect to the BIA BGC, it forms a TAD in the chromatin of immature leaves, while its chromatin structure is unknown in stem tissues[59]. In the case of the thalianol triterpenoid BGC in Arabidopsis, the formation of distinct 3D domains correlates with the expression state of the cluster (Fig. 2a, b). Interestingly, the thalianol cluster interacts with active euchromatin in the nucleoplasm in roots, whereas it interacts with silent heterochromatin at the border of the nuclei space in leaves (Figure 2c)[60]. Likewise, BGCs of other triterpenoids such as the marneral and arabidiol/baruol BGCs also form TADs.

In Arabidopsis, 3D domain formation at silenced BGCs is linked to the histone modification H3K27me3 that can be fine-tuned by long non-coding RNAs (lncRNA)[61], notably by the capacity of lncRNAs to decoy chromatin factors required for H3K27me3 patterning[61]. In the case of the marneral BGC, the *Marneral Silencing* (*MARS*) lncRNA modulates the levels of H3K27me3 in response to abscisic acid, which in turn allows the formation of a chromatin loop between the promoter of *Marneral synthase 1* with a distal enhancer for transcriptional activation[62].

In oat, the antifungal avenacin molecule is involved in plant immunity, and the avenacin BGC provides another interesting example with respect to the implication of chromatin in the regulation and/or evolution of BGCs. The avenacin cluster is localized close to the telomere of chromosome 1 in black oat (*Avena strigosa*) and does not share synteny with other cereals. As telomeres are highly dynamic chromosomal regions, these specific sites may be an appropriate landing spot for the rapid evolution of such metabolic genes. Interestingly, avenacin cluster genes are somewhat colinear with the biosynthetic pathway, the late pathway genes being furthest away from the telomeric extremity. It is tempting to speculate that this specific genomic organization may have evolved to avoid telomeric gene deletion of late genes and subsequent toxic intermediate accumulation[63]. In addition, the avenacin BGC is particularly well-expressed in the root tip where extensive cell division and differentiation occur, which is consistent with the high plasticity of plant telomeres and their involvement in the regulation of stress responses and development[64–66]. Earlier work has shown that expression of the avenacin BGC is associated with chromatin decondensation in the root epidermis as observed by molecular cytology[67]. Collectively, these studies show that TAD formation is important for BGCs expression and metabolite production, and therefore represents an attractive target for the biotechnological improvement of specialized metabolites production. For this purpose, we propose chemical-induced TAD formation as a potential strategy to engineer BGC expression. Similar to abscisic acid-induced TAD merging in animal cells[68], this approach could be based on rapamycin-induced dimerization of FKBP and FRB[69] proteins, each in fusion with a CRISPR-Cas9 protein guided to distinct genes of the same BGC (Fig. 2d).

It is striking that in microbes, BGCs also form dynamic TAD structures[70], suggesting that the physical insulation of BGCs in the 3D genomic space is an ancient co-opted mechanism for the regulation of BGC expression. In bacterial and fungal genomes, most BGCs are silent (i.e. not expressed) under regular laboratory conditions[71]. In bacteria and similar to the decoying effect of the *MARS* lncRNA on an H3K27me3-modulating factor as described above, a Streptomyces BGC transcriptional repressor has been tethered away by overexpression of its target sequence, allowing for BGC transcriptional activation and metabolite accumulation[72] (Fig. 3, left panel). This strategy requires prior characterization of the *cis-trans* regulatory modules involved in the regulation of the BGC under investigation. This information has been obtained at the genome-wide level in *Streptomyces albidoflavus* J1074 by screening the transcriptional output of a library of gene *cis*-regulatory elements[73], for which *trans*-interacting factors could then be identified by yeast one hybrid assays, for instance. Also in Streptomyces, CRISPR-Cas9-mediated promoter swapping has been used to replace conditionally active promoters by constitutive promoters to transcriptionally activate a BGC and discover a novel specialized metabolite[74] (Fig. 3, right panel).

Interestingly, in the fungus *Aspergillus nidulans*, stressful growth conditions induce chromatin changes that correlate with BGC induction[75], and specialized metabolism can be influenced by treatment with inhibitors of histone post-translational modification[76]. Although chromatin mutants and chemical treatments have been useful to discover novel specialized metabolites in Aspergillus[76,77], they often induce pleiotropic defects, and so specific tools are desirable. This hurdle has recently been overcome by using a transcriptional activator fused to CRISPR-cas9. It is worth noting that in the context of the eukaryotic chromatin of *A. nidulans*, the successful targeting of BGCs has been achieved by the careful design of guide RNAs in accessible chromatin[78]. To do so for plant specialized metabolites, the main

bottlenecks to date are the lack of chromosome-scale genomes for medicinal plants and the scarcity of knowledge about the mechanisms that control plant BGC expression.

