Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Harnessing plant metabolic diversity

Charlie Owen[1], Nicola J Patron[2], Ancheng Huang[1] and Anne Osbourn[1]

Advances in DNA sequencing and synthesis technologies in the twenty-first century are now making it possible to build large-scale pipelines for engineering plant natural product pathways into heterologous production species using synthetic biology approaches. The ability to decode the chemical potential of plants by sequencing their transcriptomes and/or genomes and to then use this information as an instruction manual to make drugs and other high-value chemicals is opening up new routes to harness the vast chemical diversity of the Plant Kingdom. Here we describe recent progress in methods for pathway discovery, DNA synthesis and assembly, and expression of engineered pathways in heterologous hosts. We also highlight the importance of standardization and the challenges associated with dataset integration in the drive to build a systematic framework for effective harnessing of plant metabolic diversity.

Addresses
[1] Department of Metabolic Biology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK
[2] Engineering Biology, The Earlham Institute, Norwich Research Park, Norwich NR4 7UZ, UK

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

Current Opinion in Chemical Biology 2017, **40**:24–30

This review comes from a themed issue on **Synthetic biology**

Edited by **Tom Ellis** and **Michael Jewett**

For a complete overview see the Issue and the Editorial

Available online 17th May 2017

http://dx.doi.org/10.1016/j.cbpa.2017.04.015

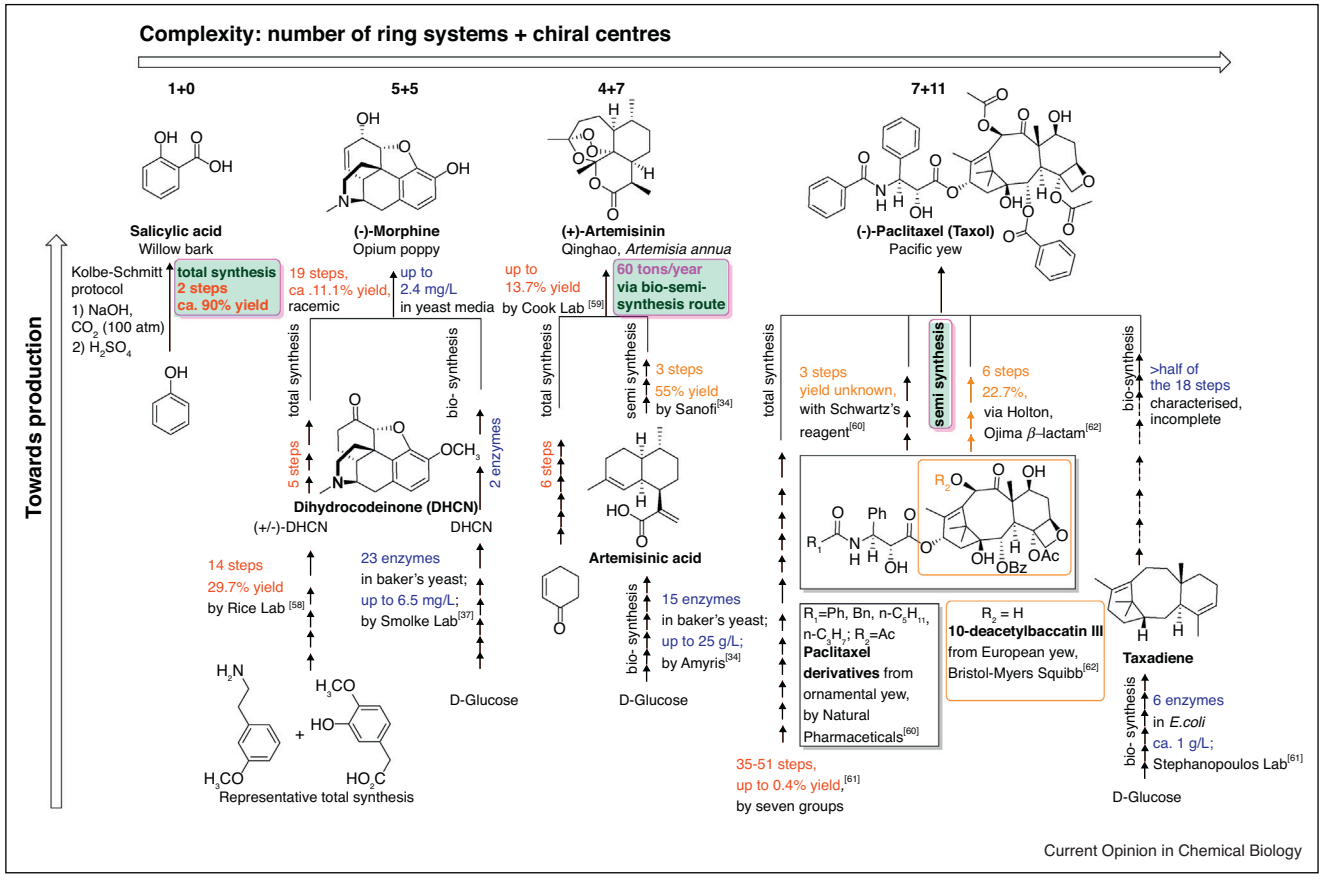
Introduction

Plants have long been recognized as a rich source of drugs and traditional medicines. During the nineteenth century, advances in chemistry enabled many bioactive natural products to be purified from plants and their structures determined, paving the way for the emergence of the pharmaceutical industry. The first naturally derived pure medicine to be commercialized was morphine, extracted from opium by H.E. Merck (Darmstadt, Germany) in 1826. Subsequently efforts were made to make natural products by chemical synthesis to facilitate high quality low cost production. The first natural product to be made in this way was salicylic acid, produced commercially by Bayer in 1899 (Figure 1). Despite these early examples and other subsequent successes, the vast majority of the metabolic diversity harboured within the Plant Kingdom has remained untapped due to the problems of accessing and cultivating source species, purifying low-abundance compounds from plant extracts, and making quantities of structurally complex molecules by chemical synthesis. Advances in genomics and bioinformatics are now greatly accelerating the discovery of new natural product pathways in plants, opening up the possibility of accessing new medicines and other valuable compounds by expression of these pathway genes in heterologous production species (*i.e.* biosynthesis, rather than chemical synthesis). Coupled with the catalyst of synthetic biology, there are now unprecedented opportunities for reading, writing and modifying the chemical information that determines plant metabolic diversity for medicinal, agricultural and industrial biotechnology applications.

Mining plant metabolic diversity

