Charting plant interactomes: possibilities and challenges

Mustafa Morsy1, Satyanarayana Gouthu1, Sandra Orchard2, David Thorneycroft2, Jeff F. Harper1, Ron Mittler1,3 and John C. Cushman1

1 Department of Biochemistry and Molecular Biology, MS200, University of Nevada, Reno, NV 89557, USA
2 European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge, CB10 1SD, UK
3 Department of Plant Science, Hebrew University of Jerusalem, Jerusalem 91904, Israel

Protein–protein interactions are essential for nearly all cellular processes. Therefore, an important goal of post-genomic research for defining gene function and understanding the function of macromolecular complexes involves creating ‘interactome’ maps from empirical or inferred datasets. Systematic efforts to conduct high-throughput surveys of protein–protein interactions in plants are needed to chart the complex and dynamic interaction networks that occur throughout plant development. However, no single approach can build a complete map of the interactome. Here, we review the utility and potential of various experimental approaches for creating large-scale protein–protein interaction maps in plants. Bioinformatics approaches for curating and assessing the confidence of these datasets through inter-species comparisons will be crucial in achieving a complete understanding of protein interaction networks in plants.

Introduction: the importance of interacting

Protein–protein interactions (PPIs) are essential for almost all biological functions mediated by macromolecular machinery including replisomes, spliceosomes, ribosomes, proteasomes, signalosomes, catalytic enzyme modules, signal recognition complexes, and specialized protein complexes unique to plants such as light harvesting and photosystem complexes. Thus, identifying, quantifying, localizing, and modeling entire PPI networks (or ‘interactomes’) is a key prerequisite for understanding the biological basis of all cellular processes and for creating a framework to characterize the function of all proteins [1].

Plant proteomes of 30 000-40 000 proteins [2] are expected to have an estimated 75 000-150 000 interaction pairs, extrapolating from estimates of the size of the yeast interactome [3,4]. This illustrates the magnitude of the challenge for creating comprehensive interactome maps. Although several small- to mid-sized protein interaction studies have been published (see Supplementary Table 1), no large-scale experimental interactome mapping efforts have been reported in plants to date. The purpose of this review is to survey the frequently applied techniques for detecting and characterizing PPIs in plants and to highlight established and newly emerging high-throughput technologies that will become increasingly important to gaining a comprehensive, functional understanding of plant interactomes Table 1.

The challenge of studying protein–protein interactions

A wide range of approaches have been developed to characterize PPIs. These include both in vitro and in vivo experimental methods [5,6], and in silico approaches that are based on literature curation [7], evolutionarily conserved complexes [8–10], and network integration to infer functional modules (Figure 1) [1,11]. However, no single experimental approach can determine all PPIs. This is due, in part, to the diverse physiochemical properties, abundance, and subcellular localization patterns of individual proteins and the wide range of equilibrium dissociation constants of different protein complexes. Single approaches can supply only limited information about the key parameters of interactions or complexes, such as the oligomeric state of the interacting partner, the stoichiometric ratio within a complex, the relative affinity of interacting partners towards one another, the location of interaction sites, and the dependency of interactions on post-translational modifications [5]. Although comparative interactome approaches are useful among animal models, they provide an incomplete view of the plant interactome. This is because plants have evolved unique gene repertoires in order to accommodate their photoautotrophic and sessile lifestyle, and these functions are not present in other model organisms [12]. Therefore, it is crucial to generate plantspecific PPI interaction data using a variety of complementary approaches (Figure 1).

In vivo identification of protein–protein interactions

The ability to study PPIs within their native cellular context is crucial. In vivo approaches offer the ability to discover and study known or novel PPIs, often in real-time in living cells.

Yeast two-hybrid strategies

The yeast two-hybrid (Y2H) screening method provides a simple, robust binary readout of protein-protein interactions that has made it one of the most popular methods for the compilation of proteome-wide interactome maps [13]. The Y2H approach exploits the modular nature of transcription factors containing DNA-binding and activation domains. When these domains are split, the factor
is functionally disabled. However, if each domain is fused to two interacting proteins as hybrid proteins (thus the name ‘two-hybrid’), the function of the factor can be restored to transcribe one or more selectable marker or reporter genes (two-hybrid), the function of the factor can be restored to transcribe one or more selectable marker or reporter genes (‘two-hybrid’), the function of the factor can be restored to transcribe one or more selectable marker or reporter genes (‘two-hybrid’).

