Systems-Based Approach to the Analyses of Plant Functions: Conceptual Understanding, Implementation, and Analysis

Brijesh Singh Yadav, Amit Kumar Singh, and Sandeep K. Kushwaha

Contents

1 Introduction of Plant Systems Biology

Plant structure and function is a complex biological system determined by its molecular constituents such as DNA, RNA, proteins, metabolites, and macro- and microelements. Plants are enormously investigated in various aspects, like molecular genetics, breeding, genomics, and proteomics; however, we still have limited knowledge about plant genetic architecture and functioning mechanisms. Various solitary and mixed approaches have been developed in the past decades which have enhanced our knowledge about the role of genetics in plant. But, a holistic research approach requires a complete understanding of the plant structure and function at the molecular level which needs a lot of computational resources, different kinds of data generation, and integration algorithmic approaches (Sheth and Thaker 2014). The most widely identified challenges are the data integration and management of large datasets from various sources such as genomic sequences, phenotype images,

B.S. Yadav (✉)
Department of Bioengineering, The University of Information Science and Technology (UIST), St. Paul, Republic of Macedonia
e-mail: brijeshbioinfo@gmail.com

A.K. Singh
Department of Molecular Biology and Ecology of Plants, Tel Aviv University, Tel Aviv, Israel

S.K. Kushwaha
Department Plant Breeding, Swedish University of Agriculture Sciences, Alnarp, Sweden

K.R. Hakeem et al. (eds.), *Plant Bioinformatics*,
https://doi.org/10.1007/978-3-319-67156-7_2

protein 3D structures, and -omics data. Data heterogeneity is another major challenge due to different data syntactic (formats, schemas, and query interfaces) and semantic (data formalisms and abstraction, scope-specific naming conventions and inconsistencies). The recent advancements of computational techniques, resources, and high-throughput sample processing technologies have broken the technical and implementation barriers and facilitated the evolution and integration of "-omics" fields such as genomics, transcriptomics, proteomics, metabolomics, and phenomics for the implementation of systems biology paradigm in plant science. A large number of comprehensive and quantitative datasets have been generated in numerous targeted and system-wide studies facilitating the development of databases, software, data formats, and multivariate approaches for the integration of multi-omics data. The high-throughput genetic and molecular tactics adopted to generate -omics data that can be analyzed and used in mathematical and computational models for revealing the networks on a global scale in the same platform is termed as systems biology (Fig. 1). Systems biology approaches successfully categorized the key molecules and their roles in complex biological events in the recent period. It reveals the large complex set of transcription factor with protein, primary metabolite (carbohydrates and lipid), and secondary metabolite (glucosinolates, phenols, and many more) association which regulate physiology, growth and development, and response to the environment requires the identification of networks on a genome and proteome scale. These interactions can be either physical or functional and often can be inferred from available data (Yadav et al. 2016; Malik et al. 2014).

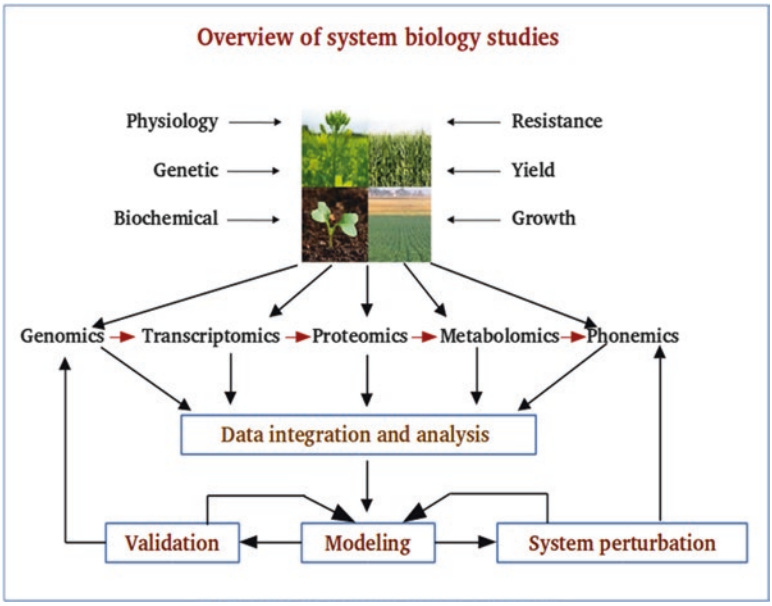

Fig. 1 Outline of systems biology approach

Systems biology-based approaches are applied to reconstruct and model metabolic networks which identify significant limiting steps in seed development of both monocot and dicot plants (Sreenivasulu and Wobus 2013). A comparative co-expression network examines to fortitude the seed size—a significant yield-related characteristics (Sreenivasulu and Wobus 2013). Malik et al. (2014) explored biological networks to study protein-carbohydrate interactions in plant lectins using glycan microarray data. The integrated systems and network methods have helped to understand the complex behaviors of nitrogen use efficiency (NUE) which provide modalities in the form of products, process, and practices for improving agricultural productivity (Bi et al. 2009). While working on photosynthetic parameters, -omics-based investigations have generated huge amounts of proteomics and genomics data which has helped to identify important traits for the improvement of photosynthetic efficiency in crop plants (Bräutigam et al. 2014; Kanwal et al. 2013). Systems-based studies on model plants characterized genes such as CIR1, RPS2, RPM1, WRKY31, MPK9, WRKY33, FLS2, RPS4, and RPP13 which play important roles in biotic stress tolerance (Carstens et al. 2014; Chinchilla et al. 2007; Leal 2013). Mishra et al. (2011) identified chlorotic toxin as a cyclodepsipeptide which affects many components of MAP kinase machinery, hypersensitive response, and systemic acquired response. Understanding the mechanisms of plant responses to multiple simultaneous abiotic stresses is therefore crucial in providing opportunities for the development of broad-spectrum stress-tolerant crops. Recently, cellular and metabolic response in abiotic stress was studied in *Arabidopsis* using systems biology and network approach (Yadav et al. 2016) which represents the effect of metabolites in osmotic, wound, genotoxic, drought, salinity, heat, and UV-B stresses.

2 High-Throughput Experimental Techniques in Plant Systems Biology

Parallelization is the central theme of high-throughput experiment (HTE), which has featured to carry out many experiments simultaneously. High-throughput experimentation technique and methods have evolved significantly in the last two decades which enabled to probe the plant circadian clock, a key coordinator of vital biological processes for whole system and targeted studies.

2.1 DNA Microarray Technology

Microarray allows to study the expression of thousands of genes simultaneously through the hybridization of probe sequences to nucleic acid sequences in mixture. In microarrays, probe sequences are fixed on solid surface and hybridizations are detected

through fluorescent detection signal of labeled samples. Microarray technology was developed in the late 1970s and revolutionized in the beginning of new century due to high growth in genomic sequences, genomic sequencing projects, and availability of publicly curated and non-curated databases. Glass spotted arrays, in situ synthesized arrays, and self-assembled arrays are basic types of arrays used in the time frame (Bumgarner 2013). Microarrays were used to measure gene expression levels and differential gene expression studies, comparison of expression patterns across samples, trait associations, etc. These studies improved our understanding of the cellular physiology and dynamics, interconnection of gene networks, and products for environmental input processing and phenotypic regulation which facilitates the global gene expression studies at systems level (Yamaguchi-Shinozaki and Shinozaki 2005; Chen and Zhu 2004). Deciphering of gene network for soybean and *Arabidopsis* seed compartments from pre-globular to early maturation stage, seeds are the excellent and widely used examples of microarray application in plants genotyping (http://seedgenenetwork.net/). Microarrays are used for single-nucleotide-polymorphism genotyping and most commonly used for allele discrimination by hybridization, allele-specific extension, and ligation (Hirschhorn et al. 2000). Microarray is a popular technique because of its simple application and cost- and time-effective process which suits best for routine practice. Limitations of microarray technology are information of known sequences required for probe, no molecular resolution for hybridization, biased detection of novel and rare transcripts, constrain of hybridization, and relatively low specificity and sensitivity.

