Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Biotechnology

Gene clustering in plant specialized metabolism

Hans-Wilhelm Nützmann and Anne Osbourn

Physically linked clusters of genes that encode the enzymatic information for the synthesis of specialized metabolites are a well-established feature of microbial secondary metabolism. In contrast, the biosynthesis of plant specialized metabolites has until recently been thought to be almost exclusively encoded by genes that are randomly scattered in the genome. However, recent reports highlight the growing number of examples of gene clusters for specialized metabolic pathways in plants. Numerous gene clusters that encode for the biosynthesis of different classes of metabolite have now been discovered in a variety of plant species. Comparison of these characterized clusters now enables us to begin to define their salient features and to exploit plant biosynthetic gene clusters for synthetic biology applications.

Addresses
Department of Metabolic Biology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom

Corresponding author: Osbourn, Anne (anne.osbourn@jic.ac.uk)

Current Opinion in Biotechnology 2014, **26**:91–99

This review comes from a themed issue on **Plant biotechnology**

Edited by **Birger Lindberg Møller** and **R George Ratcliffe**

For a complete overview see the Issue and the Editorial

Available online 16th November 2013

0958-1669
http://dx.doi.org/10.1016/j.copbio.2013.10.009

Introduction

The plant kingdom has a tremendous capacity to synthesize diverse low-molecular weight compounds. These specialized metabolites have important functions in interactions between plants and the environment (e.g. as pest and pathogen defense compounds and UV protectants). The suites of different compounds that are produced by individual plant accessions and species are likely to reflect adaptation to particular environmental niches. Plants are a rich source of valuable compounds including traditional medicines, pharmaceuticals and agrochemicals. However, the vast majority of the plant metabolite reservoir is still uncharacterized, leaving potentially disease-curing compounds undiscovered and hindering biotechnological progress of synthetic approaches to meet the demands for higher value and cheaper chemicals for medicine, agriculture and industry.

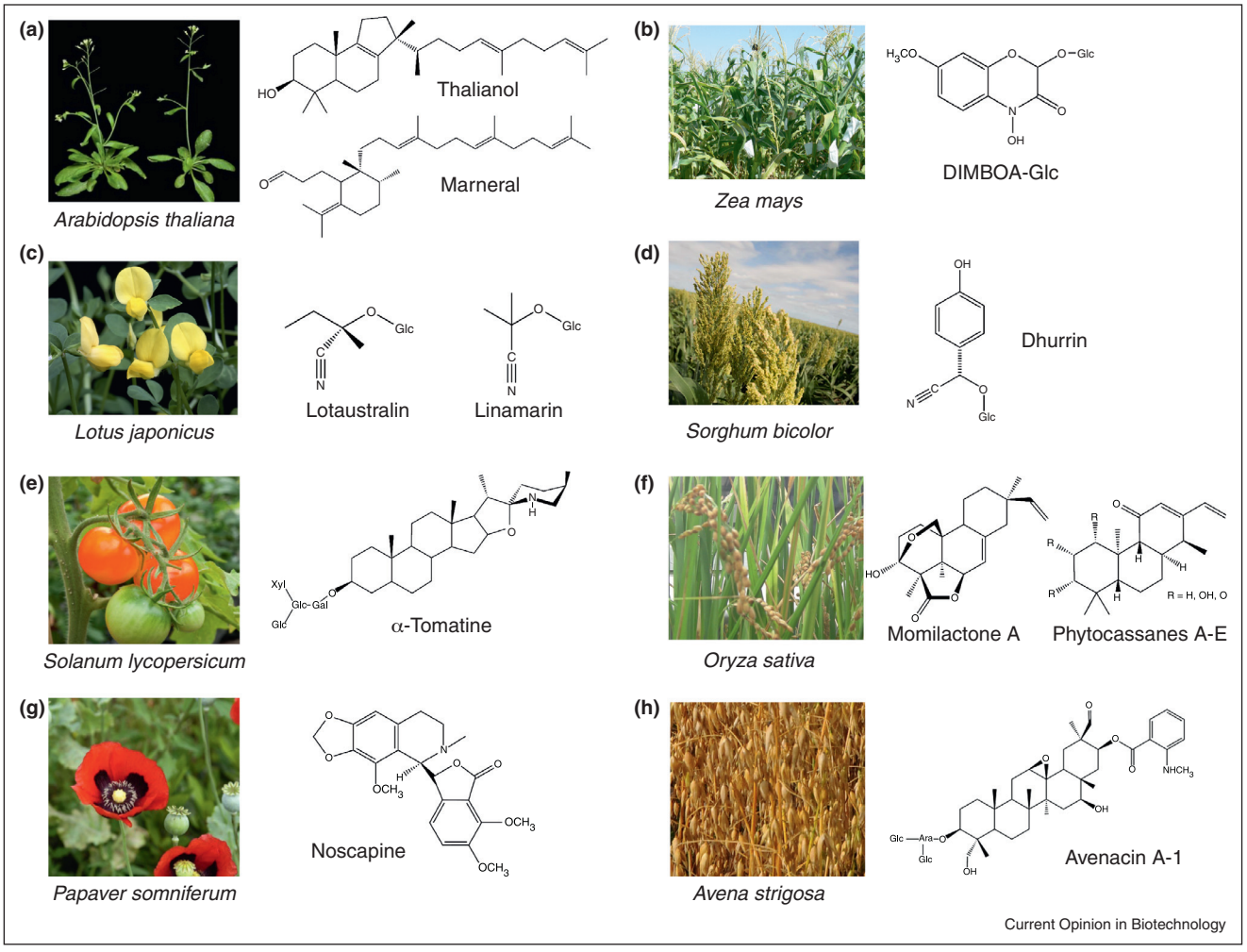
Recent genetic and biochemical studies have highlighted an intriguing facet of plant secondary metabolism, namely the physical clustering of genes for specialized metabolic pathways in plant genomes. It is not yet clear whether clustering of genes for secondary metabolic pathways predominates in plants, as it does in fungi and bacteria; certainly there are well-characterized examples of plant metabolic pathways (e.g. anthocyanins and glucosinolates) for which the genes are not linked [1•]. Nevertheless, the rapidly growing number of reports of metabolic gene clusters for synthesis of diverse classes of compounds from different plant species suggests that this form of genomic organization is common. In contrast the number of pathways for which the genes are known to be dispersed is very limited. It is important to remember that the vast majority of plant specialized metabolic pathways remain as yet undiscovered and their genomic organization is unknown.