Until recently, discovery and elucidation of pathways for the production of plant specialized metabolites has relied on characterization of individual pathway steps. Now, however, the avalanche of available sequence data from plants is enabling the process of identifying and delineating entire new natural product pathways to be fast-tracked. Massive amounts of transcriptomic data are starting to be made available for diverse plant species through large consortia endeavours such as the Medicinal Plants Genome Resource [1], PhytoMetaSyn [2] and the 1000 plants (1KP) project [3•]. Natural product biosynthesis in plants is usually tightly controlled, occurring in particular cell types, at specific development stages and/or in response to environmental triggers (*e.g.* pathogen attack or elicitor treatment). This feature can be exploited to identify sets of co-expressed candidate genes implicated in the synthesis of target compounds. For example, monoterpene indole alkaloid biosynthesis in Madagascan periwinkle is induced by the elicitor methyl jasmonate, enabling the identification of new pathway genes using RNAseq [4,5]. The delineation of the ten-step pathway for biosynthesis of the aglycone of the chemotherapy drug etoposide in the medicinal plant mayapple (*Podophyllum hexandrum*) was similarly enabled by exploiting the fact that this pathway is wound-inducible [6••]. Where one or more characterized genes are already available for a pathway for the synthesis of a known compound, these can be used as bait in co-expression analysis to identify the missing steps in the pathway,

Figure 1



Some examples of drugs derived from plant natural products and efforts to develop routes to generating these [58–62]. Numbers of steps required and yields are shown: chemical synthesis (red), semi-synthesis (orange), bioengineering (blue). Commercialized routes to production (*i.e.* other than extraction from the native producing plants) are shown in the green boxes.

as for the two above examples. It is also possible, however, to identify entirely new pathways without prior knowledge of the genes or pathway end-product through transcriptomics, as exemplified by the discovery of a new defense-related cyanogenic glycoside pathway in the model plant *Arabidopsis thaliana* by co-expression analysis and untargeted metabolomics using a pathogen-induced cytochrome P450 gene as a starting point [7•].

Generation of sequence resources for natural product pathway discovery in plants has so far concentrated on transcriptome analysis. Plant genome sequencing is more challenging and expensive, particularly for large genome species such as mayapple (genome size ~16 Gb) [6••], and so transcriptomics enables more cost-effective harvesting of the 'low hanging fruit'. Perhaps more pertinently, until recently there was no obvious advantage to having a genome sequence for identifying new natural product biosynthetic genes and pathways since the genes for such pathways were generally assumed to be distributed around the genome. However it has recently emerged that the genes for many plant natural product pathways are organised as biosynthetic gene clusters (BGCs) in plant genomes, with around 30 examples reported to date [8•]. This presents some interesting parallels with bacteria and fungi, where it has long been known that the genes for specialized metabolic pathways are physically clustered. It is clear that the plant BGCs have not originated by horizontal gene transfer from microbes, but rather have formed by reorganization and neofunctionalization of genes within plant genomes [8•]. Nevertheless, this form of genomic organization opens up opportunities for discovery of new natural product pathways in plant genomes using systematic cluster mining algorithms, a strategy that has revolutionized pathway discovery in microbes [9•].