Chromosome-scale genome assemblies of medicinal plants unlock the study of BGCs

One of the best genome assemblies among medicinal plants was recently disclosed for *Ophiorrhiza pumila* and offers new opportunities to decipher the mechanisms that control BGC expression. Along with most Apocynaceae plants, *O. pumila* synthetizes MIAs with well-known pharmacological properties, such as anti-cancer and anti-malarial activities[79]. The availability of a highly contiguous genome assembly for *O. pumila* has enabled the prediction of ~40 MIA BGCs[80]. The genomes of additional MIA-producing plants have also been assembled at chromosome-scale, allowing the prediction of ~60 BGCs in *Neolamarckia cadamba*, for instance[81,82]. Importantly, the recent chromosome-scale genome and chromatin contact map of *C. roseus* allowed the identification of an important MIA transporter that was part of a previously incomplete, TAD-forming BGC[83]. Likewise, three highly-contiguous chromosome-scale genomes have been assembled for Papaver species[59]. Although these new high-quality genome resources greatly enable investigations of MIA and BIA pathway evolution and diversification, the chromatin landscape of medicinal plant BGCs is mostly uncharted. It is now feasible to map histone modifications using chromatin immunoprecipitation followed by deep sequencing, and to correlate this epigenomic map with chromatin accessibility and conformation in a tissue cell type-specific manner. In addition, the application of genome-wide RNA-DNA mapping approaches [e.g. RNA and DNA interacting complexes ligated and sequenced (RADICL-seq) and global RNA interactions with DNA by deep sequencing (GRID-seq)[84]] in these species may be expected to yield regulatory ncRNAs involved in BGC regulation, so opening up opportunities to improve MIA biosynthesis upon elucidation of the underlying mechanisms. Notably, the identification of ncRNAs associated with TAD formation at BGCs could support new strategies for the genetic engineering of specialized metabolites in medicinal plants.

Conclusion and outlook

This perspective article highlights recent advances in the understanding of mechanisms involved in plant specialized metabolism evolution and regulation, which have so far been mostly studied in Arabidopsis and crops. In particular, multi-omics and spatial genomics have uncovered the influence of TF networks, epigenomics and epigenetics on expression of the genes for specialized metabolic pathways and many other cell type-specific functions in these species. Given the spatial segregation of specific parts of specialized metabolic pathways in distinct cell types as shown by spatially-resolved metabolomics[83,85], RNA-seq, chromatin accessibility assay, and ribosome profiling of plant organs at the single-cell resolution could be exploited. Such strategies could help to comprehensively identify *cis* RNA or DNA motifs whose chromatin or ribosome accessibility regulate specialized metabolic genes. The *trans* protein factors that bind the identified *cis* elements could then be screened by yeast one hybrid, searched through TF consensus binding site databases[86], or isolated by RNA-centric affinity purification methods[87]. Likewise, the *trans* ncRNA factors potentially regulating metabolic genes could be identified by RADICL-seq/GRID-seq. Upon discovery of *cis-trans* modules and their mechanisms of

action, manipulation of plant specialized metabolism through the elaboration of bioinspired synthetic genetic circuits[88] may be a promising approach for a more sustainable plant-based food production and alleviate the rarity of some plant-derived drugs. In the case of medicinal plants for which a growing number of chromosome-scale genomes become available, single-cell transcriptomics and metabolomics will facilitate the discovery of new specialized metabolism enzymes and cell type-specific regulatory mechanism[83]. Together with the development and improvement of genetic engineering tools tailored for medicinal plants, high-quality genomes and multi-omics will thus accelerate not only the elucidation of specialized metabolic pathways but will also provide relevant routes for improving metabolic engineering strategies for the production of pharmacologically invaluable plant natural products. Importantly, prior knowledge at the level of networks may help in further optimizing genetically engineered specialized pathways, notably to avoid competition between distinct metabolic pathways for common precursors and thus to favor metabolic channeling.

Acknowledgements

We acknowledge funding from the ANR MIACYC project (ANR-20-CE43-0010), the ARD CVL Biopharmaceutical program of the Région Centre-Val de Loire (ETOPOCentre project), and from Le Studium Loire Valley Institute for Advanced Studies (consortium fellowship). A.O. acknowledges funding from the John Innes Foundation and the BBSRC Institute Strategic Programme Grant 'Molecules from Nature – Products and Pathways' (BBS/E/J/000PR9790). HWN acknowledges support from the University of Bath and the Royal Society (UF160138).

Author contributions

LVM, HWH, NP, AO and VC conceived the project and wrote the article.

Competing interests

The authors declare no competing interests.