In this review we will summarize current knowledge of the plant metabolic gene clusters that have been described so far, define their common features and highlight the similarities and differences. We will also discuss the potential for exploiting plant metabolic gene clusters for biotechnology and synthetic biology applications.

Metabolic gene clusters in plants — no longer the exception to the rule

In 1997 Frey *et al.* reported the first example of physical clustering of the genes for a plant specialized metabolite pathway — for the synthesis of defence compounds in maize (*Zea mays*) [2]. The maize gene cluster was originally defined as a group of five adjacent genes (*Bx1–Bx5*) that encode enzymes for successive steps in the biosynthesis of the cyclic hydroxamic acid 2,4-dihydroxy-1,4-benzoxazin-3-one (DIBOA). Further investigations revealed four more biosynthetic genes (*Bx6–Bx9*) that are required for the conversion of DIBOA to 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA) and subsequent glycosylation, three of which (*Bx6-8*) are either within or genetically linked to the cluster [3–5]. Seven years later two further biosynthesis gene clusters were described, the avenacin cluster in oat (*Avena* spp.) and the phytocassane cluster in rice (*Oryza sativa*) [6,7]. By the beginning of 2012 the number of identified plant secondary metabolite gene clusters had increased to nine [8••,9,10,11••], and within the last year four more clusters have been reported [12••,13••,14••,15••].

Figure 1



Examples of different types of specialized compounds that are the products of plant metabolic gene clusters. **(A)** The triterpenes thalianol and marneral (*A. thaliana*); **(B)** the cyclic hydroxamic acid 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA) (maize; *Zea mays*); **(C, D)** the cyanogenic glucosides lotaustralin and linamarin (*L. japonicus*) **(C)** and dhurrin (*S. bicolor*) (D); **(E)** the steroidal glycoalkaloid α-tomatine (tomato; *S. lycopersicon*); **(F)** the diterpenes momilactone A and phytocassanes A–E (rice; *O. sativa*); **(G)** the alkaloid noscapine (poppy; *P. somniferum*); **(H)** the triterpene avenacin A-1 (oat; *Avena* spp.). Other examples (not shown) include a terpene cluster from *S. lycopersicon*, a cluster for synthesis of cyangenic glucosides in cassava (*M. esculenta*), a triterpene cluster in *L. japonicus* and an α-chaconine/α-solanine gene cluster in potato (*S. tuberosum*). The images of plants are reproduced with the kind permission of the John Innes Centre Photographic Services (A, C, E); Paul Cristou, Institució Catalana de Recerca I Estudis Avançats, Lleida, Spain; (B), Arthur Mostead, Murray-Darling Basin Authority, Australia (D); Uta Paszkowski, University of Cambridge, UK (F); Tanja Niggendijker/Creative Commons (G); Anthony Pugh, Institute for Biological, Environmental and Rural Sciences, Aberystwyth, UK (H).

These biosynthetic gene clusters have been found in diverse plant species, including monocots and dicots, and are required for the synthesis of different classes of molecules, including terpenes, alkaloids and cyanogenic glycosides (Figure 1). A common feature is the location of at least three non-homologous biosynthetic genes for a distinct chemical pathway adjacent to one another in the genome. One gene encodes the signature enzyme that defines the scaffold of the specialized metabolite, and a variable number of additional genes encode the tailoring enzymes that modify this initial scaffold to catalyze the formation of the pathway end-product [16]. The signature genes within these plant gene clusters appear to have evolved directly or indirectly from genes for primary metabolism by gene duplication and neofunctionalisation [17]. The newly formed signature gene then seeds the formation of a metabolic gene cluster through recruitment of additional genes encoding tailoring enzymes [8••,18•]. Comparative genomics is beginning to shed light on mechanisms of cluster formation [8••,11••,13••,18•]. Interestingly, as shown for the cyanogenic glucoside gene clusters, in some cases specialized metabolic gene clusters for similar metabolites have evolved several times independently in different plant species [11••].

The currently described gene clusters span regions of ~35–270 kb and consist of three to ten genes (Figure 2). Some gene clusters, such as the cyanogenic glucoside cluster in *Lotus japonicus*, contain additional genes with no obvious function in secondary metabolism, whereas other clusters (such as the oat avenacin cluster and the *A. thaliana* thalianol cluster) are compact and do not contain intervening genes [6,9,11••,13••]. The majority of the genes within each cluster are co-expressed, so enabling co-ordinate production of the pathway enzymes in a tissue-specific and time-specific manner. However, although all cluster genes show co-expression in at least one highly specific set of conditions, individual cluster genes may also be transcribed separately under other conditions [11••,13••,19]. The multifunctional phytocassane gene cluster in rice is exceptional. Two partly overlapping gene clusters form one giant cluster that shows differential gene transcription profiles for its sub-cluster-specific genes [20,21]. Cluster-independent gene expression presumably enables synthesis of a pathway intermediate rather than the end-product, which may be desirable in certain tissues/under certain conditions. Furthermore, it may allow the utilization of enzymes encoded by clustered genes in other pathways. Interestingly, the steroidal alkaloid gene clusters in tomato (*Solanum lycopersicum*) and potato (*Solanum tuberosum*) are each split into two gene clusters that reside on different chromosomes yet are co-regulated [15••]. The ability to synthesize DIBOA is not restricted to maize. While most plant specialized metabolic gene clusters are likely to have arisen relatively recently in evolutionary time, the DIBOA cluster is believed to have formed in an ancestral monocot. Wheat (*Triticum aestivum*) and rye (*Secale cereale*) are also able to synthesize this compound but the DI(M)-BOA cluster is split into two in these species, most likely due to a translocation event that occurred after a common wheat/rye ancestor diverged from the maize lineage. Nevertheless the pathways are functional, providing further examples of split clusters [22,23]. Some metabolic clusters are able to synthesize more than one major product, although the reasons for this differ. For example, the enzymes encoded by the *L. japonicus* cyanogenic glucoside cluster are able to use different precursor amino acids as the starting point, so catalyzing the formation of linamarin and lotaustralin [11••]. The main products of the steroidal alkaloid gene cluster in *S. tuberosum* are α-solanine and α-chaconine. These metabolites differ only in a sugar moiety, exemplifying the formation of two different products due to variable tailoring of the scaffold [15••]. The variety of different terpenes originating from a single gene cluster in *Solanum* species are formed due to the existence of several related terpene synthase genes within the gene cluster. These genes are most likely duplicates generated from an ancient terpene synthase gene that is still represented within the cluster, the activities of the duplicated terpene synthases subsequently diverging [13••].