The classical Y2H system has also been adapted for use in Arabidopsis protoplasts and has the potential of detecting protein interactions that are undetectable in heterologous systems [18]. Although this technique is suitable for the detection of nuclear protein interactions and can be adapted to a high-throughput format, protein interactions within other discrete subcellular compartments might be undetectable using this system. The production of high quality protoplasts can also make this methodology technically challenging [19]. Aside from the Y2H system, a variety of other options are now available to plant researchers to study PPIs in the context of living plant cells.

**Protein (fragment) complementation strategies**

Protein (fragment) complementation assays (PCAs) split an enzyme or a fluorescent reporter protein into two fragments that are fused to two interaction candidate proteins. Upon interaction, the split reporter protein is reconstituted (Figure 2). In contrast to classical Y2H strategies,

<table>
<thead>
<tr>
<th>Methods</th>
<th>Throughput</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In vivo approaches</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yeast two-hybrid (Y2H)</td>
<td>High</td>
<td>Mature; robust; widely used; can be automated</td>
<td>High false-positive rate, interactions limited to nucleus in classical system</td>
<td>[15,16]</td>
</tr>
<tr>
<td>Protoplast Y2H</td>
<td>Medium</td>
<td>Plant cell environment, potentially more sensitive than yeast system</td>
<td>High false-positive rate, interactions limited to nucleus; protoplasts</td>
<td>[18]</td>
</tr>
<tr>
<td>Split-ubiquitin system (SUS)</td>
<td>Medium</td>
<td>Interactions can be detected in diverse cellular compartments</td>
<td>High false-positive rate; high background of enzymatic reporters</td>
<td>[20,22–24]</td>
</tr>
<tr>
<td>Bimolecular fluorescence complementation (BiFC)</td>
<td>Medium</td>
<td>Interactions can be detected in diverse cellular compartments</td>
<td>Multiple fusion orientations should be tested; non-specific interactions can be reported; background fluorescence</td>
<td>[26–29]</td>
</tr>
<tr>
<td>Split-luciferase system</td>
<td>Medium–high</td>
<td>Extremely low background</td>
<td>No subcellular localization information provided</td>
<td>[33–35]</td>
</tr>
<tr>
<td>Fluorescence resonance energy transfer (FRET)</td>
<td>Low</td>
<td>Quantitative; subcellular localization (time-resolved interactions possible)</td>
<td>Laborous; technically demanding; sophisticated imaging system required</td>
<td>[30]</td>
</tr>
<tr>
<td>Bioluminescence resonance energy transfer (BRET)</td>
<td>Low</td>
<td>Avoids energy donor photo-bleaching</td>
<td>Laborous; technically demanding; very sensitive charge-coupled device (CCD) camera required</td>
<td>[37,38]</td>
</tr>
<tr>
<td><strong>In vitro approaches</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Native chromatography or electrophoretic purification</td>
<td>Medium</td>
<td>Simple</td>
<td>Potentially poor resolution of complexes; high sample complexity</td>
<td>[39]</td>
</tr>
<tr>
<td>Immunopurification</td>
<td>Low</td>
<td>Reduced complexity sufficient for MS–MS identification</td>
<td>Specific antibody required, expensive; non-specific protein contamination</td>
<td>[41]</td>
</tr>
<tr>
<td>Single affinity purification-tagging Tandem affinity purification (TAP) tagging</td>
<td>Medium–High</td>
<td>Efficient, flexible, low cost</td>
<td>Non-specific protein contamination</td>
<td>[42]</td>
</tr>
<tr>
<td>Stable-isotope labeling of amino acids in cell culture (SILAC)</td>
<td>High</td>
<td>High sample purity; reduces non-specific proteins</td>
<td>Weak interactions can be disrupted; promiscuous protein interactors</td>
<td>[44,47–48]</td>
</tr>
<tr>
<td>Chemical crosslinking-MS using protein interaction reporter (PIRs)</td>
<td>High</td>
<td>Uniform direct isotopic labeling without chemical differences</td>
<td>Incomplete (~80%) isotopic labeling; expensive; uses cell cultures</td>
<td>[53,54]</td>
</tr>
<tr>
<td>Protein microrarrays</td>
<td>High</td>
<td>Inexpensive, complete labeling</td>
<td>Reliable identification of interactions within complex protein mixtures</td>
<td>[55,56]</td>
</tr>
<tr>
<td>Surface plasmon resonance</td>
<td>Low–medium</td>
<td>Simple; large-scale identification of interactions possible</td>
<td>Technically challenging; requires large collections of expression ready cDNA clones</td>
<td>[57,58]</td>
</tr>
<tr>
<td><strong>In silico approaches</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predicted interactomes</td>
<td>High</td>
<td>Experimental data from plants not essential</td>
<td>Limited by interolog annotations; incomplete</td>
<td>[74,85]</td>
</tr>
</tbody>
</table>