Figure Legends

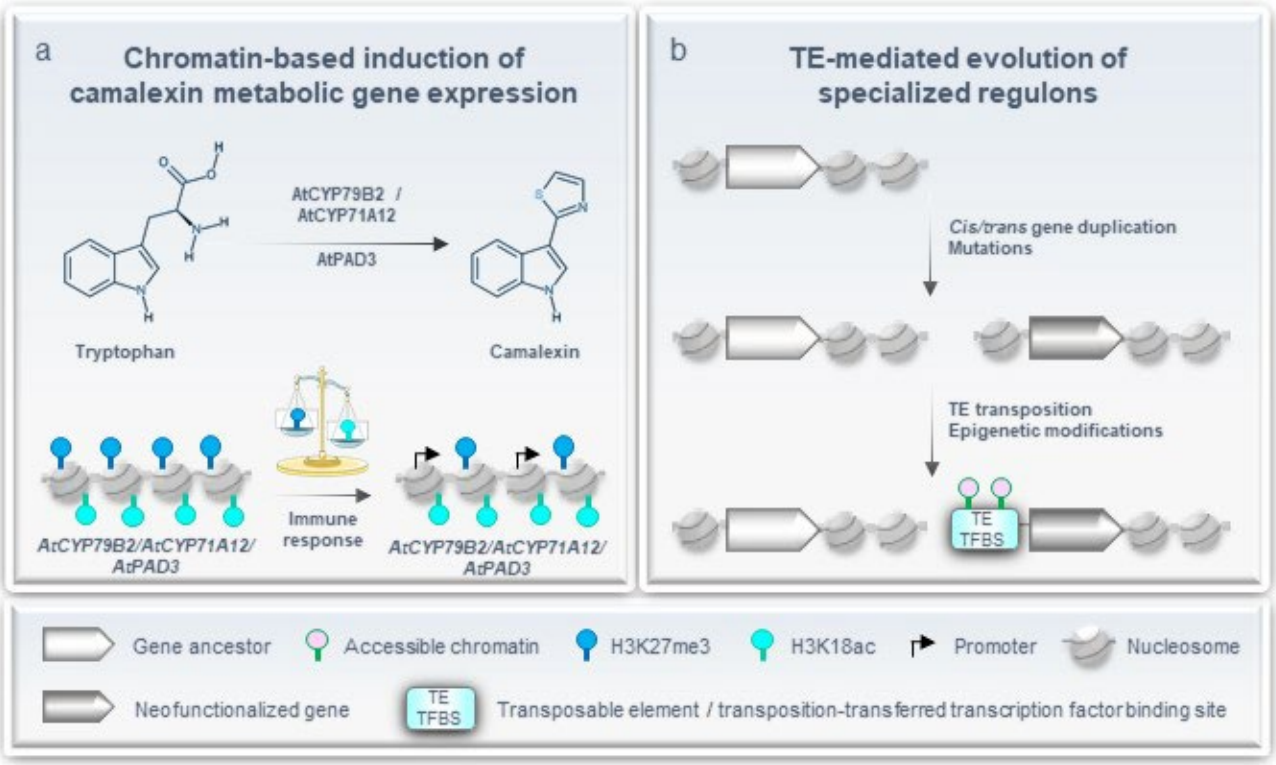

Figure 1. Mechanisms of specialized metabolism gene regulation and evolution. a, Tryptophan is transformed by several enzymes into the camalexin alkaloid, an important plant defense chemical. The chromatin environment of these enzymes is enriched with H3K27me3 and H3K18ac. Upon immune responses activation, the balance between H3K27me3 (transcriptional repression) and H3K18ac (transcriptional activation) changes in favor of H3K18ac, consistent with camalexin gene induction and camalexin synthesis. AtCYP79B2: *AT4G39950*; AtCYP71A12: *AT2G30750*; AtPAD3: *AT3G26830*. **b**, Specialized metabolic genes have evolved from gene ancestors by *cis* and *trans* gene duplication and accumulation of mutations. TE transposition transferred TF-binding sites and epigenetic-mediated regulation to newly evolved genes, thus providing novel regulatory properties.