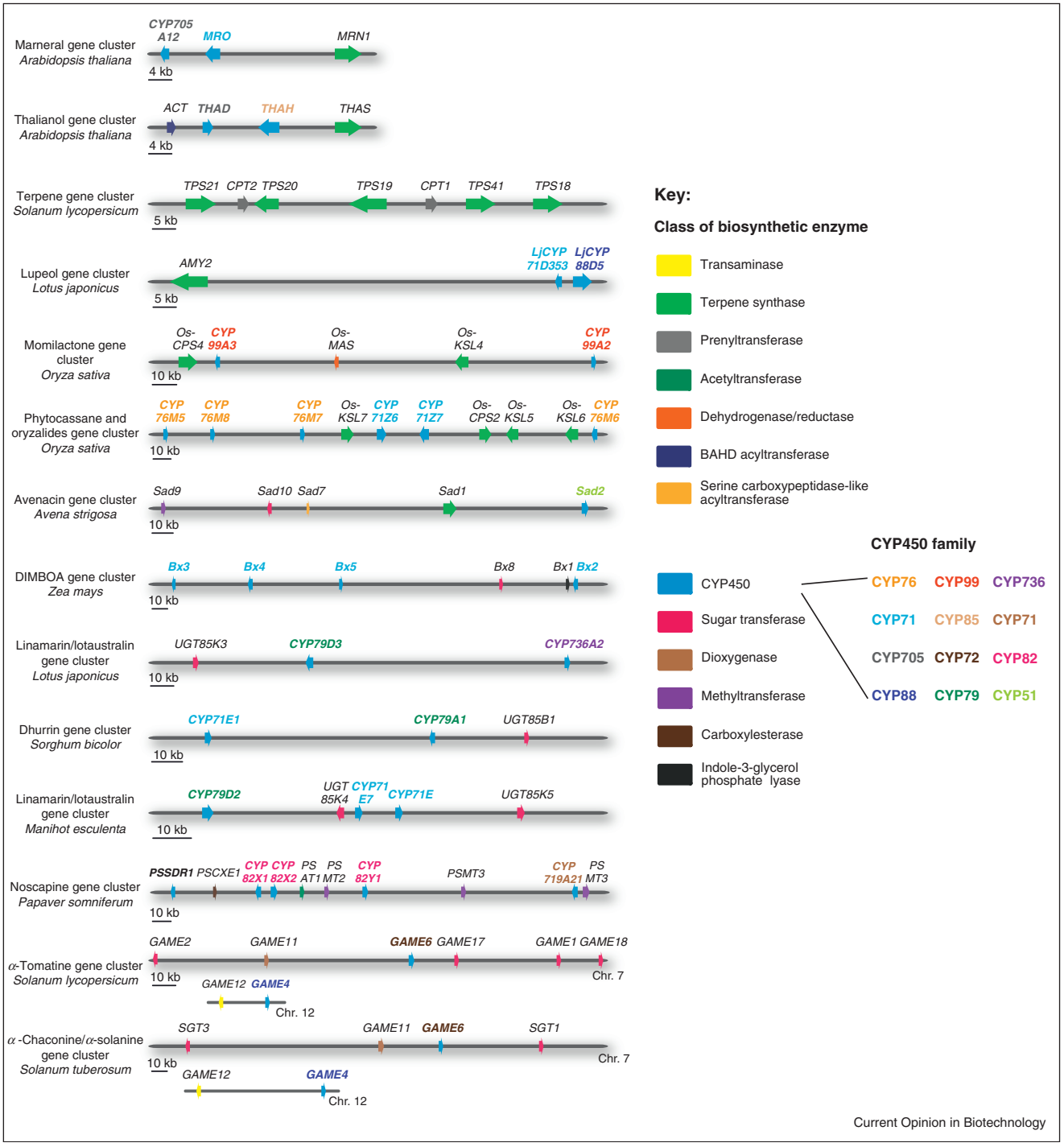
Metabolic gene clusters in plants — tools for synthetic biology

The discovery of gene clusters for synthesis of specialized metabolites in plants is reminiscent of earlier findings in bacteria and fungi [17]. A typical feature of bacterial genomes is the organization of genes for multi-step processes in operons and clusters. The discovery several decades ago of biosynthetic gene clusters in bacteria has paved the way for the rational manipulation of pathways for the synthesis of antibiotics and other bioactives, and for the discovery of novel metabolites. Gene clusters for the synthesis of specialized metabolites are also a common feature of the genomes of filamentous fungi. The advent of affordable genome sequencing techniques has enabled microbial genomes to be mined for their full complement of candidate biosynthetic gene clusters [24], so allowing the discovery of new pathways and novel metabolites in previously untapped microbes [25–28,29•].

Most of the metabolic gene clusters that have been reported in plants to date have been discovered by serendipity, using a combination of genetics and biochemistry. However, it is now becoming possible to exploit genome sequence information for the discovery of new clustered metabolic pathways in plants [8••,9,30•,31•]. In the future this is likely to be accelerated by the development of customized bioinformatics pipelines for analysis of plant genomes along similar lines to those established for microbes (e.g. antiSMASH, SMURF and ClusterMine360; [32••,33,34]), thus allowing the identification of regions of plant genomes that contain clusters of genes for predicted signature and tailoring enzymes and so have the hallmarks of candidate specialized metabolic gene clusters. A drawback for plant researchers is the fact that plant genomes are significantly larger than microbial genomes. The small size of bacterial genomes facilitates both genome sequencing and subsequent genome mining for genes and pathways of interest. However, rapid advances in sequencing technology coupled with development of appropriate genome-mining tools will position plant researchers to use strategies similar to those taken with microbes for gene cluster discovery. The body of available plant genome sequence information is increasing rapidly and genome sequencing projects are now underway even for the exceptionally large gymnosperm genomes [35,36]. The identification of candidate biosynthetic gene clusters based on genome sequence analysis offers access to complete biosynthetic pathways for new specialized metabolites.