2.2 Next-Generation Sequencing (NGS)

After microarray, there were revolutionary advances in DNA sequencing technologies with the advent of next-generation sequencing (NGS) techniques. As techniques and datasets continue to improve and grow, we are also rapidly moving to the point where every organism, not just selected "model organisms," is open to the power of NGS. For transcriptome analysis RNA sequencing is important because it can show the expressed sequences in specific tissues at a specific time and is rapidly replacing other methods for studying gene expression such as microarrays. For experimental setup it needs (1) organism-specific features, such as level of heterozygosity and availability of a reference genome and the consequences of organism choice on analysis; (2) tissue treatment and selection of tissue types to obtain the desired transcriptome and the desirability of replicates; (3) techniques for efficiently selecting transcripts for sequencing; (4) normalizing transcripts to avoid overrepresentation of highly abundant transcripts; (5) choice of sequencing platforms; and (6) methods of data assembly, with a useful table of assembly programs.

Evolution in DNA sequencing technologies has transformed the biological science research and outreach. In 1987, capillary electrophoresis-based first sequencing instrument, i.e., AB370, was introduced by Applied Biosystems, and a decade later, AB3730xl was introduced as "first-generation" working horse for high-

throughput sequencing. From 2005, sequencing techniques have revolutionized and transformed into short read, massively parallel sequencing and established as the "next-generation" sequencing technique (Metzker 2010).

In early second generation sequencer, Roche 454 GS FLX+ pyrosequencer was well known for the sequencing of long read length (approx. 700 bp) and low output (approx. 1 Million reads/run). It was widely used for 16S variable region sequencing as well as targeted amplicon sequencing. Illumina reversible terminator sequencing is popular for versatile sequencing potential, i.e., read length (50–300 bp) and read number (25 million–6 billion per run) vary from platform to platform. Illumina application is well known in small- and big-size genomic sequencing and resequencing projects, transcriptome, SNP detection, and metagenomic studies (http://www.illumina.com/technology/next-generation-sequencing/sequencing-technology.html). Evolution in sequencing techniques is continued for next-generation sequencers due to high demand for a low-cost technology. Many sequencers are already in active practices from the third generation sequencer. Ion PGM (Personal Genome Machine) and ion proton are semiconductor-based sequencing platforms with automated workflow from sample preparation to analysis and having chip-based wide range of sequencing potentials, i.e., Ion 314™ Chip v2, Ion 316™ Chip v2, and Ion 318™ Chip v2 (read length, 200–400 bp; reads/run, 500 K–5 million). These sequencers are well used in targeted, small-scale studies and routine practices (https://www.thermofisher.com/). PacBio RS is gaining popularity due to high long read lengths through single molecule real-time (SMRT) sequencing technology. It produces reads from 1 to 60 Kb. Each SMRT (single molecule real-time sequencing) cell can generate approximately 50,000 reads. PacBio RS is ideal for the sequencing of small genomes, regions of high G/C content, and DNA methylation, resequencing projects due to longer read length feature (http://www.pacb.com/products-and-services/pacbio-systems/rsii/). Oxford nano-pore, based on nano-pore conductivity, is under active development. Current sequencing technologies offered several common benefits which are perfect for exploitative studies, i.e., no need of prior molecular information, molecular level sequence resolution, discovery of novel and rare transcripts, broad dynamic range, etc. But it imposed common technical problems as well, such as short read assembly and mapping, low coverage for GC content enriched region, sequencing error, etc. High coverage and deep sequencing can overcome these problems up to certain extent (Hui 2012).

2.3 Proteomics

Proteomics is known for the study of quantitative measurement of proteins expressed by genome to characterize organism or biological processes which explain the mechanism of gene expression control. Field of protein research is continuously evolving since the isolation of individual proteins of *E. coli* from protein complexes through two-dimensional polyacrylamide gel electrophoresis (2-DE). Later, mass spectrometry (MS) was coupled with 2-DE gels for the identification of large

number of proteins which evolved as proteomics (Anderson and Anderson 1998). Proteomics studies have three main components, i.e., expression proteomics, bioinformatics analysis, and functional proteomics. Expression proteomics is about sample preparation through gel-based or gel-free methods and protein identification analysis. Protein extraction methods can be different from protein to protein due to different physical and chemical properties of proteins. After protein extraction, 2-DE or liquid chromatography (LC) is used to isolate individual proteins from sample mixtures, and later these proteins are identified through MS-based methods. Quantitative proteomics have been used to identify proteins expressed in a specific cell or tissue, comparison of protein expression profile in differential experimental conditions or disease states to explore physiology and pathogenic mechanisms. Due to recent advancement, bioinformatics is extending its outreach to guide the future direction of functional proteomics studies. Bioinformatics analysis has been used in quantitative proteomics after the protein identification to gather different kinds of information, such as sequence alignment and homology; motifs and domains; transmembrane regions; protein structure, i.e., primary, secondary, tertiary, and quaternary; protein interactions and networks; as well as subcellular locations (Bencharit et al. 2013). Functional proteomics is a protein characterization approach of proteomics to understand the role of targeted proteins in cellular functions which requires high-throughput comprehensive analyses of protein-protein interactions, protein complexes, and transmembrane of organism. Bait- and prey-based approaches are very popular in functional proteomics like protein chip, co-immunoprecipitation techniques coupled with 2-DE, affinity chromatography, biomolecular interaction analysis mass spectrometry (BIA-MS), etc. (Chandramouli and Qian 2009).

2.4 Metabolomics

Metabolome is the collection of all kinds of metabolites in a biological cell, tissue, organ, or organism as the end products of cellular processes. Metabolomics profiling is getting popularity due to instant snapshot of the cell physiology, whereas gene expression and proteomics analyses are limited up to genes and their products being produced in the cell. Nowadays, metabolomics is integrated into "-omics" family along with genomics, proteomics, and transcriptomics to provide a better understanding of cellular biology (Johnson et al. 2014). Cellular metabolomics is a cohesive network of metabolite and biochemical interactions which have not yet been fully characterized for products, reactants, intermediate steps, and regulatory molecules. Metabolite profiling and metabolic fingerprinting are the major approaches used in metabolomics. Metabolite profiling is used to identify and quantify metabolites of plant cell. Metabolic fingerprinting is the high-throughput approach of metabolomics which is used for tissue comparison and discrimination analysis. Metabolomics is also used on the metabolic response of organisms to physiological stimuli or genetic modification (Johnson et al. 2014).

3 Data Analysis

3.1 Gene Identification and Expression Analysis

In both microarrays and NGS methods, we need to identify the expressed genes for further downstream analysis. There are several methods and tools already available for microarray based identification of expressed gene. This technology allows widespread changes in expression patterns to be probed in a single experiment. Gene expression is normalized in reference to control genes on a chip. Further, t-test and false discovery rate are applied to detect differentially expressed genes between treatment and control groups (Yadav et al. 2016).

Statistical analysis is an essential component for RNA-seq data, but due to the short history of the technology and its continuous development, there are no standard methods available yet to detect and analyze differentially expressed genes based on NGS data. Analytical programs for these data are just emerging and need to be evaluated. There are freely available R software packages which provide method to detect differentially expressed genes. Kvam et al. (2012) describe four recently proposed statistical methods (edgeR, DESeq, baySeq, and a method with two-stage Poisson model [TSPM]) on significance ranking of genes and false discovery rate control through simulation studies under various settings mimicking real data. Cufflinks tool has been developed to estimate transcript-level expression by tackling the problem of related transcripts' sharing most of their reads (Trapnell et al. 2010) and TopHat software (Li and Dewey 2011) used for the estimates transcript abundances. The results show that the performances of different methods vary and that baySeq performs best in terms of significance ranking of genes. The false discovery rate may not be controlled well in practice, and they suggest applying a relatively stringent level to avoid too many false positives. In addition, the flexibility of handling different experimental design varies among the current versions of the different packages. Plant biologists may want to choose the one that best fits their experimental design and goal.