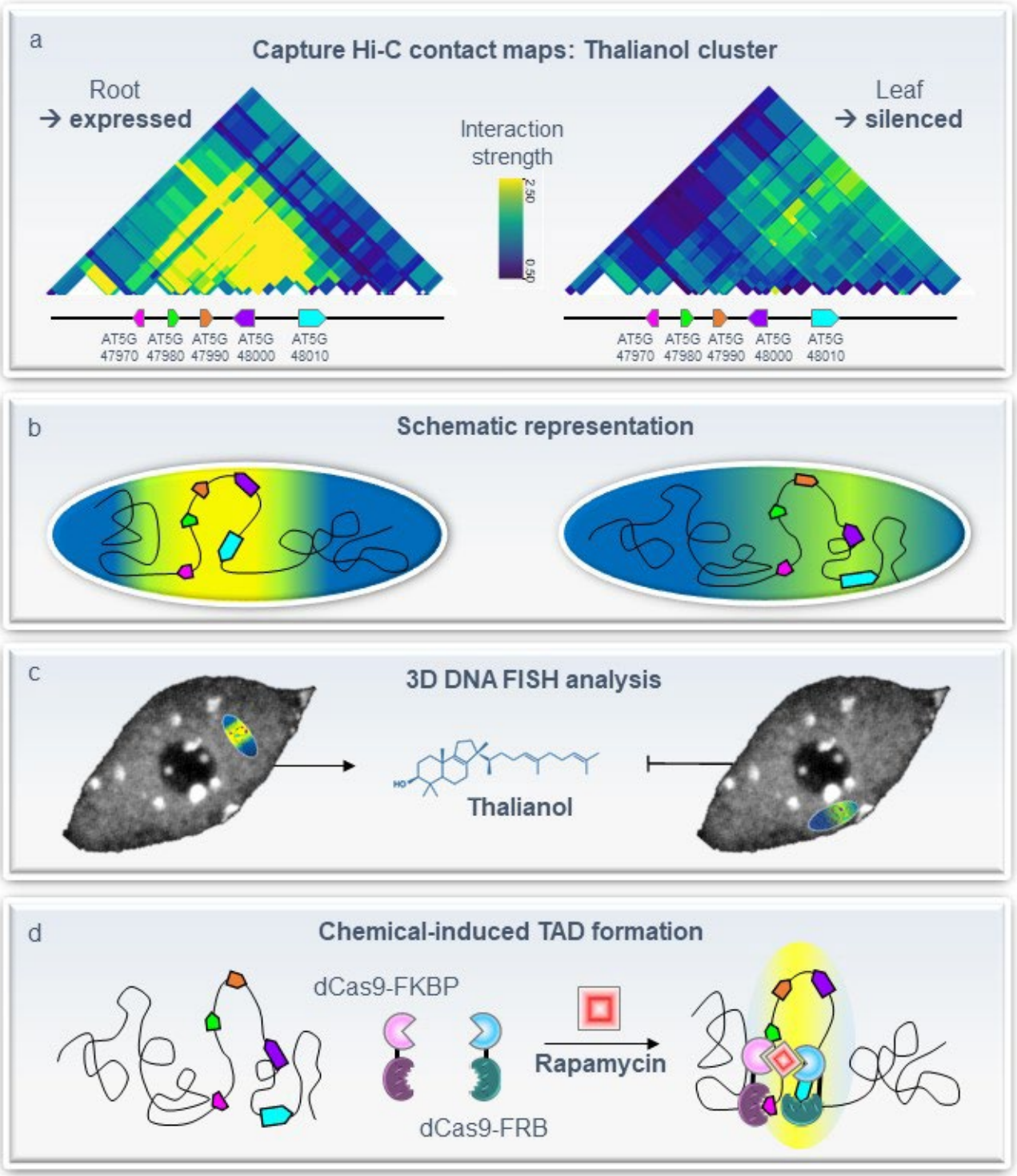

Figure 2. Chromatin architecture regulates thalianol accumulation in Arabidopsis roots. **a**, High resolution chromatin contact maps (Capture Hi-C) revealed the formation of a topologically-associated domain consisting of the thalianol gene cluster. In yellow, the thalianol BGC is clearly insulated away from the neighboring chromatin in root nucleus but not in leaves (adapted from[60]). **b**, Schematic representation of the chromatin organization of the thalianol TAD, as revealed in **a**. **c**, 3D DNA fluorescence *in situ* hybridization experiment have validated Hi-C results. The thalianol TAD is localized in the transcriptionally active nucleoplasm of root nucleus, whereas it is mostly associated to silent heterochromatin at the nuclear periphery in leaf nuclei. d, Chemical-induced TAD formation

relies on the expression of dCas9 proteins fused to FKBP and FRB proteins. Upon rapamycin treatment, FKBP and FRB heterodimerize thus bridging dCas9-targeted biosynthetic genes of the same BGC to potentially form a TAD.