Standard genome mining is common practice in plant science, often relying on manual inspection of outputs of homology search tools such as the Basic Local Alignment

Search Tool (BLAST) [10] and Hidden Markov Modelling (HMMER) [11]. This process can be partially automated via custom scripts/software and has been successfully applied for systematic mining and analysis of multiple plant genomes for all predicted terpene synthase and cytochrome P450 genes, enabling the discovery of known and new BGCs in the process [12•]. However this strategy requires curated collections of research-specific genes of interest and subsequent manual annotation of candidate gene function. Genome annotation pipelines have been very successful in the automation of these steps in general, but such tools are not particularly targeted for unpicking the biochemical potential of plant species. The rapidly growing numbers of plant BGCs have catalyzed the development of new BGC mining tools customized for use in plants, such as plantiSMASH [13] and PhytoClust [14]. Both of these tools use a framework similar to that of antiSMASH [15], an algorithm that is widely used for mining for BGCs in microbes and that relies on pre-defined, curated profile hidden Markov models (pHMMs) to identify genomic loci encoding multiple different families of enzymes associated with natural product biosynthesis. The plantiSMASH and PhytoClust pipelines also draw on transcriptome data where available to provide additional support for BGC predictions based on co-expression data. Although these BGC mining methods are prescriptive in that they rely on a suite of pre-specified pHMM profiles, as more plant BGCs are experimentally characterized these search methods will become iteratively more refined. An ultimate aspiration would be to be able to not only predict BCGs in plant genomes but to also accurately predict the chemistry encoded by these clustered pathways. Analysis of BGCs from *A. thaliana*, oat maize and rice has revealed that these clusters have pronounced chromatin markings, providing further hallmarks for cluster discovery using genome-wide chromatin immunoprecipitation data where available [16].

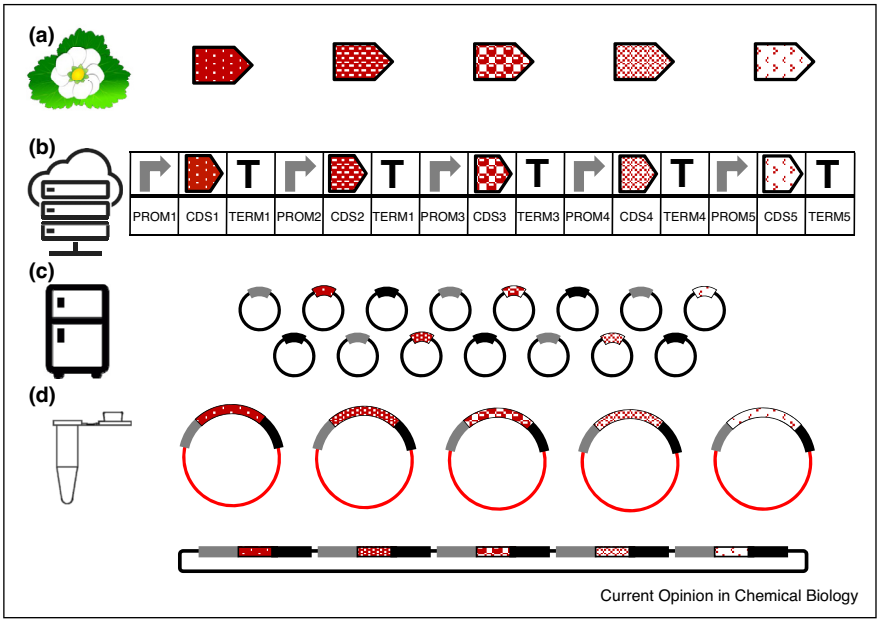
The use of standards and centralized resources for reporting and sharing data will be critical for effective harnessing of plant metabolic diversity. Numerous databases and consortia exist that collate information about enzyme families and their functions (*e.g.* the Carbohydrate-Active Enzymes database [17], the Cytochrome P450 Homepage [18], BioCatNet [19]), transcriptome databases for medicinal plants [1,2], and for linking known metabolic and chemical diversity with medicinal and/or industrial relevance and the wider academic literature (*e.g.* KNApSAcK for cataloguing bioactivities [20], and PubChem [21,22], Reaxys® [23] and SciFinder®[24] for recording the reported chemical space and results from biological assays). A major challenge in the creation of data standards lies in integrating this range of biological, chemical and functional data. The Minimum Information on Biosynthetic Gene clusters (MIBiG) standard and database [25•] makes an important first step in this regard for BGCs from microbes and plants.

Making molecules

As technologies for DNA sequencing and data mining continue to develop, the numbers of candidate genes available from plants with predicted functions in specialized metabolism will inevitably grow exponentially. The translation of this genetic information into small molecules for evaluation as drug leads and other applications already represents a major bottleneck. As the body of functionally characterized enzymes and pathways increases the predictive capacity of transcriptome and genome mining will improve, and it will become possible to eliminate likely functional redundancy at the bioinformatics stage. However an enormous amount of practical biochemical analysis will be required in order to reach that point. The establishment of pipelines that enable high-throughput characterization of suites of enzymes and pathways will therefore be critical in moving towards the vision of linking genes with pathways and chemicals.