3.2 Gene Enrichment Functional Analysis

To understand the biological context of DE (differentially expressed) genes, pathway enrichment analysis ensues. Once if the list of DE genes is available then we can start gaining biological insights into experimental systems, developmental stages, or understanding of disease or molecular mechanisms. In gene enrichment analysis experiments, gene transforms information from gene expression profiling into a pathway summary (Subramanian et al. 2005). Gene function enrichment analyses depend upon various annotation databases, for instance, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Kanehisa and Goto 2000), Gene Ontology (GO) (Botstein et al. 2000), DAVID (Huang et al. 2009), etc. One traditional

analysis starts with a gene list of interest, identified from differential RNA-seq or microarray analyses, and applies statistical methods, such as the Fisher's exact test to test for enrichment of each annotated gene set, network, and pathway. DE genes are quite often involved in the same biological pathways, and GSE (Gene Set Enrichment) results offer greater biological interpretability over individual gene analysis. GSVA (Gene Set Variation Analysis) (Hänzelmann et al. 2013) extends the current GSE methods to RNA-seq data, and provides increased power to detect subtle pathway activity changes, and constitutes a starting point to build pathway-centric models of systems biology. SeqGSEA (Wang and Cairns 2014) is a new open-source Bioconductor package for GSVA, which can detect more biologically meaningful gene sets without biases toward longer and highly expressed genes. Previous pathway analysis methods have been developed based on algorithms considering pathways as simple gene lists and ignoring pathway structures. Recently, few methods have been developed to incorporate various aspects of pathway topology. For example, Yadav et al. (2016) have done the experiment to analyze the abiotic stress data in plants. Functional annotations were retrieved from the MapMan that is structured in the form of a hierarchical tree and was specifically designed to cover plant-specific pathways and processes. *Arabidopsis* genes were assigned into functional categories within the hierarchy of MapMan pathway scheme (Thimm et al. 2004). Pathway enrichment in each experiment was determined by calculating the cumulative hypergeometric *p* value for the probability that a group of genes are overrepresented within a functional bin at a rate higher than chance expectation.

3.3 Biological Network Topology and Characteristics

Nodes and edges are the basic element for network building. In biological science, genes and proteins works as nodes, and functional or physical relationships between them are known as edges. In general, network topologies are defined as the order of nodes and edges to determine the functional aspects of the network. Node degree (i.e., number of edges connected to a node) and degree distribution (i.e., overall distribution of node degrees in a network) determine the nature of networks. The degree distribution of networks is often used to differentiate different classes of networks, whereas a number of edges are used to measure distances between networks. Navigability of network is measured through minimum path length (minimum number of edges) or mean path length (average of shortest path between all pairs of nodes). Node centrality or hub node is measured through the shortest path between all pairs of nodes in a network. In contrast to hub node, bottleneck node which does not necessarily have higher interaction works as linker between different subnetworks. A node can work as hub and bottleneck node both. Network robustness can also be characterized through network redundancy and degeneracy. Nodes' redundant connectivity through multiple paths is important for network sustainability and integrity in the absence of other connections, whereas network degeneracy is a special type of redundancy which leads to both overlapping and separate effects of network.

Network models are very important for the understanding of complex networks and explanation of observed network characteristics. Here, we described some basic models which are necessary for critical understanding of biological networks. Random network model assumes that all nodes are connected to approximately the same number of links, i.e., average and node degrees follow a Poisson distribution which indicates that nodes that significantly deviate from the average are extremely rare. Scale-free networks are well known to follow power-law degree distribution. It assumes that probability of highly connected node is statistically more significant than in a random graph. Network hub properties are determined by a relatively small number of highly connected nodes. Most of biological networks are very close to scale-free network. A substructure of large network is formed through densely connected node, and long-distance connection with other nodes is known as module or subnetwork. These modules are specialized for particular function like co-expression network where most of the nodes are organized in modules having limited connections between subparts of the network (Hu et al. 2016). Network module graphically represents integrative genomics strategies used in current research that successfully identify candidate genes taking advantage of gene co-expression networks.

4 Networking Analysis

4.1 Networking and Data Integration for System Level Understanding

Gene set enrichment analysis provides information about all the genes which are significantly involved in cellular or metabolic pathway level processes at a given condition and time point. This plasticity generally involves changes at the level of DNA, RNA, protein, and metabolites, resulting in complex phenotypes governed by multiple genes. The major challenges in plant systems biology are to elucidate the genotype-phenotype relationship in plant cellular systems. The integrated network analysis tries to find correlation between genes and pathways which with the help of mathematical models have received particular attention.

4.2 Scale Within Network System

One of the most key characteristic of system is the set of interactions existing between its constitutive elements. To understand complex interactions, the behavior of system needs to be modeled because it cannot be understood by direct approaches. Network approach is one of the common ways to model systems as a collection of its interacting elements. All the constitutive items of the system are represented as

set of nodes that are linked by edges which indicate the interactions between those items by means of mathematics. To represent systems as a unified whole into given network, all the nodes must be linked to other node by at least one edge; the mode of such linkage is called topology of the network (McCormack et al. 2016). There are three main steps for modeling of a biological system: (1) We have to define which component will be considered as a node in the network and associate variables of interest and evolution rules to those nodes based on properties of the considered biological components. (2) Describe the association between the components of the system as edges associated with interaction functions in the system network. (3) Study the nature of the system under different conditions, changing evolution rules and network topology as dictated by experiment until a perfect prediction of the modeled system behavior can be achieved (Hu et al. 2016).

4.3 Analysis of Network

There are two methods used to interpret the complexity of biology: one is "top-down" and another is "bottom-up" approach. In "top-down" method, large-scale datasets are analyzed to decode relationships between different levels of transcripts and proteins. On the other hand in "bottom-up" approach, properties of genes or proteins with available quantitative information such as kinetics data and transcription rates are used to construct a model of well-characterized components that can be simulated computationally to identify and predict the behavior of system in different conditions (Bassel et al. 2012).

Approaches have been developed to identify functional modules in the plant science such as metabolic and regulatory modules. In the network system, characterization and recognition of the community structure is one of the major issues. At present time one highly effective method is the optimization of the quality function known as "modularity" over the possible divisions of a network. It can be expressed in terms of the eigenvectors of a characteristic matrix for the network which leads to a spectral algorithm for community detection (Zhao 2016).

Co-expression of gene network method uses statistical matrices to create correlations between gene expression profiles for various samples on the basis of guilt-by-association rule (Bhardwaj and Lu 2005). The generation of co-expression networks is a feasible top-down approach to generate genome-wide cofunctional network models in plants. The Pearson correlation coefficient is generally used in co-expression networks to create linear pairwise correlations between enriched gene pairs in an adjacent matrix. For nonlinear correlations, another associative matrix that can be used is the Spearman correlation coefficient which enables nonlinear correlations between genes to be uncovered. A modified graphical Gaussian model that takes into account partial correlations between genes after removing the effects of other adjacent genes has also been used (Ma et al. 2007). Following the establishment of gene associations, a cutoff threshold is then set, and pairwise interaction values exceeding this selected threshold are kept. The strength of the correlations between gene pairs can be considered as edge

weights indicating the strength of co-regulation between gene pairs. The end result of such an approach is a network consisting of nodes which represent genes connected by edges showing significant similarity in their common expression pattern. It is important to note that both positively and negatively acting components of a biological process can be co-expressed (Lee et al. 2011). Network analysis seeks to map and understand as systems-level views of cell behavior. In the context of gene, protein, and metabolites interactions, network biology provides the tools to answer questions such as the survival of plant during different abiotic stress (Yadav et al. 2016), the effects of diverse environmental conditions on the flow of biological information between genes and proteins, and the phenotypic results of perturbations of protein communities.