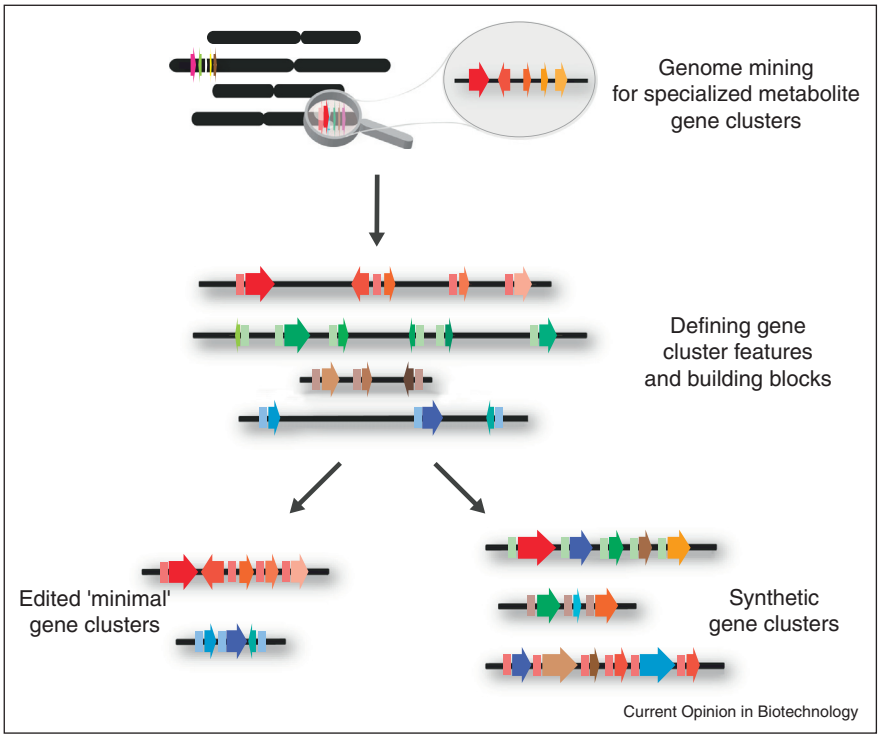
Clearly the genomics-based discovery of candidate metabolic gene clusters is only the starting point. Biochemical and chemical analyses will be essential in order to delineate these new predicted pathways and identify their end-products. We also need to understand how amenable these metabolic gene clusters are to engineering. The

Figure 2



The organization and architecture of characterized gene clusters for the synthesis of specialized metabolites in plants. The arrows representing the genes within the clusters are colour-coded according to the class of biosynthetic enzyme that they encode; the labels above the CYP450 genes (blue arrows) are also colour-coded to indicate the family of CYP450 to which the gene products belong (see key). For the marneral, thalianol, avenacin and α-tomatine gene clusters no other genes are evident other than those shown. Intervening genes lacking predicted functions in secondary metabolism are present in the other clusters but are not shown in the figure due to uncertainties about precise genome annotation. The maize DIMBOA pathway includes three genes that are not shown in the figure, namely the methyltransferase gene *Bx7*, which is separated from the core cluster by an intervening region of 15 Mb; the sugar transferase gene *Bx9*, which is located on a different chromosome; finally, a further gene *Bx6* is not shown because its genomic location has not yet been established. Note that the structure of this cluster has been revised since our previous review [1] in response to the increased genome sequence information now available for this region. Gene clusters similar to the terpene gene cluster shown for *Solanum lycopersicon* in this figure (third from the top) are also present in *Solanum pimpinellifolium*, *Solanum pennellii* and *Solanum habrochaites* [13].

Figure 3



Towards synthetic clusters. The phenomenon of clustering of genes for specialized metabolic pathways is now opening up exciting opportunities for large-scale mining of multiple plant genomes for the discovery of new pathways and chemistries. Characterization of the components of plant metabolic gene clusters (promoters, coding sequences, regulatory sequences, intergenic regions) coupled with biochemical characterization of the cognate enzymes, modules and pathways will enable the establishment of an inventory of parts that can be used in synthetic biology applications. These applications may include synthesis of streamlined minimal clusters that are optimized for transfer into plants. There is also the potential to generate synthetic clusters with novel functions by combining the building blocks of different gene clusters.

biosynthetic genes form only a small part of these clusters and the function of the intervening DNA sequences in cluster function and regulation is not yet known. This raises two important questions. Firstly, what are the requirements for a 'minimal' functional cluster? Secondly, can functional clusters be built up from defined components? Consequently, two key goals can be identified (Figure 3): (i) the construction of minimal functional clusters that have been edited to remove all non-essential sequences. This will be important for the engineering of cluster-encoded multi-gene traits in plants; (ii) the construction of synthetic clusters that combine promoter sets, terminators and coding sequences for suites of signature and tailoring enzymes from different gene clusters. This will enable the generation of designer clusters for expression in heterologous hosts, which will be important both for plant engineering and for the production of high value specialized metabolites in plant or microbial systems. Introduction of designer clusters into heterologous hosts will enable temporal and spatial control of specialized metabolite production and the discovery of new molecules through combinatorial biosynthesis, as has been amply demonstrated for microbes [37].

New recombination and DNA assembly techniques now offer routes to the rapid, reliable and precise construction of large DNA fragments [38–40,41•,42••,43]. Synthetic biology approaches for production of plant-derived specialized metabolites by metabolic engineering have so far been carried out primarily in yeast (*Saccharomyces cerevisiae*) and to lesser extent in *Escherichia coli* [44•,45•,46]. The potential of yeast for production of plant specialized metabolites has been highlighted by the genetic engineering of strains that provide the precursor of artemisinin, a major antimalarial drug, on an industrial scale [47••,48]. Heterologous expression platforms have been established for the production of specialized metabolites from various different sources. For example, genetically engineered *Streptomyces coelicolor* and *Aspergillus nidulans* strains provide convenient hosts for expression of multiple biosynthetic genes from *Actinomycetes* and filamentous fungi, respectively, and facilitate the biochemical analysis of the introduced biochemical pathways [49,50]. The tobacco species *Nicotiana tabacum* and *Nicotiana benthamiana* have emerged as hosts for the heterologous expression of biosynthetic genes and production of specialized metabolites in plants [51–58]. This can be achieved by generation of stable

transformants [51,52,55,56]. This is, however, a very slow process. *Agrobacterium*-mediated transient expression in *N. benthamiana* leaves can be achieved within a matter of days and minimizes any problems associated with detrimental effects of heterologous metabolites on the plant host [53,57,58,59•]. The Cow Pea Mosaic Virus HyperTrans (CPMV-*HT*) expression system has proven to be a highly effective tool for the rapid, transient expression of a variety of proteins, including plant biosynthetic enzymes in *N. benthamiana* leaves [14,53,54,59•,60].

