10 Advancement in Sustainable Agriculture: Computational and Bioinformatics Tools

Abstract

Sustainable agricultural production is an urgent issue in response to global climate change and population increase. Furthermore, recent increased demand for biofuel crops has created a new market for agricultural commodities. One potential solution is to increase plant yield by designing plants based on a molecular understanding of gene function and on the regulatory networks involved in stress tolerance, development and growth. Recent progress in plant genomics has allowed us to discover and isolate important genes and to analyze functions that regulate yields and tolerance to environmental stress.

Keywords

Transcriptomics · Proteomics · Informatics · Genomics · Integration

10.1 Bioinformatics and Its Applications in Plant Biology

Bioinformatics refers to the study of biological information using concepts and methods in computer science, statistics and engineering and can be divided into two broad categories: Biological information management and computer biology. The boundaries of these categories are becoming more diffuse and other categories will no doubt surface in the future as this field matures. In our society, our economy and our global environment, plant life plays important and diverse roles. For modern plant biotechnology, feeding the growing world population is a challenge. Crop yields have increased during the last century and will continue to improve as agronomy re-assorting the enhanced breeding and develop new biotechnological-engineered strategies. The onset of genomics is providing massive information to improve crop phenotypes. Accumulating sequence data enables detailed genome analysis through the use of friendly access to database and retrieval of information. Genetic and molecular genome co linearity allows efficient transfer of data

J. A. Parray et al., *Sustainable Agriculture: Biotechniques in Plant Biology*,
https://doi.org/10.1007/978-981-13-8840-8_10

revealing extensive conservation of genome organization between species. Genome research's goals are to identify sequenced genes and deduct their functions through metabolic analysis and reverse genetic screens from gene knockouts. More than 20% of the predicted genes occur as a cluster of related genes that generate a significant proportion of gene families. Multiple alignments provide a method for estimating gene numbers in gene families to identify previously described genes. This information allows for new strategies in plants to study patterns of gene expression. Available news technology information, as the DNA microarray expression data stored in the database, will assist functional genomics of plant biology. Expressed sequence tags also provide an opportunity to compare digital northern gene expression levels that provide initial clues to unknown regulatory phenomena. Collections of databases and bioinformatics resources for crop plant genomics were built on crop plant networks to harness the extensive genome mapping work. This resource facilitates the identification of ergonomically important genes through comparative analyzes between crop plants and model species, enabling genetic engineering of selected crop plants by the quality of the resulting products. Resources in bioinformatics have evolved beyond expectations, developing new biotechnologies in nutritional genomics biotechnology tools to genetically modify and improve food supply, for an ever-increasing world population. Bioinformatics can now be leveraged to speed up the translation into agriculture of basic discovery. Farming will be affected by predictive manipulation of plant growth at a time when food security, land reduction available for agricultural use, environmental stewardship, and climate change are all issues of growing public concern.

10.1.1 Sequence Analysis

Biological sequence such as DNA, RNA, and protein sequence is the most fundamental object for a biological system at the molecular level. Several genomes have been sequenced to a high quality in plants, including Arabidopsis thaliana (The Arabidopsis Genome Initiative 2000) and rice (Goff et al. 2002). Draft genome sequences are available for poplar (http://genome.jgi-psf.org/Poptr1/) and lotus (http://www.kazusa.or.jp/lotus/), and sequencing efforts are in progress for several others including tomato, maize, *Medicago truncatula*, sorghum (Bedell et al. 2005) and close relatives of *Arabidopsis thaliana*. Researchers also generated expressed sequence tags (ESTs) from many plants including lotus, beet, soybean, cotton, wheat, and sorghum (http://www.ncbi.nlm.nih.gov/dbEST/). Genome Sequencing Advances in sequencing technologies provide opportunities in bioinformatics for managing, processing, and analyzing the sequences. Shotgun sequencing is currently the most common method in genome sequencing: pieces of DNA are sheared randomly, cloned, and sequenced in parallel. Software has been developed to piece together the random, overlapping segments that are sequenced separately into a coherent and accurate contiguous sequence (Gibbs and Weinstock 2003). Numerous software packages exist for sequence assembly (Pop et al. 2004), including Phred/Phrap/Consed (http://www.phrap.org), Arachne (http://www.broad.mit.edu/wga/), and GAP4 (http://staden.sourceforge.net/overview.html). TIGR developed a modular, open-source package called AMOS (http://www.tigr.org/software/AMOS/), which can be used for comparative genome assembly (Patil et al. 2001). Current limitations in shotgun sequencing and assembly software largely remain in assembling highly repetitive sequences, although the sequencing cost is another limitation. Recently developed methods continue to reduce sequencing costs, including sequencing using differential hybridization of oligonucleotide samples, polymorphism ratio sequencing (Blazej et al. 2003), four-color chip-based DNA sequencing and the "454 method" based on high-density micro-fabricated picoliter reactors (Margulies et al. 2005). In terms of experimental design, data interpretation and analysis, each of these sequencing technologies poses significant analytical challenges for bioinformatics in terms of experimental design, data interpretation, and analysis of the data in conjunction with other data (Di et al. 2005). Gene finding and Genome Annotation Gene finding refers to prediction of introns and exons in a segment of DNA sequence. Dozens of computer programs for identifying protein-coding genes are available (Zhang 2002). Some of the well-known ones include Genscan (http://genes.mit.edu/GENSCAN.html), GeneMarkHMM (http://opal.biology. gatech.edu/GeneMark/), GRAIL (http://compbio.ornl.gov/Grail-1.3/), Genie (http://www.fruitfly.org/seqtools/genie.html), and Glimmer (http://www.tigr.org/softlab/glimmer). Several new gene-finding tools are tailored for applications to plant genomic sequences (Schlueter et al. 2003). Ab initio gene prediction remains a challenging problem, especially for large-sized eukaryotic genomes. For a typical *Arabidopsis thaliana* gene with five exons, at least one exon is expected to have at least one of its borders predicted incorrectly by the ab initio approach (Brendel and Zhu 2002). Transcript evidence from full-length cDNA or EST sequences or similarity to potential protein homologs can significantly reduce uncertainty of gene identification (Zhu et al. 2003). Several software packages have been developed for structural annotation (Allen et al. 2004). In addition, one can use genome comparison tools such as SynBrowse (http://www.synbrowser.org/) and VISTA (http://genome.lbl.gov/vista/index.shtml) to enhance the accuracy of gene identification. Current structural annotation limitations include accurate transcript start sites prediction and identification of small genes encoding less than 100 amino acids, non-coding genes) and alternative splicing sites. The analysis of repetitive DNAs, which are copies of identical or almost identical sequences present in the genome (Lewin 2003), is an important aspect of genome annotation. There are repetitive sequences in nearly any genome and abundant in most plant genomes (Jiang et al. 2004). Identifying and characterizing repeats is essential for shedding light on the evolution, function and organization of genomes and for filtering many types. A small library of plant specific repeats can be found at ftp://ftp.tigr.org/pub/data/TIGRPlant Repeats/; this is likely to grow substantially as more genomes are sequenced. One can use Repeat Masker (http://www.repeatmasker.org/) to search repetitive sequences in a genome. Repeats with poorly conserved patterns or short sequences are hard to identify using RepeatMasker due to the limitations of BLAST. To identify novel repeats, various algorithms were developed. Some widely used tools include Repeat Finder (http://ser-loopp.tc.cornell.edu/cbsu/repeatfinder.htm) and

RECON (http://www.genetics.wustl.edu/eddy/recon/). However, due to the high computational complexity of the problem, none of the programs can guarantee finding all possible repeats as all the programs use some approximations in computation, which will miss some repeats with less distinctive patterns. Inevitably, a combination of repeat finding tools is required to obtain a satisfactory overview of repeats found in an organism. Comparing sequences provides a foundation for many bioinformatics tools and may allow inference of the function, structure, and evolution of genes and genomes. For example, sequence comparison provides a basis for building a consensus gene model like UniGene (Boguski and Schuler 1995). Also, many computational methods have been developed for homology identification (Wan and Xu 2005). Although sequence comparison is highly useful, it should be noted that it is based on sequence similarity between two strings of text, which may not correspond to homology, especially when the confidence level of a comparison result is low. Also, homology may not mean conservation in function. Methods in sequence comparison can be largely grouped into pair-wise, sequence profile, and profile-profile comparison. For pair-wise sequence comparison, FASTA (http://fasta.bioch.virginia.edu/) and BLAST (http://www.ncbi.nlm.nih.gov/blast/) are popular. To assess the confidence level for an alignment to represent homologous relationship, a statistical measure was integrated into pair-wise sequence alignments (Karlin and Altschul 1990). Remote homologous relationships are often missed by pair-wise sequence alignment due to its insensitivity. Sequence-profile alignment is more sensitive for detecting remote homologs. A protein sequence profile is generated by multiple sequence alignment of a group of closely related proteins. A multiple sequence alignment builds correspondence among residues across all of the sequences simultaneously, where aligned positions in different sequences probably show functional and/or structural relationship. A sequence profile is calculated using the probability of occurrence for each amino acid at each alignment position. PSI-BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) is a popular example of a sequence-profile alignment tool. Some other sequence-profile comparison methods are slower but even more accurate than PSI-BLAST, including HMMER (http://hmmer.wustl.edu/), SAM (http://www.cse.ucsc.edu/research/compbio/sam. html), and META-MEME (http://metameme.sdsc.edu/). In the detection of remote homologues, a profile-profile alignment is more sensitive than sequence-based search programs (Yona and Levitt 2002). Because of its high false positive rate, however, the comparison between profile and profile is not widely used. It is helpful to correlate the sequence comparison results with the relationship observed in functional genomic data, especially the widely available microarray data as discussed in the transcriptome analysis section below, given potential false positive predictions. For example, if a gene is predicted to have a particular function by sequence comparison, the prediction can be trusted if the gene has a strong correlation in gene expression profile with other genes known to have the same function. Proteins can be generally classified based on sequence, structure, or function. Several sequence-based methods were developed based on sizable protein sequence (typically longer than 100 amino acids), including Pfam (http://pfam.wustl.edu/), ProDom (http://protein.toulouse.inra.fr/prodom/current/html/home.php), and Clusters of Orthologous Group (COG) (http://www.ncbi. nlm.nih.gov/COG/new/).

Other methods are based on fingerprints of small conserved motifs in sequences, as with PROSITE (http://au.expasy.org/prosite/), PRINTS (http://umber.sbs.man.ac.uk/dbbrowser/PRINTS/), and BLOCKS (http://www.psc.edu/general/software/packages/blocks/blocks.html). The false positive rate of motif assignment is high due to high probability of matching short motifs in unrelated proteins by chance. Other sequence-based protein family databases are built from multiple sources. InterPro (http://www.ebi.ac.uk/interpro/) is a database that integrates domain information from multiple protein domain databases. Using protein family information to predict gene function is more reliable than using sequence comparison alone. On the other hand, very closely related proteins may not guarantee a functional relationship (Noel et al. 2005). One can use structure or function-based protein families (when available) to complement sequence-based family for additional function information. SCOP (http://scop.mrc-lmb.cam.ac.uk/scop/) and CATH (http://cath-www.bio chem.ucl.ac.uk/) are the two well-known structure-based family resources. ENZYME (http://us.expasy.org/enzyme/) is a typical example of a function family. A protein family can be represented in a phylogenetic tree that shows the evolutionary relationships among proteins. Phylogenetic analysis can be used in comparative genomics, gene function prediction, and inference of lateral gene transfer among other things (Doolittle 1999). The analysis typically starts from aligning the related proteins using tools like ClustalW (http://bips.u-strasbg.fr/fr/Documentation/ClustalX/). Among the popular methods to build phylogenetic trees are minimum distance, maximum parsimony, and maximum likelihood trees. Some programs provide options to use any of the three methods, e.g., the two widely used packages PAUP (http://paup.csit.fsu.edu), and PHYLIP (http://evolution.genetics.washington.edu/phylip.html). Although phylogenetic analysis is a research topic with a long history and many methods have been developed, various heuristics and approximations are used in constructing a phylogenetic tree, as the exact methods are too computationally intense.

10.1.2 Transcriptome Analysis

The primary goal of transcriptome analysis is to learn how an organism's growth and development and response to the environment changes in transcript abundance control. DNA microarrays have been shown to be a powerful technology for gene-wide gene transcription profile observation (Schlueter et al. 2003). Microarray data is also combined with other information to infer coregulated processes such as regulatory sequence analysis, gene ontology, and pathway information. Whole-genome tiled arrays are used to detect transcription without prejudice to known or predicted structures of genes and alternative variants of splices. Other types of analysis include the analysis of ChIP-chip (chromatin immune precipitation (ChIP) and microarray chip, combining microarrays with methods for detecting chromosomal locations where protein-DNA interactions occur across the genome (Buck and Lieb 2004). DNA immune precipitation (DIP-chip) is used by a related technique to predict DNA-binding sites (Liu et al. 2005; Brenner et al.

2000). Microarray analysis makes it possible to measure transcript simultaneously measurement of transcript abundance for thousands of genes (Zhu and Wang 2000). Two general types of microarrays are high-density oligonucleotide arrays containing a large number of relatively short (25–100-mer) samples synthesized directly on the surface of the arrays, or arrays of amplified polymerase chain reaction products or cloned DNA fragments mechanically located directly on the surface of the array. Many different technologies are being developed (Meyers et al. 2004). Competition between microarray platforms has resulted in lower costs and higher gene numbers per array. Unfortunately, the variety of array platforms makes it difficult to compare microarray results between microarray formats that use different probe sequences, RNA sample labeling, and data collection methods (Woo et al. 2004). Even for standardized arrays like those from Affymetrix, the optimal statistical treatment for the sets of samples designed for each gene still has arguments. The Affycomp software, for example, compares Affymetrix results using two spike-in experiments and a dilution experiment for different standardization methods under different evaluation criteria (Cope et al. 2004). You can use this information to select the appropriate methods for normalization. There are many tools available for conducting a variety of analysis on large data sets of microarrays. Examples include commercial software such as Gene Traffic, GeneSpring (http://www.agilent.com/chem/genespring), Affymetrix's GeneChip Operating Software (GCOS), and public software such as Cluster (Eisen et al. 1998), CaARRAY (http://caarray.nci.nih.gov/), and BASE (Saal et al. 2002). A notable example is Bioconductor (http://www.bioconductor.org), which is an open-source and open-development set of routines written for the open-source R statistical analysis package (http://www.r-project.org). Observing transcriptional activity patterns that occur under various conditions, such as genotypes or time courses, reveals genes that have highly correlated patterns of expression. The correlation, however, cannot distinguish between genes under common regulatory control and those whose patterns of expression merely correlate. Recent microarray analysis efforts have focused on experimental analysis of microarray data (Mockler and Ecker 2005). A Toxico genomics research consortium study indicates "microarray results can be comparable across multiple laboratories, particularly when using a common platform and set of procedures"(Bard and Rhee 2004). Meta-analysis can examine the effect of the same treatment on different studies in order to arrive at a single estimate of the true effect of treatment (Rhodes et al. 2004). Tiling arrays Known and predicted genes are typical microarray samples. Tiling arrays cover the genome at regular intervals to measure unbiased transcription to known or predicted gene structures, polymorphism discovery, alternative splicing analysis, and transcription factor-binding sites identification (Mockler and Ecker 2005). Whole-genome arrays (WGAs) cover the entire genome with regular gaps overlapping samples or samples. The WGA ensures that the experimental results are not dependent on the level of current genome annotation, and those new transcripts and unusual forms of transcription are discovered. Similar studies for the entire genome of Arabidopsis (Stolc et al. 2005) and parts of the rice genome have been performed in plants (Toyoda and Shinozaki 2005). These studies identified thousands of novel transcription units including centromer genes, significant transcription of antisense genes, and transcription activity in intergenic regions. Tiling array data can also be used to validate the predicted boundaries intron/exon boundaries (Toyoda and Shinozaki 2005). Further work is needed to establish the best practices for determining when transcription has occurred and how to normalize array data across the different chips. Visualization of the output from tiling arrays requires viewing the probe sequences on the array together with the sequence assembly and the probe expression data. The Arabidopsis Tiling Array Transcriptome Express Tool (also known as ChipViewer) (http://signal.salk.edu/cgibin/atta) displays information about what type of transcription occurred along the Arabidopsis genome (Yamada et al. 2003). Another tool is Affymetrix's Integrated Genome Browser (IGB), a Java program that investigates genomes and combines annotations from multiple sources of data. Another option to view such data is collaborations such as those between Gramene (Ware et al. 2002) and PLEXdb (Shen et al. 2005), allowing users to overlay probe array information to a comparative sequence viewer. The major limitations of WGAs include a sequenced genome requirement, the large number of chips required for complete genome coverage, and recent duplicated (and thus highly homologous) gene analysis. Regulatory sequence analysis Discoverin includes the interpretation of the results of microarray experiments involves discovering why genes with similar expression profiles behave in a coordinated fashion. Regulatory sequence analysis approaches this question by extracting motifs that are shared between the upstream sequences of these genes (van Helden 2003). Comparative genomics studies of retained non-coding sequences (CNSs) may also help to identify key motives (Inada et al. 2003). There are several methods on the upstream of coregulated genes to search for over-represented motifs. Approximately two classes can be categorized: oligonucleotide-based frequency (van Helden 2003) and probabilistic sequence-based models (Roth et al. 1998). The frequency-based method of oligonucleotides calculates the statistical significance of a site based on the frequency tables of oligonucleotides observed in all non-coding regions of the genome of the specific organism. The oligonucleotide length usually varies from 4 to 9 bases. Hexanucleotide (6-length oligonucleotide) analysis is most widely used. It is then possible to group the significant oligonucleotides as longer consensus motifs. Frequency-based methods tend to be simple, effective and comprehensive. The main limitation is the problem of identifying complex patterns of motifs. Regulatory Sequence Analysis Tools (RSAT), the public web resource, performs sequence similarity searches and analyzes the genome non-coding sequences (van Helden 2003). The motif is represented as a position probability matrix for probabilistic-based methods, where the motifs are supposed to be hidden in the noisy background sequences. One of the strengths of probabilistic methods is the ability to identify motifs with complex patterns. It is possible to identify many potential motives; however, separating unique motives from this large pool of potential solutions can be difficult. Also, probabilistic-based methods tend to be computationally intense, as they must be run multiple times in order to obtain an optimal solution. AlignACE, Aligns Nucleic Acid Conserved Elements (http://atlas.med.harvard.edu/), is a popular motif finding tool first developed for yeast but expanded to include other species (Roberts et al. 2000).