**Table 1. Summary of protein–protein interaction methodologies and their utility in low-, medium- and high-throughput studies**

- **Throughput**: High (≥ 35% of interactions detected), Medium (10–35% of interactions detected), Low (< 10% of interactions detected).
- **Advantages**: Mature; robust; widely used; can be automated; plant cell environment, potentially more sensitive than yeast system; interactions can be detected in diverse cellular compartments; extremely low background; quantitative; subcellular localization (time-resolved interactions possible).
- **Disadvantages**: High false-positive rate, interactions limited to nucleus in classical system; high false-positive rate, interactions limited to nucleus; protoplasts technically challenging; multiple fusion orientations should be tested; non-specific interactions can be reported; background fluorescence.
- **Reference**: [15,16] (Yeast two-hybrid), [18] (Protoplast Y2H), [20,22–24] (Split-ubiquitin system), [26–29] (Bimolecular fluorescence complementation), [33–35] (Split-luciferase system), [30] (Fluorescence resonance energy transfer), [37,38] (Bioluminescence resonance energy transfer), [39] (Native chromatography or electrophoretic purification), [41] (Immunopurification), [42] (Single affinity purification-tagging), [44,47–48] (Tandem affinity purification), [53,54] (Stable-isotope labeling of amino acids in cell culture), [55,56] (Chemical crosslinking-MS), [57,58] (Protein microrarrays), [61–63] (Surface plasmon resonance).

- **Reference**: [26–29] (Bimolecular fluorescence complementation), [20,22–24] (Split-ubiquitin system), [33–35] (Split-luciferase system), [30] (Fluorescence resonance energy transfer), [37,38] (Bioluminescence resonance energy transfer), [39] (Native chromatography or electrophoretic purification), [41] (Immunopurification), [42] (Single affinity purification-tagging), [44,47–48] (Tandem affinity purification), [53,54] (Stable-isotope labeling of amino acids in cell culture), [55,56] (Chemical crosslinking-MS), [57,58] (Protein microrarrays), [61–63] (Surface plasmon resonance).
PCAs permit protein interactions to be surveyed in diverse subcellular compartments [20]. Examples of this approach include the split-ubiquitin system (SUS) [21], which has been used in tomato [22] and *Arabidopsis* [23,24]. Since the initial use of ubiquitin, a wide variety of reporter enzymes have been used for split-protein-based screening systems [5,20,25].

To overcome limitations of high background signals and false positive rates associated with enzymatic reporters, bimolecular fluorescent complementation (BiFC) assays have been developed and used in plant cells [26,27]. BiFC is based on the formation of a fluorescent complex from two non-fluorescent fragments of auto-fluorescent proteins (AFPs) by the interaction between the proteins to which these fragments are fused (Figure 2). By using different combinations of AFP variants, numerous spectrally distinct complexes can form, and this enables multiplexed BiFC assays to be performed within a single living cell [28]. BiFC assays have been used successfully in a variety of plant species, including *Arabidopsis thaliana*, *Nicotiana* spp., and *Allium* sp. [27]. Versatile vector and delivery systems have been developed, enabling detection of specific PPIs in various subcellular compartments [29], and delivery of constructs by particle bombardment, agroinfiltration, or protoplast transfection [30].

Although the BiFC method is simple, rapid and suitable for medium-throughput screens, researchers should be aware of its potential drawbacks. Multiple fusion orientations should be tested, and the fusion proteins should be expressed at concentrations comparable to those of their endogenous counterparts [26,27]. Some very transient interactions can go undetected if the dissociation rate of a particular complex is faster than the rate of association between the fluorescent-protein fragments [31] or if the stability or molecular properties of the chimeric fusion proteins differ from those of the native proteins [26]. By contrast, if the intracellular concentrations of a split-AFP half is sufficiently high, then non-specific interactions might be reported, owing to the ability of some AFP fragments to form fluorescent chromophores [e.g. large N-terminal fragment of split-green fluorescent protein (GFP)] or complexes in the absence of fusions to specific interaction partners [e.g. yellow fluorescent protein (YFP)] [27].

