Synthetic biosystems for the production of high-value plant metabolites

Facchini, Peter J.; Bohlmann, Joerg; Covello, Patrick S.; De Luca, Vincenzo; Mahadevan, Radhakrishnan; Page, Jonathan E.; Ro, Dae-Kyun; Sensen, Christoph W.; Storms, Reginald; Martin, Vincent J. J.

Publisher's version / Version de l'éditeur:
https://doi.org/10.1016/j.tibtech.2011.10.001
Trends in Biotechnology, 30, 3, pp. 127-131, 2011-12-29

Synthetic Biosystems for the Production of High-Value Plant Metabolites

Peter J Facchini[1,*], Joerg Bohlmann[2], Patrick S Covello[3], Vincenzo De Luca[4], Radhakrishnan Mahadevan[5], Jonathan E Page[3], Dae-Kyun Ro[1], Christoph W Sensen[6], Reginald Storms[7], Vincent J J Martin[7,*]

[1] Department of Biological Sciences, University of Calgary, 2500 University Drive N.W., Calgary, Alberta, T2N 1N4, Canada

[2] Michael Smith Laboratories, University of British Columbia, 321-2185 East Mall, Vancouver, British Columbia, V6T 1Z4, Canada

[3] National Research Council – Plant Biotechnology Institute, 110 Gymnasium Place, Saskatoon, Saskatchewan, S7N 0W9, Canada

[4] Department of Biological Sciences, Brock University, 500 Glenridge Avenue, St. Catharines, Ontario, L2S 3A1, Canada

[5] Department of Chemical Engineering and Applied Chemistry, University of Toronto, 200 College Street, Toronto, Ontario, M5S 3E5, Canada

[6] Department of Biochemistry and Molecular Biology, University of Calgary, 3330 Hospital Drive N.W., Calgary, Alberta, T2N 4N1, Canada

[7] Department of Biology, Concordia University, 7141 rue Sherbrooke ouest, Montréal, Québec, H4B 1R6, Canada

* Joint first authors, Project Leaders and corresponding authors: Facchini, P.J. (pfacchin@ucalgary.ca) and Martin, V.J.J. (vmartin@alcor.concordia.ca)

Abstract

Plants display an immense diversity of specialized metabolites, many of which have been important to humanity as medicines, flavors, fragrances, pigment, insecticides and other fine chemicals. Apparently, much of the variation in plant specialized metabolism evolved through events of gene duplications followed by neo- or sub-functionalization. Most of the catalytic diversity of plant enzymes is unexplored since previous biochemical and genomics efforts have focused on a relatively small number of species. Interdisciplinary research in plant genomics, microbial engineering and synthetic biology provides an opportunity to accelerate the discovery of new enzymes. The massive identification, characterization and cataloguing of plant enzymes coupled with their deployment in metabolically optimized microbes provide a high-throughput functional genomics tool and a novel strain engineering pipeline.

Introduction

Plants produce a bewildering array of specialized metabolites based on myriad skeletal structures and functional group combinations [1, 2]. Economically, many of these compounds are among the most valuable bioproducts in their respective markets (Figure 1). Given their enormous structural diversity and the equally staggering numbers of species in which specialized metabolites are produced, the impressive biosynthetic potential of plant metabolism has long been recognized. As with other organisms, the implementation of genomics has expedited our understanding of many basic biological processes in plants [3]. Initial genome-sequencing and genome-assembly efforts

focused on a few model species or major crops, such as Arabidopsis, rice, poplar, grapevine, tobacco and *Medicago truncatula* [4]. Although an increasing number of plants from all major taxonomic lineages are currently targeted for genome sequencing, the selection of species is generally not based on maximizing coverage of diverse specialized metabolism. As a result the diversity of genes encoding enzymes involved in plant specialized metabolism is not well represented, or at least not well annotated in public DNA sequence databases. Tapping into the biochemistry that is often specific to particular plant species via the characterization and cataloguing of potentially novel enzymes requires the establishment of genome or transcriptome sequence resources for a large number of plants producing a sufficient diversity of specialized metabolites [5]. The discovery of genes encoding previously unknown biosynthetic enzymes will also require comprehensive metabolite profiling of the same plant species, and an effective integration of genomics and metabolomics datasets [6].