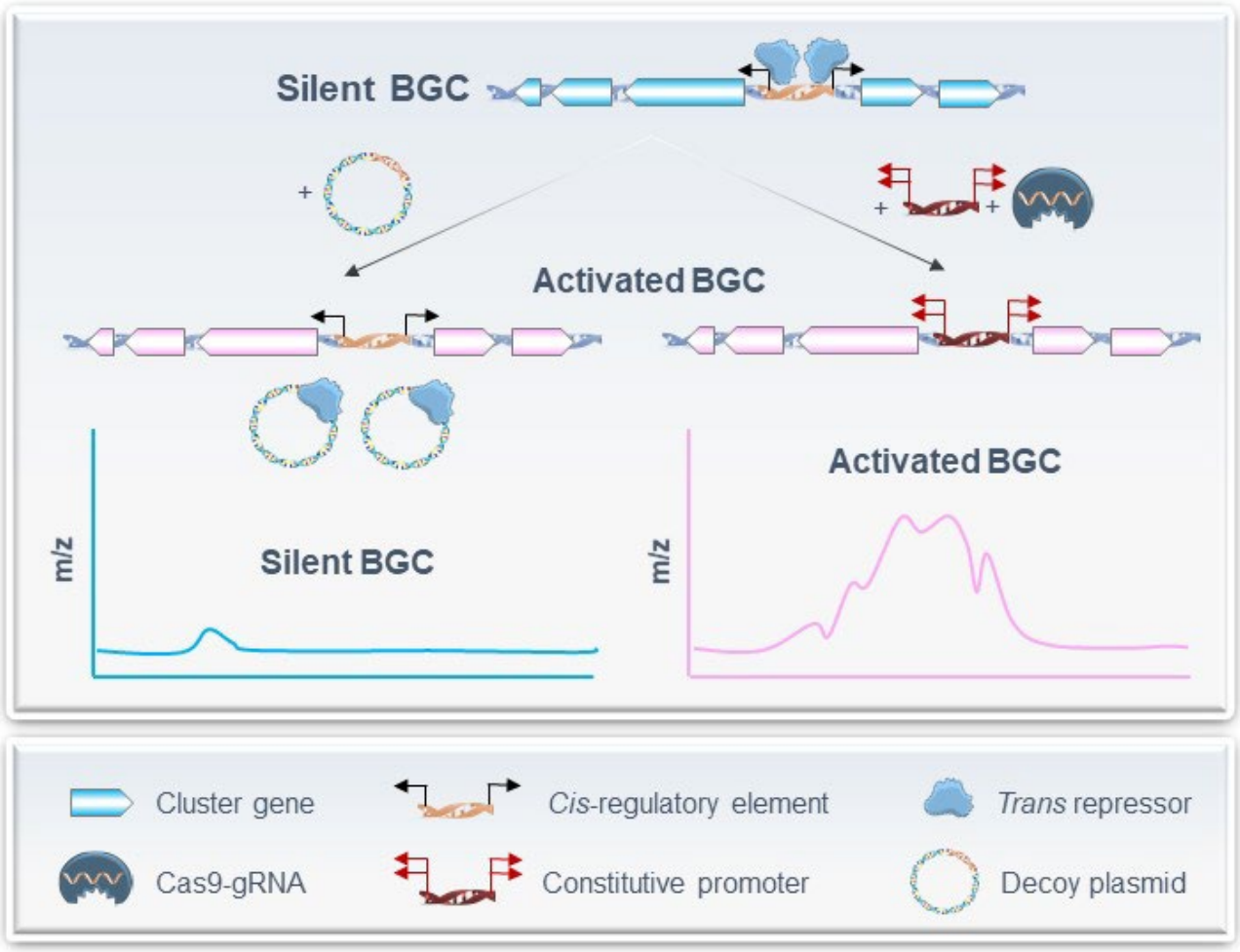

Figure 3. Transcriptional repressor decoying and promoter swapping for BGC activation. In microbes, BGCs are often transcriptionally inactive due to the presence of transcriptional repressors or absence of transcriptional activators in *cis*-regulatory elements. Left panel: A strategy of repressor decoying by plasmid-delivery of repressor target has been successfully used for BGC activation. Right panel: weak or inducible BGC promoters can be replaced by constitutive promoter for enhanced BGC expression. Decoying and promoter swapping approaches have allowed the activation of silent BGCs and discovery of novel specialized metabolites by LC-MS analysis.

Bibliography

1. Carrington, Y. *et al.* Evolution of a secondary metabolic pathway from primary metabolism: shikimate and quinate biosynthesis in plants. *Plant J.* **95**, 823–833 (2018).
2. Barra, L., Awakawa, T., Shirai, K., Hu, Z. & Bashiri, G. β-NAD as a building block in natural product biosynthesis. *Nature* **2021**, (2021).
3. Garagounis, C., Delkis, N. & Papadopoulou, K. K. Unraveling the roles of plant specialized metabolites: using