An intrinsic drawback of the use of fluorescence-based detection systems in plants is the high degree of cellular auto-fluorescence. This can be partially overcome through the use of newly developed red fluorescent proteins (RFPs) that have long excitation and emission wavelengths and have been used successfully in BiFC systems [32]. Alternatively, PCAs that use split luciferase derived from either coral (*Renilla*) or firefly offer the ability to monitor PPIs using bioluminescence in living cells with extremely low background fluorescence [33]. In addition, they are useful for conducting large-scale PPI screens [34], as reported using *Arabidopsis* protoplasts in a 96-well plate format [35]. The disadvantage of this system is that it does not provide information about the subcellular location of the interaction. Although not a common occurrence, the possibility of ‘bridging’ proteins, which might mediate the interaction between two query proteins, should be considered when using PCA and BiFC approaches as screening assays, because these could result in apparent interactions that are not truly binary.

Fluorescence or bioluminescence resonance energy transfer

The availability of a wide variety of spectral variants of AFPs [36] has enabled the use of Förster (or fluorescence) resonance energy transfer (FRET) to detect PPIs in living plant cells (Figure 2). FRET is the radiation-less energy transition from a donor (e.g. GFP) to an acceptor fluorophore [e.g. red fluorescent protein (RFP)] that are in very close proximity (<10 nM) to one another and that have the appropriate steric orientation [30]. FRET and its variations, such as fluorescence life-time imaging microscopy (FLIM), are very powerful approaches for determining time-resolved, quantitative, subcellular localization of specific PPIs within single living cells. However, these techniques have not been used as widely as BiFC approaches, owing to their being more labor-intensive and technically demanding [30].

The limitations of high-background auto-fluorescence within plant cells, and the need for direct excitation of a fluorescence acceptor can be overcome by bioluminescence resonance energy transfer (BRET). Instead of using light emitted from an excited AFP, BRET employs a bioluminescence ‘energy donor’ such as blue-light emitting *Renilla* luciferase and either GFP or YFP as the ‘energy acceptor’. BRET has been used successfully to monitor PPIs in
Arabidopsis and tobacco cells [37,38]. BRET is more amenable to high-throughput screening than FRET because no internal excitation light source is needed and the energy donor does not become photo-bleached. However, light emission from BRET is generally rather dim, and detection of light emission by the fluorescence acceptor requires a very sensitive charge-coupled device (CCD) camera [38]. Although in vivo studies are often of primary importance for understanding PPI in the context of living cells, such approaches often require validation by complementary in vitro methodologies. In vitro approaches also offer very accessible and informative means of performing high-throughput PPI analyses.

In vitro identification of protein–protein interactions

The most direct approach towards elucidating PPIs in vitro is the preparative isolation and fractionation of native protein complexes using a variety of one-dimensional and multidimensional electrophoretic or chromatographic methods coupled with matrix-assisted laser desorption ionization-time-of-flight (MALDI–TOF) tandem mass spectrometry (MS–MS) [39]. Although this simple approach has been reported in only a small number of studies of plant protein complexes, it should be explored as a potentially useful complement to affinity-based purification strategies.

Affinity purification strategies

Affinity-based purification (AP) methods typically result in protein preparations of suitable purity and reduced complexity that they can be interrogated efficiently by MS–MS-based protein identification approaches [40]. Although direct AP approaches have been used in plants [41], these methods typically rely upon purification of a bait protein and interacting prey proteins using a custom antibody (preferably monoclonal) specific for the bait or a commercially available antibody to a peptide or protein epitope tag fused to the N or C termini of the bait (Figure 3). Epitope tagging strategies afford researchers greater flexibility and reduce costs associated with the production of antibodies to specific proteins [42]. The use of multiple affinity tags permits two consecutive or ‘tandem’ affinity purification (TAP) steps to be performed, usually under very mild and selective elution conditions [43]. Although the TAP approach reduces the amount of contaminating proteins, the increased processing involved can disrupt weaker interactions, lowering the range and abundance of proteins observed. TAP-tagging and MS–MS has been performed successfully in Synechocystis sp. PCC6803 [44], A. thaliana [44–46], rice [47], and to complement the Y2H approach in barley [48]. TAP-tagging systems with cleavage sites for a more specific and low-temperature active protease [49], and Gateway® compatible vectors have been developed to facilitate this approach in plants [50,51]. Using formaldehyde or glutaraldehyde stabilizes weak or transient protein interactions that might be lost during purification [42]. Epitope tagging approaches can be limited by various factors: the efficiency with which some fusion proteins compete with endogenous proteins for interaction with the bait [47]; the recurrent purification of artifactual
proteins with promiscuous binding properties [47]; or N- or C-terminal fused tags disrupting PPIs or protein targeting of tagged proteins [49].