10.1.3 Computational Proteomics

Proteomics is a leading technology for protein qualitative and quantitative characterization and genome-scale interactions. The proteomics goals include large-scale identification and quantification of all protein types in a cell or tissue, post-translation modification analysis and association with other proteins, and characterization of protein activities and structures. Proteomics application in plants is still in its initial phase, mostly in the identification of proteins (Newton et al. 2010). Other proteomic aspects such as protein-protein interaction identification and prediction, protein activity profiling, local subcellular protein localization, and protein structure, have not been widely used in plant science. However, recent efforts such as the structural genomic initiative that includes Arabidopsis (http://www.uwstructuralgenomics.org/) are encouraging. Electrophoresis Analysis Electrophoresis analysis can qualitatively and quantitatively investigate expression of proteins under different conditions (Gorg et al. 2000). Several bioinformatics tools have been developed for two-dimensional (2D) electrophoresis analysis (Mao et al. 2005). SWISS-2DPAGE can locate the proteins on the 2D PAGE maps from SwissProt (http://au.expasy.org/ch2d/). Melanie (http://au.expasy.org/melanie/) can analyze, annotate, and query complex 2D gel samples. Flicker (http://open2dprot.sourceforge.net/Flicker/) is an open-source stand-alone program for visually comparing 2D gel images. PDQuest (http://www.proteomeworks.bio-rad.com) is a popular commercial software package for comparing 2D gel images. Some software platforms handle related data storage and management, including PEDRo (http://pedro.man.ac.uk/), a software package for modeling, capturing, and disseminating 2D gel data and other proteomics experimental data. Limited ability to identify proteins and low accuracy in detecting protein abundance are the main limitations of electrophoresis analysis. Protein Identification by Mass Spectrometry following protein separation using 2D electrophoresis or liquid chromatography and protein digestion using an enzyme (trypsin, pepsin, glu-C, etc.), proteins are typically identified using mass spectrometry (MS). MS provides a high-throughput approach for large-scale protein identification, unlike other protein identification techniques, such as Edman degradation microsequencing. The data generated from mass spectrometers are often complicated and the interpretation of computational analysis is critical in interpreting the data for protein identification (Gras and Muller 2001). The lack of open-source software is a major limitation in MS protein identification. Expensive commercial packages are the most widely used tools. Furthermore, current statistical models are generally oversimplified for matches between MS spectra and protein sequences. Consequently, confidence assessments are often unreliable for the results of computational protein identification. Two types of protein identification methods are available for MS: peptide mass fingerprinting (PMF) and tandem mass spectrometry (MS/MS).

10.1.4 Peptide Mass Fingerprinting

Identification of PMF peptides/proteins compares the masses of peptides derived from experimental spectral peaks with each of the possible protein computationally digested peptides in the sequence database. The proteins in the sequence database are considered candidates for the proteins in the experimental sample, with a significant number of peptide matches. MOWSE (Pappin et al. 1993) was a previous PMF protein identification software package, and Emowse (http://emboss.sourceforge.net/) is the latest MOWSE algorithm implementation. Several other computational tools for PMF protein identification have also been developed. MS-Fit in the Protein Prospector (http://prospector.ucsf.edu/) uses a variant of the MOWSE scoring scheme that incorporates new features, including restrictions on the minimum number of peptides to match for a possible hit, the number of missed cleavages and the molecular weight range of the target protein. The MOWSE algorithm extension is Mascot (http://www.matrixscience.com/). It incorporates the same scoring scheme with a probability-based score being added. A limitation of the identification of PMF protein is that it can sometimes not identify proteins because multiple proteins in the database can fit the spectra of PMF. In this case, further experiments with MS/MS are necessary to identify the proteins.

10.1.5 Tandem Mass Spectrometry

MS/MS further breaks each digested peptide into smaller fragments, whose spectra provide effective signatures of individual amino acids in the peptide for protein identification. Many tools have been developed for MS/MS-based peptide/protein identification, the most popular ones being SEQUEST (http://fields.scripps.edu/sequest/) and Mascot (http://www.matrixscience.com/). Both rely on the comparison between database-derived theoretical peptides and spectrometric tandem spectra of experimental mass. One of the earliest tools developed for this, SEQUEST produces a list of possible assignments of peptide / protein in a protein mixture based on a correlation scoring scheme (Yates et al. 1995). Mascot uses a similar algorithm to identify MS / MS peptide/protein together with its PMF protein identification capacity as SEQUEST. The limitations of these programs are that due to various factors, including sequencing and annotation errors in the search database, a significant portion of MS / MS spectra cannot be assigned. Furthermore, computational approaches are not currently used to handle post-translation modifications well. An active research area (Dancik et al. 1999) is the de novo sequencing approach based on MS / MS spectra. The algorithms typically match peak separations by the mass of one or more amino acids and infer the likely peptide sequences consistent with matched amino acids (Chen et al. 2001). Several popular peptide de novo sequencing software packages are available using MS/MS data, including Lutefisk (http://www.hairyfatguy.com/lutefisk/) and PEAKS (http://www.bioinformaticssolutions.com/products/peaks). One limitation of the current methods is that they are frequently used.

10.1.6 Metabolomics and Metabolic Flux

Metabolomics is the analysis at any given time of a cell's complete pool of small metabolites. Because of the proliferation of secondary metabolites, metabolomics may be particularly important in plants (van Helden et al. 2000). Metabolites are extracted from tissues, separated, and analyzed in a high-throughput manner in a metabolite profiling experiment (Dancik et al. 1999). Metabolic fingerprinting examines a few metabolites to help differentiate samples by phenotype or biological relevance (Shanks 2005). Technology has now advanced to quantify >1000 compounds from a single leaf extract semi-automatically (Ware et al. 2002). The key challenge in metabolite profiling is to identify metabolites from complex plant samples quickly, consistently, and unambiguously (Sriram et al. 2004). Identification is routinely carried out using time-consuming standard additional experiments using commercially available or purified preparations for metabolites. For gas chromatography-mass spectrometry (GC–MS) profiles from various biological sources, a publicly accessible database is needed that contains the evidence and underlying metabolite identification. Experimental metadata standards and metabolomi data quality standards are still in a very early stage and a large-scale public repository is not yet available. The ArMet (metabolomics architecture) proposal (Harris et al. 2005) provides a description and results of plant metabolomics experiments along with a database scheme. MIAMET (Minimum Information on a Metabolomics Experiment) (Gorg et al. 2000) provides reporting requirements with a view to standardizing descriptions of experiments, especially in publications. The Working Group on Standard Metabolic Reporting Structures (SMRS Working Group, 2005) developed standards to describe the biological sample origin, analytical technologies, and methods used in a metabolite profiling experiment. Metabolite data were used to build networks of metabolic correlation (Steuer et al. 2003). Such correlations may reflect the net partitioning of carbon and nitrogen through transcriptional or biochemical processes resulting from direct enzymatic conversions and indirect cell regulation. Metabolic correlation matrices, however, cannot infer that a change in one metabolite in a metabolic reaction network led to a change in another metabolite (Steuer et al. 2003). The steady-state flow between metabolites is measured by metabolic flux analysis. However, fluxes are even more difficult to measure than metabolite levels because of complications in intracellular metabolite transport modeling and incomplete knowledge of in vivo pathway topology and location (Shanks 2005). The most basic approach to metabolic flux analysis is stoichiometric analysis, which calculates the quantities of reactants and chemical reaction products to determine each metabolite's flux (Edwards and Palsson 2000). However, for large networks, this method is numerically difficult to solve and it has problems when there are parallel metabolic pathways, metabolic cycles, and reversible reactions (Wiechert et al. 2001). FluxAnalyzer is a MATLAB package that integrates metabolic network path and flux analysis (Klamt et al. 2003). Flux analysis using 13C carbon labeling data attempts to overcome some of the disadvantages of the above-mentioned stoichiometric flux analysis (Sriram et al. 2004). In the 13C restricted flux analysis and the stoichiometric and isotopomer balances, more rigorous analysis is needed to fully determine fluxes from all experimental data. Iterative methods were used to solve the resulting matrix of isotopomer balances, with the measurements of nuclear magnetic resonance or gas chromatography being used for consistency purposes. As more reliable data are collected, ordinary differential equations can be used for metabolic network dynamic simulations, combining information on connectivity, concentration balances, flux balances, metabolic control, and pathway optimization.

Ultimately, one may integrate all of the information and perform analysis and simulation in a cellular modeling environment like E-Cell (http://www.e-cell.org/) or CellDesigner (http://www.systems-biology.org).

10.1.7 Ontologies

Ontology is a set of vocabulary terms with explicit meanings and relationships with other terms used to annotate data (Ashburner et al. 2000). Bio-Ontology Types A growing number of common ontologies are being constructed and used in biology. Examples include ontologies to describe gene and protein function (Harris et al. 2004), cell types (Bard et al. 2005), anatomies and organism developmental phases (Garcia-Hernandez et al. 2002), microarray experiments (Stoeckert et al. 2002), and metabolic pathways (Mao et al. 2005). The Open Biological Ontologies Web site (http://obo.sourceforge.net/) provides a list of open-source ontologies used in biology. A lot of ontology is under development on this site and is subject to frequent changes. Gene Ontology (GO) (www.geneontology.org) is an example of bio-ontology, which has gained acceptance from the community. It is a set of more than 16,000 controlled vocabulary terms for the biological domains of molecular function, subcellular compartment, and biological process. GO is organized as a directed acyclic graph, a type of hierarchy tree that allows a term to exist as a specific concept belonging to more than one general term. Other examples of ontologies currently in development are the Sequence Ontology (SO) project (Eilbeck et al. 2005) and the Plant Ontology (PO) project (www.plantontology.org). The SO project aims to explicitly define all the terms needed to describe features on a nucleotide sequence, which can be used for genome sequence annotation for any organism. The PO project aims to develop shared vocabularies to describe anatomical structures for flowering plants to depict gene expression patterns and plant phenotypes. A few challenges in the development and use of ontologies remain to be addressed, including redundancies in the ontologies, minimal or lack of formal, computer comprehensive definitions of the terms in the ontologies, and general acceptance by the research and publishing community (Bard and Rhee 2004). An international repository of ontology standards is available to oversee the development and maintenance of ontologies. Ontology applications are mainly used to annotate data such as sequences, clusters of gene expression, experiments, and strains. Ontologies that have such annotations of data in databases can be used in numerous ways, including connecting different databases, refining search, providing a framework for interpreting the results of functional genomics experiments, and inferring knowledge (Bard and Rhee 2004). For example, one can ask which functions and processes in an expression cluster of interest are statistically significantly over-represented in an expression cluster of interest compared to the functions and processes carried out by all of the genes from a gene expression array. Since GO is one of the more well-established ontologies, this section focuses on GO to illustrate ontology applications in biology. Many model organism databases (http://www.geneontology.org/GO.current.annotations.shtml, http://www.geneontology.org/GO.biblio.shtml#annots) used ontologies to annotate genes and gene products. Function annotations of genes using GO have been used primarily in two ways: predicting protein functions, processes, and patterns of localization from different data sources (http://www.geneontology.org/GO.biblio.shtml#predictions) and providing a biological framework or benchmark set for interpreting large-scale sampling results such as genes expression profiles and protein-protein interactions (http://www.geneontology.org/GO.biblio.shtml#geneexp). Furthermore, GO annotations were used to test the robustness of search methods for semantic similarity (Lord et al. 2003) and to study adaptive evolution. Using GO annotations to predict function and use them as a benchmark for large-scale data has several problems. One is the misuse or lack of use of evidence codes, providing the kind of evidence used to make the annotation (http://www.geneontology.org/GO.evidence.shtml). Only approximately half of the codes of evidence refer to direct experimental evidence. In addition, several codes of evidence are used for indirect evidence, indicating less certainty in annotation assertion than those made with direct evidence. Other codes are used for computationally derived annotations and do not have experimental support and are more likely to be incorrect. Researchers and computer programs using the annotations to provide knowledge or analyze functional genomics data should be familiar with these codes of evidence to minimize data misinterpretation. For example, methods for evaluating the relationship between sequence conservation and gene co-expression and using GO annotations to validate their results should ensure that no annotations are used to avoid circular arguments using ISS and IEA evidence codes. Similarly, studies seeking to define biological processes and functions from gene expression data using the GO annotations should ensure that no annotation with inferred from expression pattern (IEP) evidence code is used. The other caveat is that annotations to GO are not equivalently represented throughout GO. When looking for statistical over-representation of GO terms in genes of an expression cluster, there is low statistical power for detecting deviations from expectation for terms that are annotated with a small number of genes (Khatri and Draghici 2005).

10.1.8 Emerging Areas in Bioinformatics

In this section the main focus will be on text mining, biology of systems, and semantic web. Some other emerging areas, such as image analysis (Sinha et al. 2002), grid computing (Foster 2002), directed evolution (Dalby 2003), rational protein design (Looger et al. 2003), microRNA-related bioinformatics (Brown and Sanseau 2005), and modeling in epigenomics (Fazzari and Greally 2004) are not covered due to the limitation of space. The Medline 2004 database had 12.5 million entries and is expanding at a rate of 500,000 new citations each year (Cohen and Hersh 2005). The goal of text mining is to allow researchers to identify needed information and shift the burden of searching from researchers to the computer. Without automated text mining, much of biomolecular interactions and biological research archived in the literature will remain accessible in principle but underutilized in practice. One key area of text mining is relationship extraction that finds relationships between entities such as genes and proteins. Examples include MedMiner at the National Library of Medicine (Tanabe et al. 1999), PreBIND (Donaldson et al. 2003.), the curated BIND system (Alfarano et al. 2005), PathBinderH (Ding et al. 2005), and iHOP (Hoffmann and Valencia 2004). Results on real-world tasks such as automatic extraction and assignment of GO annotations are promising, but they are far from achieving the required performance required by applications in the real world (Blaschke et al. 2005). A key challenge that needs to be addressed in this field is the complex nature of names and terminology such as the wide range of variants in free text for protein names and GO terms. The current system generation is beginning to combine statistical methods with machine learning to capture expert knowledge about how genes and proteins are referred to in scientific papers in order to create usable systems with high precision and to recall specialized tasks in the future. Computational Systems Biology Classical systems analysis in engineering treats a system as a black box whose internal structure and behavior can be analyzed and modeled by varying internal or external conditions and studying the effect of variation on external observables. The result is a comprehension of the system's internal makeup and working mechanisms (Kell et al. 2005). Biology of systems is applying this theory to biology. The observables are measurements of what the organism are doing, ranging from descriptions of phenotypes to detailed metabolic profiling. A critical issue is how different data types, such as sequence, gene expression, protein interac, can be effectively integrated and phenotypes to infer biological knowledge. Some areas that require more work include creating coherent validated data sets, developing common formats for pathway data (SBML) (Hucka et al. 2004) and BioPAX (http://www.biopax.org)), and creating ontologies to define complex interactions, curation, and linkages with textmining tools. The Systems Biology Workbench project (http://sbw.kgi.edu/) aims to develop an open-source software framework for sharing information between different types of pathway models. Other issues are that biological systems are underdefined (not enough measurements are available to characterize the system) and samples are not taken often enough to capture time changes in a system that may occur at vastly different time scales in different networks such as signaling and regulatory networks (Papin et al. 2004). The long-term goal of creating a cell's complete silico model is still a long way off; however, the tools being developed to integrate information from a wide variety of sources will be of short-term value. Semantic Web Semantic Web is a model for "creating a universal mechanism for exchanging information by giving meaning to the content of documents and data on the Web in a machine-interpretable manner" (Neumann 2005). This model will enable the development of search tools that know what type of information can be obtained from which documents and understand how the information in each document relates to another, allowing the use of reasoning and logic by software agents to make decisions automatically based on the constraints provided in the query (e.g., automatic travel agents, phenotype prediction) (Berners-Lee et al. 2001). Bioinformatics could greatly benefit from the successful implementation of this model and should play a leading role in its implementation (Papin et al. 2004). Current efforts have focused on the development of standards and specifications for the identification and description of data such as Universal Resource Identifier (URI) and Resource Definition Framework (RDF) respectively (http://www.w3c.org/2001/sw). While implementation of web-based applications is scarce at this point, some useful examples are being developed, such as Haystack (a browser that retrieves data from multiple databases and allows users to annotate and manage the information to reflect their understanding) (http://www-db.cs.wisc.edu/cidr/cidr2005/papers/P02.pdf) and BioDash (a drug development user interface that associates diseases, drug progression stages, molecular biology, and pathway knowledge for users) (http://www.w3.org/2005/04/swls/BioDash/Demo/). Cellular Localization and Spatially Resolved Data Research in nanotechnology and electron microscopy enables researchers to select specific cell and tissue areas and to picture spatiotemporal distributions of signaling receptors, gene expression and proteins. Laser microdissection capture allows specific tissue types to be selected for detailed analysis (Emmert-Buck et al. 1996). In Arabidopsis, confocal imaging is used to model the patterns of auxin transport and gene expression (Heisler et al. 2005). Methods are applied in electron microscopy to image spatiotemporal signaling distribution of signaling receptors. Improved methods in laser scanning microscopes may allow measurements of fast diffusion and dynamic processes in the microsecond-to-millisecond time range in live cells (Digman et al. 2005). These emerging capabilities will lead to new understanding of cell dynamics.

Bioinformatics integration will influence plant science and will lead to crop improvements in the following areas:

(a) Identifying important genes through genomics, analysis of expression and functional genomics.
(b) The design of agrochemicals based on the analysis of signal perception and transduction pathways components to identify targets and chemin-formatics compounds that may be used as herbicides, pesticides or insecticides.
(c) Use of plant genetic resources to preserve genetic diversity in farming species.
(d) Efficient use of biological clone, cell, organism and seed repositories.

10.1.9 Software's for Microarray Data Analysis

Most statistical packages used in microarray experiment data analysis i.e. In gene expression studies, SAS, SPSS, MATLAB and R are not entirely unique. To analyze such experiments, many tools are available and much easier than the aforementioned pure statistical programs.