The PhytoMetaSyn Project represents a consortium of researchers from across Canada with the following principal objectives: (1) the establishment of a genomics pipeline that integrates massively parallel DNA sequencing, targeted metabolomics, advanced bioinformatics and “plug-and-play” functional genomics in yeast to efficiently identify, characterize and catalogue a continuously expanding collection of biosynthetic genes responsible for the immense chemical diversity of plant metabolism, (2) the development of a framework for the commercial production of valuable plant natural products in microbial systems through the optimized production of primary metabolic precursors, and (3) a demonstration of the feasibility of synthetic biology as a platform for the production of six prototype plant natural products in engineered yeast (Figure 2).

Gene discovery: Accelerated expansion of the parts catalogue

Genome and transcriptome mining has become the strategy of choice for the discovery of genes encoding enzymes responsible for the multistep formation of plant specialized metabolites. Until recently, the most common approach involved generating an expressed sequence tag (EST) database based on traditional Sanger sequencing for a plant producing one or more natural products of interest. Generally, candidate genes are expressed in a microorganism and the recombinant enzymes are functionally characterized *in vitro*. The ability to isolate and identify a novel enzyme is largely dependent on two factors: (1) the representation of gene candidates in often relatively modest EST collections and (2) the availability of substrates, which are typically not available commercially. Recently, massively parallel sequencing technologies have dramatically extended the opportunity to establish deep transcriptome databases containing a high-percentage of full-length open reading frames for any organism [7, 8]. Owing to the accumulation of different, yet often structurally similar specialized metabolites in related plant taxa, high-quality sequence databases for a variety of related species will yield a large repository of biosynthetic gene candidates with different biochemical functions (e.g. substrate- or regio-specificity).

The effectiveness of selecting gene candidates encoding enzymes with predictable metabolic functions is clearly proportional to the amount of information available about the catalytic steps in a particular biosynthetic pathway. To facilitate the rapid identification of novel biosynthetic genes and the establishment of empirical evidence for hypothetical metabolic pathways, genomics resources must be supported

by effective bioinformatics tools beyond simple BLASTX annotations. Key tools include in-depth functional domain identification through more sensitive database comparisons and modeling to identify candidate genes, and the hierarchical clustering of homologous genes from various plants to allow the seamless inclusion of the sequence databases from new species as they become available. Convenient platforms for the unambiguous identification of plant metabolites are also provided through recent advances in mass spectrometry. The integration of deep transcript and targeted metabolite profiles from corresponding plant tissues is a key component in establishing an integrated platform to selection biosynthetic gene candidates involved in diverse natural product metabolism.

The PhytoMetaSyn Project is currently targeting compounds that belong to three categories of plant specialized metabolites: terpenoids, alkaloids and polyketides [9-11]. These compound classes reflect the structural diversity (e.g. approximately 40,000 terpenoids, 12,000 alkaloids and thousands of polyketides) and the commercial importance of plant natural products (Figure 1). The collective and collaborative expertise available within the project is key to the development of a computer infrastructure for analyzing metabolic pathway information [12], and adapting available analyses and visualization tools (e.g. the Madison Metabolite Database, KEGG, MetaCyc, and PubChem). Plant materials from 75 source species producing terpenoids, alkaloids and polyketides were rationally selected based on the following criteria: (1) the natural product composition of the species; (2) the commercial importance of the plant and/or the natural products; and (3) the availability of sufficient tissue for RNA and metabolite isolation. Isolated RNA from each system was reverse transcribed to cDNA, which was sequenced using massively parallel Roche-454 GS-FLX Titanium and

Illumina GAiix 108 paired-end technologies, that combined facilitate the assembly of full-length contigs. Targeted metabolomics involves the use of GC- and LC-MS/MS instrumentation coupled to the use of available authentic standards together with published reference spectra. A Web portal (www.phytometasyn.com) provides project team members and the research community at large access to the extensive, unique and valuable genomics and metabolomics resources established by this Project in accordance with our data release policy.