In bacteria and fungi, gene clusters for the synthesis of specialized metabolites are controlled at multiple levels [61,62]. Manipulation of these regulatory mechanisms using genetic and chemical approaches can result in activation of these microbial clusters with associated production of metabolites of interest [63–67]. The identification of regulatory processes that govern the expression of plant metabolite gene clusters will enable similar approaches to be taken in plants. So far, only one transcriptional regulator has been described for a plant metabolic gene cluster [68]. Interestingly, overexpression or deletion of this transcription factor had substantial effects on the metabolite production level of the targeted biosynthetic pathway [68]. Plant metabolic gene clusters are also likely to be regulated at the level of chromatin [8,9,19], opening up opportunities to activate/repress cluster expression following methods similar to those used in filamentous fungi [64,69].

Conclusions

The growing number of reports of clustered genes for biosynthesis pathways in plants has established a new avenue of research in plant biology and natural product discovery. These clusters together provide a critical mass of information that is now beginning to enable the commonalities and unique features of plant clusters to be defined. Increased knowledge of plant metabolic gene clusters will enhance future genome mining efforts for discovery of new pathways and chemistries and the development of biotechnological pipelines to exploit the output of this. Important tasks to tackle will be the generation of broadly applicable search engines for metabolic gene clusters in the increasing number of sequenced plant genomes, the identification of the regulatory mechanisms governing gene cluster expression, the definition of essential building blocks and the uncovering of the evolutionary forces behind the formation and maintenance of metabolic gene clusters.

Acknowledgements

This work was supported by the UK Biotechnological and Biological Sciences Research Council (BBSRC) Institute Strategic Programme Grant 'Understanding and Exploiting Plant and Microbial Secondary Metabolism' (BB/J004561/1), the John Innes Foundation, and Engineering and Physical Sciences Research Council grant EP/K03459/1 (AO), and by Marie Curie Actions and an EMBO Long-Term Fellowship to H.-W.N.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. • Kliebenstein DJ, Osbourn A: **Making new molecules — evolution of pathways for novel metabolites in plants**. *Curr Opin Plant Biol* 2012, **15**:415-423.

A recent review on mechanisms of metabolic diversification in plants that summarizes the literature on clustered genes for specialized metabolic pathways up to 2012, including earlier seminal papers such as the discovery of the first metabolic gene cluster in plants — for the synthesis of DIBOA in maize [2].

2. Frey M, Chomet P, Glawischnig E, Stettner C, Grun S, Winklmair A, Eisenreich W, Bacher A, Meeley RB, Briggs SP *et al.*: **Analysis of a chemical plant defense mechanism in grasses**. *Science* 1997, **277**:696-699.

3. von Rad U, Huttl R, Lottspeich F, Gierl A, Frey M: **Two glucosyltransferases are involved in detoxification of benzoxazinoids in maize**. *Plant J* 2001, **28**:633-642.

4. Frey M, Huber K, Park WJ, Sicker D, Lindberg P, Meeley RB, Simmons CR, Yalpani N, Gierl A: **A 2-oxoglutarate-dependent dioxygenase is integrated in DIMBOA-biosynthesis**. *Phytochemistry* 2003, **62**:371-376.

5. Jonczyk R, Schmidt H, Osterrieder A, Fiesselmann A, Schullehner K, Haslbeck M, Sicker D, Hofmann D, Yalpani N, Simmons C *et al.*: **Elucidation of the final reactions of DIMBOA-glucoside biosynthesis in maize: characterization of *Bx6* and *Bx7***. *Plant Physiol* 2008, **146**:1053-1063.

6. Qi X, Bakht S, Leggett M, Maxwell C, Melton R, Osbourn A: **A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants**. *Proc Natl Acad Sci U S A* 2004, **101**:8233-8238.

7. Wilderman PR, Xu M, Jin Y, Coates RM, Peters RJ: **Identification of syn-pimara-7,15-diene synthase reveals functional clustering of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis**. *Plant Physiol* 2004, **135**:2098-2105.

8. •• Field B, Fiston-Lavier AS, Kemen A, Geisler K, Quesneville H, Osbourn AE: **Formation of plant metabolic gene clusters within dynamic chromosomal regions**. *Proc Natl Acad Sci U S A* 2011, **108**:16116-16121.

This paper reports the discovery and characterization of a gene cluster for triterpene biosynthesis (the marneral cluster) in *A. thaliana*. A different *A. thaliana* triterpene biosynthetic gene cluster (the thalianol cluster) had previously been reported (see Ref. [9]). These clusters were both predicted to be new metabolic gene clusters using genome mining approaches and subsequently validated experimentally. Both clusters have strong repressive histone H3 lysine 27 trimethylation markings suggestive of chromatin-level regulation. The clusters formed after the α whole-genome duplication event within the Brassicales and are located in dynamic chromosomal regions that are significantly enriched in transposable elements. A model for cluster formation is presented.

9. Field B, Osbourn AE: **Metabolic diversification—independent assembly of operon-like gene clusters in different plants**. *Science* 2008, **320**:543-547.

10. Shimura K, Okada A, Okada K, Jikumaru Y, Ko KW, Toyomasu T, Sassa T, Hasegawa M, Kodama O, Shibuya N *et al.*: **Identification of a biosynthetic gene cluster in rice for momilactones**. *J Biol Chem* 2007, **282**:34013-34018.

11. •• Takos AM, Knudsen C, Lai D, Kannangara R, Mikkelsen L, Motawia MS, Olsen CE, Sato S, Tabata S, Jorgensen K *et al.*: **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* 2011, **68**:273-286.

Cyanogenic glucosides have traditionally been regarded as an ancient family of plant specialized metabolite. However, this paper shows that the genes for the synthesis of these compounds are clustered in three different plant species (*L. japonicus*, sorghum and cassava) and provides evidence that these clusters appear to be examples of repeated (or convergent) evolution.

12. •• Winzer T, Gazda V, He Z, Kaminski F, Kern M, Larson TR, Li Y, Meade F, Teodor R, Vaistij FE *et al.*: **A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine**. *Science* 2012, **336**:1704-1708.
Genetic analysis of poppy (*P. somniferum*) accessions differing in their ability to produce the alkaloid noscapine indicated that the genes for the High Noscapine 1 (HN1) phenotype were tightly linked, suggesting that they might occur as a gene cluster. Noscapine levels are much lower in heterozygotes than would be expected for a semi-dominant trait, suggesting some form of repression. Bacterial artificial chromosome sequencing revealed a cluster of ten physically linked, co-expressed genes for noscapine synthesis.