10.1.9.1 FlexArray

FlexArray is a Windows software package designed to simplify expression micro array data analysis. FlexArray's target audience is biological scientists. Currently the software supports Affymetrix Gene Chips, Nimble Gen, Illumina Bead Chips and various types of one-color and two-color arrays of expression. FlexArray is well suited to projects of small and medium size. FlexArray is on the programming language of R. FlexArray is a tool that generates gene lists that is not suitable for data mining. This tool is suitable for algorithms for standardization, statistical testing and other complex data processing tasks. It is also an exploration tool for methods and algorithms of analysis (Blazejczyk et al. 2007). This software can be found at http://genomequebec.mcgill.ca/FlexArray. As an example, Khojasteh et al. (2017) used the FlexArray software in order to identify responsive genes against two main pathovars of *Xanthomonas oryzae* in different rice varieties.

10.1.9.2 BioConductor

The Bioconductor package in R (http://www.r-project.org/) is an open source and open software project that is used to analyze and understand genomics data, particularly microarray data. Bio conductor is primarily based on the language of R programming although it is friendly with different programming languages. For different types of microarray analysis, a large number of different packages are available. Readers should follow http://www.bioconductor.org/ (Drăghici 2011) for more details on the Bioconductor. There are many studies in which various Bioconductor packages have been used to analyze microarray data. For instance, in the expression study of the Dof1 transcription factor in wheat and sorghum, the Affy package from Bio conductor was used for to normalize microarray data (Peña et al. 2017). Furthermore, for gene expression study of wheat leaves infected by Xanthomonas translucens, the DESeq2 package of Bioconductor was used to identify differentially expressed genes based on the Negative Binomial distribution. The Bioconductor q-value package was also used for p value correction (Garcia-Seco et al. 2017).

10.1.9.3 Gene ARMADA

For both cDNA and oligonucleotide (Affymetrix) microarray data, Gene ARMADA can provide a complete, open-source, flexible and handy platform. It was implemented in the program for MATLAB. Gene ARMADA is an independent platform that can be used either as a MATLAB tool or as an application on its own. This software specializes in the visualization, standardization and statistical testing of data. Gene ARMADA has been successfully used to process several datasets of microarrays (http://www.grissom.gr/armada) (Chatziioannou et al. 2009).

10.1.9.4 Babelomics

Babelomics is an integrative web-based platform that includes a complete suite of methods for the analysis of gene expression data i.e. normalization, pre-processing, gene expression analysis, predictors, clustering and large-scale genotyping assays. Currently, Babelomics has an average of more than 200 experiments analyzed per day, (http://bioinfo.cipf.es/webstats/babelomics/awstats.babelomics.bioinfo.cipf.es.html), distributed among many different countries (http://bioinfo.cipf.es/toolsusage). The current version of Babelomics (Babelomics 5.0) is freely available at: http://www.babelomics.org (Alonso et al. 2015). In a comparison study of the transcriptomic and metabolomic profiles of six rice cultivars leaves under high night temperature conditions, the background correction of microarray signal intensities and differential expression investigation were performed by class comparison methods in Babelomics tool (Glaubitz et al. 2017).

10.1.9.5 Maanova

MAANOVA refers to the Micro Array Variance Analysis. MAANOVA is suitable for the analysis of microarray experiments on both small and large scale. MAANOVA, implemented in Matlab, is the statistical language R add-on package. Friendly, to run this software on any platform that supports these packages. This package provides a complete workflow for different aspects of microarray data analysis i.e. data quality checks, ANOVA model fitting (both fixed and mixed effect models), statistical testing (F and Fs statistics), p-value (using sampling and residual shuffling permutation approach) and summarizes the results in tables and graphics including. Volcano plot and bootstrapping-based tree cluster. Functions in MAANOVA have been developed and tested in Matlab Release 12 for Windows and Linux Redhat 7.0. Executable software and source code of mannova can be downloaded from the link http://churchill.jax.org/software/archive/maanova.shtml.R/maanovapackage is also available from http://churchill.jax.org/software/rmaanova.shtml (Wu et al. 2011). This software has been used in different plant species for microarray data analysis (Calla et al. 2009).

10.1.9.6 HD BStat

HD BStat (High-Dimensional Biology-Statistics) is a package for data analysis of microarrays. It was initially developed to analyze data on gene expression, but proteomics and other aspects of genomics can also be used. This software uses a variety of methods to analyze microarray data for standardization, transformation, statistical, and quality control analysis. HD BStat can also perform a test of hypothesis. The results of preprocessing methods of data, analysis of quality control and test methods of hypothesis can be displayed in the form of Excel CSV tables, graphs and Html. HDBStat, please! It is freely available platform-independent software. The link for more details http://www.ssg.uab.edu/hdbstat/documentation.html is addressed to readers. This software has been used in a study with respect to gene expression in tomato (Windram et al. 2012).

10.1.9.7 Expander

Another useful integrated software platform for analyzing data on gene expression is EXPANDER (Expression Analyzer and Displayer). It is intended to support data preprocessing and standardization and identification of differentially expressed genes, clustering; downstream enrichment analyzes of (GO functional categories, TF binding sites in promoter regions, 3’-UTR micro RNA sites and biological pathways and chromosomal locations) and network-based expression data analysis. Expander operates on platforms such as Windows and Linux and provides analysis for various types of organisms such as humans, animals, pests, plants and microorganisms. This package is available free of charge for academic use at http://acgt.cs.tau.ac.il/expander/ (Ulitsky et al. 2010). In a case study, Expander software was used for the hierarchical clustering of transcriptomic data of *Lotus japonicus* (Regel) Larsen cv. Gifu (B-129-S9) (Pérez-Delgado et al. 2016).

10.2 Integration of Transcriptomics, Proteomics and Metabolomics

Proteomics, defined as a high-throughput protein study, has taken the lead in plant biological research and response to stress, particularly due to the growing number of plant genomes being sequenced and released (Jorrín-Novo et al. 2015). Additionally, advances in mass spectrometry (MS), quantitative methods and bioinformatics approaches have enabled a wide range of proteins from specific organ/tissue/cells to be identified, quantified, validated and characterized (Glinski and Weckwerth 2006). The information obtained through these advanced approaches is useful for the deciphering of protein structure and complex mechanisms such as enzymatic and regulatory mechanisms functions of proteins coded by specific genes (Abdallah et al. 2012). In addition, proteomics approaches provide valuable information such as levels of protein associated with stress tolerance, changes in stressed proteomes that link transcriptomics and metabolomics analyzes, as well as the role of expressed genes in functionally translated genome regions associated with traits of interest (Kosová et al. 2011). Many proteomics-based publications, particularly related to plant development and other biological phenomena such as leguminous symbiosis, can be found in model legumes and Arabidopsis thaliana, as well as in some plants such as rice (*Oryza sativa*), *Triticum aestivum*, *Zea mays*, *Solanum lycopersicum* and *Nicotiana tabacum* (Jorrín-Novo et al. 2015).

The evolution of the plant immune response has resulted in a highly effective defense system that can withstand microbial pathogens from potential attacks. The primary immune response is known as pathogen-related molecular pattern (PAMP) triggered immunity and has evolved to recognize common characteristics of microbial pathogens (Janeway and Medzhitov 2002). In response to pathogen effector protein delivery, plants acquired Resistance (R) proteins to combat pathogen attack. R-dependent defense responses are important in understanding biochemical and cellular mechanisms and underlying these interactions will allow increased molecular and transgenic approaches for crops. A new research area, i.e. the analysis of more complex microbial communities and their interaction with plants, has been initiated by recent developments in the field of proteome analysis. Such areas have great potential to elucidate not only the interactions between bacteria and their host plants, but also the interactions between bacteria and bacteria between various bacterial taxa, symbiotic, pathogenic and commensal bacteria. Plant hormonal signaling pathways give priority to defense over other cellular functions during biotic stress. Some plant pathogens use the hormone-dependent regulatory system to mimic hormones that interfere with the immune response of the host to promote virulence (vir) (Woodward et al. 2010). The majority of bacteria are exposed to a constantly changing physical and chemical environment, unlike plant and animal cells. Phylogenetic diversity of plant-associated bacteria (PAB) can categorize them into commensals (acquire plant nutrients without harm), mutualists (influencing plant health positively) and pathogens (damaging plant). Notably pathogenic bacteria, commensals, or mutualists have developed strategies for interacting with overlapping plants, an exceptionally modified physiology that reflects individual needs (Ozinsky et al. 2000). Bacteria respond to changes in their environment by adapting structural protein patterns, transporting proteins, toxins and enzymes which adapt them to a particular habitat (Torres et al. 2004). Enzymes are either constitutive (always produced by cells independently of the medium's composition) or inducible (produced in cells in response to a pathway's end product). Regulation of enzyme activity, which is primarily used to regulate biosynthetic pathways and repression of catabolites is considered a form of positive control as it affects an increase in protein transcription rates. Plant immunity recognizing membrane protein pathogens is referred to as pattern recognition receptors (PRRs), which recognize pathogen-associated molecular pattern (PAMP) and is the basis of plant-inborn immunity(Witte et al. 2012). PAMP recognition also results in plant systemic acquired resistance and resistance (R) protein production leading to effector-triggered immunity (ETI), often accompanied by hypersensitive response (HR), and cell death programmed. Many R genes that confer resistance to various pathogens, including viruses, bacteria, fungi, or nematodes, have been isolated over the past 10 years. These R-gene products are divided into intracellular protein kinases (Pto), proteins with an extracellular leucine-rich repeat (LRR) domain and a cytoplasmic protein kinase region (e.g. Xa21), intracellular proteins containing a region of LRRs and a nucleotide bin, based on predicted protein sequences (e.g., Cf-4, Cf-9) (Zhang et al. 2012). Proteomic analyzes made it possible to analyze complex microbial communities that had great potential to elucidate not only the interactions between bacteria and their host plants, but also the interactions between bacteria and bacteria. For various PABs, proteomic reference data sets were established using two-dimensional polyacrylamide gel electrophoresis (2-DE) gels, resulting in a few hundred identified proteins or multi-dimensional liquid chromatography-tandom mass spectrometry (LC-MS/MS) techniques leading to the detection of over 1000 proteins (Anderson et al. 2006). Era is followed by gel-free proteomics, but before gel-based proteomics, quantitation procedures must be optimized before the gel-based proteomics can be replaced by gel free procedures. Complete genome sequence of a *Xylella fastidiosa* is available which can be very helpful in genomics and proteomic studies of plant-bacterium interactions (Bagnarol et al. 2007). In order to understand the molecular signaling pathways involved in plant-bacterial interactions, more genomic data are needed for pathogenic and symbiotic bacteria. Because of the agricultural importance and intensity of scientific research, P. syringae and Xanthomonas campestris are important PABs. On the model plant Arabidopsis, both are pathogenic (Andrade et al. 2008). There has been extensive study of pathogenic and mutualist

PAB (Jacobs et al. 2012). The mutualism process involved a significant change in the metabolism of the mutualists as well as the host, which involves a change in the metabolism of plant cells to support the mutualist's ATP synthesis and nitrogen fixation for nodule development (Delmotte et al. 2010). Transcriptomics data shows that pathogenic bacteria involve the hypersensitive reaction and pathogenicity (hrp) gene and different secretion systems (SS) for colonization and damaging host cells (Buttner and Bonas 2002). They typically exchange signals with their hosts and have a range of specific plant colonization adaptations. To understand the molecular mechanism by which bacteria adapt to live in association with plants for symbiosis and pathogenesis evolution is explored the importance of proteomics. This will open up new research areas on protein-based plant-microbe communication and provide important information on manipulating gene expression of specific proteins to modify plant behaviour associated with compatible or incompatible interactions. The use of proteomics for crop plant analyzes has increased rapidly over the last decade. While proteomic techniques are routinely used in plant laboratories around the world and are powerful study tools, considerable room for improvement still exists (Komatsu et al. 2013). The fraction of the plant proteome that can be detected using current approaches is significantly lower than that of other "Omics" techniques and therefore does not fully represent the cellular proteins. The predominant technique used for separating proteins is the two-dimensional electrophoresis (2-DE) gel. However, proteome analyzes based on liquid chromatography (LC) are increasing in many common laboratories. Both techniques of protein separation have specific benefits. Protein modification and degradation can be quickly visualized with a standard 2-DE approach, whereas LC-based methods require much lower starting material quantities. The limited availability of genomic information has hindered the application of crop proteomics. However, with the successful development of "next-generation" sequencing technologies, identification and annotation of proteins and their isoforms in a particular crop species is becoming much more straightforward (Komatsu et al. 2013). A specific advantage of proteomics over other "Omics" techniques is the capacity to reveal post-translational modifications (PTMs), which is a prerequisite to determine the functional impact of protein modification on crop plant productivity. To date, proteomic analyzes have identified approximately 300 PTMs. However, major efforts are needed to develop reliable tools and strategies to assess the impact of this growing number of different crop PTMs. Lastly, crop proteomics should become an essential part of integrated "Omics" approaches. A major challenge for crop proteomics, however, will be to keep pace with other "Omics" techniques' throughput capacity. Advances in plant phenotyping will benefit the application of proteomics for plant functional analysis. In particular, improved techniques for automated, non-invasive plant collection phenotyping will help in the selection of appropriate genotypes for proteomics-based functional analyses aimed at characterizing the relevant traits for future crop breeding (Wang et al. 2013).

Numerous proteins that play crucial roles in plant growth and development have been identified in proteomic studies. However, it is a major challenge to determine how this wealth of information can be applied to agriculture and artificial crop regulation. As seed viability is related to crop yields, seed is one of the most important factors in crop production. He and Yang (2013) applied proteomics to the study of rice seed germination regulation and demonstrated that starch is degraded in endosperm and subsequently biosynthesized in the embryo during germination, a process that appears to promote the gradual use of nutritional reserves. Wide spread use of heterosis in crop production where sterile male line is critical for hybrid breeding. Identifying the proteins involved in the regulation of male sterility represents a major target in crop proteomic studies (Wang et al. 2013). Unlike traditional breeding methods, transgenic techniques are becoming increasingly popular to obtain crops with desired qualities quickly. It is essential to evaluate these GM crops using proteomic methods (Gong and Wang 2013). Maintaining food safety is a serious challenge worldwide due to impending changes in global climate and ongoing industrialization. Effective methods to increase the efficiency of sunlight conversion are needed to sustainably feed the world population (Driever and Kromdijk 2013). In light conversion, C4 plants are more efficient than C3 plants because they contain two different chloroplasts. Comparative proteomic analyses of C4 chloroplastsmight help to determine the key components that influence the efficiency of sunlight conversion (Manandhar-Shrestha et al. 2013). An important factor that influences crop growth and eventual crop yield is the interaction between crops and other organisms. For example, the pathogen *Fusarium graminearum* causes small grain cereal head blight and dramatically reduces grain yield and quality, which has a major economic impact on the cereal industry. Proteomic analysis is expected to complement traditional approaches to molecular genetics to study the mechanisms by which this pathogen attacks cereal crops(Yang et al. 2013). It has also been demonstrated the use of proteomics to analyze the interaction between crops and bacteria, especially the symbiotic interactions in legume root nodules. Most studies have been carried out on whole organs or tissues that do not allow spatial information to be collected. The use of MS imaging techniques, which have been successfully applied in the field of medicine, is therefore expected to help in obtaining information on the spatial distribution of metabolites and proteins (Matros and Mock 2013). In addition to proteomics, metabolomics is another important approach to functional genomics in which the identification and quantification of metabolomes (collection of metabolites or small molecules) within a cell, tissue or organism produced through cellular metabolism connects the cellular biochemical activity with the phenotype. Major approaches to plant metabolomics include metabolic fingerprints, metabolite profiling, and targeted analysis (Weckwerth 2003). Different metabolomic approaches or a combination of approaches are applied depending on the study objective. In addition, the use of MS, bioinformatics tools and software enables rapid measurement of metabolites at the same time, which are spatially localized within the biological material (Bhalla et al. 2005). As metabolites are closer to the phenotype, they more accurately reflect gene expressions and various regulatory processes, and metabolomics is a powerful tool for studying molecular phenotypes of plants in response to stress. For instance, the plant metabolism is affected under abiotic stress conditions due to factors such as metabolic enzyme inhibition, substrate shortage, extreme demand for specific compounds and a combination of these factors. The plant is thus subjected to metabolic reprogramming to adapt to the predominant stress conditions by producing anti-stress components such as compatible solutes, antioxidants and stress-responsive proteins (Wienkoop et al. 2008). The use of metabolites as selection biomarkers has been of great interest in crop breeding programs, as metabolites integrate the complex interaction between genotype and environment (Wienkoop et al. 2008). With proteomics and metabolomics emerging as state-of – the-art functional biology disciplines for understanding plant adaptation mechanisms to stresses in different plant systems at cellular and developmental stages, there was great interest in applying knowledge to understand responses in different crop plants. These approaches, integrated with data obtained from genomics, enable accurate identification of candidate genes and pathways involved in important agronomic traits that can be applied in crop breeding programs (Langridge and Fleury 2011). PAMPs are the first layer of plant innate immunity and failure to recognize them may lead to increased susceptibility to disease. PAMPs are ideal elicitors for "non-self" surveillance systems such as chitin, ergosterol and fungal transglutaminase, and/or bacterial lipopolysaccharides and flagellin, which stimulate PAMP receptors encoded by plants (Chisholm et al. 2006). Intracellular responses related to PAMP-triggered immunity (PTI), including rapid ion fluxes across the plasma membrane, kinase activation of mitogen-activated protein (MAP), production of reactive oxygen species (ROS), and rapid changes in gene expression and reinforcement of the cell wall. Suppression of PTI can be achieved through the pathogens 'secretion of virulence (vir) effectors or plant signaling suppression. ETI is accompanied by R protein or HR production, which illustrates the dynamic co-evolution between plants and pathogens (Jones and Dangl 2006). Flagellin, elongation factor (EF) Tu, peptidoglycan, lipopolysaccharide, and bacterial cold shock proteins are important PAMPS and their induced plant responses are called "basal" defenses. Once the highly preserved amino terminus of flagellin (flg22) is recognized, flagellin sensing 2 (FLS2) induces a series of defense responses, including MAP kinase signaling, transcriptional activation and callose deposition a putative physical barrier at the site of infection (Gomez-Gomez et al. 1999). EF Tu potent bacterial PAMP in *Arabidopsis* and other members of the Brassicaceae family, serves as an adhesion factor at the bacterial surface, in addition to its primary role in translation. Asparate oxidase is required for PAMP-triggered RBOHD-dependent (responsible for stomatal closure) ROS burst and stomatal immunity against the *P. syringae* (Macho et al. 2012). The LRR receptor kinases, EF-Tu receptor and FLS2 are PRRs, contributing to disease resistance against the hemibiotrophic bacterium *P. syringae* (Roux et al. 2011). The plant hormones, salicylic acid (SA), jasmonic acid (JA) and ethylene, have emerged as key players in the signaling networks involved in plant immunity. Rhamnolipids are glycolipids produced by bacteria and are involved in surface motility and biofilm development and are considered as PAMPS. Ethylene is found to be involved in rhamnolipid-induced resistance to *H. arabidopsidis* and to *P. syringae* whereas JA is essential for the resistance to *B. cinerea*. SA participates in restriction of all bacterial and fungal pathogens, so involving in broadly rhamnolipid-mediated immunity (Sanchez et al. 2012). PAMPS are sometimes succeeded and sometimes fails to induce PTI depending upon the type of compatible and non-compatible interactions. Flagellin is capable of suppressing HR via PTI induction during an incompatible interaction (Wei et al. 2012). Type III secretion system (T3SSs) were essential components of two complex bacterial machineries: the flagellum, which drives cell motility and the non-flagellar T3SS (NF-T3SS), which delivers effectors into eukaryotic cells. *P. syringae* use T3SS to deliver up to 40 effector proteins into host cells, inhibiting basal host defense responses, such as HR. PAMP induced PTI serves as a primary plant defense response against microbial pathogens, with MAP kinase cascade downstream of PAMP receptors. LRR-RLKs including PSKR1 act as PTI against pathogenic bacteria, and plants expressing this gene show enhanced PAMP responses and less lesion formation after infection with the bacterial pathogen *P. syringae* via jasmonate signaling pathway (McCann and Guttman 2008). Peptidoglycan, an important PAMP from *Staphylococcus aureus* results in PTI, such as medium alkalinization, elevation of cytoplasmic calcium concentrations, nitric oxide, and camalexin production, and the post-translational induction of MAP kinase activities. PAMP recognition also results in plant systemic acquired resistance and production of R proteins such as SUMM2 that becomes active when the MAP kinase cascade is disrupted by pathogens, leading to ETI (Zhang et al. 2012). In rice, the LRR-RK Xa21 confers resistance to *Xanthomonas oryzae* pv. oryzae strains carrying the Avr gene AvrXa21. AvrXa21 as a type I secreted sulfated peptide, is conserved among all *Xanthomonas* strains sequenced (P. Ronald, pers. communication), suggesting that *AvrXa21/Xa21* constitutes a PAMP/PRR perception system (Lee et al. 2008). Although many PAMPs recognized by plants have been described, number of known PRR and PTI is still in its infancy, constituting a highly active and competitive field of research. Protein analysis of asscoaited bacteria (PAB) either they are pathogenic or symbiotic bacteria adhere to plant surfaces, invade the intercellular space of the host tissue, counteract plant defense systems and acquire nutrients. However either there is establishment of a pathogenic interaction or mutualist relationship develops. Cell surface proteins such as adhesions, polysaccharides, lipopolysaccharides, and degradative enzymes enable the degradation of the plant cell wall and also result in basal plant defenses (Newton et al. 2010). Proteins of PAB are studied either in planta, by means of bacterial responses to selected biomolecule or plant extracts, synthetic media, or secretome analysis to study the vir factor of the bacterial pathogens (Gourion et al. 2006). Analysis of proteomas is very tricky when dealing with bacteria separation from infected plants and additional steps are needed to avoid changes in the map of proteomas. Bacterial separation protocols for density gradient centrifugation using percoll or saccharose gradients had been proposed(Gourion et al. 2006). The transcriptomics profile of R was discussed by Jacobs et al. (2012). Solanacearum *in vitro* and the importance of T3SS in the Ralsotonia Vir Cascade (45% HR and Hrp gene regulated). Pathogenic bacteria X proteome analysis. Pv campestris. Campestris in conjunction with *B. Oleracea* and savastanoipv Pseudomonas. Savastanoi led to a comprehensive analysis of expression analysis including stress and metabolic proteins (Andrade et al. 2008). Increased protein levels associated with xanthan biosynthesis, stress response, and induced metabolism in X. Unlike in vitro grown cells, campestris in plant conditions Chaperonin is reported to be involved in stress responses, and EF, which acts in plants as an important PTI, is the key component of bacteria's translation machinery. Xanthan is an extracellular polysaccharide likely to cause disease symptoms in planta growth through the mucoid appearance of bacterial colonies and host plant wilting by blocking water flow in xylem vessels (Andrade et al. 2008).