Synthetic biology: Assembly of plant pathways in yeast

Biosynthetic genes identified from extensive sequence databases are used for the stepwise reconstruction of plant specialized metabolic pathways in *Saccharomyces cerevisiae*. Yeast is the organism of choice for the reconstitution of complex plant pathways, owing partly to the availability of extensive genome-wide metabolic models and genetic resources [13], but mainly because many plant enzymes such as cytochrome P450s, prenyltransferases or FAD-dependent oxidoreductases require a eukaryotic cell environment for optimal expression and activity. The development of an *in vivo* system for the functional characterization of novel plant enzymes circumvents many of the problems frequently associated with protein purification, the availability of substrates, and the *in vitro* analysis of recombinant proteins (Figure 3). Microbial hosts are engineered to produce sufficient levels of metabolic precursors to support targeted terpenoid, alkaloid and polyketide biosynthesis. The ability of yeast to perform multistep plant metabolism has also been demonstrated [14-16]. These platform strains are the starting point for our “plug-and-play” strategy involving the stepwise introduction of plant

genes encoding known and unknown enzymes to: (1) establish strains producing central intermediates in each of the targeted pathways, (2) discover novel enzymes capable of catalyzing the conversion of available intermediates, (3) reconstruct prototype plant natural product pathways and (4) create novel metabolic diversity based on combinatorial biochemistry [17-19].

As expected, the number of previously unknown genes generated by massively parallel sequencing is large and an effective triage and binning strategy is necessary to identify high-priority targets. The broad selection of gene targets is based on (1) identifying missing enzymes in prototype biosynthetic pathways and (2) maximizing the catalytic properties of key enzymes. The specific selection of gene candidates is initially based on sequence similarity searches for putative enzymes catalyzing reactions potentially occurring via the same mechanism in hypothetical or empirical pathways derived via targeted metabolite profiling. With the aid of gene expression profiles determined by RT-PCR, 5 to 10 candidate genes corresponding to a particular enzyme category and plant species typically emerge from the initial triage process and are functionally evaluated. An example of the effectiveness of this strategy is revealed among the enzymes in the biosynthesis of approximately 2,500 known benzylisoquinoline alkaloids (BIAs), including the narcotic analgesics codeine and morphine produced in opium poppy (Figure 1). Over the past two decades, cDNAs corresponding to most of the enzymes involved in the conversion of tyrosine to morphine, and a few others functioning in branch pathways leading to related BIAs, have been identified [20]. The first committed step in BIA metabolism is catalyzed by a unique member of the pathogenesis-related (PR)10 protein family. All other known

enzymes belong to one of only nine enzyme families including cytochrome P450s, *O*- and *N*-methyltransferases, FAD-linked oxidoreductases, 2-oxoglutarate-dependent dioxygenases, acyl-CoA-dependent acyltransferases and three different subgroups of NADPH-dependent reductases. Candidate genes encoding enzyme variants responsible for the immense diversity of BIAs can be readily selected from the deep transcriptome databases of various plant species shown to produce these compounds. Moreover, the profile of specific BIAs in each species provides guidance in the assignment of potential catalytic function. The use of eukaryotic cells for the reconstitution of multistep plant metabolic pathways and the *in vivo* functional characterization of unknown genes reduces the limitations associated with the commercial unavailability of substrates required for *in vitro* enzyme assays.

Initial efforts to integrate plant biochemical genomics and yeast functional genomics have revealed several expected and some unexpected challenges [21]. Emerging DNA assembly methodologies have improved the speed and efficiency of pathway reconstitution in yeast [22], but much remains to be learned about the selection of appropriate promoters and transcription terminators to promote efficient recombinant gene expression for optimized pathway flux. Enzyme levels are clearly important to maximize flux through reconstituted metabolic pathways, yet the key parameters affecting the steady-state levels of foreign proteins in yeast are not well understood. Synthetic genes provide an increasingly cost-effect opportunity for codon optimization [23-25], but the testing of natural variants representing functional orthologs identified in plants provide an alternative to empirical protein engineering to address post-

transcriptional issues. These uncertainties further validate the utility of a parts catalogue that includes both functionally unique and redundant enzymes (Figure 3).