13. •• Matsuba Y, Nguyen TT, Wiegert K, Falara V, Gonzales-Vigil E, Leong B, Schafer P, Kudrna D, Wing RA, Bolger AM *et al.*: **Evolution of a complex locus for terpene biosynthesis in solanum**. *Plant Cell* 2013, **25**:2022-2036.
This paper reports a cluster of genes for synthesis of terpenes in tomato (*S. lycopersicon*) and investigates the evolution of terpene biosynthetic genes and gene clusters within the *Solanaceae*. A model for the evolution of a functional gene cluster for terpene biosynthesis in several *Solanum* species is presented. This elegant combination of genomic, phylogenetic, and biochemical analyses indicates dynamic processes of gene accretion and divergent biochemical evolution associated with metabolic diversification.

14. •• Krokida A, Delis C, Geisler K, Garagounis C, Tsikou D, Pena-Rodriguez LM, Katsarou D, Field B, Osbourn AE, Papadopoulou KK: **A metabolic gene cluster in *Lotus japonicus* discloses novel enzyme functions and products in triterpene biosynthesis**. *New Phytol* 2013.
A triterpene biosynthetic gene cluster that is expressed in the roots and nodules of the model legume *L. japonicus* is reported. The genes within this cluster are coordinately expressed in response to developmental and environmental cues and treatment with plant growth hormones. Co-expression of the triterpene synthase signature enzyme with a candidate tailoring enzyme (a cytochrome P450) encoded by another gene within the cluster using the CPMV-*HT* transient expression system in *N. benthamiana* enabled the activity of the tailoring enzyme towards the triterpene scaffold to be demonstrated. As is the case for other plant triterpene biosynthesis gene clusters [6,8,9], this cluster has arisen *de novo* within recent evolutionary history; the clusters do not share a common origin. Silencing of the triterpene synthase gene resulted in short, stunted roots, suggestive of a role for this pathway in development. Interestingly, hairpin-mediated gene silencing of any of the genes in the cluster induced DNA methylation and thus repression of gene expression in the adjacent genes within the cluster, suggestive of epigenetic regulation.

15. •• Itkin M, Heinig U, Tzfadia O, Bhide AJ, Shinde B, Cardenas PD, Bocobza SE, Unger T, Malitsky S, Finkers R *et al.*: **Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes**. *Science* 2013, **341**:175-179.
A combination of high-throughput transcriptomics coupled with chemical profiling and genomic analysis has led to the discovery of clustered genes for the synthesis of steroidal glycoalkaloids in tomato and potato. Most of these genes are organised as one large cluster, with two other pathway genes clustered elsewhere in the genome. This finding paves the way to rational manipulation of the levels of these toxic substances in potato tubers and tomato fruit.

16. Osbourn A: **Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation**. *Trends Genet* 2010, **26**:449-457.

17. Chu HY, Wegel E, Osbourn A: **From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants**. *Plant J* 2011, **66**:66-79.

18. • Dutartre L, Hilliou F, Feyereisen R: **Phylogenomics of the benzoxazinoid biosynthetic pathway of Poaceae: gene duplications and origin of the *Bx* cluster**. *BMC Evol Biol* 2012, **12**:64.
The origins of the DIMBOA gene cluster in the Poaceae are investigated using a phylogenetics-based approach. Clustering of the genes encoding the first two enzymes in the pathway is proposed to be a key founding event in cluster formation, followed by elongation events that involved recruitment of additional genes to the pathway.

19. Wegel E, Koumproglou R, Shaw P, Osbourn A: **Cell type-specific chromatin decondensation of a metabolic gene cluster in oats**. *Plant Cell* 2009, **21**:3926-3936.

20. Okada A, Shimizu T, Okada K, Kuzuyama T, Koga J, Shibuya N, Nojiri H, Yamane H: **Elicitor induced activation of the methylerythritol phosphate pathway toward phytoalexins biosynthesis in rice**. *Plant Mol Biol* 2007, **65**:177-187.

21. Wang Q, Hillwig ML, Okada K, Yamazaki K, Wu YS, Swaminathan S, Yamane H, Peters RJ: **Characterization of CYP76M5-8 indicates metabolic plasticity within a plant biosynthetic gene cluster**. *J Biol Chem* 2012, **287**:6159-6168.

22. Nomura T, Ishihara A, Imaishi H, Endo TR, Ohkawa H, Iwamura H: **Molecular characterization and chromosomal localization of cytochrome P450 genes involved in the biosynthesis of cyclic hydroxamic acids in hexaploid wheat**. *Mol Genet Genomics* 2002, **267**:210-217.

23. Nomura T, Ishihara A, Imaishi H, Ohkawa H, Endo TR, Iwamura H: **Rearrangement of the genes for the biosynthesis of benzoxazinones in the evolution of Triticeae species**. *Planta* 2003, **217**:776-782.

24. Winter JM, Behnken S, Hertweck C: **Genomics-inspired discovery of natural products**. *Curr Opin Chem Biol* 2011, **15**:22-31.

25. Lincke T, Behnken S, Ishida K, Roth M, Hertweck C: **Closthioamide: an unprecedented polythioamide antibiotic from the strictly anaerobic bacterium *Clostridium cellulolyticum***. *Angew Chem Int Ed Engl* 2010, **49**:2011-2013.

26. Feng Z, Kallifidas D, Brady SF: **Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites**. *Proc Natl Acad Sci U S A* 2011, **108**:12629-12634.

27. Franke J, Ishida K, Hertweck C: **Genomics-driven discovery of burkholderic acid, a noncanonical, cryptic polyketide from human pathogenic *Burkholderia* species**. *Angew Chem Int Ed Engl* 2012, **51**:11611-11615.

28. Young TS, Walsh CT: **Identification of the thiazolyl peptide GE37468 gene cluster from *Streptomyces* ATCC 55365 and heterologous expression in *Streptomyces lividans***. *Proc Natl Acad Sci U S A* 2011, **108**:13053-13058.