Methylobacterium extorquens differential proteome analysis of 45 metabolic proteins and proteins involved in stress response such as extracellular protease, SOD, catalases, and DNA protein (Gourion et al. 2006) resulted in the identification of 45 metabolic proteins and proteins involved in stress response. The protein analysis of symbiosis-living cyanobacteria revealed several adjustments to a symbiotic lifestyle, including an increase in proteins involved in the production of energy and fixation of nitrogen. On the other hand, under symbiotic conditions, proteins involved in photosynthesis were reduced, pointing to a heterotrophic lifestyle. Bacteroids 'general proteome analysis is compared with *in vitro* grown cells in order to identify nodule specific adaptations, over time or when plants were exposed to drought stress (Nomura et al. 2010). ABC-type transporters was present in nodule bacteria for transport of amino acids and inorganic ions along with proteins involved in vitamin synthesis, fatty acid, nucleic acid, cell surface synthesis, and stress-related processes. Integrated proteomics and transcriptomics data for *B. japonicum* bacteroids resulted in 2315 proteins involved in carbon and nitrogen metabolism, including a complete set of tricarboxylic acid cycle enzymes, gluconeogenesis and pentose phosphate pathway enzymes, along with other proteins important in symbiosis. Amino acids (Asn, Gln, Pro), organic acids (threonic acid), sugars (Rib, maltose), and polyols (mannitol) were reported to be more abundant in symbiotic roots (Delmotte et al. 2010).

Plant metabolites are involved in many responses to resistance and stress and also contribute to fruit and flowers colour, taste, aroma, and scent. Since an organism's biochemical phenotype is the final result of genotype-environmental stimuli interactions, it is also modulated by intracellular physiological fluctuations that are part of homeostasis. Therefore, it is necessary to simultaneously identify and quantify metabolites to understand the metabolome dynamics analyze fluxes in metabolic pathways and decipher the role of each metabolite after different stimuli. Metabolomics 'challenge is to find changes in biochemical pathways and metabolic networks that may correlate with a cell, tissue, or organism's physiological and developmental phenotype (Bino et al. 2004). The completion of the entire genome sequences of model plants like Arabidopsis thaliana and rice is one of the greatest achievements of plant biology. In Arabidopsis ~27,000 genes were predicted based on information about nucleotide sequence; however, only half of these genes were functionally annotated based on sequence similarity to known genes, and only ~11% of these genes were confirmed with direct experimental evidence (Pichersky and Gang 2000). Therefore, elucidating the function of unknown genes is currently a major challenge in plant research. Because there is very little information about the number of genes in a specific gene family of a non-model plant, the profile of expression of these genes under different conditions and stimuli becomes necessary. Integrating metabolomics with transcription profiles can provide clues for identifying the functions of the unknown genes, irrespective of whether they are from model or non-model plants (Fiehn 2001). Plants produce over 200,000 metabolites, many of which have specific roles in adapting to specific ecological niches (Fiehn 2001). Therefore, the main problems encountered when characterizing the metabolome of the plant are due to the fact that the metabolome is highly complex in nature compared to the proteome or transcriptome due to the huge chemical diversity of the compounds. Additionally, there is a wide range of concentrations of metabolites that can vary in magnitude over nine orders (pM to mM). These wide variations in the nature and concentration of analytes to be studied pose challenges to all the analytical technologies used in metabolomics strategies. Using metabolomics, it is possible to identify pathways that are responsible for producing important food metabolites that could be important to human health improvement. There are several examples where the modification of some metabolic pathways has resulted in plant production with an increased nutritional value. This is the case with Golden Rice (GR), a genetically modified rice accumulating β-carotene in endosperm (Ye et al. 2000). Production of this rice variety has helped alleviate vitamin A deficiency, a major global nutritional issue. GR’s nutritional value was subsequently enhanced by the overexpression of a phytoene synthase gene leading to obtaining GR2 variety, which accumulates higher amounts of carotenoids (84% of the total is β-carotene) (Paine et al. 2005). Mehta et al.(2002) expressed a S-adenosylmethionine decarboxylase gene in tomato under the inducible E8 promoter. The transgenic variety shows higher levels of various polyamines during fruit maturation, including spermidine and spermine, leading to an increase in metabolite lycopene, which prolonged the life of the vine and increased fruit juice and nutrient quality(Mehta et al. 2002). Other examples include plant engineering to improve anthocyanin content. Anthocyanins are flavonoids, a pigment class that contributes to the plant’s colors and antioxidant properties. In addition, these metabolites were linked to protection against several human diseases, but their natural levels in plants are inadequate to confer optimal benefits. It has also been reported that the expression of two transcription factors in tomatoes has resulted in the accumulation of higher anthocyanin concentrations at concentrations comparable to those founded in high antocyanin-containing plants such as blackberries and blueberries (Yeager 1927). The new variety has an intense purple coloring and an increased antioxidant capacity of 3 times. Plant metabolomics is increasingly being used to understand other processes like cellular responses to stress conditions. An example of this is the metabolic adjustment to sulfur deficiency. There was a close relationship between the metabolism of sulfur, nitrogen, lipids and purine metabolism and enhanced photorespiration. Metabolomics has also been applied to the study of the cold stress response (Blake-Kalff et al. 1998 and Miyagi et al. 2010). Other applications include metabolic engineering of biochemical pathways, gene function discovery, and engineering pathways for pharmaceuticals production.

10.3 Omic Plant Development Database Approaches

Technological advances in each research area of omics have become essential resources for gene function research in association with phenotypic changes. Some of these advances include the development of high-throughput methods to profile thousands of gene expressions, identify modification events and interactions in the plant proteome, and simultaneously measure the abundance of many metabolites. Furthermore, large-scale bioresource collections such as mass-produced mutant lines and full-length cDNA clones and their integrative relevant databases are now available (Brady and Provart 2009). Arabidopsis thaliana's entire genome sequencing was completed in 2000 (The Arabidopsis Genome Initiative 2000). Subsequently, the National Science Foundation (NSF) Arabidopsis 2010 project in the USA was launched with the stated goal of determining the functions of 25,000 genes of Arabidopsis by 2010. The genome sequencing project of *japonica* rice was completed in 2005, and the Rice Annotation Project (RAP), which was orchestrated via 'jamboree-style' annotation meetings, aimed to provide an accurate annotation of the rice genome (International Rice Genome Sequencing Project 2005, Itoh et al. 2007). In conjunction with the rice genome sequence and its related genomic resources, advanced development of mapping populations and molecular marker resources has allowed researchers to accelerate the isolation of agronomically important quantitative trait loci (QTLs) (Ma et al. 2007).

The aforementioned recent advances in high-throughput technology have provided opportunities for specific organisms to develop collections of sequence-based resources and related resource platforms. Each biological element comprehensively measured by a high-throughput method is depicted in a corresponding plane in a conceptual model with layers ranging from genome to phenome, a model called 'omic space'. Such comprehensive models often provide an excellent starting point for experiment design, hypothesis generation, or conceptualization based on the integrated knowledge found in a particular organism's omic space. Such comprehensive models often provide an excellent starting point for experiment design, hypothesis generation, or conceptualization based on the integrated knowledge found in a particular organism's omic space. In addition, the development of such omic resources and data sets for different species allows for the comparison of omic properties between species, which promises to be an effective way of finding collateral evidence for preserved gene functions that could be evolutionarily supported. Bioinformatics platforms have become essential tools for accessing omics data sets for efficient mining and biologically important knowledge integration (Table 10.1).

Examples of resources related to each omics instance are presented as the model plant in Arabidopsis, rice and soybean, as well as a monocotype model and a sequenced crop, and as an important crop recently sequenced. These resources can be accessed from the above-mentioned URLs or links (Mochida et al. 2011).

An overview of several representative resources available for use in omics plant research is given above, with particular emphasis on recent progress related to crop species in addition to sequence related resources such as whole genome, protein coding and non – coding transcripts, and updates of sequencing technology.

Table 10.1 Omic space and related resources in plants

1. http://www.arabidopsis.org/,
2. http://www.gramene.org/,
3. http://soybase.org/,
4. http://nazunafox.psc.database.riken.jp,
5. http://rarge.gsc.riken.jp/dsmutant/index.pl,
6. http://signal.salk.edu/tabout.html
7. http://tilling.fhcrc.org/,
8. http://www.postech.ac.kr/life/pfg/risd/,
9. http://tos.nias.affrc.go.jp/,
10. http://www.soybeantilling.org/psearch.jsp,
11. http://mulch.cropsoil.uga.edu/~parrottlab/Mutagenesis/acds/index.php,
12. http://arabidopsis.org.uk/home.html,
13. http://abrc.osu.edu/,
14. http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp,
15. http://www.irri.org/grc/GRChome/home.htm,
16. http://www.legumebase.agr.miyazaki-u.ac.jp/index.jsp,
17. http://www.plantcyc.org:1555/ARA/server.html,
18. http://pathway.gramene.org/gramene/ricecyc.shtml,
19. http://www.plantcyc.org/,
20. http://mediccyc.noble.org/,
21. http://prime.psc.riken.jp/,
22. http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html,
23. http://ppdb.tc.cornell.edu/,
24. http://phosphat.mpimp-golm.mpg.de/,
25. http://cdna01.dna.affrc.go.jp/RPD/main_en.html,
26. http://proteome.dc.affrc.go.jp/Soybean/,
27. http://oilseedproteomics.missouri.edu/,
28. http://bioinfo.esalq.usp.br/cgi-bin/atpin.pl,
29. http://atpid.biosino.org/,
30. http://suba.plantenergy.uwa.edu.au/, 32. http://proteomics.arabidopsis.info/,
31. http://www.brc.riken.go.jp/lab/epd/catalog/cdnaclone.html,
32. http://rarge.gsc.riken.jp/,
33. http://cdna01.dna.affrc.go.jp/cDNA/,
34. http://rsoy.psc.riken.jp/,
35. http://www.arabidopsis.org/portals/expression/microarray/ATGenExpress.jsp,
36. https://www.genevestigator.com/gv/index.jsp,
37. http://bioinformatics.med.yale.edu/riceatlas/,
38. http://bioinformatics.towson.edu/SGMD/Default.htm,
39. http://soyxpress.agrenv.mcgill.ca/cgi-bin/soy/soybean.cgi,
40. http://mpss.udel.edu/at/,
41. http://mpss.udel.edu/rice/,
42. http://signal.salk.edu/,
43. http://rapdb.dna.affrc.go.jp/,
44. http://rice.plantbiology.msu.edu/,
45. http://www.phytozome.net/,

(continued)

Table 10.1 (continued)

46. http://walnut.usc.edu/,
47. http://www.oryzasnp.org/,
48. http://www.soymap.org/,
49. http://1001genomes.org/,
50. http://rarge.gsc.riken.jp/rartf/,
51. http://arabidopsis.med.ohio-state.edu/,
52. http://datf.cbi.pku.edu.cn/,
53. http://drtf.cbi.pku.edu.cn/,
54. http://grassius.org/,
55. http://soybeantfdb.psc.riken.jp,
56. http://legumetfdb.psc.riken.jp/

Comprehensively collected sequence data provide essential genomic resources to accelerate molecular understanding of biological properties and to promote the use of such knowledge. The recent accumulation of model plant nucleotide sequences, as well as applied species such as crops and domestic animals, has provided basic information for the design of functional genomics sequence-based research applications. Species-specific collections of nucleotide sequences also offer opportunities to identify the genomic aspects of phenotypic characters based on genome-wide comparative analysis and model organism knowledge (Tanaka et al. 2008).

10.3.1 Genome Sequencing Projects

The first genome sequence of a plant was completed for *A. thaliana*, which is now used as a model species in plant molecular biology due to its small size, short generation time and high efficiency of transformation. The Arabidopsis genome sequence project was performed as a cooperative project among scientists in Japan, Europe and the USA (Bevan 1997). The genome sequencing was completed and published in 2000 by the Arabidopsis Genome Initiative (AGI) (The Arabidopsis Genome Initiative 2000). The draft genome sequence of rice, *japonica* and *indica*, an important staple food as well as a model monocotyledon, was published in 2002 (Goff et al. 2002). Subsequently, the genome sequence of *japonica* rice was completed and published by the International Rice Genome Sequencing Project in 2005 (International Rice Genome Sequencing Project 2005).

There are a number of providers for sequences and annotations of plant genomes. Phytozome is a web-accessible resource of information that provides genome sequences and annotations of different plant species. This resource is a joint project of the Joint Genome Institute (DOE–JGI) of the Department of Energy and the Center for Integrative Genomics to facilitate comparative genomic studies among green plants (http://www.phytozome.net/Phytozomeinfo.php). Phytozome's current version (ver. 5.0, January 2010) consists of 18 plant species sequenced by JGI and other sequencing projects. Gramene (http://www.gramene.org/) is a website-based information resource for grass species, and it provides various kinds of information related to grass genomics, including genome sequences (Liang et al. 2008). The current version of Gramene (#30, October 2009) provides genome sequence information for 15 plant species, including five wild rice genome assemblies. The release of sequenced genomes is expected to accelerate with ongoing innovations in sequencing technologies of the next generation (Ansorge 2009). Whole-genome sequence information allows us to derive sets of important genomic features including identification of protein coding or non-coding genes and constructs such as gene families, regulatory elements, repetitive sequences, simple sequence repeats (SSRs) and guanine–cytosine (GC) content. These data sets have become the primary sequence material for designing genome-based sequence platforms such as microarrays, tiling arrays or molecular markers, as well as reference data sets for integrating omics elements into a genome sequence. Chromosome-scale comparisons identifying conserved similarities of gene coordinates facilitate documentation of segmental and tandem duplications in related species (De Bodt et al. 2005). Whole-genome comparisons identifying chromosomal duplication and conserved synteny among related species provide evidence for hypotheses on comparative evolutionary histories with regard to the diversification of species in a related lineage (Paterson et al. 2009). ESTs are created through the partial one-pass sequencing of randomly selected gene transcripts converted into cDNA (Adams et al. 1993). Since cDNA and EST collections can be acquired irrespective of genomic complexity, due to polyploidy and/or their number of repetitive sequences this approach has been applied not only to model species but also to a number of applied species with large genome sizes. As of November 2009, dbEST, a public domain EST database (http:/www.ncbi.nlm.gov/dbEST/), which includes a number of plant species, has >63 million ESTs in the National Center for Biotechnology Information (NCBI) (Table 10.2) (Boguski et al. 1993).