**Quantitative MS strategies**

An exciting development within the field of PPI studies has been the application of quantitative MS strategies, involving the incorporation of stable isotopes into proteins using metabolic labelling, and subsequent comparison of natural versus heavy isotopic ratios of the labeled proteins under control and test conditions [52].

Stable-isotope labeling of amino acids in cell culture (SILAC) involves growing cells in a suitable medium containing isotopically labeled amino acids [53]. Incorporation of isotopically labeled arginine into cell cultured proteomes and subsequent identification of 14-3-3 protein phosphorylation-specific interactions with bait peptides was demonstrated [54]. 14-3-3 proteins, named for their characteristic migration pattern following electrophoretic separation, participate in signaling by binding to a functionally diverse set of signaling proteins through recognition motifs that contain phosphorylated Ser or Thr residues [48]. Although plants can synthesize amino acids directly, cell cultures can be used in SILAC experiments, but incorporation of only 80% was achieved owing to the autotrophic nature of plant cells [54]. The advantage of SILAC-based methods is the efficient direct incorporation of the isotopic amino acid counterparts, which have no chemical differences, into the target proteome and the analysis of these in the same experiment. However, the technique is largely restricted to the use of cell cultures at present, primarily because of the high cost of isotopically labeled amino acids. These limitations can be overcome by the complete metabolic labeling of whole *Arabidopsis* seedlings in shaking liquid cultures using 15N-enriched inorganic ammonium and nitrate salts as the sole nitrogen source. This approach appears to be a more robust and affordable strategy for isotopic labeling studies than SILAC [55,56] and should provide an effective tool for quantitative interactome studies in the future. However, one of the key advantages of SILAC is that peptides from labeled and unlabeled samples have the same mass and can be easily detected, whereas, in the N labeling approach, they have different masses and can be more difficult to identify accurately.

Interactome studies that rely on mass spectrometry-identifiable chemical cross-linkers called protein interaction reporters (PIRs) have enabled the large-scale identification of interacting proteins [57]. The main limitation to applying this approach on a proteome-wide scale is the complexity of the mass spectral information generated. However, recent simulations of PIR-type experiments have demonstrated the feasibility and constraints of the PIR strategy. Very complex protein mixtures may result in false discovery of PIRs due to random matching between peptides that coincidentally share similar masses [58].

**Protein microarrays**

Protein microarrays have great potential for high-throughput analysis of protein-protein interactions. Many different types of protein microarrays have been developed that
can be classified as either analytical, functional or reverse-phase [59,60]. Functional arrays suitable for assessing PPIs consist of full-length proteins or protein domains immobilized onto the surface of a glass slide. This permits the detection of protein interactions using fluorescence or chemiluminescent probes. A protein array displaying a collection of 1133 A. thaliana ORFs was fabricated and used to investigate the interactions of calmodulin (CaM) and CaM-like (CML) proteins [61]. Although multiple CaM or CML proteins bound to multiple partners, most interactions were specific to one or a few proteins, indicating that the CaM or CML network is more extensive than previously predicted [61].

Self-assembling or Nucleic Acid Programmable Protein Arrays (NAPPA) provide an alternative to protein microarrays that are fabricated by printing of proteins generated ‘off-chip’. In these arrays, a collection of expression-ready cDNA clones [62] or unpurified polymerase chain reaction (PCR) products containing a protein expression cassette [63] is printed onto the array. The array is then bathed in a linked in vitro transcription-translation system to produce the protein array. Epitope tags fused to the expressed proteins enable them to remain immobilized on the surface of the array. With the development of large collections of expression-ready full-length cDNA clones for a growing number of plant species, the size of plant protein microarrays and their use can be expected to increase rapidly in the future.