29. • Chang FY, Brady SF: **Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening**. *Proc Natl Acad Sci U S A* 2013, **110**:2478-2483.
This paper reports the screening of soil metagenomes for discovery of gene clusters for the synthesis of rare, biomedically relevant natural products. Screening of large environmental DNA (eDNA) libraries from diverse sources is opening up opportunities for finding new enzymes and pathways.

30. • Osbourn A, Papadopoulou KK, Qi X, Field B, Wegel E: **Finding and analyzing plant metabolic gene clusters**. *Methods Enzymol* 2012, **517**:113-138.
This methods article describes approaches for the identification of secondary metabolic gene clusters in plants through forward and reverse genetics, map-based cloning, and genome mining and gives examples of methods used for the analysis and functional confirmation of new clusters.

31. • Castillo DA, Kolesnikova MD, Matsuda SP: **An effective strategy for exploring unknown metabolic pathways by genome mining**. *J Am Chem Soc* 2013, **135**:5885-5894.
Heterologous expression and 2D NMR is used to study the enzymatic oxidation of three *A. thaliana* triterpenes, thalianol, marneral and arabidiol. Gene clusters for the synthesis and modification of thalianol and marneral were previously identified using genome mining approaches [8,9]. This paper reports a functional connection between arabidiol synthase and a cytochrome P450 encoded by a co-regulated physically linked gene, so providing experimental support for clustering of genes for arabidiol metabolism.

32. •• Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T: **antiSMASH 2.0 — a versatile platform for genome mining of secondary metabolite producers**. *Nucleic Acids Res* 2013, **41**:W204-W212.
This paper describes a comprehensive bioinformatics pipeline for systematic mining of microbial genomes for gene clusters for secondary metabolic pathways.

33. Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND: **SMURF: genomic mapping of fungal secondary metabolite clusters**. *Fungal Genet Biol* 2010, **47**:736-741.

34. Conway KR, Boddy CN: **ClusterMine360: a database of microbial PKS/NRPS biosynthesis**. *Nucleic Acids Res* 2013, **41**:D402-D407.

35. Mackay J, Dean JF, Plomion C, Peterson DG, Canovas FM, Pavy N, Ingvarsson PK, Savolainen O, Guevara MA, Fluch S *et al.*: **Towards decoding the conifer giga-genome**. *Plant Mol Biol* 2012, **80**:555-569.

36. Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A *et al.*: **The Norway spruce genome sequence and conifer genome evolution**. *Nature* 2013, **497**:579-584.

37. Menzella HG, Reid R, Carney JR, Chandran SS, Reisinger SJ, Patel KG, Hopwood DA, Santi DV: **Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes**. *Nat Biotechnol* 2005, **23**:1171-1176.

38. Schmid-Burgk JL, Xie Z, Frank S, Winter SV, Mitschka S, Kolanus W, Murray A, Benenson Y: **Rapid hierarchical assembly of medium-size DNA cassettes**. *Nucleic Acids Res* 2012, **40**:e92.

39. Wingler LM, Cornish VW: **Reiterative recombination for the *in vivo* assembly of libraries of multigene pathways**. *Proc Natl Acad Sci U S A* 2011, **108**:15135-15140.

40. Xu P, Vansiri A, Bhan N, Koffas MAG: **ePathBrick: a synthetic biology platform for engineering metabolic pathways in *E. coli***. *ACS Synth Biol* 2012, **1**:256-266.

41. • Ellis T, Adie T, Baldwin GS: **DNA assembly for synthetic biology: from parts to pathways and beyond**. *Integr Biol* 2011, **3**:109-118.

Substantial progress has been made over the last few years in the development of methods for assembly of DNA from small fragments into large constructs and even into synthetic chromosomes and genomes. As the cost of commercial DNA synthesis has gone down, DNA assembly is increasingly becoming a pivotal technology for synthetic biology applications. This article provides an accessible overview of the various DNA assembly strategies available and their relative merits.

42. •• Werner S, Engler C, Weber E, Gruetzner R, Marillonnet S: **Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system**. *Bioeng Bugs* 2012, **3**:38-43.

The Golden Gate cloning method, which is based on the use of type IIs restriction enzymes and restriction-ligation, allows high efficiency assembly of multiple fragments from different plasmids in a one-pot reaction. To eliminate the need for construct-specific cloning strategies a modular cloning system (MoClo) has been developed that allows any multigene construct of choice to be made by using a defined set of pre-made vectors and a defined assembly strategy. As an example, this paper demonstrates the assembly of 68 DNA fragments encoding basic genetic elements using three one-pot cloning steps, resulting in a 50 kb construct containing 17 eukaryotic transcription units.

43. Engler C, Marillonnet S: **Generation of families of construct variants using Golden Gate shuffling**. *Methods Mol Biol* 2011, **729**:167-181.

44. • Cyr A, Wilderman PR, Determan M, Peters RJ: **A modular approach for facile biosynthesis of labdane-related diterpenes**. *J Am Chem Soc* 2007, **129**:6684-6685.

Labdane-related diterpenoids have a characteristic core structure that is generated from the universal diterpenoid precursor geranylgeranyl diphosphate (GGPP). Cyclization of GGPP to a specific stereoisomer of labdadienyl/copalyl diphosphate (CPP) is catalyzed by class II diterpene cyclases. This core structure is then further cyclized and/or rearranged by class I diterpene synthases. Here, a modular approach is used in which base vectors capable of making any one of three stereoisomers of CPP in *E. coli* can be co-expressed with diterpene synthases that convert these CPP intermediates to different diterpene skeletons. As proof of concept, *E. coli* was engineered to produce eight different diterpene skeletons, which can collectively serve as precursors to thousands of distinct natural products.

45. • Nakagawa A, Minami H, Kim JS, Koyanagi T, Katayama T, Sato F, Kumagai H: **A bacterial platform for fermentative production of plant alkaloids**. *Nat Commun* 2011, **2**:326.

Here an *E. coli* fermentation system is developed for the production of plant alkaloids, using selected enzymes to construct a tailor-made synthetic pathway.

46. Siddiqui MS, Thodey K, Trenchard I, Smolke CD: **Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools**. *FEMS Yeast Res* 2012, **12**:144-170.

47. •• Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D *et al.*: **High-level semi-synthetic production of the potent antimalarial artemisinin**. *Nature* 2013, **496**:528-532.