Since EST data collected from a particular organism's cDNA libraries consist of redundant sequence data derived from the same gene locus or transcription unit, it is often necessary to perform EST grouping by transcription units and assemble these groups to create a consolidated alignment and representative sequence of each transcript prior to further analysis. Such steps are carried out in a computational manner: a typical workflow consists of 'base calling,' i.e. converting the output trace of a sequencer to identified nucleotide data, followed by a cleaning step involving identification and removal of contaminated sequences, masking of cloning vector sequences, clustering of contaminated sequences, the masking out of cloning vector sequences, clustering of identical sequences and alignment of clustered sequences (Masoudi-Nejad et al. 2006). Then, the obtained data sets of representative transcripts can be used as unified transcript data. There are several data resources that provide such unified data sets of plants, such as NCBI-UniGene, PlantGDB, TIGR Plant Gene Index and HarvEST (Duvick et al. 2008).

In addition to the mass volume data sets of their sequence tags, the comprehensive and rapid accumulation of cDNA clones has become significant resources for functional genomics. ESTs derived from various tissue types, including tissues from

Table 10.2 Numbers of ESTs and unified transcripts in plants

Species	No. of ESTs (dbEST)	No. of entries (UniGene)
Physcomitrella patens	*382,584*	*18,870*
Picea glauca (white spruce)	*299,455*	*22,472*
Picea sitchensis (Sitka spruce)	*175,662*	*18,838*
Pinus taeda (loblolly pine)	*328,628*	*18,921*
Aquilegia Formosa×Aquilegia pubescens	*85,039*	*8046*
Arabidopsis thaliana (Thale cress)	*1,527,298*	*30,579*
Artemisia annua (sweet wormwood)	*85,402*	*9462*
Brassica napus (rape)	*643,601*	*26,733*
Brassica oleracea	*59,946*	*5617*
Brassica rapa (field mustard)	*44,570*	*14,497*
Capsicum annuum	*116,541*	*8868*
Citrus Clementina	*118,365*	*9123*
Citrus sinensis (Valencia orange)	*208,909*	*15,808*
Glycine max (soybean)	*1,422,604*	*33,001*
Gossypium hirsutum (upland cotton)	*268,786*	*21,738*
Gossypium raimondii	*63,577*	*3297*
Helianthus annuus (sunflower)	*133,682*	*12,216*
Lactuca sativa (garden lettuce)	*80,781*	*7940*
Lotus japonicas	*195,385*	*14,493*
Malus × domestica (apple)	*324,308*	*23,731*
Medicago truncatula (barrel medic)	*269,237*	*18,098*
Nicotiana tabacum (tobacco)	*317,190*	*24,069*
Populus tremula × Populus tremuloides (hybrid aspen)	*76,160*	*9652*
Populus trichocarpa (western balsam poplar)	*89,943*	*14,965*
Prunus persica (peach)	*79,203*	*7620*
Raphanus raphanistrum (wild radish)	*164,119*	*18,788*
Raphanus sativus (radish)	*83,034*	*17,649*
Solanum lycopersicum (tomato)	*296,848*	*18,228*
Solanum tuberosum (potato)	*236,568*	*18,784*
Theobroma cacao	*159,320*	*24,958*
Vigna unguiculata (cowpea)	*187,443*	*15,740*
Vitis vinifera (wine grape)	*357,856*	*22,083*
Selaginella moellendorffii	*93,806*	*8810*
Hordeum vulgare (barley)	*501,614*	*23,595*
Oryza sativa (rice)	*1,249,110*	*40,978*
Panicum virgatum (switchgrass)	*436,535*	*20,973*
Saccharum officinarum (sugarcane)	*246,892*	*15,594*
Sorghum bicolor (sorghum)	*209,814*	*13,899*
Triticum aestivum (wheat)	*1,067,290*	*40,349*
Zea mays (maize)	*2,018,798*	*97,123*
Chlamydomonas reinhardtii	*204,076*	*11,310*
Volvox carteri	*132,038*	*5638*

Source: (Boguski et al. 1993)

organisms at various stages of development or under stress, could significantly facilitate gene discovery as well as structural gene annotation, large-scale analysis of expression, intra-specific and interspecific genome-scale comparative analysis of expressed genes and the design of expressed gene-oriented molecular markers and probing for microarrays.

10.3.2 Full-Length cDNA Projects

While partial cDNAs are useful in the rapid development of catalogs of expressed genes, they are not suitable for further gene function study. This is because the most popular method of preparing a cDNA library does not provide a complete cDNA that includes the sequence of the capped site. Hayashizaki's RIKEN group developed the biotinylated cap trapper method, which uses reverse transcriptase trehalose-thermostabilized and is an efficient method of building full-length cDNA-enriched libraries about 10 years ago. Full-scale cDNA libraries and large-scale clone data sets have become invaluable resources for life science projects that study different species (Tanaka et al. 2008). The sequence resources derived from full-length cDNAs can also help to identify transcribed regions in completed or draft genome sequences substantially. Full-length cDNA sequences were used in Arabidopsis and rice to identify genomic structural features such as transcription units, transcription starting sites (TSSs) and transcriptional variants (Yamamoto et al. 2009). Full-length cDNA clones were sequenced to help consolidate genomic infrastructure in species for which we have draft genomes, such as Physcomitrella, soybean and poplar; this should also contribute to gene discovery (Umezawa et al. 2008). Also full-length cDNAs are useful for the three-dimensional determination (3D) structures of proteins by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy and for functional biochemical analyses of expressed proteins in the molecular interactions of protein–ligands, protein–proteins and protein–DNAs. In addition, recent advances in proteomics infrastructure require extensive data sets on the full-length amino acid sequences used to assign a protein to peptides. These advances also require functional annotations to support systematic knowledge extracted from proteins to identify peptides and residues modified by, for example, phosphorylation for use in combination with comparative analysis of modified events between species. By creating overexpressors used in reverse genetics, the full-length cDNA library also contributed significantly to functional analysis. Systems such as full-length cDNA overexpressor (FOX) gene hu have developed function-based gene discovery by systems such as full-length cDNA overexpressor (FOX) gene hunting, which use full-length cDNA transgenic plants as overexpressors, has provided an effective approach to high-throughput discovery of functional genes associated with phenotypic changes (Kondou et al. 2009). A full-length enriched cDNA libraries have been constructed for non-sequenced crops or forestry species, such as wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), cassava (*Manihot esculenta*), Japanese cedar (*Cryptomeria japonica*) and Sitka spruce (*Picea sitchensis*), as well as for plant species showing specific biological characters such as salt tolerance in salt cress (*Thellungiella halophila*). These full-length cDNA libraries have been used to identify biological features through comparisons of target sequences with those of model organisms such as Arabidopsis, rice and poplar. These libraries also serve as primary sequence resources for designing microarray probes and as clone resources for genetic engineering to improve crop efficiency (Taji et al. 2008).

10.3.3 Emerging Layers in Plant Omics

10.3.3.1 Plant Epigenome Analysis

The epigenome, the genome-scale properties of epigenetic modifications, has attracted attention as a new area of omics that NGS technology-based solutions have advanced (Schmitz and Zhang 2011). Small RNAs (sRNAs) may direct and mediate epigenetic modifications in plants (Matzke et al. 2009). For the interpretation of genetic information, the epigenomic regulation of chromatin structure and genome stability is crucial (He et al. 2011). NGS-based cytosine methylome (methylC-seq), transcriptome (mRNA-seq) and small RNA transcriptome (small RNA-seq) sequencing in inflorescences of arabidopsis revealed patterns of genome-scale methylation and a direct relationship between the location of sRNAs and DNA methylation (Lister et al. 2008). At the entire plant level, bisulfite sequencing using NGS technologies, the so-called BS-seq, has generated a genome-scale map of methylated cytosine in Arabidopsis (Cokus et al. 2008). These analyzes based on NGS methylome enabled a holistic understanding of patterns of genome-scale methylation at a single base resolution. In addition to DNA methylation, histone N-terminal tail modifications such as acetylation are crucial in plant development (He et al. 2011) and mechanisms of defense (Kim et al. 2008). A genome-wide nucleosome positioning analysis combined with DNA methylation profiles revealed 10 basis periodicities in nucleosome-bound DNA methylation status of nucleosome-bound DNA (Chodavarapu et al. 2010). The Epigenomics of Plants International Consortium web site (https://www.plant-epigenome.org/) provides hyperlinks to plant epigenome data resources. In fact, numerous efforts have been made to acquire epigenome information from plant species (He et al. 2010).

10.3.3.2 Plant Interactome Analysis

Interactions between proteins are essential to nearly all cellular processes. The interactome, a comprehensive set of all protein-protein interactions within an organism, is crucial to our understanding of the cellular system's molecular networks (Morsy et al. 2008). Analysis of interactomes was used to characterize plant cellular functions such as cell cycle, Ca2+/calmodulin-mediated signaling, auxin signaling and membrane protein-signaling interactions in Arabidopsis (Vernoux et al. 2011). Recently, the Arabidopsis Interactome Mapping Consortium presented a proteome-wide binary protein–protein interaction map of Arabidopsis with around 6200 highly reliable interactions interactions between about 2700 proteins (Arabidopsis Interactome Mapping Consortium 2011).

To generate the large-scale Arabidopsis interactome map, the consortium prepared approximately 8000 open reading frames of Arabidopsis protein-coding genes and then analyzed all pairly protein combinations encoded by these constructs using an enhanced binary interactome-mapping pipeline based on the two-hybrid (Y2H) yeast system. A large-scale plant pathogen effector interactome network has been created using the Y2H pipeline method (Mukhtar et al. 2011). In rice, biotic and abiotic stress responses were addressed by a focused interactom analysis (Seo et al. 2011). A number of databases were available on the web for protein-protein interaction data sets (Aranda et al. 2010). In addition to the curated data sets, predicted protein–protein interaction data sets are a valuable complement to experimental approaches (Li et al. 2011).

10.3.3.3 Plant Hormonome Analysis

Plant hormones play a critical role in regulating plant development and environmental responses as signaling molecules. A number of low molecular weight plant hormones, including auxin, ABA, cytokinin, gibberellins, ethylene, brassinosteroids, jasmonates and salicylic acid, have been identified to date (Davies 2004). In addition, strigolactone, a novel plant hormone, has recently been identified as an inhibitor of shooting branching (Umehara et al. 2008). A special issue of strigolactone plant and cell physiology has been published to gather current knowledge on the topic (Yamaguchi and Kyozuka 2010). Small peptides (peptide hormones) also work in plan regulation as signaling molecules in the regulation of plant growth and development (Fukuda et al. 2007). A special issue on peptide hormones was also published to describe recent advances in the area of plant peptide research (Fukuda and Higashiyama 2011).

Many remarkable advances have been made in our understanding of the molecular basis of plant hormones over the past decade, including biosynthesis, transportation, perception and response (Santner and Estelle 2009). Umezawa et al. 2010 reported that ABA response of recent exciting advances in understanding the molecular basis of regulatory networks. A remarkable recent advance was the discovery of receptors for several important phytohormones, including auxin, gibberellins, ABA and jasmonates (Santner and Estelle 2009). Structural analysis of each complex revealed the structural basis of the interaction of each receptor with phytohormone and its signaling mechanisms (Sheard et al. 2010). According to a number of recent mutant analyses, it is almost certain that all plant hormones cross-talk with one or more other hormones depending on the tissue, developmental stage and environmental changes (Depuydt and Hardtke 2011). A comprehensive analysis known as hormonal analysis based on high-throughput, high-sensitivity and simultaneous profiling of plant hormones is a key approach to a holistic understanding of the plant hormone network and its association with biological phenomena because of the interplay between multiple plant hormones. A recently developed analytical platform for high-sensitivity, high-throughput plant hormone measurements enables 43 molecular species of cytokinin, auxin, ABA and gibberellin to be measured simultaneously (Kojima et al. 2009). The platform was used to acquire hormonal profiles of plant hormones in rice organ distribution patterns. The hormonomal analysis of endogenous levels of cytokinin, gibberellin, ABA and auxin in wild type and gibberellin signaling mutants indicated that cytokinin, ABA and auxin metabolism cross talk with the gibberellin signaling system. In Arabidopsis, comprehensive hormone profiling was used to analyze the accumulation of ABA, gibberellins, IAA, cytokinins, jasmonates and salicylic acid in wild-type Arabidopsis seeds and an ABA-deficient mutant. The hormonal approach results suggested that ABA interacts with other hormones to regulate the development of seed (Kanno et al. 2010).

10.3.3.4 Plant Metabolome Analysis

In system approaches to plant functional analysis and applied plant biotechnology, plant metabolomics now plays a significant role. There are many applications for functional genomics, system biology and molecular breeding, driven by advances in related technologies including metabolite measurement instruments, analytical methodologies and information resources. A number of excellent metabolomics reviews describing emerging methodologies and attractive applications have been published (Sumner 2010). Here we will review recent developments in analytical platforms briefly and describe examples of practical applications for understanding plant metabolism systems. Analysis of metabolomes involves chemically diverse compounds. The metabolome of the plant consists of extremely large metabolite varieties with different dynamic ranges of concentration. Consequently, the integration of combined analytical techniques and data set from heterogeneous instruments was key to a comprehensive understanding of various metabolites. Simplified analytical platforms integrating analytical steps such as sample preparation, data acquisition and data analysis allow us to address the complex plant metabolome (Saito and Matsuda 2010). Improved coverage and performance to detect large numbers of metabolites simultaneously expanded the practical application significantly. Ultra-performance liquid chromatography–tandem quadrupole mass spectrometry is used by a widely targeted metabolomics platform that provides coverage and throughput (Sawada et al. 2009a). The approach allows us to simultaneously acquire hundreds or more metabolites of accumulation patterns for large numbers of samples. The platform allows us to address complex metabolic plant systems and develop practical genetic and breeding approaches. For metabolic profiling of the Arabidopsis knockout mutants for methionine chain elongation enzymes, the widely targeted metabolomics approach was used. The results suggest that the metabolism of these enzymes ranges from methionine to primary and related secondary metabolites (Sawada et al. 2009b).

Metabolome profiling provides a snapshot of metabolite accumulation patterns in response to different types of biological conditions such as treatments, tissues, and genotypes. For example, approaches to metabolome profiling were used to monitor changing accumulation of metabolites in response to stress conditions (Kusano et al. 2011). Metabolome profiling was also used to evaluate genetic resources, not only for Arabidopsis and rice model plants, but also for metabolic phenotyping in different crop species (Fujimura et al. 2011). Metabolome profiling was also used to evaluate natural and/or segregation populations 'metabolic phenotypes. Several approaches in metabolite quantitative trait locus (mQTL) analysis have been performed in various plant species in recent years.

10.3.4 Omics Resources in Emerging Plant Species

In a number of plant species, the development of genomic resources has progressed. At Phytozome (ver. 7.0, http://www.phytozome.net/), genome sequence data sets of 25 plant species are available as a typical example. Recent sequenced plant species have typically been nominated for the development of genomic resources because: I they have specific systems not covered by' conventional' model plants; (ii) they are of evolutionary importance; or (iii) they provide a commodity resource such as food and energy.

10.3.4.1 Solanaceae Species

The Solanaceae family includes a number of important crops, including tomato, potato, pepper, paprika, petunia, and tobacco. Tomato (*Solanum lycopersicum*) is a representative crop species for which genomic resources have made significant progress. Tomatoes are an important crop sold fresh and used in processed foods. Because of its small genome size and shared conserved synteny with other Solanaceae genomes, the tomato is a model plant to study Solanaceae species in addition to its agricultural importance. The tomato has also become a model plant to study fruit development, maturation, maturation, and metabolic systems. The International Tomato Genome Sequencing Project was initiated in 2004. Following the initial BAC-by-BAC approach for the euchromatic regions, a whole-genome shotgun approach was initiated in 2009. The International Tomato Annotation Group provided the official annotation of the tomato genome assembly (http://solgenomics.net/organism/Solanum_lycopersicum/genome). A full-length cDNA resource from the tomato cultivar Micro-Tom was recently launched (Aoki et al. 2010; http://www.pgb.kazusa.or.jp/kaftom/). Transcriptome data from 296 samples of 16 series using the Affymetrix GeneChip tomato genome array can be found in NCBI GEO (September 8, 2011). The tomato GeneChip data deposited in NCBI GEO includes, for example, data sets acquired for co-expression analysis using cultivar Micro-Tom (Ozaki et al. 2010), for comparative transcriptome analysis between salt-tolerant and salt-sensitive wild tomato species (Sun et al. 2010) and for examining the transcriptome of the ripening process of an orange ripening mutant (Nashilevitz et al. 2010).

Significant progress has been made with metabolome analyses such as metabolome profiling and mQTL analysis (Enfissi et al. 2010). Metabolome analysis information resources are available and updated, providing data archives for tomato metabolome data sets and analytical platforms such as Plant MetGenMAP, Metabolome Tomato Database (MotoDB) (Moco et al. 2006), KaPPA-View4 SOL (Sakurai et al. 2011) and KOMICS (Iijima et al. 2008). The Tomato Functional Genomics Database provides data on gene expression, metabolites and microRNAs (miRNAs) via a web interface as an integrative information resource. TOMATOMA was launched as a tomato-mutant resource as a web-based database for a phenotypically classified Micro-Tom EMS mutant collection and a Targeting Induced Local Lesions IN Genomes (TILLING) resource (Saito et al. 2011). The potato genome was recently sequenced using a homozygous doubled-monoploid potato clone and

86% assembly of the 844 Mb genome revealed 39,031 protein-coding genes predicted (Xu et al. 2011).