The implementation of microbial engineering at the pathway level for the discovery of biosynthetic genes involved in plant specialized metabolite biosynthesis is an essential step in the development of a versatile functional genomics strategy. Our emerging technology will expedite the discovery of a multitude of novel catalysts that can be compiled into a biochemical parts catalogue. Combined with emerging “plug-and-play” engineering technologies in yeast, a compendium of biological resources will also create genuine synthetic biology potential for combinatorial biochemistry and the commercial production of high-value specialized plant metabolites [26].

Acknowledgements

The authors acknowledge the outstanding contributions of numerous PhytoMetaSyn Project administrative and research staff, especially Michael O’Connell, Ada Sijerčić, Dr. Andy Ekins, Dr. Isabel Desgagné-Penix and Dr. Jillian Hagel. The latter also contributed the original watercolor artwork used in the figures. We recognize the skilled technical work of next-generation sequencing personnel at the McGill University-Génome Québec Innovation Centre, especially Fanny Chagnon, Alexandre Montpetit, Louis Letourneau and Gary Levesque, and the dedicated state-of-the-art bioinformatics support provided by Ye Zhang and Mei Xiao. This work was supported by Genome Canada, Genome Alberta, Genome Québec, Genome Prairie Genome British Columbia, the Canada Foundation for Innovation, the Ontario Ministry of Research and Innovation, the National Research Council of Canada and other government and private sector partners. P.J.F.,

V.J.J.M., D.K.R., V.D.L. were also supported by the Canada Research Chair program. J.B. was also supported by the Distinguished University Scholar program at the University of British Columbia.

References

[1] Kroymann, J. (2011) Natural diversity and adaptation in plant secondary metabolism. *Curr Opin Plant Biol* 14, 246-251
[2] Wink, M., ed (2010) *Biochemistry, physiology and ecological functions of secondary metabolites*. Wiley-Blackwell
[3] Borevitz, J.O. and Ecker, J.R. (2004) Plant genomics: the third wave. *Annu Rev Genomics Hum Genet* 5, 443-477
[4] Feuillet, C.*, et al.* (2011) Crop genome sequencing: lessons and rationales. *Trends Plant Sci* 16, 77-88
[5] Fridman, E. and Pichersky, E. (2005) Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. *Curr Opin Plant Biol* 8, 242-248
[6] Saito, K.*, et al.* (2008) Decoding genes with coexpression networks and metabolomics - 'majority report by precogs'. *Trends Plant Sci* 13, 36-43
[7] Mardis, E.R. (2008) Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 9, 387-402
[8] Schuster, S.C. (2008) Next-generation sequencing transforms today's biology. *Nat Methods* 5, 16-18
[9] Bohlmann, J. and Keeling, C.I. (2008) Terpenoid biomaterials. *Plant J* 54, 656-669
[10] Facchini, P.J. and De Luca, V. (2008) Opium poppy and Madagascar periwinkle: model non-model systems to investigate alkaloid biosynthesis in plants. *Plant J* 54, 763-784
[11] Yu, O. and Jez, J.M. (2008) Nature's assembly line: biosynthesis of simple phenylpropanoids and polyketides. *Plant J* 54, 750-762
[12] Faust, K.*, et al.* (2011) Prediction of metabolic pathways from genome-scale metabolic networks. *Biosystems* 105, 109-121
[13] Krivoruchko, A.*, et al.* (2011) Opportunities for yeast metabolic engineering: Lessons from synthetic biology. *Biotechnol J* 6, 262-276
[14] Hawkins, K.M. and Smolke, C.D. (2008) Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae. *Nat Chem Biol* 4, 564-573
[15] Nakagawa, A.*, et al.* (2011) A bacterial platform for fermentative production of plant alkaloids. *Nat Commun* 2, 326
[16] Ro, D.K.*, et al.* (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440, 940-943
[17] Julsing, M.K.*, et al.* (2006) Combinatorial biosynthesis of medicinal plant secondary metabolites. *Biomol Eng* 23, 265-279