Here synthetic biology is used to develop strains of yeast (*S. cerevisiae*) for high-yielding production of artemisinic acid, a precursor of the antimalaria drug artemisinin. The transformation of artemisinic acid to artemisinin was then achieved by chemical conversion. The developments in yeast strain engineering, fermentation and synthetic chemistry reported here pave the way for industrial bioengineering of microbes for production of plant-inspired high-value chemicals.

48. Westfall PJ, Pitera DJ, Lenihan JR, Eng D, Woolard FX, Regentin R, Horning T, Tsuruta H, Melis DJ, Owens A *et al.*: **Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin**. *Proc Natl Acad Sci U S A* 2012, **109**:E111-E118.

49. Gomez-Escribano JP, Bibb MJ: **Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters**. *Microb Biotechnol* 2011, **4**:207-215.

50. Chiang YM, Oakley CE, Ahuja M, Entwistle R, Schultz A, Chang SL, Sung CT, Wang CC, Oakley BR: **An efficient system for heterologous expression of secondary metabolite genes in *Aspergillus nidulans***. *J Am Chem Soc* 2013, **135**:7720-7731.

51. Farhi M, Marhevka E, Ben-Ari J, Algamas-Dimantov A, Liang Z, Zeevi V, Edelbaum O, Spitzer-Rimon B, Abeliovich H, Schwartz B *et al.*: **Generation of the potent anti-malarial drug artemisinin in tobacco**. *Nat Biotechnol* 2011, **29**:1072-1074.

52. Kumar S, Hahn FM, Baidoo E, Kahlon TS, Wood DF, McMahan CM, Cornish K, Keasling JD, Daniell H, Whalen MC: **Remodeling the isoprenoid pathway in tobacco by expressing the cytoplasmic mevalonate pathway in chloroplasts**. *Metab Eng* 2012, **14**:19-28.

53. Mugford ST, Louveau T, Melton R, Qi X, Bakht S, Hill L, Tsurushima T, Honkanen S, Rosser SJ, Lomonossoff GP *et al.*: **Modularity of plant metabolic gene clusters: a trio of linked genes that are collectively required for acylation of triterpenes in oat**. *Plant Cell* 2013, **25**:1078-1092.

54. Sainsbury F, Saxena P, Geisler K, Osbourn A, Lomonossoff GP: **Using a virus-derived system to manipulate plant natural product biosynthetic pathways**. *Methods Enzymol* 2012, **517**:185-202.

55. Wu S, Jiang Z, Kempinski C, Eric Nybo S, Husodo S, Williams R, Chappell J: **Engineering triterpene metabolism in tobacco**. *Planta* 2012, **236**:867-877.

56. Zhang Y, Nowak G, Reed DW, Covello PS: **The production of artemisinin precursors in tobacco**. *Plant Biotechnol J* 2011, **9**:445-454.

57. Liu Q, Majdi M, Cankar K, Goedbloed M, Charnikhova T, Verstappen FW, de Vos RC, Beekwilder J, van der Krol S, Bouwmeester HJ: **Reconstitution of the costunolide biosynthetic pathway in yeast and *Nicotiana benthamiana***. *PLoS ONE* 2011, **6**:e23255.

58. van Herpen TWJM, Cankar K, Nogueira M, Bosch D, Bouwmeester HJ, Beekwilder J: ***Nicotiana benthamiana* as a production platform for artemisinin precursors**. *PLoS ONE* 2010:5.

59. • Geisler K, Hughes RK, Sainsbury F, Lomonossoff GP, Rejzek M, Fairhurst S, Olsen CE, Motawia MS, Melton RE, Hemmings AM *et al.*: **Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants**. *Proc Natl Acad Sci U S A* 2013, **110**:E3360-E3367.

Here the CPMV-*HT* transient plant expression system was used to co-express the first and second enzymes in the oat avenacin pathway — the triterpene synthase signature enzyme SAD1 and an unusual multifunctional cytochrome P450 enzyme that carries out epoxidation and hydroxylation of the triterpene scaffold, and the structure of the modified triterpene determined. This approach, coupled with expression of later pathway enzymes using the same system [9], opens up opportunities for synthetic biology-based approaches for engineering known and novel triterpene pathways and has broad relevance for metabolic engineering more widely.

60. Mugford ST, Qi XQ, Bakht S, Hill L, Wegel E, Hughes RK, Papadopoulou K, Melton R, Philo M, Sainsbury F *et al.*: **A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats**. *Plant Cell* 2009, **21**:2473-2484.

61. Brakhage AF A.: **Regulation of fungal secondary metabolism**. *Nat Rev Microbiol* 2013, **11**:21-32.

62. van Wezel GP, McDowall KJ: **The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances**. *Nat Prod Rep* 2011, **28**:1311-1333.

63. Bergmann S, Schumann J, Scherlach K, Lange C, Brakhage AA, Hertweck C: **Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans***. *Nat Chem Biol* 2007, **3**:213-217.

64. Bok JW, Chiang YM, Szewczyk E, Reyes-Dominguez Y, Davidson AD, Sanchez JF, Lo HC, Watanabe K, Strauss J, Oakley BR *et al.*: **Chromatin-level regulation of biosynthetic gene clusters**. *Nat Chem Biol* 2009, **5**:462-464.

65. Scherlach K, Nützmann HW, Schroeckh V, Dahse HM, Brakhage AA, Hertweck C: **Cytotoxic pheofungins from an engineered fungus impaired in posttranslational protein modification**. *Angew Chem Int Ed Engl* 2011, **50**:9843-9847.

66. Aigle B, Corre C: **Waking up *Streptomyces* secondary metabolism by constitutive expression of activators or genetic disruption of repressors**. *Methods Enzymol* 2012, **517**:343-366.

67. Gao C, Hindra, Mulder D, Yin C, Elliot MA: **Crp is a global regulator of antibiotic production in *Streptomyces***. *MBio* 2012:3.

68. Okada A, Okada K, Miyamoto K, Koga J, Shibuya N, Nojiri H, Yamane H: **OsTGAP1 a bZIP transcription factor, coordinately regulates the inductive production of diterpenoid phytoalexins in rice**. *J Biol Chem* 2009, **284**:26510-26518.

69. Nützmann HW, Reyes-Dominguez Y, Scherlach K, Schroeckh V, Horn F, Gacek A, Schumann J, Hertweck C, Strauss J, Brakhage AA: **Bacteria-induced natural product formation in the fungus *Aspergillus nidulans* requires Saga/Ada-mediated histone acetylation**. *Proc Natl Acad Sci U S A* 2011, **108**:14282-14287.