10.3.4.2 Poaceae (Gramineae) Species

The Poaceae family includes staple food crops such as rice, maize, wheat and barley, as well as grasses used for lignocellulose biomass production, such as switchgrass and *Miscanthus* (Lobell et al. 2011). Since the completion of the japonica rice (*Oryza sativa*) genome project (International Rice Genome Sequencing Project 2005), whole-genome sequences have been completed in sorghum (*Sorghum bicolor*), maize (*Zea mays*) and *Brachypodium* (*Brachypodium distachyon*) (International Brachypodium Initiative 2010). Rice is a model species of monocot plants as well as one of the three major staple cereals in the world. To date, the japonica rice genome sequence with high-quality gene annotations has played an important role in promoting the development of a number of genomic resources for the discovery and isolation of important genes for molecular breeding application. The sorghum genome has been sequenced as a representative Saccharinae species that includes starch, sugar and cellulose plants from the source of biomass. Maize is another of the major staple food and feed cereals and is a model organism for basic research into complex heritage and genomic properties such as domestication, epigenetics, evolution, chromosome structure and transposable elements (Walbot 2009). Accompanying the release of the maize B73 genome sequence, the '2009 Maize Genome Collection' was edited (Walbot 2009). Brachypodium is an emerging plant species of the Pooideae subfamily, a model plant for Triticeae crops such as wheat and barley, as well as for grass species understanding systems for the production of cellulose biomass. Brachypodium has attracted attention since the release of the Brachypodium Bd21 genome sequence, and a number of genomic resource projects have been initiated at different institutions (Brkljacic et al. 2011). Therefore, we have access to published reference genome sequences of four species from each of the three major subfamilies of Poaceae. Wheat and barley were the subjects of ongoing attempts to regenerate their highly complex genomes through genome sequencing. A tentative linear order of 32,000 barley genes was recently regenerated by incorporating chromosome sorting, NGS, array hybridization and preserved syntenia with the brachypodium genome (Mayer et al. 2011). To introduce genetic and genomic resources and their applications, a special issue on barley has recently been published (Saisho and Takeda 2011). For several species of Poaceae, data sets of transcriptome profiles are available. For example, Genevestigator's current version provides processed transcriptome data from data sets of GeneChip hybridization (1.626 for barley, 1275 for rice, 1000 for wheat and 458 for maize; https://www.genevestigator.com/gv/). The transcriptome from primordial development through pollination / fertilization to zygote formation throughout the reproductive process was analyzed in rice using an oligomicroarray as an atlas for rice expression (Fujita et al. 2010). Accumulated transcriptome data sets have been made for co-expression analysis of the transcriptome in Poaceae species. The Rice ArrayNet and Oryza Express databases provide web-accessible co-expression data for rice. The ATTED-II database also provides co-expression data sets for rice in addition to those for Arabidopsis. A co-expressed barley gene network was recently generated and then applied to comparative analysis to discover potential Triticeae-specific gene expression networks (Mochida et al. 2011). To analyze transcriptomes of the male gametophyte and tapetum in rice, microarray analyzes coupled with laser microdissection were used (Hobo et al. 2008). Recently, transcriptome data from rice grown in field environments has been collected. For example, the plant proteome database (http://ppdb.tc.cornell.edu/) provides information about the proteomes of maize and arabidopsis as a proteomics resource in Poaceae. The RIKEN Plant Phosphoproteome Database (RIPP-DB, http://phosphoproteome.psc.database.riken.jp) has been updated with a large-scale rice phosphorylated protein identification data set. The OryzaPG-DB was launched as a shotgun-based rice proteome database on proteomics (Helmy et al. 2011).

10.3.4.3 Fabaceae (Leguminosae) Species

The large family of Fabaceae includes economically significant crops such as soybean, common bean and alfalfa. A biological phenomenon, especially in legume species, is the symbiotic nitrogen fixation produced by the communication between plants and nitrogen-fixing bacteria. For plant and microbial biology as well as for agriculture, understanding this symbiosis is important. Recent progress in plant-microbe symbiosis research, including nitrogen-fixing symbiosis in legume plants, was presented in a special issue on plant-microbe symbiosis by Ikeda et al. (2010) and Kouchi et al. (2010) and (Kawaguchi and Minamisawa 2010). *Lotus japonicus* and *Medicago truncatula* served as models for molecular genetics and functional genomics studies to investigate the symbiotic system and carry out gene discovery in legume species (Stacey et al. 2006). In 2008, the genome sequence of *L. japonicus* was released with a 315.1 Mb sequence corresponding to 67% of the genome, covering 91.3% of the gene space. The TILLING resource for *L. japonicus* is used to identify allelic series for symbiosis genes. Proteome analyses on pod and seed development were performed in *L. japonicus* (Dam et al. 2009,).In *M. truncatula*, a *number* of genome resources have become available in recent years (Young and Udvardi 2009). For example, a gene expression atlas provides an information resource for the transcriptome (Benedito et al. 2008). Insertional mutagenesis by the *Tnt1* transposon system and the flanking sequence data set has provided a reverse genetics resource (Tadege et al. 2008). The web site of the *Medicago truncatula* Genome Project in JCVI/TIGR (http://medicago.jcvi.org/cgi-bin/medicago/annotation.cgi) is an information resource that provides the current version of pseudomolecules (ver. 3.5) and an annotation of the *M. truncatula* genome. The web page of the *Medicago truncatula* HapMap Project (http://medicagohapmap.org/index.php) provides not only a reference genome sequence but also NGS re-sequencing data as a GWAS resource. sRNAs expressed in roots and nodules were analyzed using 454 pyrosequencing, and the MIRMED database (http://medicago.toulouse.inra.fr/Mt/RNA/MIRMED/LeARN/cgi-bin/learn.cgi) was constructed as an informative resource for *M. truncatula* miRNAs (Lelandais-Briere et al. 2009). Large-scale analysis of the phosphoproteome in *M. truncatula* roots was performed using immobilized metal affinity chromatography and MS/MS followed by launch of the *Medicago* PhosphoProteinDatabase (http://www.phospho.medicago.wisc.edu/db/index.php) (Grimsrud et al. 2010). The world's largest legume crop, soybean (Glycine max), is widely grown for food and biofuel. A draft assembly of soybean genomes was released in 2010 and the first version of the Glyma1.0 genome annotation was based on homology and gene predictions based ab initio (Schmutz et al. 2010). Also in favor of homology-based gene prediction was applied a sequence data set of soybean full-length cDNAs (Umezawa et al. 2008). In order to improve productivity and stress tolerance, several genomic resources are available for soybean genomics as well as molecular breeding (Manavalan et al. 2009). Genome-scale gene exploration by using the soybean genome sequence and annotated gene models genome-scale exploration of gene families and those functional analyses have been performed to identify genes for molecular breeding (Mochida et al. 2010). A soybean transcriptome atlas (http://digbio.missouri.edu/soybean atlas/) was developed using a NGS platform to perform sample RNA-seq from 14 different conditions (Libault et al. 2010). A genome-scale survey of sRNAs also included NGS-based approaches (Song et al. 2011). Soybase (http://soybase.org/) has played a significant role as an information portal for soybean research in integrating various soybean research resources and analytical platforms (Grant et al. 2010).

10.3.5 Systems Analysis of Plant Functions

Analysis of systems based on a combination of multiple omics analyzes was an effective approach to determining the cellular system's global image. From the early stages of plant metabolomics research, we have made significant progress in our understanding of gene function in metabolic systems through the integration of metabolome analysis with genome and transcriptome resources (Okazaki et al. 2009). Multi-omics-based system analyzes have improved our understanding of cellular plant systems following these successes. For example, recently integrated metabolome and transcriptome analyzes were used to analyze rice developing caryopses under high temperature conditions (Yamakawa and Hakata 2010), molecular events underlying pollination-induced and pollination-independent fruit sets (Wang et al. 2009) and the effects of *DE-ETIOLATED1*down-regulation in tomato fruits (Enfissi et al. 2010). Integrated metabolome and transcriptome analysis has also been applied to investigate changing metabolic systems in plants growing in field conditions, such as the rice Os-*GIGANTEA* (Os-*GI*) mutant and transgenic barley (Izawa et al. 2011). In addition, a system approach combined hormonomal, metabolome and transcriptome analysis in transgenic lines of Arabidopsis, showing increased leaf growth to gain insight into the molecular mechanisms controlling leaf size (Gonzalez et al. 2010). To compare the differences in response to anoxia between rice and wheat coleoptiles, an integrated proteome and metabolome analysis was applied (Shingaki-Wells et al. 2011). To characterize the cascading changes in UV-B-mediated responses in maize, an integrated transcriptome, proteome and metabolome analysis was conducted (Casati et al. 2011). These illustrative examples show the power to understand multi-omics-based systems analysis for understanding the key components of cellular systems underlying various plant functions. GWAS identified genetic loci associated with enzyme activity, metabolome profiles, and biomass using a large set of accessions to Arabidopsis and data sets of genome-scale variation (Sulpice et al. 2010). The hormonal responses of natural variations were addressed in order to find relationships between physiological hormonal response variations and other variations, such as in the genome and transcriptome (Delker et al. 2010). A combinatorial approach to population genomics using hormonome profiling would allow us to identify the link between genomic polymorphisms and plant hormone abundance as quantitative features that could be closely linked to environmental adaptation. Recently, in human population genomics, relational instances of epigenomic modification, gene transcription coding and non-coding RNAs were coupled with genome-scale nucleotide polymorphism data sets (Shoemaker et al. 2010). In a comparable manner, plant epigenome analysis can also be integrated with genome-scale variations to provide important clues to phenotypic diversity-related epigenetic and genetic regulation.

10.3.5.1 Ultrahigh-Throughput DNA Sequencing

The Sanger sequencing method has been used to complete microbial and higher eukaryote genomes sequencing over the past decade. A number of alternative technologies have become available in recent years, which are method adaptations such as pyrosequencing procedures, massively parallel DNA sequencing or single molecule sequencing (Ansorge 2009). In the fields of comparative genomics, metagenomics and evolutionary genomics, such new sequencing technologies have provided us with new opportunities to address the entire genome level (Varshney et al. 2009).

10.3.5.2 Whole-Genome Re-Sequencing

Next-generation sequencing technology coupled with reference genome sequence data enables us to detect variations between individuals, strains and/or populations. Effectively identifies nucleotide polymorphisms by mapping sequence fragments to a specific reference genome data set, a capability of immense importance in all genetic research. A full-genome resequencing project to detect all-genome sequence variations in 1001 Arabidopsis strains (accessions) will result in a data set that will become a fundamental resource for the promotion of future genetic studies to identify allles associated with phenotypic diversity throughout the genome and throughout the species (http://1001genomes.org/) (Weigel and Mott 2009). In rice, an Illumina Genome Analyzer generated high-throughput method for genotyping recombinant populations was performed (Huang et al. 2009). The application to whole-genome de novo sequencing is one of the most anticipated innovations for next-generation sequencers. Although this approach has only been realized in bacterial genomes to date (Moran et al. 2009), there are several attempts to realize this progress in higher species.

10.3.5.3 Comprehensive Discovery of Small RNAs (sRNAs)

SRNAs, including microRNAs (miRNAs), short-interfering RNAs (siRNAs) and trans-acting siRNAs (ta-siRNAs), also play roles in plants as key components of epigenetic processes and gene networks involved in development and homeostasis (Ruiz-Ferrer and Voinnet 2009). These RNA molecules are important targets to be fully identified and their expression should be analyzed using genomic technologies of the next generation (Chellappan and Jin 2009). In maize, deep-sequencing sRNAs in the wild type and isogenic mop1–1 loss-of-function mutant were analyzed using Illumina's sequencing-by-synthesis (SBS) technology to characterize maize complement sRNA (Nobuta et al. 2008). In poplar, expressed sRNAs from leaves and vegetative buds were also discovered using Roche 454 high-throughput pyrosequencing, then genes from miRNA families were identified, including the novel ones (Barakat et al. 2007). Deep sequencing of Brachypodium sRNAs was also performed at the global genome level, resulting in the identification of miRNAs involved in the response to cold stress. The miRNA plant database (PMRD) is a useful plant miRNA information resource available on the Web (http://bioinformatics.cau.edu.cn/PMRD/).

10.3.6 Resources for Variation Analysis

Recent innovations related to DNA sequencing technology and the rapid growth of genome and cDNA sequence resources allow us to design various types of molecular markers covering entire genomes (Feltus et al. 2004). For high-throughput genotyping, a number of platforms have been developed that have been applied to genetic map construction, marker-assisted selection and QTL cloning using multiple segregation populations (Hori et al. 2007). Such genotyping systems have also been used in post-genome sequencing projects such as genotyping of genetic resources, accessions to evaluate population structure and association studies to identify genetic loci involved in phenotypic changes of species.

10.3.6.1 Molecular Markers

The accumulation and saturation of available genetic markers contributes to progress in marker-assisted genetic studies and is an important resource with a wide range of uses. Genetic markers designed to cover a genome extensively enable not only the identification by QTL analyzes of individual genes associated with complex traits, but also the exploration of genetic diversity in natural variations (Caicedo et al. 2007). These sequence data sets have become quite efficient sequence resources for designing molecular markers with the advancement of genome sequencing and large-scale EST analysis in different species. A number of attempts to design polymorphic markers from accumulated sequence data sets have been made for various species. Several genome-wide rice (*Oryza sativa*) DNA polymorphism data sets have been constructed based on alignment between *japonica* and *indica* rice genomes (Shen et al. 2004). Large-scale EST data sets are also important resources for sequence polymorphism discovery, particularly for allocating expressed genes to a genetic map. The computational discovery of ESTbase single-nucleotide polymorphisms (SNPs) and/or EST-SNP markers for the identification of sequence-tagged site (STS) markers has therefore progressed for many species, including barley, wheat, maize, melon, brassica, common bean and sunflower (Li et al. 2009). Several databases provide information about plant species molecular markers. PlantMarkers is a genetic marker database containing predicted molecular markers from different plant species, such as SNP, SSR and preserved orthology set (COS) markers (Heesacker et al. 2008). GrainGenes is a popular genomics site for Triticeae; it also provides genetic markers and mapping information on wheat, barley, rye and oat (Carollo et al. 2005). Gramene is a comparative plant genomics database that provides genetic maps of different species of plants (Liang et al. 2008). The Triticeae Mapped EST (TriMEDB) database provides information on mapped cDNA markers related to barley and their wheat homologs (Mochida et al. 2008).

Analysis of high-throughput polymorphism is a key tool for facilitating any approach based on genetic maps. To date, genome-wide genotyping using a hybridization-based SNP typing method has been used to analyze representative ecotypes of Arabidopsis and rice strains, and data sets have been released for each species containing the calculated genome-wide variation pattern. As typified by the project Arabidopsis 1001, the study of genome-wide variation is a key analysis that should be carried out for a particular reference strain after genome sequencing has been completed. Therefore, the demand for high-performance and cost-effective platforms for comprehensive analysis of variation or also known as variome analysis.

The whole-genome resequencing approaches are already being implemented in species whose reference genome sequence data are available as a direct solution for variable analysis. Diversity Array Technology (DArT) is a high-throughput genotyping system developed on the basis of a microarray (http://www.diversityarrays.com/index.html) platform (Wenzl et al. 2007). DArT markers were used together with conventional molecular markers in various crop species such as wheat, barley and sorghum to construct denser genetic maps and/or conduct association studies (Mace et al. 2009).

Affymetrix GeneChip Arrays has been used in barley and wheat to discover nucleotide polymorphisms as single-function polymorphisms based on the differential hybridization of GeneChip samples (Bernardo et al. 2009). The Illumina Golden Gate Assay allows simultaneous analysis of up to 1536 SNPs in 96 samples and was used to analyze genotypes of segregation populations to construct genetic maps allocating SNP markers in crops such as barley, wheat and soybean (Close et al. 2009).

10.3.7 Transcriptome Resources in Plants

Comprehensive and high-throughput gene expression analysis, called transcriptome analysis, is also a major approach to screening candidate genes, predicting gene function, and discovering cis-regulatory motives. The method of hybridization, such as that used in microarrays and GeneChips, has been well established to acquire large-scale gene expression profiles for different species. The recent rapid accumulation of data sets containing profiles of large-scale gene expression and the ability of related databases to support the availability of such large data repositories has given us access to large amounts of public domain information. This public domain data are an efficient and valuable resource for many secondary uses, such as co-expression and comparative analyses. Furthermore, as a next-generation DNA sequencing application, deep sequencing of short fragments of expressed RNAs, including sRNAs, is quickly becoming an efficient tool for use with genome-sequenced species (de Hoon and Hayashizaki 2008).

10.3.7.1 Sequence Tag-Based Platforms in Transcriptomics

An early approach to the acquisition of transcriptome profiles was the large-scale sequencing of ESTs from cDNA libraries. In this approach, sequencing and/or assembly methods are used to classify randomly sequenced ESTs in an unbiased cDNA library into clusters of transcript sequences. The abundance of transcripts expressed in each tissue is then estimated by counting the number of ESTs for each cDNA library and/or sequence cluster with identifiers. Human and mouse have applied the same methodological principle in the form of a' body map' to derive the transcriptome in different organs (Ogasawara et al. 2006). In addition, this principle was also applied in the digital field of differential display (DDD), which is a component of NCBI's UniGene database and has been applied in large-scale cDNA projects for various species, including plants (Zhang et al. 2004). While this approach, coupled with clone resources from cDNA, has facilitated gene discovery and profiling of expression, its disadvantages include cost and limited resolution due to large-scale sequencing. Serial gene expression analysis (SAGE) is a method based on deep sequencing of short cDNA read tags. SAGE allows a large number of transcripts present in tissues to be identified and allows a quantitative comparison of transcriptomes (Velculescu et al. 1995). SAGE is designed to generate a short specific tag (13–15 bp) from the 3′ end of each sample mRNA, after which >10 tags are concatenated and cloned to generate a SAGE library. The sequencing of selected clones from the SAGE library allows efficient collection of transcript tag sequences. A data set of genome sequences or large-scale ESTs is required to identify genes corresponding to each SAGE tag. Some derivatives of the original protocol (MAGE, SADE, microSAGE, miniSAGE, longSAGE, superSAGE, deepSAGE, 5′ SAGE, etc.) have been developed to improve and expand the utility of SAGE (Anisimov 2008). For example, superSAGE is an improved version of SAGE that produces 26 bp fragment tags from cDNAs. This method has been applied to simultaneous and quantitative gene expression profiling of both host cells and their eukaryotic pathogens in rice (Matsumura et al. 2003). The 26 bp superSAGE tags have also been used to design probes directly for oligo microarrys (Matsumura et al. 2008).