synthetic biology to design molecular biosensors. *New Phytol.* **231**, 1338–1352 (2021).
4. Wurtzel, E. T. & Kutchan, T. M. Plant metabolism, the diverse chemistry set of the future. *Science (80-.).* **353**, 1232–1236 (2016).
5. Kulagina, N., Méteignier, L., Papon, N., O'Connor, S. E. & Courdavault, V. More than a Catharanthus plant: A multicellular and pluri-organelle alkaloid-producing factory. *Curr. Opin. Plant Biol.* **67**, 102200 (2022).
6. Ozber, N. & Facchini, P. J. Phloem-specific localization of benzylisoquinoline alkaloid metabolism in opium poppy. *J. Plant Physiol.* **271**, 153641 (2022).
7. Nielsen, J. Systems biology of metabolism. *Annu. Rev. Biochem.* **86**, 245–275 (2017).
8. Lacchini, E. & Goossens, A. Combinatorial Control of Plant Specialized Metabolism: Mechanisms, Functions, and Consequences. *Annu. Rev. Cell Dev. Biol.* **36**, 291–313 (2020).
9. Colinas, M. & Goossens, A. Combinatorial Transcriptional Control of Plant Specialized Metabolism. *Trends Plant Sci.* **23**, 324–336 (2018).
10. Zhang, Y. & Fernie, A. R. Metabolons, enzyme–enzyme assemblies that mediate substrate channeling, and their roles in plant metabolism. *Plant Commun.* **2**, 100081 (2021).
11. Zhang, Y. *et al.* A moonlighting role for enzymes of glycolysis in the co-localization of mitochondria and chloroplasts. *Nat. Commun.* **11**, (2020).
12. Mucha, S. *et al.* The formation of a camalexin-biosynthetic metabolon. *Plant Cell* **31**, tpc.00403.2019 (2019).
13. Gaudinier, A. *et al.* Transcriptional regulation of nitrogen-associated metabolism and growth. *Nature* **563**, 259–264 (2018).
14. Tang, M. *et al.* A genome-scale TF–DNA interaction network of transcriptional regulation of Arabidopsis primary and specialized metabolism . *Mol. Syst. Biol.* **17**, 1–19 (2021).
15. Ihmels, J., Levy, R. & Barkai, N. Principles of transcriptional control in the metabolic network of Saccharomyces cerevisiae. *Nat. Biotechnol.* **22**, 86–92 (2004).
16. Xiong, Y. *et al.* Glucose-TOR signalling reprograms the transcriptome and activates meristems. *Nature* **496**, 181–186 (2013).
17. Dong, Y. *et al.* Sulfur availability regulates plant growth via glucose-TOR signaling. *Nat. Commun.* **8**, (2017).
18. Salazar-Díaz, K. *et al.* TOR senses and regulates spermidine metabolism during seedling establishment and growth in maize and Arabidopsis. *iScience* **24**, (2021).
19. Meteignier, L.-V. *et al.* Translatome analysis of an NB-LRR immune response identifies important contributors to plant immunity in Arabidopsis. *J. Exp. Bot.* **68**, 2333–2344 (2017).
20. De Vleesschauwer, D. *et al.* Target of rapamycin signaling orchestrates growth-defense trade-offs in plants. *New Phytol.* (2017) doi:10.1111/nph.14785.
21. Groth, M. *et al.* MTHFD1 controls DNA methylation in Arabidopsis. *Nat. Commun.* **7**, 11640 (2016).
22. Samo, N., Ebert, A., Kopka, J. & Mozgová, I. Plant chromatin, metabolism and development – an intricate crosstalk. *Curr. Opin. Plant Biol.* **61**, 1–11 (2021).
23. Escaray, F., Felipo-Benavent, A. & Vera, P. Linking plant metabolism and immunity through methionine biosynthesis. *Mol. Plant* **15**, 6–8 (2022).
24. González, B. & Vera, P. Folate Metabolism Interferes with Plant Immunity through 1C Methionine Synthase-Directed Genome-wide DNA Methylation Enhancement. *Mol. Plant* **12**, 1227–1242 (2019).
25. Zhai, K. *et al.* NLRs guard metabolism to coordinate pattern- and effector-triggered immunity. *Nature* **601**, (2021).
26. Méteignier, L.-V. *et al.* Topoisomerase VI participates in an insulator-like function that prevents H3K9me2 spreading. *Proc. Natl. Acad. Sci.* **119**, 1–40 (2022).
27. Latrasse, D. *et al.* Dual function of MIPS1 as a metabolic enzyme and transcriptional regulator. *Nucleic Acids Res.* **41**, 2907–2917 (2013).
28. Meng, J. *et al.* METHIONINE ADENOSYLTRANSFERASE4 Mediates DNA and Histone Methylation. *Plant Physiol.* **177**, 652–670 (2018).
29. Venturelli, S. *et al.* Plants release precursors of histone deacetylase inhibitors to suppress growth of competitors. *Plant Cell* **27**, 3175–3189 (2015).
30. Zhao, K., Kong, D., Jin, B., Smolke, C. D. & Rhee, S. Y. A novel form of bivalent chromatin associates with rapid induction of camalexin biosynthesis genes in response to a pathogen signal in arabidopsis. *Elife* **10**, 1–15 (2021).