[18] Nguyen, K.T.*, et al.* (2006) Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc Natl Acad Sci U S A* 103, 17462-17467
[19] Zhang, W. and Tang, Y. (2008) Combinatorial biosynthesis of natural products. *J Med Chem* 51, 2629-2633
[20] Ziegler, J. and Facchini, P.J. (2008) Alkaloid biosynthesis: metabolism and trafficking. *Annu Rev Plant Biol* 59, 735-769
[21] Prather, K.L. and Martin, C.H. (2008) De novo biosynthetic pathways: rational design of microbial chemical factories. *Curr Opin Biotechnol* 19, 468-474
[22] Gibson, D.G. (2011) Gene and genome construction in yeast. *Curr Protoc Mol Biol* Chapter 3, Unit3 22
[23] Fox, J.M. and Erill, I. (2010) Relative codon adaptation: a generic codon bias index for prediction of gene expression. *DNA Res* 17, 185-196
[24] Hamann, T. and Moller, B.L. (2007) Improved cloning and expression of cytochrome P450s and cytochrome P450 reductase in yeast. *Protein Expr Purif* 56, 121-127
[25] Klein-Marcuschamer, D.*, et al.* (2010) De novo metabolic engineering and the promise of synthetic DNA. *Adv Biochem Eng Biotechnol* 120, 101-131
[26] Carothers, J.M.*, et al.* (2009) Chemical synthesis using synthetic biology. *Curr Opin Biotechnol* 20, 498-503

Figure legends

Figure 1. Six natural products and selected source plants. The compounds represent a sesquiterpene (nootkatone), a diterpene (abietic acid), a triterpene (betulinic acid), a benzylisoquinoline alkaloid (morphine), a monoterpenoid indole alkaloid (strictosidine) and a polyketide (xanthohumol).

Figure 2. Pipeline for the discovery of novel biosynthetic genes from plants, their reassembly in yeast, and the subsequent commercial production of high-value metabolites in scalable fermentation systems.

Figure 3. Synergy between plug-and-play biochemical genomics and synthetic biology. Transcript profiling of selected plants producing identical or related natural products provides a source of candidate genes (*i.e.* parts), which are used to assemble synthetic metabolic pathways in yeast. Pathway assembly in yeast is used: (1) as a tool to discover gene function, which is particularly useful when pathway intermediates are not available for in vitro enzyme characterization, (2) to optimize pathway efficiency using (natural or synthetic) orthologous genes, and control elements of gene expression, and (3) as a combinatorial biochemistry platform to expand on the diversity of molecules produced in plants.

