Phytochemical Genomics on the Way

Toshiya Muranaka and Kazuki Saito
1Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka, 565-0871 Japan
2RIKEN Center for Sustainable Resource Science, Suehiro-cho 1-7-22, Tsurumi-ku, Yokohama, 230-0045 Japan
3Graduate School of Pharmaceutical Sciences, Chiba University, Inohana 1-8-1, Chuo-ku, Chiba, 260-8675 Japan

This is a really exciting time for phytochemistry research. Plant scientists studying specialized plant metabolism (previously often known as ‘secondary metabolism’) can foresee tremendous advances in their understanding of the mechanisms underlying the production and function of an enormous variety of plant metabolites, and of their regulation and evolution. Such remarkable progress would never have been expected 10 years ago and has taken place in just the last few years (Yonekura-Sakakibara and Saito 2009).

Two technological breakthroughs seem to have been the key triggers for this unforeseen progress: metabolomics and mass DNA sequencing. Metabolomics is <15 years old, yet the concept was immediately applied to the plant field as scientists realized its tremendous potential for understanding the capacity of plants for making a huge array of compounds (Bino et al. 2004). Metabolomics now makes possible comprehensive profiling of nearly all the metabolites that accumulate in plant cells (Saito and Matsuda 2010). The second technological breakthrough was ultra-high-throughput DNA sequencing, with so-called ‘next-generation DNA sequencers’ (Wang et al. 2009, Ozsolak and Milos 2011). Until the realization of this technology, no one would have anticipated genomics-based studies of hundreds of medicinal or exotic plant species—genomics was only for model plant species such as Arabidopis and rice. Reconstructed from omics sequence data have been mainly genomics only for model plants or major crops, whose genome sequences could only be revealed by large international research consortia (or by huge well-funded undertakings). However, newly developed mass DNA sequencers have made comprehensive genome and transcriptome sequencing possible for ordinary phytochemical researchers who are investigating the synthesis of particular plant metabolites out of their own curiosity. By combining data sets from metabolomics and genomics/transcriptomics, one can understand relationships between metabolites and genes that allow the generation of testable hypotheses about the functions of genes and metabolites (Higashi and Saito 2013, Hur et al. 2013, Yonekura-Sakakibara et al. 2013). During this integration and hypothesis-generation phase, databases and bioinformatics play indispensable roles in reducing this ‘systems biology’ workload. Such studies will provide us with an understanding of the evolution of genes and pathways and hence in-depth fundamental insights into plant life. This special focus issue (SFI) on ‘Phytochemical Genomics’ is intended to serve as a synthesis of the most up to date information in this field, indicating not only current trends but also its future prospects.

For this SFI of Plant and Cell Physiology, we invited one review article and six original articles reporting on the exciting and growing field of phytochemical genomics. The huge numbers of specialized metabolites produced in plants are mainly classified into three basic groups: alkaloids, terpenoids and phenolic compounds. Among alkaloids, benzylisoquinoline alkaloids (BIAs), including the narcotic analgesics codeine and morphine, are one of the groups of compounds most exploited for their medicinal properties. In this SFI, Hagel and Facchini (2013; see pp. 647–672) review recent advances in biosynthetic studies of BIAs including transcriptomics, proteomics and metabolomics; they also discuss the application of synthetic biology to the development of production by microbes as an alternative to plants as a potential commercial source of valuable BIAs.

Another three deep transcriptome analyses are reported in this SFI. Catharanthus roseus synthesizes numerous terpenoid indole alkaloids, such as the anticancer drugs vinblastine and vincristine. So far, pathway databases and metabolic networks reconstructed from omics sequence data have been mainly built for model plant species such as Arabidopsis and rice. Van Moerkercke et al. (2013; see pp. 673–685) conducted a deep transcriptome analysis in the medicinal plant, C. roseus, and constructed a detailed metabolic pathway database ‘CathaCyc’ that contains 390 pathways with more than a thousand assigned enzymes. Yamazaki et al. (2013; see pp. 686–696) coupled the deep transcriptome analysis with untargeted metabolic profiling in Ophiopilum plumula, which produces the anti-cancer alkaloids camptothecin and anthraquinones. Yamazaki and colleagues compared both transcriptome and metabolome data sets in the alkaloid-producing hairy root with a cell suspension that does not produce alkaloids. Glycyrrhiza plants produce various phytochemicals, among them many different terpenoids and flavonoids. Among the phytochemicals, glycyrrhizin exhibits several sorts of pharmacological activity and is also used as a natural sweetener. Two cytochrome P450 monoxygenase (P450s) genes for glycyrrhizin biosynthesis have been recently identified (Seki et al. 2008, Seki et al. 2011). Ramiłowski et al. (2013; see pp. 697–710) conducted deep transcriptome analysis in G. uralensis, and
extracted not only the P450s but also vacuolar transporters in the biosynthesis of glycyrrhizin.

Two bioinformatics articles are included in this issue: the species–metabolite relationship database, KNApSaCK Core, has been widely utilized and cited in metabolomics research (Afendi et al. 2012). Ikeda et al. (2013; pp. 711–727) extended the KNApSaCK database by incorporating it in a specialized plant metabolic pathway database and examined the huge amount of enzyme diversity. They mention that data-intensive approaches are needed for systematizing these protein sequences. Metabolomics analysis tools can provide quantitative information about the concentration of metabolites in an organism. However, the number of parameters to be determined for this is huge, and such parameters cannot be directly ascertained from the experimental data. Katsu&agi et al. (2013; see pp. 728–739) proposed a tool to extract a subsystem of the metabolic network automatically and simulate a time course of concentration of metabolites within the subsystem.

The model legume Medicago truncatula has an interesting profile of triterpenoid saponins. In terpenoid biosynthesis, different subfamilies of P450s are involved. Fukushima et al. (2013; see pp. 740–749) have identified candidate P450s for triterpenoid biosynthesis by co-expression analysis. Interestingly, their combination with the candidate genes expressed in yeast transgenic systems produced not only the natural (expected) triterpenoids but also rare (unexpected) triterpenoids, and indicates the potential for combinatorial synthesis of diverse triterpenoid structures for new drug discovery.

Statistical analysis suggests that the total number of plant metabolites lies between 200,000 and 1,000,000 compounds (Afendi et al. 2012). Needless to say, the ability to produce such a huge variety of metabolites is considered a unique strategy of plants as sessile organisms for defense against biotic and abiotic stresses. As human beings, we very much depend on this metabolite-producing plant function for food, drugs, flavors, cosmetics, energy and industrial materials. Phytochemical genomics unveils the secrets of how these specialized plant products are produced, what genes are involved, how their pathways are regulated and how these functions have evolved. The answers to these questions will provide the basis for the applications of plant specialized metabolic pathways in our future lives.

**Funding**

This work was supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan [a Grant-in-Aid for Scientific Research on Innovative Areas]; the Bio-oriented Technology Research Advancement Institution, Japan.

**References**