31. Catoni, M. & Cortijo, S. EpiRILs: Lessons from Arabidopsis. in *Advances in Botanical Research* vol. 88 87–116 (Elsevier Ltd, 2018).
32. Kooke, R. *et al.* Epigenetic mapping of the Arabidopsis metabolome reveals mediators of the epigenotype-phenotype map. *Genome Res.* **29**, 96–106 (2019).
33. Shirai, K. *et al.* Positive selective sweeps of epigenetic mutations regulating specialized metabolites in plants. *Genome Res.* **31**, 1060–1068 (2021).
34. de Bernonville, T. D. *et al.* Developmental methylome of the medicinal plant catharanthus roseus unravels the tissue-specific control of the monoterpene indole alkaloid pathway by dna methylation. *Int. J. Mol. Sci.* **21**, 1–26 (2020).
35. Zhan, C. *et al.* Plant metabolic gene clusters in the multi-omics era. *Trends Plant Sci.* (2022) doi:10.1016/j.tplants.2022.03.002.
36. Zhan, C. *et al.* Selection of a subspecies-specific diterpene gene cluster implicated in rice disease resistance. *Nat. Plants* **6**, 1447–1454 (2020).
37. Polturak, G. *et al.* Pathogen-induced biosynthetic pathways encode defense-related molecules in bread wheat. *Proc. Natl. Acad. Sci.* **119**, (2022).
38. Shen, S. *et al.* An Oryza-specific hydroxycinnamoyl tyramine gene cluster contributes to enhanced disease resistance. *Sci. Bull.* **66**, 2369–2380 (2021).
39. Chen, W. *et al.* Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* **46**, 714–721 (2014).
40. Fang, H. *et al.* A monocot-specific hydroxycinnamoylputrescine gene cluster contributes to immunity and cell death in rice. *Sci. Bull.* **66**, 2381–2393 (2021).
41. Grützner, R. *et al.* High-efficiency genome editing in plants mediated by a Cas9 gene containing multiple introns. *Plant Commun.* **2**, 1–15 (2021).
42. Gardiner, J., Ghoshal, B., Wang, M. & Jacobsen, S. E. CRISPR–Cas-mediated transcriptional control and epi-mutagenesis. *Plant Physiol.* **188**, 1811–1824 (2022).
43. Smit, S. J. & Lichman, B. R. Plant biosynthetic gene clusters in the context of metabolic evolution. *Nat. Prod. Rep.* (2022) doi:10.1039/d2np00005a.
44. Zhou, X. & Liu, Z. Unlocking plant metabolic diversity: A (pan)-genomic view. *Plant Commun.* **3**, 100300 (2022).
45. Laverty, K. U. *et al.* A physical and genetic map of Cannabis sativa identifies extensive rearrangements at the THC/CBD acid synthase loci. *Genome Res.* **29**, 146–156 (2019).
46. Conart, C. *et al.* Duplication and Specialization of NUDX1 in Rosaceae Led to Geraniol Production in Rose Petals. *Mol. Biol. Evol.* **39**, 1–20 (2022).
47. Kominek, J. *et al.* Eukaryotic Acquisition of a Bacterial Operon. *Cell* **176**, 1356-1366.e10 (2019).
48. Xia, J. *et al.* Whitefly hijacks a plant detoxification gene that neutralizes plant toxins. *Cell* **184**, 1693-1705.e17 (2021).
49. Wu, D., Jiang, B., Ye, C.-Y., Timko, M. P. & Fan, L. Horizontal transfer and evolution of the biosynthetic gene cluster for benzoxazinoids in plants. *Plant Commun.* **3**, 100320 (2022).
50. Barco, B., Kim, Y. & Clay, N. K. Expansion of a core regulon by transposable elements promotes Arabidopsis chemical diversity and pathogen defense. *Nat. Commun.* **10**, 3444 (2019).
51. Schweizer, F. *et al.* An engineered combinatorial module of transcription factors boosts production of monoterpenoid indole alkaloids in Catharanthus roseus. *Metab. Eng.* **48**, 150–162 (2018).
52. Brophy, J. A. N. *et al.* Synthetic genetic circuits as a means of reprogramming plant roots. *Science (80-.).* **377**, 747–751 (2022).
53. Bharadwaj, R., Kumar, S. R., Sharma, A. & Sathishkumar, R. Plant Metabolic Gene Clusters: Evolution, Organization, and Their Applications in Synthetic Biology. *Front. Plant Sci.* **12**, 1–23 (2021).
54. Guo, L. *et al.* The opium poppy genome and morphinan production. *Science (80-.).* **362**, 343–347 (2018).
55. Liu, Z. *et al.* Formation and diversification of a paradigm biosynthetic gene cluster in plants. *Nat. Commun.* **11**, 1–11 (2020).
56. Zhao, S. *et al.* Plant HP1 protein ADCP1 links multivalent H3K9 methylation readout to heterochromatin formation. *Cell Res.* **29**, 54–66 (2019).
57. Zhong, Z. *et al.* DNA methylation-linked chromatin accessibility affects genomic architecture in Arabidopsis. *Proc.*