The advent of commercial DNA synthesis now makes it possible to rapidly generate single and multi-gene constructs for expression in heterologous systems. In the past few years, further growth and innovation in the DNA synthesis industry has continued to drive down the price of synthetic genes to the extent that purchasing DNA is often competitive with the cost of PCR cloning, particularly when subsequent sequence manipulation is required to obtain the exact sequence desired for experimentation [26]. One of the foundations of synthetic biology is the concept of modular, interoperable genetic parts, which are units of DNA with defined and modular function. The assembly of new metabolic pathways from such parts has been facilitated by advances in parallel DNA assembly technologies that can assemble multiple fragments of DNA into a desired order in a single reaction and with high efficiency. Parallel DNA assembly can be mediated by site-specific recombinases, by regions of sequence overlap between adjacent fragments, or by restriction enzymes (for recent reviews see Refs. [27,28]). For the latter, the use of Type IIS restriction enzymes in a process commonly known as Golden Gate Cloning [29] has become particularly widely adopted in plant synthetic biology. In 2015, a common syntax was proposed for Type-IIS-mediated DNA assembly that defines standard parts for plant synthetic biology [30]. The standard enables the exchange and re-use of interoperable DNA parts with different molecular toolkits for facile assembly of multigene constructs including MoClo (Golden Gate Modular Cloning; [31]) and GB2.0 (Golden Braid; [32]) (Figure 2). Requiring nothing more than mixing purified plasmid DNA containing the parts to be assembled with an enzyme cocktail prior to incubation, Type IIS-mediated assembly methods are particularly amenable to miniaturization and automation; reactions have been

Figure 2



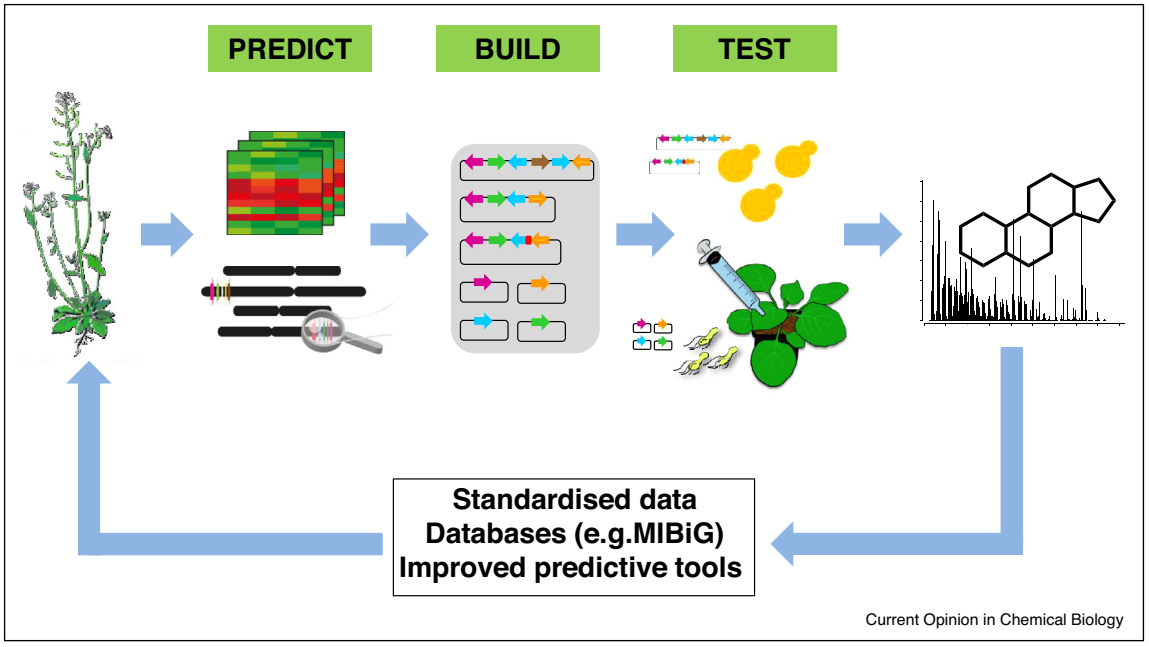
Standardization and assembly of plant genes for metabolic engineering. **(a)** Coding sequences of pathway enzymes; **(b)** modularization and standardization for chosen parallel DNA assembly method. The primary sequence may be refactored and/or codon-optimized. A synthetic gene to express each coding sequence is designed by the inclusion of characterized regulatory sequences; **(c)** the corresponding physical DNA samples for each part are retrieved from repositories or synthesized according to the design; **(d)** standard parts are assembled into transcriptional units and, if desired, into multigene constructs.

automated in volumes as low as 50 nL within DNA foundries, suites of laboratory automation for engineering organisms [33].

These advances in DNA synthesis and assembly technologies now make it possible to progress from genome sequencing and gene discovery to modularization and reassembly of proposed pathways for expression in heterologous hosts within weeks. Microorganisms, particularly yeasts, have been used for expression of several plant natural product pathways including the artemisinin precursor artemisinic acid [34], opioids [35,36•,37•,38•] and monoterpene indole alkaloids [39•,40•]. Metabolic engineering in yeast can enable industrial-scale production of high-value plant products, although optimization is likely to require extensive engineering of the host. For example, the yeast platform for artemisinic acid production is estimated to have taken >150 person years to develop [41]. Making microbes into efficient factories for the production of heterologous metabolites using the design-build-test principles of synthetic biology is hindered by lack of fundamental knowledge about cellular processes. Nielsen and Keasling [42] make the point that it will be necessary to expand the list of platform cell factories available for industrial production of different types of target compounds since not all pathways will express well in yeast and *Escherichia coli*, a strategy that has been embraced by the custom organism design approach of the biotech company Gingko Bioworks [43].