Surface plasmon resonance
Surface plasmon resonance (SPR) exploits changes that occur in the local refractive index at a metal surface when immobilized target molecules interact with analyte containing potential interactors [64]. Proteins are unlabelled, which avoids many processing steps associated with other approaches. The technique also enables kinetic studies to be carried out, providing valuable information about the binding affinities, and association and dissociation kinetics of PPIs. Recent improvements in SPR imaging technologies have significantly improved the potential utility of this approach for studying PPIs. Such improvements include the use of charge-coupled device (CCD) cameras to monitor light scattering [65], high-throughput array formats [64], and coupling of arrays with MALDI-TOF MS analysis [66]. Although SPR has not been widely applied to investigate plant PPIs, these recent technical advances will make this approach increasingly useful for plant scientists.

Despite the utility of experimental methods in discovering novel PPIs or in the validation of known PPIs, all are prone to experimental error. Therefore, it is imperative that PPI datasets be validated rigorously by one or more complementary experimental techniques or that the quality of the reported interactions is evaluated using in silico approaches that can predict or assess the relative validity of the PPI in question.

In silico analysis of protein interactions
Various confidence-scoring systems have been developed to assess the reliability of PPI datasets derived from high-throughput approaches. These systems are based on one or more of the following metrics: co-expression, co-localisation, co-evolution, functional similarity, occurrence of orthologs or interologs (i.e. interactions conserved across species), number of observations of the interaction, type of interaction (e.g. direct, physical association, genetic interaction, etc.), topological characteristics of protein networks across species, or reciprocity of protein interactions [1,11,67–69].

Comparison and assessment of these different approaches has been difficult because many of them apply different combinations of confidence assignment metrics or false-positive identification rates. Furthermore, they are confounded by the lack of experimental validation of ‘untested’ or ‘tested but not proven interactions’ [67]. Another major caveat of in silico analyses is that protein function is frequently assigned using gene ontology (GO) terms, which are widely accepted, controlled vocabulary terms defining gene and gene product attributes [70]. In the case of Arabidopsis, the majority of annotations are inferred from sequence or structural similarity, and not from direct experimental evidence. Thus, the reliability of these assignments is questionable. Approximately 45% of the proteins in Arabidopsis are annotated by the term ‘unknown’ [71], indicating that coverage and depth of GO annotation in Arabidopsis is far from complete. In addition, some proteins have multiple functions [72,73], which can complicate GO assignment. Several recent studies have used homology-based predictive approaches to determine possible interaction networks in Arabidopsis.

Predicted interactomes
Predicted interologs identification has been used as a predictor of protein interaction on the premise that orthologous proteins, which are known to interact in one organism, can interact in the organism under study [8–10]. This approach was used recently to predict the interactome of Arabidopsis [74]. The study used confidence scoring based on the number of observations and experimental approaches that detected the interaction, together with subcellular location and co-expression data, to indicate the reliability of the interaction. For the 3617 conserved Arabidopsis proteins identified in the study, a total of 1159 ‘high confidence’, 5913 ‘medium confidence’ and 12 907 ‘low confidence’ interactions were determined. A related study reported the prediction of 23 396 interactions in Arabidopsis, using several approaches, including interolog identification. These data are available through the A. thaliana protein interaction database (AtPID, http://atpid.biosino.org/) [75].

Predicted interactomes of this type can be useful in identifying networks of proteins that are highly conserved among species and for suggesting possible roles for unknown proteins [69]. However, the interolog approach is limited by the number of predicted orthologs found for the organism studied. In Arabidopsis, approximately 10 000 orthologs are detected in at least one reference species (i.e. in yeast, Caenorhabditis elegans, Drosophila or human). This amounts to less than half of the Arabidopsis proteome. Furthermore, the assignment of interologs is based on global protein sequence similarity, which could be insensitive to residue-specific requirements for interaction
specificity and affinity [76]. Genome duplication events have also been a pervasive force in plant evolution. These phenomena have resulted in expanded and plant-specific gene families involved in transcriptional regulation, signal transduction, and development [2]. Gene families are often composed of members differing from each other by a few amino acid residues. This close similarity can make unambiguous assignment of interologs problematic and can result in a contrived view of biological complexity. In addition, highly disordered proteins or proteins that lack recognized secondary structural motifs can play important roles in PPIs. However, low sequence similarity within disordered regions imply that homology-based approaches are unsuitable for the detection of functionally equivalent disordered proteins across species [77]. Predictive studies are still at an early stage. However, they are expected both to improve rapidly as more interactome data become available in plants and to provide powerful tools for generating hypotheses for future experimental investigations [67].

