Review Article

Metabolomics for Functional Genomics

M.K.Samota[1*], L.Bhatt[2], D.K. Yadav, N. Garg and R. Bajiya

[1]ICAR-IARI, New Delhi-110012, India
[2]G.B. Pant University of Agriculture and Technology - 263145, Pantnagar, India
*Corresponding author

A B S T R A C T

Metabolomics approaches enable the parallel assessment of the levels of a broad range of metabolites and have been great value in both phenotyping and diagnostic analyses in plants and other organisms. Analysis of metabolites was done by using different method like HPLC, GC, CE, MS etc but due to problem in their efficiency, accuracy and speed they are replaced by new throughput instrument like GC-MS, HPLC-MS, CE-MS, NMR, GC-MS-TOF etc. These high through instrument gives better result over traditional one in respect to their efficiency, accuracy and speed for the analysis of huge number of metabolite in the cell. Metabolomic approach nowadays effectively used to identify mQTL in plant which benefical in breeding programme for the selection of progeny in short period of time that of MAS selection. Metabololomics not only used in breeding programme but also make effective role in diagnosis and prediction of different disease in human. Comprehensive study of human metabolites "human metabolome project" was lunched jan-2005 in Canada. Metabolomics technologies are needed to cover the wide array of phytochemicals and increase the spatial and temporal resolution of metabolome analysis. Strengthening of spectral databases and sophisticated informatics for more accurate annotation of metabolites is necessary, and these efforts must be internationally harmonized and publicly available.

Introduction

In organism information store in DNA and which are pass through different stages for development of phenotype. The different stages are gene, transcript, protein and metabolite, study of each stage can be under participate genomics, transcriptomics, proteomics and metabolomics respectively (Gigolashvili *et al.,* 2007). Metabolomics is developed field for analysing samples to detect difference between them. Metabolomics is proved powerful tool analysis of difference of an organism and it have certain benefits over another 'omics'

technology in relevance to sample complexity, cost and speed. Only analysis of metabolites is not given sufficient information for modification of crop to get desired result and hence simultaneous study of all 'omics' are very important for understanding of whole information pathway of an organism and manipulation to get desired phenotype(Hirai *et al.,* 2007).

Metabolomics

Metabolomics is the "systematic study of the unique chemical fingerprints that specific

cellular processes leave behind" specifically, the study of their small-molecule metabolite profiles (Fernie *et al.,* 2009). While mRNA gene expression data and proteomic analyses do not tell the whole story of what might be happening in a cell, metabolic profiling can give an instantaneous snapshot of the physiology of that cell. One of the challenges of systems biology and functional genomics is to integrate proteomic, transcriptomics, and Metabolomic information to give a more complete picture of living organisms. Metabolome refers to the complete set of small-molecule metabolites (such as metabolic intermediates, hormones and other signalling molecules, and secondary metabolites) to be found within a biological sample, such as a single organism. Metabolites are the intermediates and products of metabolism having less than 1 kDa in size. There are some exceptions to this depending on the sample and detection method. For instance macromolecules such as lipoproteins and albumin are reliably detected in NMR-based metabolomics research of blood plasma. (Kazuki *et al*., 2010) In plant-based metabolomics, it is two type "primary" and "secondary" metabolites. A primary metabolite is directly played role in growth, development, and reproduction. A secondary metabolite is not directly involved in those physiological processes, but usually has important ecological function (Carrari *et al.,* 2006).

Analysis of metabolites

The metabolomics experiment (sampling, sample preparation, instrumental analysis, data processing and data interpretation) provides unique challenges which fulfil the aim of improving the current status of biological information related to the metabolome and more generally functional genomics. The metabolome has been defined as the qualitative and quantitative collection of all low molecular weight molecules that are

participant's in general metabolic reactions and that are required for the maintenance, growth and normal function of a cell. These include organic species (e.g., amino and fatty acids, carbohydrates, vitamins and lipids). *Saccharomyces cerevisiae* contains approximately 600 metabolites, the plant kingdom has an estimated 200,000 primary and secondary metabolites and the human metabolome can be expected to be even larger in size. Metabolites constitute a diverse set of atomic arrangements when compared to the proteome (arrangement of 20 amino acids) and transcriptome (arrangement of four nucleotide bases bonded with sugar and phosphate backbone) (Jones *et al.,* 2007)and this provides wide variations in chemical (molecular weight, polarity, solubility) and physical (volatility) properties (Ettenhuber *et al.,* 2005). The degree of diversity is showed by the analysis of low molecular weight (MW), polar, volatile organic metabolites, such as ethanol or isoprene to the higher MW, polar (carbohydrates) and non-polar (terpenoids and lipids) metabolites (Eisenreich *et al.,* 2007).