Plants have a number of advantages as heterologous hosts for metabolic engineering. They require only simple inorganic nutrients, water, carbon dioxide and sunlight for efficient growth. They are also more amenable to expression of genes of plant origin than microbes since they support correct mRNA and protein processing, protein localisation and metabolic compartmentalization, and already have many of the necessary metabolic precursors and co-enzymes. As non-hosts for animal and human pathogens they are an attractive alternative to human cell cultures for the manufacture of vaccines and therapies. Since 1994 more than 100 pharmaceutical proteins have been expressed and characterized in plants, and several large-scale facilities have been constructed to enable manufacturing at commercial scales (for recent reviews see Refs. [44–46]). Species of tobacco are relatively easy and fast to transform by the integration of new genes into the genome, but high yields can also be achieved in just a few days through transient expression following infiltration into the leaves of a culture of *Agrobacterium tumefaciens* carrying genes of interest on a binary plasmid vector, a process commonly known as agro-infiltration. The wild relative of tobacco, *Nicotiana benthamiana*, is highly amenable to agro-infiltration and is attracting increasing attention as a practical expression host.

Figure 3



Translation of plant genetic information into chemicals using synthetic biology approaches. The predictive power of this iterative process will continue to grow as more plant biosynthetic genes, enzymes and pathways are characterised.

Indeed, it is currently being used for commercial production of flu vaccines [46]. Agro-infiltration is particularly useful for rapid combinatorial expression of candidate enzymes since multiple *Agrobacterium* strains, each harbouring a different expression construct, can be simultaneously infiltrated into tobacco leaves, negating the need to fabricate multiple large constructs. Various different types of plant natural products have been produced in *N. benthamiana* using transient expression technology including glucosinolates [47,48], cyanogens [7•], indole sulfur phytoalexins [49], the etoposide aglycone [6••], the monoterpene indole alkaloid strictosidine [50], the antimalarial precursor artemisinin [51], triterpenes [52,53] and indole-derived betalains [54]. This expression host therefore provides an excellent system for functional analysis of diverse natural product pathways of plant origin. Yields are not industrial-scale. However, opportunities for improving the efficiency of transient plant expression technology, for example by increasing precursor supply, balancing pathway components, altering the subcellular location of heterologously expressed enzymes and modulating the expression of endogenous plant genes using genome editing [51,55–57] have not yet been fully explored. Coupled with advances in the establishment of standards and automation protocols for the fabrication of DNA constructs for plants and the increasing investment in large-scale facilities for bioproduction, bioengineering for heterologous biosynthesis in plants is set to enter a new era.

Conclusions

This review highlights the rapid advances that are being made in deciphering the genetic information encoded within plant genomes and discovering new pathways and chemistries. Although there have been notable successes in commercial production of plant-derived natural products through bioengineering in heterologous hosts, the exemplar being the production of artemisinic acid in yeast [34], there is still much to be done before industrial-scale production of target compounds in surrogate hosts becomes routine. Multiple different types of customised platform cell factories are likely to be needed to enable the large-scale production of different types of compounds. Plant transient expression platforms are proving to be extremely powerful for rapid delineation of new pathways. Such platforms are already being used for commercial production of pharmaceutical proteins and, with optimisation, may also have potential as factories for production of small molecules. The principles of synthetic biology provide inspiration for taking a systematic approach towards defining and tackling bottlenecks in the predict-build-test cycle to make it increasingly more efficient (Figure 3), with the ultimate aim of being able to make designer molecules on demand using sustainable practices.

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 Metabolism' (BB/J004561/1), the John Innes Foundation and European Union grant KBBE-

2013-7 (TriForC) (A.O., C.O.), the joint Engineering and Physical Sciences Research Council/BBSRC-funded OpenPlant Synthetic Biology Research Centre grant BB/L014130/1 (A.O and N.P.), National Institutes of Health Genome to Natural Products Network award U101GM110699 (A.O. and A. H) and European Commission Marie Skłodowska-Curie Individual Fellowship H2020-MSCA-IF-EF-ST-702478-TRIGEM (A.H.).

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. The Medicinal Plant Genomics Resource; URL: http://medicinalplantgenomics.msu.edu/.

2. Xiao M, Zhang Y, Chen X, Lee EJ, Barber CJ, Chakrabarty R, Desgagné-Penix I, Haslam TM, Kim YB, Liu E *et al.*: **Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest**. *J. Biotechnol.* 2013, **166**:122-134 https://bioinformatics.tugraz.at/phytometasyn/.

3. • Matasci N, Hung L-H, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N, Warnow T, Ayyampalayam S, Barker MS *et al.*: **Data access for the 1,000 plants (1KP) project**. *Gigascience* 2014, **3**:17.

This ambitious large-scale transcriptomics project has increased the number of available plant gene sequences by around 100-fold and greatly expanded knowledge of the gene content of taxonomically diverse species.

