Journal of Proteomics xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Proteomics in commercial crops: An overview

Boon Chin Tan [a,*], Yin Sze Lim [b], Su-Ee Lau [a]

[a] Centre for Research in Biotechnology for Agriculture, University of Malaya, Lembah Pantai, 50603 Kuala Lumpur, Malaysia
[b] School of Biosciences, Faculty of Science, University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 27 December 2016
Received in revised form 21 April 2017
Accepted 19 May 2017
Available online xxxx

Keywords:
2-DE
Crop improvement
Mass spectrometry
Plant proteomics

ABSTRACT

Proteomics is a rapidly growing area of biological research that is positively affecting plant science. Recent advances in proteomic technology, such as mass spectrometry, can now identify a broad range of proteins and monitor their modulation during plant growth and development, as well as during responses to abiotic and biotic stresses. In this review, we highlight recent proteomic studies of commercial crops and discuss the advances in understanding of the proteomes of these crops. We anticipate that proteomic-based research will continue to expand and contribute to crop improvement.

Significance: Plant proteomics study is a rapidly growing area of biological research that is positively impacting plant science. With the recent advances in new technologies, proteomics not only allows us to comprehensively analyses crop proteins, but also help us to understand the functions of the genes. In this review, we highlighted recent proteomic studies in commercial crops and updated the advances in our understanding of the proteomes of these crops. We believe that proteomic-based research will continue to grow and contribute to the improvement of crops.

1. Introduction

The term "proteomics" describes the comprehensive identification and quantitative analysis of protein expression in an organism, cell, tissue or organelle at a specific time under certain conditions [1]. It is an extension of the word "proteome" (protein complement of the genome [2,3]) which was first coined by Marc Wilkins in 1994. Although a vast amount of information has become available as a result of genomic sequencing, researchers are realising that complete genomic sequences provide insufficient information for elucidating biological functions [4]. Furthermore, cells usually depend on multiple metabolic and regulatory pathways for survival. Because proteomes reflect the processes occurring in biological systems, an understanding of these proteome profiles would provide better insight into such metabolic processes and their interactions with other regulatory pathways in a biological system [5].

Proteomics is among the rapidly growing areas of biological research that are positively affecting plant science. This technology allows qualitative and quantitative measurements of important proteomes in specific cell types or organelles during specific developmental and physiological stages and interactions [3,6]. Exponential progress has been achieved since the first plant proteomic study in maize [7], though the potential of the plant proteomics is far from being fully exploited. Indeed, compared with human and yeast proteomics, plant proteomics lags far behind [8]. However, recent advancements in new or improved technologies and protocols or workflows have provided new possibilities for high-throughput proteome analyses and have reduced errors in protein assessment.

Translational proteomics is currently gaining much attention from plant researchers. Output from any discovery in model plants is being applied to commercial crops, to address challenges faced in the field [9,10], including improving crop tolerance to environmental stresses and enhancing the quality and yield of agricultural production to assure food safety and security [4].

In this review, we highlight certain proteomic findings in commercial crops; we also provide an update on recent progress as well as the potential of rapidly evolving proteomic approaches to crop improvement.

2. Tools and technologies

Proteomic analysis is usually performed through either gel-based analysis (involving initial protein separation through gel electrophoresis followed by quantification, protein spot digestion and mass spectrometric (MS) identification) or gel-free analysis (involving protease degradation of protein mixtures followed by liquid chromatographic (LC) separation and MS identification) [11]. Gel-based analysis includes one- or two-dimensional polyacrylamide gel electrophoresis (1- or 2-DE) [12] and differential in-gel electrophoresis (DIGE) [13]. Gel-free technologies include multidimensional protein identification technology (MudPIT) for peptide separation [14], isotope-coded affinity tagging

* Corresponding author.
E-mail address: boonchin@um.edu.my (B.C. Tan).

http://dx.doi.org/10.1016/j.jprot.2017.05.018
1874-3919/

[15], isobaric tagging for relative and absolute quantification [16], stable isotope labelling of amino acids in cell culture [17], isotope-coded protein labelling [18] for peptide quantification, and label-free methods (peak integration or spectral counting). Several reviews have provided information on these techniques [19,20].

2.1. Gel-based proteomics

Gel-based proteomic techniques are the most commonly used methods for global protein analysis [21,22] and involve a separation step (usually 2-DE) and an identification step (MS). These techniques have been extensively reviewed [23,24,25]. 2-DE resolves proteins on the basis of isoelectric point (p*I*) and molecular mass (M_r) [26]. The separated protein spots can then be stained, with Coomassie brilliant blue, silver nitrate, or SYPRO Ruby [27], among others. When combined with advanced MS techniques, 2-DE allows hundreds of proteins to be characterised in a single polyacrylamide gel [28], including the position of the protein spot (p*I* and M_r) on the gel. This capability of 2-DE has allowed for analysis of post-translational modifications (PTMs) of proteins.

DIGE was developed to improve the reproducibility of 2-DE and to overcome gel-to-gel variation [13]. Each protein sample is labelled at a lysine residue with different fluorophores, such as CyDye2, CyDye3, and CyDye5 [29], prior to mixing and separation on the same gel, and the abundance of the same protein in different samples can easily be determined by using these fluorophores [28]. This technique reduces the number of gels needed for one experiment and is able to detect as little as 150 pg of a single protein with a linear response in protein concentration of over five orders of magnitude. In comparison, silver staining can detect only 1 ng of protein with a dynamic range of less than two orders of magnitude [30]. The relatively high cost of DIGE equipment, software, and consumables, however, has limited its use. Three-dimensional separation of proteins has also been developed to avoid protein co-migration [31]. After isoelectric focusing, proteins are separated by two consecutive SDS-PAGE runs using two different buffer systems. Colignon and colleagues [31] have found that a 3-D approach is able to increase the number of spots analysed and thus improve the accuracy of protein identification and comparative quantification.