Sample preparation

The time and method of sampling can be influence the reproducibility of the analytical sample. Diurnal and dietary influences can have major effects on the composition of the metabolome, as can the section of a plant sampled. Storage of samples is important, as the continued freeze/thawing of samples can be detrimental to stability and composition. Both invasive (blood, intra-cellular metabolites in plants and microbes) and non-invasive (urine, volatile components, metabolic footprint) sampling can be performed (Kazuki *et al.,* 2010). The extraction of intra-cellular metabolites provides a snapshot of the metabolome, time consuming, and laborious found difficulties when compared to other sampling strategies.

Gas chromatography (GC)

Simplest method is chromatography used in analytic chemistry for separating and analyzing compounds that can be vaporized without decomposition. Typical uses of GC include testing the purity of a particular substance or separating the different components of a mixture. In some situations GC may help in identifying a compound. In preparative chromatography GC can be used to prepare pure compounds from a mixture. It requires high resolution, but requires chemical derivatization for bio molecules and volatile chemicals can be analysed without derivatization. Gas chromatography separates the compound on the basis of their relative vapors pressure, unsuitable for volatile and heat stable compound.

Capillary electrophoresis

Capillary electrophoresis (CE), also known as capillary zone electrophoresis (CZE), can be used to separate ionic species by their charge and frictional forces and mass. In traditional electrophoresis, electrically charged analytes move in a conductive liquid medium under the influence of an electric field. Capillary electrophoresis (CE) was designed to separate species based on their size to charge ratio in the interior of a small capillary filled with an electrolyte (Huang *et al.,* 2008) and (Monton and Soga, 2007).

Mass spectrometry

Mass spectrometry (MS) is an analytical technique for the determination of the elemental composition of a sample or molecule. Role in elucidating the chemical structures of molecules, for example peptides and other chemical compounds, the MS principle consists of ionizing chemical compounds to generate charged molecules or molecule fragments and measurement of their mass-to-charge ratios (Tolstikov and Fiehn 2002).

It is a rapid, sensitive and selective qualitative and quantitative analyses in metabolomics with the ability to identify metabolites. Mass spectrometers operate by ion formation, separation of ions according to their mass-to-charge (m/z) ratio and detection of separated ions.

GC-MS

GC-MS is a use a combined where volatile and thermally stable compounds are first separated by GC and then eluting compounds are detected traditionally by electron-impact mass spectrometers. In metabolomics, GC-MS has been described as the gold standard it is biased against non-volatile, high-MW metabolites. In derivatisation first carbonyl functional groups are converted to oximes with O-alkylhydroxylamine solutions, followed by formation of trimethylsilyl (TMS) esters with silylating reagents to replace exchangeable protons with TMS groups. Some metabolites contain a number of exchangeable protons and hence a range of derivatisation products are formed (e.g., amino acids and carbohydrates will form multiple derivatisation products, whereas organic acids often react to create only one detected product) (Kusano *et al.,* 2007). Oxime formation is required to eliminate undesirable slow and reversible silylation reactions with carbonyl groups, whose products can be thermally labile.

LC-MS

It is another combined method LC-MS, provides metabolite separation by LC followed by electrospray ionisation (ESI) or less typically atmospheric pressure chemical ionisation (APCI). This technique differs from GC-MS in distinct ways (lower analysis

temperatures, and sample volatility not required) and this simplifies sample preparation. Microbial, plant, mammalian biomarker discovery, samples are prepared after intracellular extraction and/or protein precipitation by dilution in an appropriate solvent (Chiwocha *et al.,* 2003). No need of sample derivatisation can be beneficial to improve chromatographic resolution and sensitivity or to provide ionisable groups on metabolites otherwise undetectable by ESI-MS. Quantification is performed by external calibration or response ratio in pharmaceutical applications and peak areas are currently employed in animal, disease, plant and microbial work (Katajamaa and Oresic, 2005). ESI does not result in fragmentation of molecular ions as observed in electron impact mass spectrometers, so it does not allow direct metabolite identification by comparison of ESI mass spectra, as ESI mass spectral libraries are not commonly available. Use of accurate mass measurements and/or tandem MS (MS/MS) to provide collisional induced dissociation (CID) and related mass spectra (MS/MS), metabolite identification can be performed.