4. Giddings L-A, Liscombe DK, Hamilton JP, Childs KL, DellaPenna D, Buell CR, O'Connor SE: **A stereoselective hydroxylation step of alkaloid biosynthesis by a unique cytochrome P450 in *Catharanthus roseus***. *JBC* 2011, **286**:16751-16757.

5. Geu-Flores F, Sherden NH, Courdavault V, Burlat V, Glenn WS, Wu C, Nims E, Cui Y, O'Connor SE: **An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis**. *Nature* 2012, **492**:138-142.

6. •• Lau W, Sattely ES: **Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone**. *Science* 2015, **349**:1224-1228.

Delineation of a ten-step pathway in a genetically intractable (~16 Gb genome) plant species by co-expression analysis and assembly of the entire functional pathway in *N. benthamiana*.

7. • Rajniak J, Barco B, Clay NK, Sattely ES: **A new cyanogenic metabolite in *Arabidopsis* required for inducible pathogen defense**. *Nature* 2015, **525**:376-379.

Discovery of a new defense-related metabolic pathway in the model plant species *A. thaliana* by co-expression analysis and untargeted metabolomics using a pathogen-induced cytochrome P450 gene as a starting point.

8. • Nützmann HW, Huang A, Osbourn A: **Plant metabolic clusters – from genetics to genomics**. *New Phytol.* 2016, **211**:771-789.

A recent review of all plant BGCs reported so far, summarising their architecture and features.

9. • Ziemert N, Alanjary M, Weber T: **The evolution of genome mining in microbes – a review**. *Nat. Prod. Rep.* 2016, **33**:988-1005.

A recent review of the advances made over the last ten years in the development of methods for genome mining in microbes that will inform plant pathway discovery.

10. Altschul S, Gish W, Miller W, Myers E, Lipman D: **Basic local alignment search tool**. *J. Mol. Biol.* 1990, **215**:403-410.

11. Eddy SR: **Accelerated profile HMM searches**. *PLoS Comput. Biol.* 2011, **7**:e1002195.

12. • Boutanaev AM, Moses T, Zi J, Nelson DR, Mugford ST, Peters RJ, Osbourn A: **Investigation of terpene diversification across multiple sequenced plant genomes**. *Proc. Natl. Acad. Sci. U. S. A.* 2015, **112**:E81-E88.

This paper demonstrates systematic mining and analysis of multiple plant genomes for terpene synthases and cytochrome P450s, enabling the discovery of known and new BGCs and shedding light on the evolution of terpene diversity.

13. Kautsar SA, Suarez Duran HG, Blin K, Osbourn A, Medema MH: **plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters**. *Nucleic Acids Res.* 2016 http://dx.doi.org/10.1093/nar/gkx305.

14. Toepfer N, Fuchs L-M, Aharoni A: **The PhytoClust tool for metabolic gene clusters discovery in plant genomes**. *bioRxiv* 2017 http://dx.doi.org/10.1093/nar/gkx404. [Epub ahead of print].

15. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R: **antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences**. *Nucleic Acids Res.* 2011, **39**:W339-W346.

16. Yu N, Nützmann H-W, MacDonald JT, Moore B, Field B, Berriri S, Trick M, Rosser SJ, Kumar SV, Freemont PS, Osbourn A: **Delineation of metabolic gene clusters in plant genomes by chromatin signatures**. *NAR* 2016, **44**:2255-2265.

17. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: **The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics**. *Nucleic Acids Res.* 2009, **37**:D233-D238 http://www.cazy.org/.

18. Nelson DR: **The cytochrome P450 homepage**. *Hum. Genomics* 2009, **4**:59-65 http://drnelson.uthsc.edu/cytochromeP450.html.

19. Buchholz PCF, Vogel C, Reusch W, Pohl M, Rother D, Spieß AC, Pleiss J: **BioCatNet: a database system for the integration of enzyme sequences and biocatalytic experiments**. *Chembiochem* 2016, **17**:2093-2098 https://www.biocatnet.de/.

20. Nakamura Y, Mochamad Afendi F, Kawsar Parvin A, Ono N, Tanaka K, Hirai Morita A, Sato T, Sugiura T, Altaf-Ul-Amin M, Kanaya S: **KNApSAcK metabolite activity database for retrieving the relationships between metabolites and biological activities**. *Plant Cell Physiol.* 2014, **55** e7–e7. http://kanaya.naist.jp/MetaboliteActivity/top.jsp.

21. Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Zhou Z, Han L, Karapetyan K, Dracheva S, Shoemaker BA *et al.*: **PubChem's BioAssay database**. *Nucleic Acids Res.* 2012, **40**:D400-D412 https://pubchem.ncbi.nlm.nih.gov/sources#assay.

22. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA *et al.*: **PubChem substance and compound databases**. *Nucleic Acids Res.* 2016, **44**:D1202-D1213.