Massively parallel signature sequencing (MPSS) is another technology based on sequencing. MPSS uses a unique method to quantify levels of gene expression; by sequencing 16–20 bp from the 3′ side of cDNA using a microbead array, it generates millions of short sequence tags per library (Brenner et al. 2000). Online (http://

mpss.udel.edu) databases containing MPSS data on plant species, including Arabidopsis, rice, grape and Magnaporthe grisea (the rice blast fungus). In addition, the genome-scale discovery and expression profiling of sRNAs in Arabidopsis and rice was also carried out using the MPSS method (Nobuta et al. 2007). The CT-MPSS was a recently developed method for quantitative transcript 5′end analysis coupled with cap-trapper method for full-length cDNA cloning. This method was used to carry out TSS high-density mapping in Arabidopsis to identify plant promoter genome-scale instances (Yamamoto et al. 2009). Arabidopsis CT-MPSS tags data set can be accessed from ppdb (http://www.ppdb.gene.nagoya-u.ac.jp), a plant promoter database providing promoter annotation of Arabidopsis and rice (Yamamoto and Obokata 2008).

10.3.7.2 Hybridization-Based Platforms in Transcriptomics

The DNA microarray history began with a paper from P O. Brown University Laboratory in 1995 (Schena et al. 1995). Since then, technologies related to microarray and DNA chips have advanced rapidly and their application has expanded to a wide range of disciplines in life sciences. The methodological principle of the DNA microarray or GeneChip analysis is to acquire in a given sample a comprehensive data set of the molecular abundance of each molecule based on its simultaneous hybridization with large numbers of molecular DNA species immobilized on a glass slide or on a silicon chip used as a sample set.

DNA microarrays can be classified into two main types: I the type of spotting developed at Stanford University; and (ii) the type of on-chip synthesis based on manufactured samples. During the early years of transcriptome research, spotting type was widely used. This method involved preparing microarrays of DNA by spotting a solution of cDNA on a glass slide. The on-chip (in situ) oligo synthesis method is a process of light-directed chemical synthesis combining solid-phase chemical synthesis with techniques of photolithographic manufacturing. This method was initially used only in conjunction with the GeneChip Array system manufactured by Affymetrix. In the Affymetrix GeneChip system, a known gene or potentially expressed sequence is represented on the chip by 11–20 unique oligomeric probes that are each 25 bases in length. Roche NimbleGen and Agilent Technology offer platforms to manufacture high-density DNA arrays based, respectively, on Roche's proprietary Maskless Array Synthesizer (MAS) technology and on a non-contact industrial inkjet printing process, both of which are also used for in situ oligo synthesis.

A number of DNA microarrays were also developed for transcriptome analysis in different plant species with the recent and rapid increase in the number of sequenced species in whole-genome and/or large-scale cDNA clones. For example, Seki and colleagues designed a custom DNA microarray that uses 7000 full-length Arabidopsis cDNA clones as samples and then screens genes successfully using a two-color method in response to abiotic stress (Seki et al. 2002). With the recent increase in commercially available DNA microarrays, laboratories can use a specific DNA microarray design to obtain transcriptome data from numerous experiments to accumulate more comprehensive data. AtGenExpress was a multinational effort designed to uncover the transcriptome of *A. thaliana*. The data sets collected in AtGenExpress have been one of the most comprehensive resources for the Arabidopsis transcriptome to date (Goda et al. 2008). The Gene Expression Omnibus (GEO) of the NCBI and the ArrayExpress of the European Bioinformatics Institute (EBI) were the primary public domain transcriptome archives (Barrett et al. 2009). There are also several more focused databases that provide user-friendly interfaces and annotations on probes with calculated transcriptome data. ATTED II (http:/atted.jp/) is a database that provides data calculated from publicly available data on Arabidopsis ATH1 GeneChip for co-expression analysis (Obayashi et al. 2009). Co-expression analysis data sets generated from extensively collected transcriptome data sets have become an efficient resource that facilitates the discovery of transcriptome data sets have become an efficient resource capable of facilitating the discovery of genes closely correlated in their expression patterns. Genevestigator (https://www.genevestigator.com/gv/index.jsp), which is a reference expression database and meta-analysis system, also provides summary information from hundreds of microarray experiments on various organisms, including Arabidopsis, barley and soybean, with easily interpretable results (Zimmermann et al. 2004a, b). The electronic fluorescent pictograph (eFP) browser provides gene expression patterns collected from Arabidopsis, poplar, *Medicago*, rice and barley via a user-friendly interface on the Web (http://www.bar.utoronto.ca/) (Winter et al. 2007). The Arabidopsis Gene Expression Database AREX is a database that provides data sets of high-resolution gene expression patterns of root tissues in Arabidopsis (http://www.arexdb.org/index.jsp) (Brady et al. 2007). The RICEATLAS is a database housing rice transcriptome data covering various types of tissues (http://bioinformatics.med.yale.edu/riceatlas/). Tiling arrays, which are high-density oligonucleotide samples spanning the entire genome in a particular organism, are a platform for analyzing expressed regions across a whole genome; an effective method for discovering novel genes and clarifying their structure. Seki and colleagues performed transcriptome analysis in Arabidopsis using a whole-genome tiling array under abiotic stress conditions and discovered a number of transcripts of antisense induced by abiotic stress. The *A. thaliana* Tiling Array Express (At-TAX) is a whole-genome tiling array resource for developmental expression analysis and transcript identification in Arabidopsis (Zeller et al. 2009). Coupling this platform with the immune-precipitation method has recently extended the usefulness of tiling arrays. For example, the binding sites of AGAMOU-Like15, AGL15, a MADS domain transcriptional regulator promoting somatic embryogenesis, were identified using a chromatin immunoprecipitation (ChIP) approach coupled with the Affymetrix tiling array for Arabidopsis. This method found approximately 2000 sites (Zheng et al. 2009). Using the methylcytosine immunoprecipitation (mCIP) method in combination with the Arabidopsis tiling array, a comprehensive DNA methylation map of the genome was constructed as an Arabidopsis methylome data set (Zhang et al. 2006a, b). Sequencing of co-precipitated DNAs together with a protein using the next generation sequencer, ‘ChIP-seq’, has also become an alternative approach (Park 2009).

10.3.8 Combinatorial Approaches in Metabolomics and Other Omics Resources

Metabolome approaches also support understanding global relationships between cellular metabolic systems in combination with other instances of omics such as transcriptome and proteome profiles, as well as genetic variations. In the well-studied Arabidopsis, these combinatorial approaches have been successfully demonstrated by taking advantage of the many other omics resources that currently exist, including the full-genome sequence with mature annotations, large-scale transcriptome data sets and related co-expression data, and bioresources such as mutant collections and full-length cDNA clones. A conceptual scheme for the systematic clarification of molecules from gene to metabolites molecular networks through a combinatorial approach using transcriptome and metabolome resources has been demonstrated by Saito's group in the RIKEN Plant Science Center. A batch-learning, self-organizing map (BL-SOM) was used to analyze data sets containing transcriptome and metabolome changes of Arabidopsis under stress conditions induced by sulfur and nitrogen deficiency to identify genes involved in glucosinolate biosynthesis (Hirai et al. 2004). For the investigation of an activation-tagged mutant and overexpressors of a MYB TF, PAP1 gene, an integrated approach involving metabolome and transcriptome analysis was conducted to identify genes involved in anthocyanin biosynthesis in Arabidopsis (Tohge et al. 2005). The Arabidopsis transcriptome co-expression data provided by the ATTED-II database was applied to the investigation of key genes involved in specific metabolic pathways and then to the configuration of a metabolome analysis coupled with mutant lines of the targeted genes (Obayashi et al. 2009). The ATTED-II database was used to identify novel genes involved in lipid metabolism leading to the identification of a novel gene, UDP-glucose pyrophosphorylase3 (UGP3), which is required for the first step of sulfolipid biosynthesis (Okazaki et al. 2009). Co-expression analysis has also been used to identify all genes associated with flavonoid biosynthesis, leading to further detailed analysis of two UGT78D3 and RHM1 flavonoid pathway genes (Yonekura-Sakakibara et al. 2008). Approaches that have integrated metabolome and transcriptome data also have elucidated regulatory networks that respond to plant environmental stress. Metabolome analysis using different types of MS combined with microarray analysis of gene overexpressors encoding two TFs, DREB1A/CBF3 and DREB2A, investigated the metabolic pathways that act in response to cold and dehydration conditions in Arabidopsis (Maruyama et al. 2009). Metabolomic profiling was also used under conditions of dehydration stress to investigate chemical phenotypic changes between wild-type Arabidopsis and a NCED3 gene knockout mutant. The metabolic data was then integrated into the transcriptome data to reveal ABA-dependent regulatory networks (Urano et al. 2009).

Metabolome profiling was also used simultaneously to evaluate chemical phenotypes of natural variations and/or populations of segregation. A comprehensive exploration of the association between metabolic and genomic diversity will allow key genes involved in metabolic changes to be discovered and will also help to identify genetic associations between metabolic and/or visible phenotypes (Fu et al. 2009). Analysis of metabolite QTL (mQTL) using segregated populations has been applied in a popular forward genetic approach to different plant species such as Arabidopsis, poplar and tomato (Schauer et al. 2008). In addition, together with the recent availability of genome wide variation acquired by high-throughput genotyping methods including resequencing, interest in the discovery of the genetic association between nucleotide variation and phenotypic changes has also increased, especially with regard to the identification of key genes that play significant roles in evolutionary histories. The attempts to mine correlative patterns between metabolic and genomic diversities have recently been applied to sesame and rice using seed stocks of natural variations (Mochida et al. 2009).

With the completion of genome sequencing in a number of plant species, comparative genome-scale analyzes can be used to produce and publish data sets that facilitate the identification of preserved and/or characteristic properties among plant species. Several efforts have been made to build comprehensive gene families using model proteome data sets deduced from sequenced genomes in establishing platforms to verify gene content and elucidating the process of gene duplication and functional diversification among species (Sterck et al. 2007). Comprehensive gene family data sets are usually produced through computational procedures, including a step that performs an all-against-all sequence similarity search and then a step for building protein family clusters by methods such as Markov Clustering (MCL) or consideration of protein domain structures (Hulsen et al. 2006). The results of such studies can produce databases that are useful for further phylogenetic studies (Wall et al. 2008). Correlated gene arrangements between taxa along with chromosome allocation, also known as synteny and collinearity, have become valuable frameworks for the inference of shared gene ancestry and the transfer of knowledge from species to another related species (Tang et al. 2008a). The plant genome duplication database (PGDD) provides a data set of intragenome or cross-genome syntenic relationships identified throughout genome-sequenced plant species (http://chibba.agtec.uga.edu/duplication/) (Tang et al. 2008b).

Databases that contain focused data sets together with rich annotations and well-related cross-references are also very useful for better understanding of focused issues in particular gene families and/or specific cellular processes. Sequence-specific DNA-binding TFs are key molecular switches that control or influence many biological processes, such as development or responses to environments. The genome-wide identification of gene repertories encoding Arabidopsis genome TFs was first reported in plants and comparisons with other organisms revealed the properties of plant-specific TFs (Riechmann et al. 2000). Over the past decade, we have been able to compile catalogs describing the function and organization of TF regulatory systems in a number of organisms with the availability of complete genome sequences. In many plant species, there are many databases that provide data sets of genes that putatively encode TFs; these are usually predictions based on computational methods such as sequence similarity search and/or hidden Markov model search for preserved DNA-binding domains. Recently, there has been further integration of data sets of TF-encoding genes, thereby creating an integrative, knowledge-based resource of TFs across related plant species in terms of comparative transgenomics regulatory networks. GRASSIUS provides the first step toward building a comprehensive platform for integration of information, tools and resources for comparative regulatory genomics across the grass species. The Grass Transcription Factor Database (GrassTFDB) of GRASSIUS houses integrated information on MaizeTFDB, RiceTFDB, SorghumTFDB and CaneTFDB (http://grassius.org/grasstfdb.html). The LegumeTFDB provides predicted TF- encoding genes annotated in the genome sequences of three major legume species: soybean, *L. japonicus* and *M. truncatula* (http://legumetfdb.psc.riken.jp/). This database is an extended version of the SoybeanTFDB (http://soybeantfdb.psc.riken.jp/) and is aimed at integrating knowledge on legume TFs and providing a public resource for comparative genomics of the TFs of legumes, non-legume plants and other organisms (Mochida et al. 2010).

10.4 Integration of Interdisciplinary Approaches for Solving Biological Problem with Respect to Agriculture or Crop Improvements

The information and resources generated from various omic technologies offer prospects for the production of new biological knowledge. To describe and understand complex biological phenomena, it is essential to integrate various types of biological information and large-scale omics data sets through systematic analysis. We have developed a web-based system, Plant MetGenMAP, for this purpose, which can integrate and analyze large-scale gene expression and data sets of metabolite profiles along with various biological information. Under certain conditions, significantly altered biochemical pathways and biological processes can be retrieved quickly and efficiently using this system, and transcriptional events and/or metabolic changes in a pathway can easily be visualized. The system also provides a unique function that can identify candidate promoter motifs related to regulating specific biochemical pathways. Using data sets from Arabidopsis and tomato, we demonstrate the functions and application of the system. The results obtained by Plant MetGenMAP can contribute to a better understanding of the mechanisms underlying interesting biological phenomena and provide new insights into the biochemical changes associated with them at the gene and metabolite levels (Thijs et al. 2001). Extensive insight into molecular mechanisms and the coordination of biological networks has been obtained after the application of several different methods. Our knowledge of how the cell's different molecular entities interact with each other suggests that the integration of data from different techniques could nevertheless lead to a more comprehensive understanding of the data emanating from different techniques. The main focus on the pairwise integration of large-scale metabolite data with that of the transcriptomic, proteomics, whole-genome sequence, growth- and yield-associated phenotypes, and archival functional genomic data sets (Gonzalez et al. 2010).

The advances in high-throughput analytics have enabled us to gain insights into individual biomolecules using the various omics technologies. Any single omic approach, however, may not be sufficient to characterize the complexity and behavior of biological systems as a whole (Gygi et al. 1999). Therefore, molecular research gradually shifts towards the holistic perceptions of system biology, through the integration of the individual omics datasets, in order to obtain biologically meaningful interpretation of plant systems. The integration of multiple layers of biological information will therefore provide an accurate' picture' of the' whole' plant systems. Multiple omics datasets must be integrated after preprocessing (normalization, attribution of missing value and selection of features). Data integration is a key to successful system philosophy development through the development of comprehensive plant system models. Given the enormous promise to integrate multiple omics data, there is a growing interest in logical input to designing different experiments and analyzing heterogeneous data (Choi and Pavelka 2011). The successful integration of data will depend on appropriate experimental design, sound statistical analysis and correct interpretation of the results. The various aspects of successful integration of multiple heterogeneous omics datasets are to deposit individual 'omics' data to respective public repositories, to generate relationships among various kinds of datasets, visualization of the data and application of statistical and bioinformatics resources, where and when needed. These aspects have been elaborately discussed in Joyce and Palsson (2006).

The literature contains several instances of omics data integration. There are a number of reports on gene function elucidation through the combination of metabolomic analysis with genomic and transcriptomic data (Tohge et al. 2005; Okazaki et al. 2009). In maize hybrids, an integrated approach to transcriptomics and epigenomics has recently been used (He et al. 2013). Integrated use of transcriptomic and proteomic data has been reported in several recent studies involving whole plant nitrogen maize economy, growth to transition to dormancy in white spruce stems (Galindo González et al. 2012), phytohormone crosstalk (Proietti et al. 2013) and flour quality in wheat (Altenbach et al. 2010). Similarly, integrated metabolome and transcriptome analyses were recently applied in analysis of rice developing caryopses under high-temperature conditions (Yamakawa and Hakata 2010), molecular events underlying pollination-induced and pollination-independent fruit sets, the effects of DE-ETIOLATED1 down-regulation in tomato fruits (Enfissi et al. 2010) and changing metabolic systems in plants growing in field conditions, such as the rice mutant and transgenic barley (Kogel et al. 2010). An integrated metabolome and proteome analysis was applied in wheat and rice coleoptiles to illustrate the differences in response to anoxia (Shingaki-Wells et al. 2011) and characterization of starch and raffinose metabolisms to low and high temperatures in A. thaliana (Mostafavi et al. 2008). An integrated transcriptome, proteome and metabolome approach was adopted to describe the cascading changes to UV-B in maize (Casati et al. 2011). Moreover, an integrated hormonome, metabolome and transcriptome analyses in Arabidopsis transgenic lines, displayed increased leaf growth to gain insight into the molecular mechanisms that control leaf size (Gonzalez et al. 2010) have been reported. The literature mining is also a useful approach to knowledge integration in plant biology (Winnenburg et al. 2008). Apart from single problems, more complex problems like photosynthesis have been addressed by Weston et al. (2011), where they characterized a network for heat transcriptome of three plant species (Arabidopsis, Populus and Soybean) where expression of one heat responsive module showed a negative correlation with leaf-level photosynthesis at a critical temperature. Later they proposed a conceptual model where traditional network analysis can be linked to whole-plant models (Weston et al. 2012). Also recently, Fouracre et al. (2014), threw light on the application of systems approaches in understanding the Kranz anatomy of the C4 plants. Several web-based resources like PLAN2L (Krallinger et al. 2009) and PosMed-plus (positional Medline for plant upgrading science) (Makita et al. 2009) are available to integrate literature-derived bioentities and associated information. The integration of multiple omics data has several challenges (De Keersmaecker et al. 2006; Steinfath et al. 2007). One of the problems with complex annotation and integration is the lack of agreed formats across various omics datasets due to the primary data sources 'heterogeneous repositories. The solutions to this problem include creating' data warehouses, using extensible markup language (XML), navigating hypertext, Unmediated MultiDB queries, creating a federated database and using controlled vocabulary. A Data Warehouse collects data from multiple resources, translates the formats and places them in a single database. The examples of data warehouses include: Atlas, BioMart, BioWarehouse, Columba, SYSTOMONAS, BioDWH, VINEdb, Booly, GNCPro (Turenne 2011). The XML is a general-purpose markup language that helps in sharing data across heterogeneous systems. The development of Systems Biology Markup Language (SBML) (Hucka et al. 2003) is probably the first and most successful efforts in this aspect. The Plant Ontology Consortium is a collaborative effort between plant genome model databases and plant researchers to create a controlled vocabulary (ontology) for plants to be maintained and facilitated (Avraham et al. 2008). The other problem includes statistical analysis, i.e. evaluation of integration complexity that differs from that of individual omics analysis and subsequently applying an appropriate method. Therefore, integrating omics data is far more than merely' joining the pieces;' it is actually a journey of exploring uncharted territories and transforming information into more useful biological knowledge.