Other MS-based techniques

To a lesser extent other MS based techniques have been employed in analysing the metabolome. Capillary electrophoresis mass spectrometry (CE MS) has significant potential. Both high-resolution chromatographic separation and sensitive detection methods have shown the ability to detect up to 1600 metabolites in positive and negative ion modes. (Camacho *et al.,* 2005) Area of growing interest is the application of matrix-assisted laser desorption ionisation (MALDI) laser desorption ionisation (LDI) or direct ionisation on silicon (DIOS) to provide ionisation of metabolite solutions spotted directly on a target plate so allowing minimal sample preparation and high-throughput analysis (Fraser *et al.,* 2007).

FT-IR

The principle of FT-IR based on when a sample is interrogated with light (or electromagnetic (EM) radiation), chemical bonds at specific wavelengths absorb this light and vibrate in one of a number of ways, such as stretching or bending vibrations. The mid-IR can be further broken down into what are termed spectral windows of interest, where strong absorption bands are able to be directly related to specific compounds. FT-IR spectroscopy is a valuable metabolic fingerprinting/foot printing tool owing to its ability to analyse carbohydrates, amino acids, lipids and fatty acids as well as proteins and polysaccharides simultaneously(Krishnan *et al.,.* 2005).

NMR spectroscopy

NMR spectroscopy provides a rapid, non-destructive, high-throughput method that requires minimal sample preparation. NMR spectroscopy functions by the application of strong magnetic fields and radio frequency (RF) pulses to the nuclei of atoms. For atoms with either an odd atomic number (1H) or odd mass number (e.g., 13C), the presence of a magnetic field will cause the nucleus to possess spin, termed nuclear spin (Kikuchi *et al.,* 2004) Absorption of RF energy will then allow the nuclei to be promoted from low-energy to high-energy spin states, and the subsequent emission of radiation during the relaxation process is detected. The NMR spectrum depends on the effect of shielding by electrons orbiting the nucleus. NMR spectroscopy is a high fingerprinting technique. Crude samples are mixed with a reference compound solution (e.g., tetramethylsilane dissolved in D2O for 1H NMR), added to an NMR probe (generally less than 2 ml), inserted into the instrument and analysed (Ward *et al.,* 2007).

This technique is used extensively in clinical and pharmaceutical applications for the analysis of bio fluids or tissues. Studies are based on cells responding to stress, including disease or therapeutic interventions by adjustment of their intra and extra-cellular environments to ensure homeostasis (constant intra-cellular environment). NMR spectroscopy has been employed in other fields for the analysis of plant-cell extracts, such as Arabidopsis and tobacco, to analyse cold stresses on worms, to determine disease biomarkers of environmentally stressed red abalone and to determine the mode of action of biochemical (Urano *et al*., 2009).

Gene identification

Metabolomic study benefit in gene identification. Metabolomics study help to identify particular mQTL which gives particular gene related to that particular trait. It is simplest method of gene identification because once mQTL identify then it is easy to tag gene responsible for that particular metabolite (Lisec *et al*., 2008).

Breeding programme

Human being tries to select desirable progeny from plant population. Initially selection procedure based on the phenotypic appearance of the plants but it required more than ten years to released improved variety. To reduce time duration for released variety use of enzyme based marker, marker assisted selection etc. are applied which reduce time duration up to 6 year. By using mQTL based selection we can reduce time up to 4 year because most of metabolite directly related to particular phenotype and selection of mQTL easy and faster than that of MAS (Alisdair *et al.,* 2005).

Human metabolome project

The purpose of the project is to facilitate metabolomics research to improve disease identification, prognosis and monitoring; provide insight into drug metabolism and toxicology; provide a linkage between the human metabolome and the human genome; and to develop software tools for metabolomics. More than 800 compounds have been identified and by end of 2006, it is expected that more than 1400 metabolites will have been identified, quantified.