The desired goal of any proteomic study is to identify, characterise and quantify proteins of interest, commonly by MS. The selected protein spots are digested with a site-specific protease, usually trypsin, to produce a set of peptides that are subjected to MS. A correctly identified protein will have a large number of "matching" peptides after a database search.

Mass spectrometers include an energy source to ionise a sample, a mass analyser for ion separation according to the mass/charge ratio (m/z), and a detector for detecting ions [32]. Two types of ionisation are commonly used for proteomic studies, matrix-assisted laser desorption-ionisation (MALDI) and electrospray ionisation (ESI) [33]. Four types of mass analysers are currently used: time of flight (TOF), ion trap (quadrupole ion trap, linear ion trap (LIT), or linear trap quadrupole), triple-quadrupole tandem MS (MS/MS), and Fourier transform ion cyclotron resonance [34], which differ in several aspects, including sensitivity, resolution, and mass accuracy. Hence, the choice of mass spectrometer is usually determined by the intended application. As this review does not focus on this technology, more detailed information can be found in other reviews [35,36].

2.2. Limitation of gel-based proteomics

Despite the successes of 2-DE, the method has many limitations [37]. For example, 2-DE can separate only 30–50% of the entire proteome, depending on the tissue, and it is unable to separate all the proteins present in a complex sample [38]. In practice, the total proteome coverage using this method is restricted to proteins with an M_r of 10–120 kDa and a neutral to acidic p*I*. Strongly alkaline proteins (pH > 9.5), such as ribosomal and nuclear proteins, are difficult to focus [21]. Low-abundance proteins with physiological relevance, including regulatory and signal-transducing proteins, are also rarely detected on traditional 2-DE gels, because the large amount of highly abundant proteins masks their detection [21,39]. For instance, ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), which accounts for a large percentage of total plant protein, hinders absorption of low-abundance proteins on the IPG strips and results in poor detection and identification of these proteins on 2-D gels and by MS. Because of these drawbacks, an additional RuBisCO depletion step is required to reduce sample complexity and to improve recovery of low-abundance plant proteins. The use of narrow-range (2–3 pH units) and overlapping pH gradient strips also enhances the number of proteins resolved [21]. In addition, increasing the loading capacity of narrow-range strips has enabled the visualisation and identification of previously undetected proteins. Several RuBisCO depletion/removal methods using poly-ethylene glycol (PEG), calcium and phytate, protamine sulfate, and anti-RuBisCO antibodies have been reported to enhance proteome coverage, and it is advisable that such measures be included in every plant protein extraction step [40].

Although it is labour intensive and requires trained personnel to obtain reproducible results, 2-DE undoubtedly remains a standard tool for protein separation [28,41]. Furthermore, some degree of gel-to-gel or run-to-run variability in the detection of the same protein set might occur. Thus, it can be challenging to achieve a high degree of reproducible profiles between two replicate experiments [33]. To overcome this limitation, the variability coefficients of reference spots should be as low as possible [21]. The drawbacks of 2-DE have encouraged the application of gel-free proteomics for analysing proteomes [33].

2.3. Gel-free proteomics

The challenges facing proteomic studies cannot be addressed by gel-based strategies alone, and the drawbacks of gel-based proteomic approaches have motivated the development of alternative gel-free proteomic techniques, either to overcome limitations or to entirely replace these gel-based techniques [42]. Gel-free approaches involve tag-based labelling, metabolic labelling, and label-free methods. For tag-based labelling, different mass tags such as ICAT, iTRAQ, TMT, dimethyl labelling, and ^{18}O labelling are introduced into proteins or peptides; in contrast, metabolic labelling methods, such as SILAC and ^{15}N labelling, involve stable isotope labelling of proteins in living cells. Label-free methods use multidimensional capillary LC coupled to nanoESI tandem MS to separate and identify the peptides obtained via enzymatic digestion of proteins without any labelling [43]. All these gel free-based methods have particular strengths and weaknesses (reviewed by [44]), and each should be selected according to the aim of experiments and the types of samples.

Gel-free methods are more reproducible and show far less bias than do gel-based methods [45], as exemplified by a number of studies comparing protein analysis using both approaches. Nouri and Komatsu [46] have performed a comparative proteomic analysis of the soybean plasma membrane under osmotic stress conditions. Four and eight protein spots were identified as high- and low-abundance proteins, respectively, using a gel-based method, whereas 11 and 75 proteins were identified as high- and low-abundance proteins, respectively, by nano LC-MS/MS. Using the same strategy, Cutsem et al. [47] have successfully identified 680 and 858 proteins of *Nicotiana tabacum* trichomes through gel-based and gel-free approaches, respectively. Recently, gel-free methods have been used to analyse somatic embryogenesis [48,49,50], seed germination and development [51,52,53], biotic stress [54,55,56,57,58], abiotic stress [59,60,61,62], fruit development and ripening [63,64].

2.4. Limitations of label-free quantification

Peptides shared by multiple proteins, also known as non-unique or degenerate peptides, limit reliable identification of proteins isolated

Please cite this article as: B.C. Tan, et al., Proteomics in commercial crops: An overview, J Prot (2017), http://dx.doi.org/10.1016/j.jprot.2017.05.018