4.4 Software and Tools for Network Analysis

Researchers develop and use bioinformatics software or databases for the comprehensive study of plant systems biology. Many of the tools, databases, and other resources used in the analyses of the individual -omics platforms include the tools for network visualization, modeling environments, pathway construction and visualization tools, systems biology platforms, and repositories of the models. Visualization is a means of investigative data analysis and a key method for network analysis. The purpose of large -omics data visualization should be to create clear, meaningful, and integrated resources without being besieged by the inherent complexity of data (Gehlenborg et al. 2010). Pathway databases are used for modeling systems, since they offer a clear-cut way of building network topologies by the annotated reaction system. We have listed some of the widely used tools and databases by plant research community in Table 1.

5 Application of Network System in Plant Biology

5.1 Gene-to-Metabolite Network

This network calculates the correlation and significance between differentially expressed genes which are associated with metabolic regulation at a given set of condition. In this interaction network, genes and metabolites act as node and edge, respectively. The interactions are interpreted depending on the distance between the genes and the metabolites. This type of network is highly complex and difficult to study in plants, owing to the enormous diversity and number of metabolites being produced in the cells due to their sessile lifestyle. In the area of plant science, gene-to-metabolite networks elucidate the interrelations among biological processes, gene functional annotation, discovery of new genes in biosynthesis regulation, and transport of metabolites (Yuan et al. 2008). For the various biotic and abiotic stresses in plants, researchers utilize gene-to-metabolite networks to reveal how genes

Table 1 Systems biology tools and database resources. Protein-protein interaction (1–24), metabolic pathways (25–35), signaling pathways (36–39), transcription factors/gene regulatory networks (40–48), and genetic interaction network (49–53) in plants

S.No.	Tool	URL	Description
1	AHD2.0	http://ahd.cbi.pku.edu.cn/	The main goal of the *Arabidopsis* hormone database is to provide a systematic and comprehensive view of morphological phenotypes regulated by plant hormones, as well as regulatory genes participating in numerous plant hormone responses. The database also provides interactive protein-protein interaction networks and pathway diagrams for interactions and pathways your gene of choice is involved in
2	ANAP	http://gmdd.shgmo.org/Computational-Biology/ANAP/ANAP_V1.2/	ANAP is a knowledgebase that contains information on protein interaction networks. The knowledgebase can be searched through the use of a simple keyword search, which will then search the whole of the ANAP dataset for all the protein interactions from the integration and the nearest neighbors. An interactive protein interaction network will then be generated
3	APID	http://cicblade.dep.usal.es:8080/APID/init.action	APID (Agile Protein InteractomesDataServer) is a new full redesigned biological resource that provides a comprehensive collection of protein interactomes for more than 400 organisms based in the integration of known experimentally validated protein-protein physical interactions (PPIs).The analytical and integrative effort done in APID unifies PPIs from primary databases of molecular interactions (BIND, BioGRID, DIP, HPRD, IntAct, MINT), from other original resources (like BioPlex) and also from experimentally resolved 3D structures (PDB) where more than two distinct proteins have been identified
4	AS-ALPS	http://as-alps.nagahama-i-bio.ac.jp/	The main goal of the AS-ALPS database is to provide useful information that allows users to analyze the effect of altering the protein structure, through alternative splicing, on protein structure, interactions, and interaction networks. The database also provides links to the interactive protein-protein interaction network of choice
5	AtPID	http://www.megabionet.org/atpid/webfile/	The AtPID (*Arabidopsis thaliana* protein interactome database) represents a centralized platform to depict and integrate the information pertaining to protein-protein interaction networks, domain architecture, ortholog information, and GO annotation in the *Arabidopsis thaliana* proteome
6	CCSB interactome-database	http://interactome.dfci.harvard.edu/	CCSB interactome database is a database of binary protein-protein interactions for a number of different organisms: *homo sapiens*, viruses, *Arabidopsis thaliana*, *Caenorhabditis elegans,* and *Saccharomyces cerevisiae*. All datasets are available for download and can be easily searched and visualized on the web interface

7	CORNET	https://bioinformatics.psb.ugent.be/cornet/	CORNET integrates different types of data (co-expression, protein interaction) to produce correlation networks The protein-protein interaction tool queries all available protein interaction databases for both experimental and predicted interactions, including IntAct, MINT, DIP, BIND, BioGRID, TAIR, AtPID. Results can be visualized using CytoscapeWebstart. The regulatory interactions tool gives a similar result as it queries from both confirmed and unconfirmed interactions from databases such as AGRIS, EVEX regulation, etc.
8	DIP	http://dip.doe-mbi.ucla.edu/dip/Main.cgi	The DIP (TM) database catalogs experimentally determined interactions between proteins. It combines information from a variety of sources to create a single, consistent set of protein-protein interactions. The data stored within the DIP database were curated, both, manually by expert curators and also automatically using computational approaches that utilize the knowledge about the protein-protein interaction networks extracted from the most reliable, core subset of the DIP data
9	FLOR-ID	http://www.phytosystems.ulg.ac.be/florid/	Flowering is an important topic in plant biology, and important progress has been made in *Arabidopsis thaliana* toward unraveling the genetic networks involved. Flowering Interactive Database (FLOR-ID) is a database containing detailed information on gene networks involved in the flowering-time control of *Arabidopsis thaliana*. The hand-curated database contains information on genes and links to publications gathering the work of thousands of authors. Gene/protein functions and interactions within the flowering pathways were inferred from the analysis of related publications, included in the database and translated into interactive manually drawn snapshots
10	FunCoup	http://funcoup.sbc.su.se/search/	FunCoup is a statistical framework of data integration for finding functional coupling (FC) between proteins. It transfers information from model organisms via orthologs found by InParanoid program. FunCoup derives novel functional links from mostly raw high-throughput data or large-scale database annotations and estimates each piece of information by relevance and reliability. Moreover, FunCoup employs carefully tested algorithms of across-species data transfer via orthologs, and Eukaryota-wide networks for multiple organisms are available and comparable

(continued)

Table 1 (continued)

S.No.	Tool	URL	Description
11	GeneMANIA	http://www.genemania.org	GeneMANIA finds other genes that are related to a set of input genes, using a very large set of functional association data
12	GermOnline	http://www.germonline.org/index.html	GermOnline 4.0 is a cross-species microarray expression database focusing on germline development, meiosis, and gametogenesis as well as the mitotic cell cycle 1. The database contains a unique combination of information:
			(1) High-throughput expression data obtained with whole-genome high-density oligonucleotide microarrays (GeneChips)
			(2) Sample annotation (mouse over the sample name and click on it) using the Multiomics Information Management and Annotation System (MIMAS 3.0 2)
			(3) In vivo protein-DNA binding data and protein-protein interaction data (available for selected species)
			(4) Genome annotation information from Ensembl version 50
			(5) Orthologs are identified using data from Ensembl and OMA and linked to each other via a section in the report pages
13	HINT	http://hint.yulab.org/	HINT is a database of high-quality protein-protein interactions in different organisms. These have been compiled from different sources and then filtered both systematically and manually to remove erroneous and low-quality interactions
14	HitPredict	http://hintdb.hgc.jp/htp/	HitPredict is a resource of high confidence protein-protein interactions. Interactions can be searched and downloaded with their predicted confidence level. Protein-protein interactions from IntAct, BIOGRID, and HPRD are combined, annotated, and assigned a reliability score in order to identify a high confidence subset. The reliability score is calculated as the likelihood ratio using naive Bayesian networks combining sequence, structure, and functional annotations of the interacting proteins