Natl. Acad. Sci. U. S. A. **118**, (2021).
58. Dong, P. *et al.* 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. *Mol. Plant* **10**, 1497–1509 (2017).
59. Yang, X. *et al.* Three chromosome-scale Papaver genomes reveal punctuated patchwork evolution of the morphinan and noscapine biosynthesis pathway. *Nat. Commun.* **12**, 6030 (2021).
60. Nützmann, H. W. *et al.* Active and repressed biosynthetic gene clusters have spatially distinct chromosome states. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 13800–13809 (2020).
61. Ariel, F. *et al.* R-Loop Mediated trans Action of the APOLO Long Noncoding RNA. *Mol. Cell* **77**, 1055-1065.e4 (2020).
62. Roulé, T. *et al.* The lncRNA MARS modulates the epigenetic reprogramming of the marneral cluster in response to ABA. *Mol. Plant* **15**, 840–856 (2022).
63. Li, Y. *et al.* Subtelomeric assembly of a multi-gene pathway for antimicrobial defense compounds in cereals. *Nat. Commun.* **12**, (2021).
64. Choi, J. Y. *et al.* Natural variation in plant telomere length is associated with flowering time. *Plant Cell* **33**, 1118–1134 (2021).
65. Farrell, C. *et al.* A complex network of interactions governs DNA methylation at telomeric regions. *Nucleic Acids Res.* **50**, 1449–1464 (2022).
66. Campitelli, B. E. *et al.* Plasticity, pleiotropy and fitness trade-offs in Arabidopsis genotypes with different telomere lengths. *New Phytol.* **233**, 1939–1952 (2022).
67. Wegel, E., Koumproglou, R., Shaw, P. & Osbourn, A. Cell type-specific chromatin decondensation of a metabolic gene cluster in oats. *Plant Cell* **21**, 3926–3936 (2009).
68. Wang, J. *et al.* Manipulation of TAD reorganization by chemical-dependent genome linking. *STAR Protoc.* **2**, 100799 (2021).
69. Winkler, J. *et al.* Visualizing protein–protein interactions in plants by rapamycin-dependent delocalization. *Plant Cell* **33**, 1101–1117 (2021).
70. Lioy, V. S. *et al.* Dynamics of the compartmentalized Streptomyces chromosome during metabolic differentiation. *Nat. Commun.* **12**, 5221 (2021).
71. Won, T. H. *et al.* Copper starvation induces antimicrobial isocyanide integrated into two distinct biosynthetic pathways in fungi. *Nat. Commun.* **13**, 4828 (2022).
72. Wang, B., Guo, F., Dong, S. H. & Zhao, H. Activation of silent biosynthetic gene clusters using transcription factor decoys. *Nat. Chem. Biol.* **15**, 111–114 (2019).
73. Park, J., Yim, S. S. & Wang, H. H. High-Throughput Transcriptional Characterization of Regulatory Sequences from Bacterial Biosynthetic Gene Clusters. *ACS Synth. Biol.* **10**, 1859–1873 (2021).
74. Zhang, M. M. *et al.* CRISPR-Cas9 strategy for activation of silent Streptomyces biosynthetic gene clusters. *Nat. Chem. Biol.* **13**, 607–609 (2017).
75. Gacek-Matthews, A. *et al.* KdmB, a Jumonji Histone H3 Demethylase, Regulates Genome-Wide H3K4 Trimethylation and Is Required for Normal Induction of Secondary Metabolism in Aspergillus nidulans. *PLoS Genet.* **12**, 1–29 (2016).
76. Keller, N. P. Fungal secondary metabolism: regulation, function and drug discovery. *Nat. Rev. Microbiol.* **17**, 167–180 (2019).
77. Greco, C., Pfannenstiel, B. T., Liu, J. C. & Keller, N. P. Depsipeptide Aspergillicins Revealed by Chromatin Reader Protein Deletion. *ACS Chem. Biol.* **14**, 1121–1128 (2019).
78. Schüller, A. *et al.* A novel fungal gene regulation system based on inducible VPR-dCas9 and nucleosome map-guided sgRNA positioning. *Appl. Microbiol. Biotechnol.* **104**, 9801–9822 (2020).
79. Kumar, S., Singh, B. & Singh, R. Catharanthus roseus (L.) G. Don: A review of its ethnobotany, phytochemistry, ethnopharmacology and toxicities. *J. Ethnopharmacol.* **284**, 114647 (2022).
80. Rai, A. *et al.* Chromosome-level genome assembly of Ophiorrhiza pumila reveals the evolution of camptothecin biosynthesis. *Nat. Commun.* **12**, 1–19 (2021).
81. Kang, M. *et al.* A chromosome-level Camptotheca acuminata genome assembly provides insights into the evolutionary origin of camptothecin biosynthesis. *Nat. Commun.* **12**, 1–12 (2021).
82. Zhao, X. *et al.* Chromosome-level assembly of the Neolamarckia cadamba genome provides insights into the

evolution of cadambine biosynthesis. *Plant J.* **109**, 891–908 (2022).
83. Li, C. *et al.* Single-cell multi-omics enabled discovery of alkaloid biosynthetic pathway genes in the medical plant Catharanthus roseus. *bioRxiv* 1–40 (2022) doi:10.1101/2022.07.04.498697v1.
84. Bonetti, A. *et al.* RADICL-seq identifies general and cell type–specific principles of genome-wide RNA-chromatin interactions. *Nat. Commun.* **11**, 1–14 (2020).
85. Yamamoto, K. *et al.* The complexity of intercellular localisation of alkaloids revealed by single-cell metabolomics. *New Phytol.* **224**, 848–859 (2019).
86. Tian, F., Yang, D. C., Meng, Y. Q., Jin, J. & Gao, G. PlantRegMap: Charting functional regulatory maps in plants. *Nucleic Acids Res.* **48**, D1104–D1113 (2020).
87. Gräwe, C., Stelloo, S., van Hout, F. A. H. & Vermeulen, M. RNA-Centric Methods: Toward the Interactome of Specific RNA Transcripts. *Trends Biotechnol.* **39**, 890–900 (2021).
88. Yaschenko, A. E., Fenech, M., Mazzoni-Putman, S., Alonso, J. M. & Stepanova, A. N. Deciphering the molecular basis of tissue-specific gene expression in plants: Can synthetic biology help? *Curr. Opin. Plant Biol.* **68**, 102241 (2022).