Figure 1

Figure 1

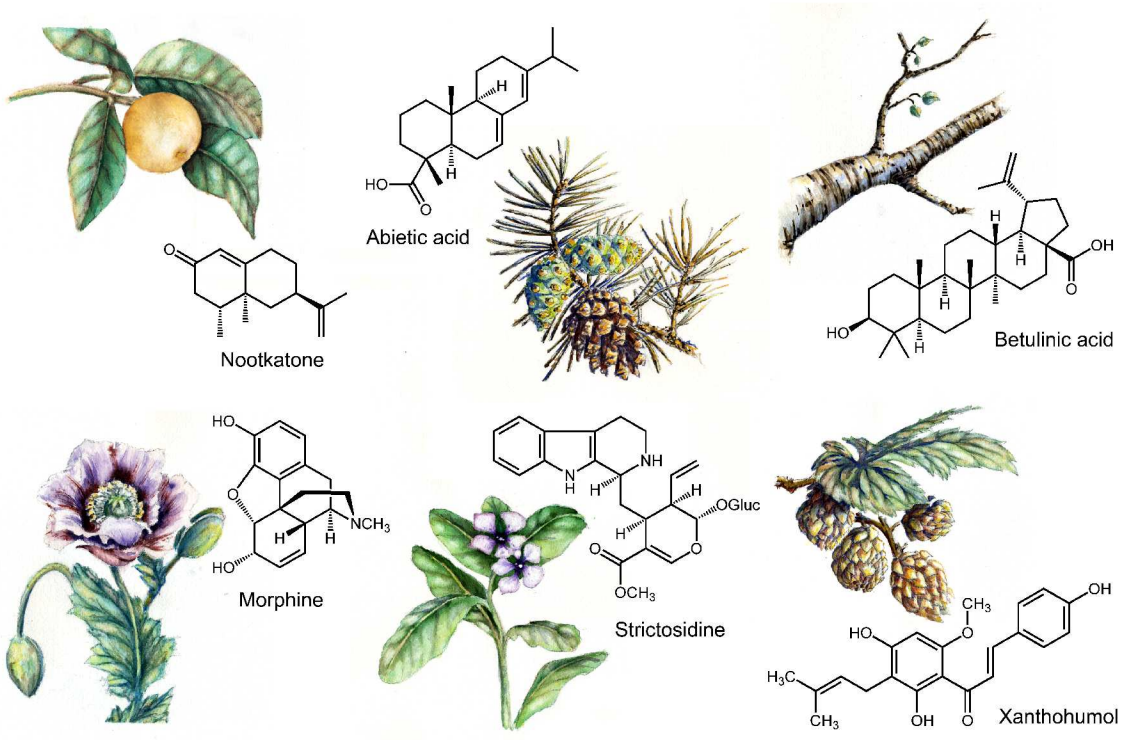

Figure 2

Figure 2

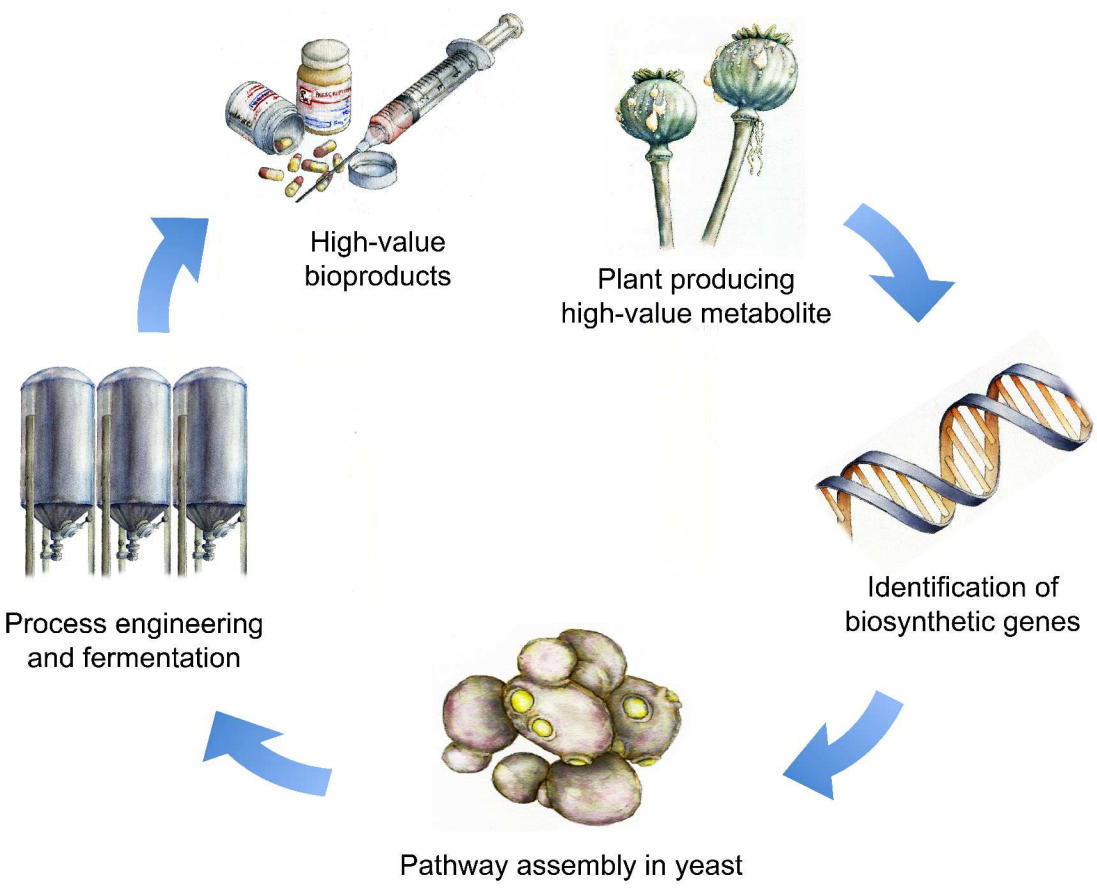

Figure 3

Figure 3