15	INstruct	http://instruct.yulab.org/	INstruct is a database of high-quality protein interactome networks annotated to 3D structural resolution. The interactions shown on this site have been curated from some of the most popular interaction databases and filtered to reflect only binary interactions that meet our strict quality criteria
16	IntAct	http://www.ebi.ac.uk/intact	A database of freely available molecular interactions maintained by the European Bioinformatics Institute (EBI)
17	Interolog	http://interolog.gersteinlab.org/	A database of predicted protein-protein interactions and transcription factor-regulated gene interactions using the interolog and regulog method. Basically, interactions in one species are used to predict orthologous interactions in another species
18	MINT	http://mint.bio.uniroma2.it/mint/	A database of molecular interactions gathered from the literature and manual input from Rome, Italy. It is also integrated with HomoMINT, a database of molecular interactions inferred from ortholog proteins in model organisms, and VirusMINT, a database of interactions between human and viral proteins
19	MitoInteractome	http://mitointeractome.kobic.kr/	MitoInteractome is a web-based portal containing information relevant to mitochondrial proteins. It also serves as a research tool for finding interacting partners and studying mitochondrial diseases. It has a comprehensive collection and organization of organelle-specific data. The data is primarily obtained by keyword search at Swissprot, MitoP, and MitoProteome
20	MitoP2	http://www.mitop.de:8080/mitop2/	The aim of this database is to provide a comprehensive list of mitochondrial proteins of yeast, mouse, *Arabidopsis thaliana*, *Neurospora*, and human. Datasets relevant to the study of the mitochondrial proteome are integrated and accessible via search tools and links. They include computational predictions of signaling sequences and summarize results from proteome mapping, mutant screening, expression profiling, protein-protein interaction, and cellular sublocalization studies. Predictive scores (score based on support vector machine prediction) are given and are calculated using the integration of the included datasets and annotated reference sets of mitochondrial proteins
21	PAIR	http://www.cls.zju.edu.cn/pair/	The PAIR database accurately predicts *Arabidopsis* interactome and facilitates these results into a user-friendly interface with detailed annotations. These interactions were predicted through the use of a Support Vector Machine (SVM). False positives were tightly controlled, and as a result, 43.52% of the identified interactions were expected to be accurate and cover 24.47% of the entire *Arabidopsis* interactome

(continued)

Table 1 (continued)

S.No.	Tool	URL	Description
22	PPIRA	http://protein.cau.edu.cn/ppira/	PPIRA is a database on protein-protein interactions between *Ralstonia solanacearum* (a plant pathogen) and *Arabidopsis thaliana* (its host). For any phytopathogenic, protein-protein interactions (PPIs) play very important roles in infecting hosts. Users can query a *Ralstonia solanacearum* or *Arabidopsis thaliana* protein, and a table containing potential interacting partners will be returned
23	PubGene	http://www.coremine.com/medical/#search	PubGene is a tool for viewing associations between genes that have been found in abstracts in PubMed using a text-mining algorithm. A commercial version makes available more data than the freely available version
24	SUBA3	http://suba.plantenergy.uwa.edu.au/	SUBA provides a powerful tool to investigate subcellular localization in *Arabidopsis* through the unification of disparate datasets and through the provision of a web accessible interface for the construction of powerful user-based queries resulting in a one-stop shop for protein localization in this model plant. Protein-protein interaction data can also be accessed
25	*Arabidopsis* reactome	http://www.arabidopsisreactome.org	The aim of *Arabidopsis* reactome is to develop a curated resource of core pathways and reactions in plant biology. The information in this database is authored by biological researchers with expertise in their field and maintained by the *Arabidopsis* reactome editorial staff. Contents are cross-referenced with the following external databases: PubMed, GO, ATIDB, TAIR, MIPS, UniProt., ChEBI, and KEGG COMPOUND. In addition to curated events (center of reaction map), imported *Arabidopsis* events from KEGG and AraCyc databases are also provided. Moreover, inferred orthologous events in five other plants including rice, grape, poplar, and moss are also available
26	AraCyc	http://www.arabidopsis.org/biocyc/introduction.jsp	AraCyc is a tool for visualizing biochemical pathways of *Arabidopsis thaliana*. It is supported by the pathway tools software developed by Peter Karp's group at SRI
27	AtIPD	http://www.atipd.ethz.ch/	AtIPD contains a manually curated list of *Arabidopsis* isoprenoid pathways and genes and allows the user to visualize pathway topology. The database was compiled using information on pathways and pathway genes from BioPathAt, KEGG, AraCyc, SUBA, and literature. Users can search or browse the database, extract the underlying data, and follow external links related to the pathway topologies, enzyme activities, or subcellular localizations

28	KaPPA-view	http://kpv2.kazusa.or.jp/kpv4/	KaPPA-view is a database that contains a variety of metabolic pathway maps. Users can search or browse the database for pathway maps and can search for genes, metabolites, and enzymes, and the database will return pathway maps that contain the gene, metabolite, or enzyme of interest. Users can also upload their own data to have it analyzed
29	MetaCrop	http://metacrop.ipk-gatersleben.de/	MetaCrop contains hand-curated information of major metabolic pathways in various crop plants with special emphasis on the metabolism of agronomically important organs such as seed or tuber. Species of both monocotyledons and dicotyledons are represented. Reactions incorporate information about involved metabolites, stoichiometry, and detailed location (species, organ, tissue, compartment, and developmental stage). Furthermore, for central metabolism (glycolysis, TCA cycle) kinetic data is available for the reactions
30	MetExplore	http://metexplore.toulouse.inra.fr/joomla3/	MetExplore is a web server that offers the possibility to link the metabolites identified in untargeted metabolomic experiments within the context of genome-scale reconstructed metabolic networks. The analysis pipeline comprises mapping metabolomics data (from masses or identifiers) onto the specific metabolic network of an organism, then applying graph-based methods and advanced visualization tools to enhance data analysis. MetExplore stores metabolic networks and information about metabolites from about 60 organisms into a relational database. Various filters can be applied in MetExplore to restrict the scope of the study, for example, by selecting only particular pathways or by restricting the network to the small-molecule metabolism
31	MetNetDB	http://metnetweb.gdcb.iastate.edu/MetNet_db.htm	Contains information on networks of metabolic and regulatory and interactions in *Arabidopsis*. This information is based on input from biologists in their area of expertise. In addition to the MetNet-curated interactions, we provide Aracyc-curated pathways and AGRIS-curated regulatory network. The data in MetNetDB is updated regularly. This information is accessible only by downloading files for import into Cytoscape
32	PeroxisomeDB	http://www.peroxisomedb.org/home.jsp	The aim of Peroxisome database is to gather, organize, and integrate curated information on peroxisomal genes, their encoded proteins, their molecular function, the metabolic pathway they belong to, and their related disorders. PeroxisomeDB contains the complete peroxisomal proteome of Homo sapiens and *Saccharomyces cerevisiae*
33	Plant reactome	http://plantreactome.gramene.org/	The plant reactome is a free, open-source, curated, and peer-reviewed database of plant metabolic and regulatory pathways. Its goal is to provide intuitive bioinformatics tools for the visualization, interpretation, and analysis of pathway knowledge to support basic research, genome analysis, modeling, systems biology, and education

(continued)

Table 1 (continued)