23. Reaxis® database; URL: https://www.elsevier.com/solutions/reaxys.

24. SciFinder®; URL: http://www.cas.org/products/scifinder.

25. • Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi YH, Claesen J, Coates RC *et al.*: **The minimum information about a biosynthetic gene cluster (MIBiG) specification**. *Nat. Chem. Biol.* 2015, **11**:625-631 http://mibig.secondarymetabolites.org/.

The MIBiG data standard provides a mechanism for consistent and systematic deposition and retrieval of data on microbial and plant BGCs.

26. Kosuri S, Church GM: **Large-scale de novo DNA synthesis: technologies and applications**. *Nat. Methods* 2014, **11**:499-507.

27. Casini A, Storch M, Baldwin GS, Ellis T: **Bricks and blueprints: methods and standards for DNA assembly**. *Nat. Rev. Mol. Cell Biol.* 2015, **9**:1-9.

28. Weyman PD, Suzuki Y: **Synthetic biology standards and methods of DNA assembly**. In *Synthetic Biology Handbook.* Edited by Nesbeth D. CRC Press; 2016:35-66.

29. Engler C, Kandzia R, Marillonnet S: **A one pot, one step, precision cloning method with high throughput capability**. *PLoS One* 2008, **3**:e3647.

30. Patron N, Orzaez D, Marillonnet S, Warzecha W, Matthewman C, Youles M, Raitskin O, Leveau A, Farré G, Rogers C *et al.*: **Standards for plant synthetic biology: a common syntax for exchange of DNA parts**. *New Phytol.* 2015, **208**:13-19.

31. Engler C, Youles M, Grüetzner R: **A golden gate modular cloning toolbox for plants**. *ACS Synth. Biol.* 2014, **3**:839-843.

32. Sarrion-Perdigones A, Vazquez-Vilar M, Palací J, Castelijns B, Forment J, Ziarsolo P, Blanca J, Granell A, Orzaez D: **GoldenBraid2.0: a comprehensive DNA assembly framework for plant synthetic biology**. *Plant Physiol.* 2013, **162**:1618-1631 http://dx.doi.org/10.1104/pp.113.217661.

33. Kanigowska P, Shen Y, Zheng Y, Rosser S, Cai Y: **Smart DNA fabrication using sound waves: applying acoustic dispensing technologies to synthetic biology**. *J. Lab. Autom.* 2015 http://dx.doi.org/10.1177/2211068215593754.

34. 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 artemisin**. *Nature* 2013, **469**:528-532.

35. Thodey K, Galanie S, Smolke CD: **A microbial biomanufacturing platform for natural and semisynthetic opioids**. *Nat. Chem. Biol.* 2014, **10**:837-844.

36. • DeLoache WC, Russ ZN, Narcross L, Gonzales AM, Martin VJ, Dueber JE: **An enzyme-coupled biosensor enables (*S*)-reticuline production in yeast from glucose**. *Nat. Chem. Biol.* 2015, **11**:465-471.

Reconsitution of the seven-step pathway from *L*-tyrosine to the codeine and morphine precursor (*S*)-reticuline in yeast. An enzyme-coupled biosensor was used to identify and improve the properties of a tyrosine hydroxylase that enabled the biosynthesis of the upstream intermediate *L*-3,4-dihydroxyphenylalanine (*L*-DOPA).

37. • Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M, Smolke CD: **Complete biosynthesis of opioids in yeast**. *Science* 2015, **349**:1095-1100.

A proof of principle of production of opioids in yeast by expression of 21 (thebaine) and 23 (hydrocodone) enzyme activities from plants, mammals, bacteria and yeast itself. The authors state that 'major hurdles remain before optimization and scale-up could be achieved.'

38. • Li Y, Smolke CD: **Engineering biosynthesis of the anticancer alkaloid noscapine in yeast**. *Nat. Commun.* 2016, **7**:12137 http://dx.doi.org/10.1038/ncomms12137.

Reconstitution of the 14-step biosynthetic pathway for the anticancer drug noscapine from opium poppy by engineering a yeast strain expressing 16 heterologous plant enzymes.

39. • Brown S, Clastre M, Courdavault V, O'Connor SE: **De novo production of the plant-derived alkaloid strictosidine in yeast**. *Proc. Natl. Acad. Sci. U. S. A.* 2015, **112**:3205-3210.

Production of strictosidine in yeast using 14 known monoterpene indole alkaloid pathway genes, along with ad additional seven genes and three gene deletions that enhance secondary metabolism.

40. • Qu Y, Easson MLAE, Froese J, Simionescu R, Hudlicky T, De Luca V: **Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast**. *Proc. Natl. Acad. Sci. U. S. A.* 2015, **112**:6224-6229.

Assembly of a seven-step pathway from production of the monoterpene indole alkaloid anticancer drug precursor vindoline from tabersonine in yeast using genes identified from Madagascan periwinkle (*Catharanthus roseus*). Tabersonine was introduced for conversion by feeding.

41. Kwok R: **Five hard truths for synthetic biology**. *Nature* 2010, **463**:288-290.

42. Nielsen J, Keasling JD: **Engineering cellular metabolism**. *Cell* 2016, **164**:1185-1197.

43. Gingko Bioworks; URL: http://www.ginkgobioworks.com/.

44. Sack M, Hofbauer A, Fischer R, Stoger E: **The increasing value of plant-made proteins**. *Curr. Opin. Biotechnol.* 2015, **32**:163-170.