In conclusion, metabolomics is particularly important in the plant field, because plants produce a huge diversity of metabolites far more than are produced by animals and microorganisms. Comprehensive coverage of metabolome analysis is achieved not by a single analytical technology but by multiparallel complementary technologies. Increasing the annotation rate of unknown signals is a big challenge. The cooccurrence principle of transcripts and metabolites, particularly transcriptome coexpression network analysis, is powerful for decoding functions of genes not only in a model plant such as Arabidopsis but also in crops and medicinal plants. Metabolome quantitative trait loci (mQTL) analysis along with scoring of gene expression and agronomical traits is beneficial for crop breeding. (Wentzell *et al.,* 2007) Metabolomics plays a key role in the evaluation of genetically modified crops. Systems biology, data-driven by metabolomics and other omics will play a key role in understanding plant systems and developing further biotechnology applications.

Future issues

Further advancements in metabolomics technologies are needed to cover the wide array of phytochemicals and increase the spatial and temporal resolution of metabolome analysis. Strengthening of spectral databases and sophisticated informatics for more accurate annotation of

metabolites is necessary, and these efforts must be internationally harmonized and publicly available.

The databases of metabolome accumulation in Arabidopsis and other plants are anticipated to be integrated more intensively with transcriptomics and proteomics in the future. On the basis of these advances, we expect the introduction of high-throughput functional genomics of crops and medicinal plants through more holistic omics analyses, which would directly lead to metabolic engineering applications (Hirai *et al*., 2007).

Metabolic genomics studies adopting mQTL would be promising for crop breeding for Improvements in agriculture and nutraceutical and pharmaceutical production (Lisec *et al*., 2008).

The development of data-driven metabolic systems biology could open new channels for high-speed construction and evaluation of hypotheses for multilayer mechanisms for cellular regulatory systems.

References

Alisdair, R., Fernie and Nicolas Schauer. 2008. Metabolomics-assisted breeding: a viable option for crop improvement? *Trends in Genetics,* 25: 39-48.

Camacho, D., dela Fuente, A., Mendes, P. 2005. The origin of correlations in metabolome data. *Meta bolomics,* 1: 53–63.

Carrari, F., Baxter, C., Usadel, B., Urbanczyk-Wochniak, E., Zanor, M.I., *et al*. 2006. Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. *Plant Physiol.*, 142: 1380–9.

Choi, H., Nesvizhskii, A.I. 2008. False discovery rates and related statistical concepts in mass spectrometry based proteomics. *J. Proteome Res.,* 7: 47–50.

Clark, R.M., Schweiker, G., Toomajian, C., Ossowski, S., Zeller, G*., et al.* 2007. Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. *Sci.,* 317: 338–42.

Codrea, M.C., Jimenez, C.R., Heringa, J., Marchiori, E. 2007. Tools for computational processing of LC-MS datasets: a user's perspective. *Comput. Methods Programs Biomed.,* 86: 281–90.

Colquhoun, I.J., Le Gall, G., Elliott, K.A., Mellon, F.A., Michael, A.J. 2006. Shall I compare thee to a GM potato? *Trends Genet.,* 22: 525–2.

Chiwocha, S.D.S., Abrams, S.R., Ambrose, S.J., Cutler, A.J., Loewen, M., *et al*. 2003. A method for profiling classes of plant hormones and their metabolites using liquid chromatography-electrospray ionization tandem mass spectrometry: an analysis of hormone regulation of thermodormancy of lettuce (Lactuca sativa L.) seeds. *Plant J.,* 35: 405–17.

Choi, H., Nesvizhskii, A.I. 2008. False discovery rates and related statistical conceptsin mass spectrometry based proteomics. *J. Proteome Res*., 7: 47–50

Eisenreich, W., Bacher, A. 2007. Advances of high-resolution NMR techniques in the structural and metabolic analysis of plant biochemistry. *Phytochem.,* 68: 2799–815.

Ettenhuber, C., Radykewicz, T., Kofer, W., Koop, H.U., Bacher, A., Eisenreich, W. 2005. Metabolic flux analysis in complex isotopolog space. Recycling of glucose in tobacco plants. *Phytochem.,* 66: 323–35.