S.No.	Tool	URL	Description
34	PlantCyc	http://www.plantcyc.org/	PlantCyc is a database containing pathways and their catalytic enzymes and genes, as well as compounds from many plant species. The majority of pathway diagrams in PlantCyc were manually extracted from the plant literature. They are either supported by experimental evidence or are based on expert hypotheses
35	TAIR	http://www.arabidopsis.org/	The *Arabidopsis* Information Resource (TAIR) maintains a database of genetic and molecular biology data for the model higher plant *Arabidopsis thaliana*. Data available from TAIR includes the complete genomic sequence along with gene structure, gene product information, metabolism, gene expression, DNA and seed stocks, genome maps, genetic and physical markers, publications, and information about the *Arabidopsis* research community
36	*Arabidopsis* reactome	http://www.arabidopsisreactome.org	The aim of *Arabidopsis* reactome is to develop a curated resource of core pathways and reactions in plant biology. The information in this database is authored by biological researchers with expertise in their field and maintained by the *Arabidopsis* reactome editorial staff. Contents are cross-referenced with the following external databases: PubMed, GO, ATIDB, TAIR, MIPS, UniProt., ChEBI, and KEGG COMPOUND. In addition to curated events (center of reaction map), imported *Arabidopsis* events from KEGG and AraCyc databases are also provided. Moreover, inferred orthologous events in five other plants including rice, grape, poplar, and moss are also available
37	DNAtraffic	http://dnatraffic.ibb.waw.pl/	DNAtraffic database is dedicated to be a unique comprehensive and richly annotated database of genome dynamics during the cell life. DNAtraffic contains extensive data on the nomenclature, ontology, structure, and function of proteins related to control of the DNA integrity mechanisms such as chromatin remodeling, DNA repair, and damage response pathways
38	PathoPlant	http://www.pathoplant.de/	PathoPlant is a database on plant-pathogen interactions and components of signal transduction pathways related to plant pathogenesis
39	TRRD	http://wwwmgs.bionet.nsc.ru/mgs/gnw/trrd/	TRRD contains information about regulatory regions including 10,135 transcription factor binding sites (DNA-protein interactions). This database is very similar to TRANSFAC. It is available via an SRS database interface freely over the web

40	AGRIS	http://arabidopsis.med.ohio-state.edu/	The *Arabidopsis* Gene Regulatory Information Server (AGRIS) is an information resource of *Arabidopsis* promoter sequences, transcription factors, and their target genes. AGRIS currently contains three databases, AtcisDB (*Arabidopsis thaliana* cis-regulatory database), AtTFDB (*Arabidopsis thaliana* transcription factor database), and AtRegNet (*Arabidopsis thaliana* regulatory network). The three databases, used in tandem, provide a powerful tool for use in continuous research
41	Athena	http://www.bioinformatics2.wsu.edu/cgi-bin/Athena/cgi/home.pl	Athena provides several features to enable exploration of the regulatory mechanisms of *Arabidopsis* gene control. Athena is a web-based application that warehouses disparate datatypes related to the control of gene expression. Accompanying this warehouse is a large set of data visualization, mining, and analysis tools
42	DATF	http://datf.cbi.pku.edu.cn/	The Database of Arabidopsis Transcription Factors (DATF) collects all *Arabidopsis* transcription factors and classifies them into 64 families. It also keeps old information items such as the uniquely cloned and sequenced information of about 1200 transcription factors, protein domains, 3D structure information with BLAST hits against PDB, predicted nuclear location signals, UniGene information, as well as links to literature reference
43	Interolog	http://interolog.gersteinlab.org/	A database of predicted protein-protein interactions and transcription factor-regulated gene interactions using the interolog and regulog method. Basically, interactions in one species are used to predict orthologous interactions in another species
44	miRBase	http://www.mirbase.org/index.shtml	miRBase contains information about microRNAs including predicted mRNA targets, which represent potential translation regulatory interactions. It is automatically updated using a computational pipeline
45	PlantTFDB	http://planttfdb.cbi.pku.edu.cn/	PlantTFDB is a database dedicated to plant transcription factors. The transcription factors are classified into 58 different families, and a brief introduction and key references are presented for each family. Comprehensive annotations are made for each identified TF, including functional domains, 3D structures, gene ontology (GO), expression information derived from EST and microarray, and annotations in various databases such as UniProt, RefSeq, TransFac, and STRING. In addition, PlantTFDB has a simple and user-friendly interface to allow users to query based on combined conditions or make sequence similarity search using BLAST

(continued)

Table 1 (continued)

S.No.	Tool	URL	Description
46	STIFDB	http://caps.ncbs.res.in/stifdb2/	STIFDB is a comprehensive collection of biotic and abiotic stress responsive genes in *Arabidopsis thaliana* and *Oryza sativa* L. with options to identify probable transcription factor binding sites in their promoters. It also contains information on involved transcription factors
47	TRRD	http://wwwmgs.bionet.nsc.ru/mgs/gnw/trrd/	TRRD contains information about regulatory regions including 10,135 transcription factor binding sites (DNA-protein interactions). This database is very similar to TRANSFAC. It is available via an SRS database interface freely over the web
48	TrSDB	http://bioinf.uab.es/cgi-bin/trsdb/trsdb.pl	TrSDB is a BioDB (Biological Relational DataBase) that through the same web interface join several predictive analyses of protein motifs, domains, cellular localization, and functional annotation around different compiled sequences, now IPI and SPTR non-redundant proteomes. Emphasis is placed upon transcription factor prediction supported by TranScout
49	AraNet	http://www.inetbio.org/aranet/	AraNet is a probabilistic functional gene network of *Arabidopsis thaliana*, constructed by a modified Bayesian integration of 24 types of "-omics" data from multiple organisms, with each data type weighted according to how well it links genes that are known to function together in *Arabidopsis thaliana*. Each interaction in AraNet has an associated log-likelihood score (LLS) that measures the probability of an interaction representing a true functional linkage between two genes
50	BioGRID	http://www.thebiogrid.org/	The Biological General Repository for Interaction Datasets (BioGRID) database was developed to house and distribute collections of protein and genetic interactions from major model organism species, as derived from both high-throughput studies and conventional focused studies

51	FLOR-ID	http://www.phytosystems.ulg.ac.be/florid/	Flowering is an important topic in plant biology, and important progress has been made in *Arabidopsis thaliana* toward unraveling the genetic networks involved. Flowering Interactive Database (FLOR-ID) is a database containing detailed information on gene networks involved in the flowering-time control of *Arabidopsis thaliana*. Gene/protein functions and interactions within the flowering pathways were inferred from the analysis of related publications, included in the database and translated into interactive manually drawn snapshots
52	FunCoup	http://funcoup.sbc.su.se/search/	FunCoup is a statistical framework of data integration for finding functional coupling (FC) between proteins. It transfers information from model organisms via orthologs found by InParanoid program. FunCoup derives novel functional links from mostly raw high-throughput data or large-scale database annotations and estimates each piece of information by relevance and reliability
53	GeneOrienteer	http://geneorienteer.org/	GeneOrienteer is a database that predicts gene-gene interactions. It does this based on correlations between genetic traits, orthology to known interacting genes, and public two-hybrid data. Inputting a gene name for a specific organism can result in a list of predicted interactions, information about the inputted gene, or details about an inputted pair of genes

regulate cellular pathways as well as primary and secondary metabolites synthesis to protect plants (Zulak et al. 2007; Yadav et al. 2016). These networks have helped in the discovery of novel candidate genes for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* (Rischer et al. 2006) in response to nitrogen deficiency and during diurnal cycles (Blasing et al. 2005) and so on.

5.2 Protein: Protein Interaction Network

Protein-protein interactions (PPIs) are one of the most significant components of biological networks. In PPI networks, the nodes are proteins which are associated by direct edges if the direction of information flow during their interaction is known or nondirect edges if there is strong evidence of their physical interaction or association without an evidence for directionality of interaction (Assmann and Albert 2009). Two types of interactions might be possible: genetic or physical. In genetic approaches, a network of genes characterized on the basis of genetic interactions to explain gene function within physiological processes (Boone et al. 2007). Still, this method is difficult to implement owing to the ploidy levels and perennial plants. While in physical methods, interaction maps have been experimentally elucidated for homo- and hetero-dimerization within two large classes of transcription factors, for example, networks between the MADS box transcription factors (Immink et al. 2003; de Folter et al. 2005) and the MYB transcription factor family (Zimmermann et al. 2004a, b).

5.3 Carbohydrate: Protein Interaction Network

People are exploring protein-carbohydrate interactions because it plays significant role in numerous physiological and pathological processes in plants. Recently, many computational methods have been suggested to study protein-carbohydrate interactions. Malik et al. (2014) developed a method to group various plant lectins and their interacting carbohydrates by the community detection analysis of a lectin-glycan network generated by the glycan microarray data. They identified three lectins having large degrees of connectivity playing the roles of hubs. This work reveals global organization of lectin-glycan interactions and helps to identify strongly correlated lectin and glycan clusters in plants.