45. Yao J, Weng Y, Dickey A, Wang KY: **Plants as factories for human pharmaceuticals: applications and challenges**. *Int. J. Mol. Sci.* 2015, **16**:28549-28565.

46. Marsian J, Lomonossoff GP: **Molecular pharming-VLPs made in plants**. *Curr. Opin. Biotechnol.* 2016, **37**:201-206.

47. Geu-Flores F, Nielsen MT, Nafisi M, Moldrup ME, Olsen CE, Motawia MS, Halkier BA: **Glucosinolate engineering identifies a γ-glutamyl peptidase**. *Nat. Chem. Biol.* 2009, **5**:575-577.

48. Crocall C, Mirza N, Reichelt M, Gershenzon J, Halkier BA: **Optimization of engineered production of the glucoraphanin precursor dihomomethionine in *Nicotiana benthamiana***. *Front. Bioeng. Biotechnol.* 2016, **114**:1910-1915 http://dx.doi.org/10.3389/fbioe.2016.00014.14.

49. Klein AP, Sattely ES: **Biosynthesis of cabbage phytoalexins from indole glucosinolate**. *Proc. Natl. Acad. Sci. U. S. A.* 2017 http://dx.doi.org/10.1073/pnas.1615625114. pii: 201615625.

50. Miettinen K, Dong L, Navrot N, Schneider T, Burlat V, Pollier J, Woittiez L, van der Krol S, Lugan R, Ilc T, Verpoorte R *et al.*: **The seco-iridoid pathway from *Catharanthus roseus***. *Nat. Commun.* 2014, **5**:3606 http://dx.doi.org/10.1038/ncomms4606.

51. Wang B, Kashkooli AB, Sallets A, Ting HM, de Ruijter NC, Olofsson L, Brodelius P, Pettier M, Boutry MJ, Bouwmeester HJ *et al.*: **Transient production of artemisinin in *Nicotiana benthamiana* is boosted by a specific lipid transfer protein from *A. annua***. *Metab. Eng.* 2016, **38**:159-169.

52. Geisler K, Hughes RK, Sainsbury F, Lomonossoff GP, Rejzek M, Fairhurst S, Olsen CE, Motawia MS, Melton R, Hemmings A, Bak S, Osbourn A: **Biochemical analysis of a multi-functional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants**. *Proc. Natl. Acad. Sci. U. S. A.* 2013, **110**:E3360-E3367.

53. Khakimov B, Kuzina V, Erthmann PO, Fukushima EO, Augustin JM, Olsen CE, Scholtalbers J, Volpin H, Andersen SB, Hauser TP *et al.*: **Identification and genome organization of saponin pathway genes from a wild crucifer, and their use for transgenic production of saponins in *Nicotiana benthamiana***. *Plant J.* 2015, **84**:478-490.

54. Polturak G, Breitel D, Grossman N, Sarrion-Perdigones A, Weithorn E, Pliner M, Orzaez D, Granell A, Rogachev I, Aharoni A: **Elucidation of the first committed step in betalain biosynthesis enables the heterologous engineering of betalain pigments in plants**. *New Phytol.* 2016, **210**:269-283.

55. Henry LK, Gutensohn M, Thomas STT, Noel J, Dudareva N: **Orthologs of the archeal isopentenyl phosphate kinase regulate terpenoid production in plants**. *Proc. Natl. Acad. Sci. U. S. A.* 2015, **112**:10050-10055.

56. Wu S, Schalk M, Clark A, Miles RB, Coates R, Chappell J: **Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants**. *Nat. Biotechnol.* 2006, **24**:1441-1447.

57. Bortesi L, Zhu C, Zischewski J, Perez L, Bassié L, Nadi R, Forni G, Lade SB, Soto E, Jin X *et al.*: **Patterns of CRISPR/Cas9 activity in plants, animals and microbes**. *Plant Biotechnol. J.* 2016, **14**:2203-2216 http://dx.doi.org/10.1111/pbi.12634.

58. Rice KC: **Synthetic opium alkaloids and derivatives. A short total synthesis of (.+-.)-dihydrothebainone, (.+-.)-dihydrocodeinone, and (.+-.)-nordihydrocodeinone as an approach to a practical synthesis of morphine, codeine, and congeners**. *J. Organic Chem.* 1980, **45**:3135-3137.

59. Zhu C, Cook SP: **A concise synthesis of (+)-artemisinin**. *J. Am. Chem. Soc.* 2012, **134**:13577-13579.

60. Ganem B, Franke RR: **Paclitaxel from primary taxanes: a perspective on creative invention in organozirconium chemistry**. *J. Org. Chem.* 2007, **72**:3981-3987.

61. Ajikumar PK, Xiao W-H, Tyo KEJ, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G: **Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli***. *Science* 2010, **330**:70-74.

62. Holton RA: **Semi-synthesis of taxane derivatives using metal alkoxides and oxazinones**. US patent 5254703 A (1993).