Fraser, P.D., Enfissi, E.M., Goodfellow, M., Eguchi, T., Bramley, P.M. 2007. Metabolite profiling of plant carotenoids using the matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Plant J.,* 49: 552–64.

Fernie, A.R., Schauer, N. 2009. Metabolomics-assisted breeding: a viable option for crop improvement? *Trends Genet.,* 25: 39–48.

Fiehn, O. 2002. Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.,* 48: 155–7.

Gigolashvili, T., Yatusevich, R., Berger, B., Muller, C., Fl¨ugge, U.I. 2007.TheR2R3-MYB transcription factor HAG1/MYB28 is

a regulator of methionine-derived glucosinolate biosynthesis in Arabidopsis thaliana. *Plant J.,* 51: 247–61.

Huhman, D.V. and Summer, L.W. 2002. Metabolic profile of saponins in medicago sativa and medicago truncatula using HPLC coupled to an electoscopy ion-trap mass spectrometer, *Phytochem.,* 59: 347-360.

Huang, W.H., Ai, F., Wang, Z.L., Cheng, J.K. 2008. Recent advances in single-cell analysis using capillary electrophoresis and microfluidic devices. *J. Chromatogr. B.,* 866: 104–2.

Hirai, M.Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi T., *et al*. 2007. Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl. Acad. Sci. USA* 104: 6478–83

Jones Gu, L., A.D., Last, R.L. 2007. LC-MS/MS assay for protein amino acids and metabolically related compounds for large-scale screening of metabolic phenotypes. *Anal. Chem.,* 79: 8067–7.

Kikuchi, J., Shinozaki, K., Hirayama, H. 2004. Stable isotope labeling of Arabidopsisthaliana for an NMR based metabolomics approach. *Plant Cell Physiol.,* 45: 1099–110.

Krishnan, P., Kruger, N.J., Ratcliffe, R.G. 2005. Metabolite fingerprinting and profiling in plants using NMR. *J. Exp. Bot.,* 56: 255–65.

Katajamaa, M., Oresic, M. 2005. Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics*, 6: 17.

Kusano, M., Fukushima, A., Kobayashi, M., Hayashi, N., Jonsson, P., *et al.* 2007. Application of a metabolomics method combining one-dimensional and two-dimensional gas chromatography-time-of-flight/mass spectrometry to metabolic phenotyping of natural variants in rice. *J. Chromatogr. B.,* 855: 71–79.

Kazuki Saito and Fumio Matsuda. 2010. Metabolomics for functional genomics, systems biology and biotechnology. *Ann. Rev. Plant. Biol.,* 61: 463-489.

Lisec, J., Meyer, R.C., Steinfath, M., Redestig, H., Becher, M., *et al*. 2008. Identification of metabolic and biomass QTL in Arabidopsis thalianaina parallel analysis of RIL and IL populations. *Plant J.,* 53: 960–7.

Monton, M.R., Soga, T. 2007. Metabolome analysis by capillary electrophoresis–mass spectrometry. *J. Chromatogr. A.,* 1168: 237–46.

Sawada, Y., Kuwahara, A., Nagano, M., Narisawa, T., Sakata, A., *et al*. 2009. Omics-based approaches to methionine side chain elongation in Arabidopsis characterization of the genes encoding methylthio alkylmalate isomerase and methylthio alkylmalate dehydrogenase. *Plant Cell Physiol*., 50: 1181–90.

Tolstikov, V.V., Fiehn, O. 2002. Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Anal. Biochem.,* 301: 298–307.

Urano, K., Maruyama, K., Ogata, Y., Morishita, Y., Takeda, M., *et al*. 2009. Characterization of the ABA regulated global responses to dehydration in Arabidopsis by metabolomics. *Plant J.,* 57: 1065–78.

Wentzell, A.M., Rowe, H.C., Hansen, B.G., Ticconi, C., Halkier, B.A., Kliebenstein, D.J. 2007. Linking metabolic QTLs with network and cis-eQTLs controlling biosynthetic pathways. *PLoS Genet.,* 3: 1687–701,

Ward, J.L., Baker, J.M., Beale, M.H. 2007. Recent applications of NMR spectroscopy in plant metabolomics. *FEBS J.,* 274: 1126–31.