5.4 Transcriptional Regulatory Network

Transcription regulatory network elucidates the regulatory interactions between transcription factors and downstream genes. To understand cellular dynamics, thorough knowledge of each regulatory network is required. In this network, nodes represent

transcription factors and regulatory genes, whereas edges represent transcriptional regulation (Babu et al. 2004). There are various methods that are applied to interpret the transcriptional regulatory networks which includes genome-wide expression profiling and RNA interference (RNAi) screens (Baum and Craig 2004). Nachman et al. presented transcription rate assessment by measurement of mRNA decay rates (Nachman et al. 2004). Other regulatory network models are evaluated based on promoter co-occupancy by pairs of transcription factors (Geisberg and Struhl 2004) and computational prediction of *cis*-elements (Beer and Tavazoie. 2004). In *Arabidopsis*, a transcriptional regulatory map was created for cold signaling mediated by the ICE1 transcription factor (Benedict et al. 2006), in rice to understand the role of oxidative signals in cold stress (Yun et al. 2010; Todaka et al. 2012), and in response to abiotic stresses in grasses (Nakashima et al. 2009).

5.5 Gene Regulatory Network

A gene regulatory network reveals role of genes in physiological processes of life, including cell differentiation, metabolism, the cell cycle, and signal transduction. In this network, the nodes correspond to genes and messengers RNAs or proteins, and the edges represent the regulatory interactions like activation, inhibition, and repression between the components of the network. Generally it's a collective network of genes, noncoding RNAs, proteins, metabolites, and signaling components (Long et al. 2008). Gene regulatory network incorporates regulation of DNA transcription, RNA translation, posttranscriptional RNA processing, as well as the posttranslational modifications like protein targeting and covalent protein modifications. Gene regulatory networks display the dynamics of the plant systems (Yuan et al. 2008). Coen and Meyerowitz presented ABC model, a first plant gene regulatory networks model, and explained the interactions among transcription factors that regulate floral pattern formation across plant species (Coen and Meyerowitz 1991). The gene regulatory model controls guard cell size in stomatal closure (Li et al. 2006) in plants, and cell fate determination during flower development in *A. thaliana* (Espinosa-Soto et al. 2004). Reconstruction of biological network and analyses has been an important method in plant systems biology. Figure 2 graphically illustrates the central part of typical systems biology method which is thoroughly discussed in this chapter.

6 Conclusion and Future Prospective

It is a scientific challenge to incorporate every piece of biological knowledge into a unified manner, but there is a requirement of an iterative process between different experimental data and mathematical modeling. In the last few years, systems

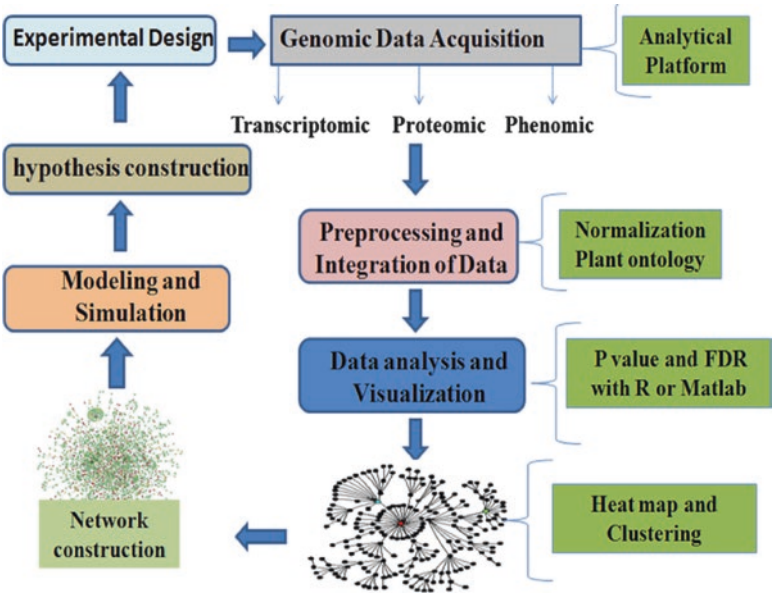

Fig. 2 Graphical representation of systems biology approach in plant biology

biology approach is employed to address key biological questions which provide crucial information and developed hypothesis not only in plant but also in the other areas of life sciences. This chapter defines how to accomplish our research; it starts with genomics and makes relationships between gene-gene, gene to protein, protein-protein, and gene to metabolites. To make this correlation, various methods calculate confidence hypothesis between entities which adds depth breadth to a network and leads to the identification of general network properties. Validation of generated biological network by the various screens and predictions, and their functional consequences in a spatiotemporal manner, is still major challenge. For example, protein-protein interaction shades were attained in systems that do not deliver the spatiotemporal environment of complex formation in addition with protein levels exceeding the native levels. To understand these networks, information is required at the cellular, tissue, and whole-plant system levels. The interpretation of biological networks is important to link an organism that can develop, grow, and reproduce. There are some studies which have nicely demonstrated gene regulation in developmental processes and functions of the cell type and tissue, to understand transcriptional dynamics using cell- and tissue-specific transcript profiling. In system and network biology to understand regulatory networks and protein-protein interactions, it will be important to implement structural features and posttranslational modifications.

References

Anderson NL, Anderson NG (1998) Proteome and proteomics: new technologies, new concepts, and new words. Electrophoresis 19(11):1853–1861

Assmann SM, Albert R (2009) Discrete dynamic modeling with asynchronous update, or how to model complex systems in the absence of quantitative information. Plant Syst Biol 553:207–225

Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA (2004) Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol 14(3):283–291

Bassel GW, Gaudinier A, Brady SM, Hennig L, Rhee SY, De Smet I (2012) Systems analysis of plant functional, transcriptional, physical interaction, and metabolic networks. Plant Cell 24(10):3859–3875

Baum B, Craig G (2004) RNAi in a postmodern, postgenomic era. Oncogene 23(51):8336–8339

Beer MA, Tavazoie S (2004) Predicting gene expression from sequence. Cell 117(2):185–198

Bencharit S, Border MB, Edelmann A, Byrd WC (2013) Update in research and methods in proteomics and bioinformatics. Expert Rev Proteomics 10(5):413–415

Benedict C, Geisler M, Trygg J, Huner N, Hurry V (2006) Consensus by democracy. Using meta-analyses of microarray and genomic data to model the cold acclimation signaling pathway in Arabidopsis. Plant Physiol 141(4):1219–1232

Bhardwaj N, Lu H (2005) Correlation between gene expression profiles and protein–protein interactions within and across genomes. Bioinformatics 21(11):2730–2738

Bi YM, Kant S, Clark J, Gidda S, Ming F, Xu J, Rochon A, Shelp BJ, Hao L, Zhao R, Mullen RT (2009) Increased nitrogen-use efficiency in transgenic rice plants over-expressing a nitrogen-responsive early nodulin gene identified from rice expression profiling. Plant Cell Environ 32(12):1749–1760

Bläsing OE, Gibon Y, Günther M, Höhne M, Morcuende R, Osuna D, Thimm O, Usadel B, Scheible WR, Stitt M (2005) Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in *Arabidopsis*. Plant Cell 17(12):3257–3281

Boone C, Bussey H, Andrews BJ (2007) Exploring genetic interactions and networks with yeast. Nat Rev Genet 8(6):437–449

Botstein D, Cherry JM, Ashburner M, Ball CA, Blake JA, Butler H, Davis AP, Dolinski K, Dwight SS, Eppig JT (2000) Gene ontology: tool for the unification of biology. Nat Genet 25(1):25–29

Bräutigam A, Schliesky S, Külahoglu C, Osborne CP, Weber AP (2014) Towards an integrative model of C4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C4 species. J Exp Bot 65(13):3579–3593