Databases and standards
Several protein interaction databases, including IntAct [78], Molecular Interaction Database (MINT) [79], AtPID [75], Database of Interacting Proteins (DIP) [80], Biomolecular Interaction Network Database (BIND) [81], and Biological General Repository for Interaction Datasets (BioGRID) [82], have emerged to organize, store and make PPI data available for analysis by the research community. Two databases are notable for their focus on plants: IntAct, which actively curates all plant protein interaction data from the literature and submitted datasets; and AtPID, which contains curated and predicted interaction data. The Arabidopsis Interactions Viewer (http://bar.utoronto.ca/interactions/cgi-bin/Arabidopsis_interactions_viewer.cgi) permits the visualization of selected interaction datasets from Geisler-Lee et al. [74] and Popescu et al. [61]. Readers should be cautioned that manually curated data are not error-free, owing to inconsistent curation standards and human error. However, manually curated data appear to be of higher quality than those obtained by automated approaches such as text-mining. Improvement in data curation is an ongoing process involving cooperation between databases, and the continued development of curation standards.

The work of the Human Proteome Organization Proteomics Standards Initiative (HUPO–PSI; http://www.psidev.info/) has greatly improved the ability to combine or compare interaction data from diverse sources. Currently, all the major public domain databases export their data in a common standardized, data descriptive Extensible Markup Language (XML) interchange format annotated by a series of controlled vocabularies [83]. The data can be accessed from the IntAct molecular interaction database [78] in both XML and text-file formats.

In addition, HUPO–PSI has published guidelines to outline the Minimum Information required for reporting a Molecular Interaction Experiment (MIMIx) [83]. MIMIx outlines the minimum information required to describe all relevant aspects of the interaction experiment while minimizing the burden placed on the scientists generating the data. The establishment of the International Molecular Exchange Consortium (IMEx) encourages data deposition and sharing by all of the participating major public domain databases, the use of standardized protein identifiers, and clear experimental descriptors of the techniques used, and it ensures maximal data availability to the scientific community (http://imex.sf.net) [84].

Conclusions
We have surveyed the major methodologies for characterizing PPIs in planta, in vitro and in silico. As high-throughput technologies are applied to plant systems, we can expect rapid progress towards a comprehensive examination of plant interactomes, especially within selected plant models such as Arabidopsis and rice, for which large ORF collections of full-length cDNAs are already available. Current models of predicted plant interactomes will become increasingly robust and a better representative of the dynamic nature of the interactome by a steady accumulation of interaction data from a variety of experimental approaches. An important application of these data will be to inform plant scientists about the function of unknown proteins by virtue of their relative context within macromolecular complexes. Such novel information will greatly improve our understanding of the mechanisms that control protein interaction and organize molecular structures in plants.

Note added in proof
A firefly luciferase PCA was described recently by Chen et al. (2008) for the detection of PPIs in transient assays using Arabidopsis protoplasts or Agrobacterium-mediated transient expression in N. benthamiana. See: Chen, H. et al. (2008) Firefly luciferase complementation imaging assay for protein-protein interactions in plants. Plant Physiology 146, 368–376.

Acknowledgements
Supported by the National Science Foundation (NSF-0420033; NSF-0420152) and the University of Nevada Agricultural Experiment Station, publication number 03077064. This publication was made possible by the IDeA Network of Biomedical Research Excellence (INBRE) Program of the National Center for Research Resources (NCRR) through its support of the Nevada Genomics Center and the Nevada Center for Bioinformatics. We acknowledge the contributions of all of the members of the IntAct research team, including Yasmin Alam-Faruque, Bruno Aranda, Alan Bridge, Cathy Derow, Emily Dimmer, Micheal Feuermann, Rachel Huntley, Jyoti Khadake, Sam Kerrien, Catherine Leroy, Luisa Montecchi-Palazzi, Sandra Orchard, David Thorneycroft, Rolf Apweiler and Henning Hermjakob.

Supplementary data

References
8 Matthews, L.R. et al. (2001) Identification of potential interaction networks using sequence-based searches for conserved protein–protein interactions or “interologs”. Genome Res. 11, 2120–2126
81 Kerrien, S. et al. (2007) Broadening the horizon - level 2.5 of the HUPO-PSI format for molecular interactions. BMC Biol. 5, 44