Bumgarner R (2013) Overview of DNA microarrays: types, applications, and their future. Curr Protoc Mol Biol Chapter 22:Unit 22.1

Carstens M, McCrindle TK, Adams N, Diener A, Guzha DT, Murray SL, Parker JE, Denby KJ, Ingle RA (2014) Increased resistance to biotrophic pathogens in the Arabidopsis constitutive induced resistance 1 mutant is EDS1 and PAD4-dependent and modulated by environmental temperature. PLoS One 9(10):e109853

Chandramouli K, Qian PY (2009) Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. Hum Genom Prot 2009:1–22

Chen WJ, Zhu T (2004) Networks of transcription factors with roles in environmental stress response. Trends Plant Sci 9(12):591–596

Chinchilla D, Zipfel C, Robatzek S, Kemmerling B, Nürnberger T, Jones JD, Felix G, Boller T (2007) A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. Nature 448(7152):497

Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. Nature 353(6339):31–37

De Folter S, Immink RG, Kieffer M, Pařenicová L, Henz SR, Weigel D, Busscher M, Kooiker M, Colombo L, Kater MM, Davies B (2005) Comprehensive interaction map of the *Arabidopsis* MADS box transcription factors. Plant Cell 17(5):1424–1433

Espinosa-Soto C, Padilla-Longoria P, Alvarez-Buylla ER (2004) A gene regulatory network model for cell-fate determination during *Arabidopsis thaliana* flower development that is robust and recovers experimental gene expression profiles. Plant Cell 16(11):2923–2939

Gehlenborg N, O'Donoghue SI, Baliga NS, Goesmann A, Hibbs MA, Kitano H, Kohlbacher O, Neuweger H, Schneider R, Tenenbaum D, Gavin AC (2010) Visualization of omics data for systems biology. Nat Methods 7:S56–S68

Geisberg JV, Struhl K (2004) Quantitative sequential chromatin immunoprecipitation, a method for analyzing co-occupancy of proteins at genomic regions in vivo. Nucleic Acids Res 32(19):e151

Hänzelmann S, Castelo R, Guinney J (2013) GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinforma 14(1):7

Hirschhorn JN, Sklar P, Lindblad-Toh K, Lim YM, Ruiz-Gutierrez M, Bolk S, Langhorst B, Schaffner S, Winchester E, Lander ES (2000) SBE-TAGS: an array-based method for efficient single-nucleotide polymorphism genotyping. Proc Natl Acad Sci 97(22):12164–12169

Hu JX, Thomas CE, Brunak S (2016) Network biology concepts in complex disease comorbidities. Nat Rev Genet 17:615–629

Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4(1):44–57

Hui P (2012) Next generation sequencing: chemistry, technology and applications. In: Chemical diagnostics. Springer, Berlin, pp 1–18

Immink RG, Ferrario S, Busscher-Lange J, Kooiker M, Busscher M, Angenent GC (2003) Analysis of the petunia MADS-box transcription factor family. Mol Gen Genomics 268(5):598–606

Johnson CH, Ivanisevic J, Benton HP, Siuzdak G (2014) Bioinformatics: the next frontier of metabolomics. Anal Chem 87(1):147–156

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28(1):27–30

Kanwal S, Ashraf M, Shahbaz M, Iqbal MY (2013) Influence of saline stress on growth, gas exchange, mineral nutrients and non-enzymatic antioxidants in mungbean [(Vigna radiata (L.) Wilczek]. Pak J Bot 45(3):763–771

Kvam VM, Liu P, Si Y (2012) A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. Am J Bot 99(2):248–256

Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM (2011) Prioritizing candidate disease genes by network-based boosting of genome-wide association data. Genome Res 21(7):1109–1121

Leal WS (2013) Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. Annu Rev Entomol 58(1):373–391

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinforma 12:323

Li S, Assmann SM, Albert R (2006) Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. PLoS Biol 4(10):e312

Long TA, Brady SM, Benfey PN (2008) Systems approaches to identifying gene regulatory networks in plants. Annu Rev Cell Dev Biol 24:81

Ma S, Gong Q, Bohnert HJ (2007) An *Arabidopsis* gene network based on the graphical Gaussian model. Genome Res 17(11):1614–1625

Malik A, Lee J, Lee J (2014) Community-based network study of protein-carbohydrate interactions in plant lectins using glycan array data. PLoS One 9(4):e95480

McCormack ME, Lopez JA, Crocker TH, Shahid Mukhtar M (2016) Making the right connections: network biology and plant immune system dynamics. Curr Plant Biol 5:2–12

Metzker ML (2010) Sequencing technologies—the next generation. Nat Rev Genet 11(1):31–46

Mishra A, Pandey D, Singh M, Kumar A (2011) Involvement of hsr203J like gene homologue, protease and protease inhibitors in triggering differential defense response against Alternaria blight in Brassica. Australas Plant Pathol 40(5):461

Nachman I, Regev A, Friedman N (2004) Inferring quantitative models of regulatory networks from expression data. Bioinformatics 20(suppl 1):i248–i256

Nakashima K, Ito Y, Yamaguchi-Shinozaki K (2009) Transcriptional regulatory networks in response to abiotic stresses in *Arabidopsis* and grasses. Plant Physiol 149(1):88–95

Rischer H, Orešič M, Seppänen-Laakso T, Katajamaa M, Lammertyn F, Ardiles-Diaz W, Van Montagu MC, Inzé D, Oksman-Caldentey KM, Goossens A (2006) Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in Catharanthus roseus cells. Proc Natl Acad Sci 103(14):5614–5619

Sheth BP, Thaker VS (2014) Plant systems biology: insights, advances and challenges. Planta 240(1):33–54

Sreenivasulu N, Wobus U (2013) Seed-development programs: a systems biology–based comparison between dicots and monocots. Annu Rev Plant Biol 64:189–217

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci 102(43):15545–15550

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M (2004) Mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37(6):914–939

Todaka D, Nakashima K, Shinozaki K, Yamaguchi-Shinozaki K (2012) Toward understanding transcriptional regulatory networks in abiotic stress responses and tolerance in rice. Rice 5(1):1

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ et al (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515

Wang X, Cairns MJ (2014) SeqGSEA: a Bioconductor package for gene set enrichment analysis of RNA-Seq data integrating differential expression and splicing. Bioinformatics 30:1777–1779. btu090

Yadav BS, Lahav T, Reuveni E, Chamovitz DA, Freilich S (2016) Multidimensional patterns of metabolic response in abiotic stress-induced growth of *Arabidopsis thaliana*. Plant Mol Biol 15:1–1

Yamaguchi-Shinozaki K, Shinozaki K (2005) Organization of cis-acting regulatory elements in osmotic-and cold-stress-responsive promoters. Trends Plant Sci 10(2):88–94

Yuan JS, Galbraith DW, Dai SY, Griffin P, Stewart CN (2008) Plant systems biology comes of age. Trends Plant Sci 13(4):165–171

Yun KY, Park MR, Mohanty B, Herath V, Xu F, Mauleon R, Wijaya E, Bajic VB, Bruskiewich R, de los Reyes BG (2010) Transcriptional regulatory network triggered by oxidative signals configures the early response mechanisms of japonica rice to chilling stress. BMC Plant Biol 10(1):1

Zhao G, Jiang K, Zhang T, Wu H, Qiu C, Deng G (2016) Specific interferon tau gene-regulation networks in bovine endometrial luminal epithelial cells. Theriogenology 105:51–60

Zimmermann IM, Heim MA, Weisshaar B, Uhrig JF (2004a) Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. Plant J 40(1):22–34

Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004b) GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. Plant Physiol 136(1):2621–2632

Zulak KG, Cornish A, Daskalchuk TE, Deyholos MK, Goodenowe DB, Gordon PM, Klassen D, Pelcher LE, Sensen CW, Facchini PJ (2007) Gene transcript and metabolite profiling of elicitor-induced opium poppy cell cultures reveals the coordinate regulation of primary and secondary metabolism. Planta 225(5):1085–1106