10.4.1 Modeling and Simulation in Plant System Dynamics

The systems interest to biological sciences dates back to the days of von Bertalanffy (1933 1968), Wiener (1948) and Forrester (1958, 1961). In the context of biology, Biochemical Systems Theory (Voit 2000) and Metabolic Control Theory (Heinrich and Schuster 1996), proposed general mathematical models of biological systems at and around a steady state (equilibrium). Successful plant modeling is the ultimate goal of biology of plant systems. In a system, in mathematics, a model (Latin mode, meaning manner/measurement) usually represents the causal relationship. Cells or higher units of biological organization are understood as interacting element systems in systems biology. The identity of the constituents, dynamic behavior and interactions between the constituents, of the biological system under study (Kitano 2002) must be known for an explanation of the system level. Ultimately this information can be combined into a model that is not only consistent with current knowledge but can also predict system behavior under new unexplored perturbations. Modeling and simulation are central to bridge the gaps between theory and experiment (Dhar et al. 2004). Experimental results usually require correct mathematical/statistical input, and model hypotheses require experimental evidence to provide meaningful biological interpretations. Modeling usually begins with building biological networks from the molecular data sets available. Network building and analysis are key components of biology of systems. In system biology, a network/graph has two basic parts: the system elements are represented as graph nodes and the interactions are represented as edges, i.e. lines connecting pairs of nodes. Edges can be directed (from a source (start node) to a sink (end node) and represent unidirectional flow of material or information) or non-directed (representing mutual interactions where the directional flow of information is not known). In biological networks, nodes represent the molecules present inside a cell (e.g., proteins, RNAs and/or metabolites) and links (or edges) between nodes represent their biological relationships (e.g., physical interaction, regulatory connections, metabolic reactions) (Blais and Dynlacht 2005). Activation or inhibition signs can be displayed on the edges to increase the network's information content. The important characteristics of biological networks are the scale-free structure (the number of nodes that connect with other nodes is much lower than the number of nodes with few connections) and the relative scarcity of hubs that connect directly with each other (Barabasi and Oltvai 2004). The interaction network nodes represent the biomolecular population whose abundance varies over time and in response to internal and environmental disturbances. The interaction network needs to visualize the changes and create a model which needs to be augmented by variables (expression, concentration, activity), thus indicate the state of each node and set of equations, signifying the how the state changes corresponding to the stimuli. Depending on their behavior in the system with time, models can be static or dynamic. The four common types of networks in plant biology systems include gene-to-metabolite networks, protein-protein interaction networks, transcriptional regulatory networks, and gene regulatory networks in which the first three types are often static, whereas the gene regulatory network is often dynamic (Yuan et al. 2008).

10.4.2 Gene-To-Metabolite Networks

Gene-to-metabolite networks are derived under a given set of conditions from the analysis of the correlation of genes and metabolites. The genes and metabolites act as nodes here, and the edges represent the interactions between regulators. Depending on the distance between genes and metabolites, interactions are interpreted. Due to the enormous diversity and number of metabolites produced in cells corresponding to their sessile lifestyle, these types of networks are highly complex and difficult to study in plants. Different new research dimensions such as interrelation between biological processes, functional gene annotation, and discovery of new genes in biosynthesis, regulation and transportation of metabolites, have been added to plant science owing to the elucidation of gene-to metabolite networks (Yuan et al. 2008). The gene-to- metabolite networks have been worked out in various studies like in stress responses, discovery of novel candidate genes for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* (Rischer et al. 2006), in the response to nitrogen deficiency and during diurnal cycles (Scheible et al. 2004) an so on.

10.4.3 Protein–Protein Interaction Networks

In protein-protein interaction networks, nodes are proteins that are connected by direct edges if the information flow direction is known during their interaction, or non-directed edges if there is strong evidence of their physical interaction or association without evidence of interaction directionality (Assmann and Albert 2009). It may be possible to have two types of interactions: genetic or physical. A protein-protein genetic interaction is a network of genes characterized by genetic interactions in order to explain gene function in physiological processes (Boone et al. 2007). However, due to ploid levels and long life cycles, this approach is difficult to implement the ploidy levels and long life cycles of plants. On the contrary, physical interactions are easier to be characterized on the plant systems. In plants, interaction maps have been experimentally elucidated for homo and heterodimerization within two large classes of transcription factors: the MADS (MCM1, Agamous, Deficiens, SRF) box transcription factors (de Folter et al. 2005) and the MYB (myeloblastosis) transcription factor family (Zimmermann et al. 2004a, b). The further details regarding interactome are furnished in a preceding section in the current review namely 'interactomics'.

10.4.4 Transcriptional Regulatory Networks

The transcription regulatory network explains the regulatory interactions between transcription factors and downstream genes. They have two types of nodes—transcription factors and regulatory genes and two types of directed edged viz. transcriptional regulation and translation (Babu et al. 2004). In addition, the regulatory edges can have two types of signs, corresponding to activation or repression. Despite the general organizational similarity of networks across the phylogenetic spectrum, there are interesting qualitative differences among the network components, such as the transcription factors (Babu et al. 2004). Transcription factors usually regulate multiple genes and hence transcriptional regulatory networks are unidirectional and do not have strongly connected components. The various approaches to deciphering transcriptional regulatory networks include genome-wide expression profiling, genome-wide RNA interference (RNAi) screens (Baum and Craig 2004), transcription rate assessment by measuring mRNA decline rates, pair evaluation of promoter co-occupancyand cis-element computational prediction. A transcriptional regulatory map for cold signaling mediated by the transcription factor of ICE1 was created in Arabidopsis (Benedict et al. 2006). Recent transcriptional regulatory network reports include the role of oxidative signals in chilling stress in rice (Yun et al. 2010), those in response to abiotic stresses in Arabidopsis and grasses (Nakashima et al. 2009) as well as rice (Todaka et al. 2012), abiotic light-regulated transcriptional networks in higher plants (Jiao et al. 2007) and so on.

10.4.5 Gene Regulatory Networks

The nodes correspond to genes, messengers or proteins in a gene regulatory network and the edges represent the regulatory interactions (activation, inhibition, repression or other functional interactions) among the network components. Complex gene regulatory networks consist of genes, non-coding RNAs, proteins, metabolites and components of signaling (Long et al. 2008). This type of network includes all stages of gene expression regulation including DNA transcription regulation, RNA translation, post-transcription RNA processing, as well as post-translation changes such as protein targeting and covalent protein modification. Unlike other static networks in nature, these networks are often used to display the dynamics of plant systems (Yuan et al. 2008). The ABC model, one of the first modeled plant gene regulatory networks, explained the interactions between transcription factors that regulate plant species-wide floral pattern formation (Coen and Meyerowitz 1991). In several studies, gene regulatory networks were reported to study plant developmental and physiological processes. The studies include the attempt to model the essential components controlling stomatal closure of the cell size, determining the cell fate during flower development in A. thaliana, microRNA (miRNA)-mediated gene regulatory networks (Meng et al. 2011) and recently in explaining land plant evolution (Pires et al. 2013).

Biological network construction and analyzes were therefore an important approach to explaining the organism or a biological process as a whole in the biology of plant systems. In modern science, high-throughput technologies provide huge quantitative data. However, in systems where knowledge of mechanical details and kinetic parameters is scarce, the use of quantitative data is obstructed. In such cases, it may be helpful to model the system with a wealth of molecular data on individual constituents as well as interactions (Assmann and Albert 2009). The system biology’s individual key components viz. Earl explained genomics, transcriptomics, proteomics, metabolomics, etc. have been explained earlier. The biological networks along with these components are chief aspects of plant systems biology. Although the models could not exactly mimic the system with pure accuracy, still are highly capable to explain the intrinsic complexity of the plant systems.

10.4.6 Softwares and Algorithms for Plant Systems Biology

Using software from bioinformatics is inevitable for the comprehensive study of biology of plant systems. In addition to the tools and resources used in the analysis of the individual omics platforms, it requires several resources to elucidate the' complete picture.' Joyce and Palsson (2006) and Turenne (2011) list detailed discussion of various algorithms and software used for system biology. These include network visualization tools, modeling environments, pathway building and visualization tools, modeling platforms for systems biology, and model repositories. Visualization is a means for analyzing research data and a key method for analyzing networks. The purpose of omics data visualization should be to create clear, meaningful and integrated resources without the inherent complexity of the data being undermined (Gehlenborg et al. 2010). There are several tools to help visualize' omics' data on a system scale such as Sungear (Poultney et al. 2007), MapMan (Thimm et al. 2004), Genevestigator (Zimmermann et al. 2004a, b), Cytoscape (Shannon et al. 2003), VirtualPlant (Katari et al. 2010), REACTOME (Joshi-Tope et al. 2005). Pathway databases are used for modeling systems as they provide a clear way to create network topologies by the annotated reaction systems. The various pathway databases for systems analyses include KEGG (Kanehisa et al. 2012), BioCyc (Caspi et al. 2010), Aracyc (Mueller et al. 2003), Pathway Interaction Database (PID) (Schaefer et al. 2009) and BioCarta (Nishimura 2001). Also, several comprehensive modeling environments are available, like Gepasi (Mendes 1997), Virtual Cell (Loew and Schaff 2001), Osprey (Breitkreutz et al. 2003), Arabidopsis eFP browser (Winter et al. 2007), COPASI (Hoops et al. 2006), R (http://www.R-project.org),MatLab and InfoBiotics workbench (Blakes et al. 2011), E-Cell (Tomita et al. 1999), Systems Biology WorkBench (Sauro et al. 2003). The Systems biology model repositories include BioModels database (Le Novere et al. 2006) or JWS (Olivier and Snoep 2004). Both are public, centralized databases of curated, published, quantitative kinetic models of biochemical and cellular systems. The core systems biology networks include SynBioWave (Staab et al. 2010), Cell Illustrator (Nagasaki et al. 2010), Moksiskaan (Laakso and Hautaniemi 2010), MEMOSys (Pabinger et al. 2011), Babelomics (Al-Shahrour et al. 2006), MetNet (Sucaet et al. 2012), etc.

10.4.7 Integrating Metabolite and Transcriptome Data

The initial integrative approaches with plant metabolism relevance included the combination of transcript data and metabolite profiling(Tohge et al. 2005). Such studies were initially limited to model species for which ESTs or oligonucleotides were available; early transcriptomic approaches in fact relied on differential hybridization of complementary DNA samples to known immobilized sequences on solid supports. However, this barrier has been removed by the advent of next-generation sequencing technologies and far more exotic species are beginning to be studied using this approach (Gechev et al. 2013). By combining transcript and metabolite, two basic questions are commonly addressed by combining transcript and metabolite data. The first concerns whether a gene functions within a given metabolic pathway. When a better characterization of the pathway is achieved, it is also essential to examine the extent of transcriptional control (except in some cases, for example, by regulating post – transcription modifications of the enzyme and by regulating positive/negative feedback by substrates/products) under different physiological conditions and how it is distributed across the different enzymatic steps.

The initial focus of these investigations was specific pathways, such as hormone, glucosinolate, and flavonoid biosynthesis. For example, differential gene expression mechanisms helped clarify the involvement of two different genes of nitrilase in auxin synthesis in Arabidopsis. Similarly, the contributions of gene duplication and inducible gene expression (differential activation of biosynthetic gene subsets) have been shown to impact glucosinolates amount and composition. An additional early evidence of the role of specific transcript accumulation on a metabolic phenotype stemmed from the clarification of the role that various regulatory mechanisms affect Trp synthase α and β had on the amount of 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one, a natural pesticide synthesized in maize (*Zea mays*) leaves. Another example of the coordination between transcripts and metabolite accumulation was the maize anther analysis, where a strong correlation was found between the expression of a structural gene (flavanone 3-hydroxylase) and the appearance of specific flavonols (mainly quercetin and kaempferol). In this case, the comparison of sweet and hot pepper varieties made it easier to identify certain placenta-specific, differentially expressed genes that were directly correlated with the accumulation of capsaicinoids. One of the first examples of this approach focused on the identification of transcripts strongly correlated with the abundance of given metabolites across tuber development, irrespective of whether the transcript was associated with the metabolic pathway under question or not (Urbanczyk-Wochniak et al. 2003).

Indeed, this approach was able to identify certain transcripts that exhibited very high correlations with the expression of certain genes and, as such, proved effective in identifying a number of biofortification candidate genes. The same approach can and has been used by corollary to elucidate the variation in gene-to-metabolite networks following short-and long-term nutritional stress in Arabidopsis or to identify gene expression metabolic regulators. For example, in Arabidopsis (Hannah et al. 2010), cryptoxanthin was found to be highly correlated with a large number of genes across diverse environmental conditions, and organic acid malate was putatively identified (Carrari et al. 2006) and subsequently confirmedto be important in mediating the ripening process in tomato (*Solanum lycopersicum*). Such current studies are all examples of the guilt-by-association approach, which in essence postulates biological entities as being functionally related if they exhibit strong correlation or coresponse across a wide range of cellular circumstances. The power of this approach is that it may have great use in identifying novel metabolic integration and/or new regulatory mechanisms, given that it does not rely on a priori pathway knowledge. The main drawback of the approach, however, is that, in the absence of subsequent rounds of experimentation, it is difficult to gain insight into the mechanistic links underlying the behavior observed, given that correlation between biological entities does not always imply causation or the existence of functional links (Tohge and Fernie 2010).

In this respect, it becomes imperative to validate the resultsof co-expression analyzes using follow-up approaches to prove the existence of putting functional links. Arguably, the greatest advances made to date following approaches to integrate transcript and metabolite data have been achieved in gene annotation and the structural elucidation of plant intermediary and secondary metabolism.

Two early studies of particular note are those from the laboratories of Saito and Dixon investigating the metabolism of Arabidopsis anthocyanin and Medicago truncatula triterpene. In the case of the anthocyanin pathway, no late biosynthetic genes involved in anthocyanin decoration steps were identified in Arabidopsis prior to the study of (Achnine et al. 2005; Tohge et al. 2005), although visible phenotype screening characterized all early biosynthetic genes. A combination of transcript and metabolite profiling on an activation-tagged line of anthocyanin pigment1 along with validation experiments involving both heterologically expressed enzymes and knockout mutations resulted in the identification of five genes and the identification of up to 11 anthocyanins. Such confirmatory experiments are essential to assign gene function unambiguously. Combining reverse genetic strategies with enzyme activity characterization when the gene is expressed in a heterologous system remains the gold standard for molecular identification of novel enzyme-catalyzed reactions. Subsequent follow-up studies have identified some six genes associated with the metabolism of flavonol, and some 24 compounds of this class (among 35 compounds found) have now been identified in Arabidopsis While the expansion of the characterized triterpenoid metabolism in M. Truncatula is not that impressive, Tohge et al. (2005)'s study enabled the functional annotation of 30 different saponins, and currently, over 70 metabolites of this compound class have been identified in *M. truncatula*. The usefulness of this approach is at its greatest for the relatively unchartered pathways of specialized metabolism; however, it should be noted that the gene encoding plant Thr aldolase(322–324) in Arabidopsis and 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one glucoside methyltransferase in maize was independently identified in this strategy(Meihls et al. 2013). The number of species and pathways for which this approach was adopted expanded massively to include several crops and medicinal plants a decade later. Strategies combining transcript and metabolite profiling have been effective in shedding light on the structure of several metabolic pathways involved in the synthesis of primary metabolites, flavonoids, terpenoids, and alkaloids (Lin et al. 2015). The combination of transcript and metabolite profiling has been commonly used on a broader level for multi-layered descriptions of plant responses, especially those to abiotic stress. This has resulted in a number of studies assessing the combined transcript and metabolite responses to water stress, temperature stress, light stress, and nutrient supply limitations (Urano et al. 2009). Though descriptive by nature, such studies can provide insight into global metabolic variations under certain conditions as well as identify which pathways are under tight control and which are under loose transcriptional control. Given the highly interconnected nature of the metabolic system and its nonlinearity of metabolic pathways in the global network structure, and even in the absence of flux profiling data, the integration of transcriptomics with wide metabolic profiling can, in any case, narrow down which metabolic steps could be active under specific conditions. Occasionally, however, more mechanistic information can also be provided. A prominent example of this is the detailed analysis of several transgenic Arabidopsis lines with altered flavonoid levels through transcriptomic and metabolomic analysis, including hormone analysis, which revealed that flavonoid overaccumulation with strong oxidative capacity in vitro also gives oxidative stress and drought tolerance (Nakabayashi and Saito 2015). Moreover, a combination of transcript and metabolite profiles followed a range of developmental processes at high resolution. Such studies are dominated by fruit maturation and leaf development studies, but they are not limited to these processes, with studies also covering the development of various organs, lignin deposition, and the establishment of arbuscular mycorrhizal symbiosis (Nakamura et al. 2014). In this respect, these approaches prove informative in clarifying the relative importance of apparently redundant biosynthesis pathways and the degradation of specific metabolites or may also help to define the role of those primary metabolites (e.g., aminobutyrate) for which a signaling role was assumed. For example, ascorbate biosynthesis has been revealed as the dominant route of ascorbate biosynthesis during tomato maturation, which is one of the well-studied metabolisms in several higher plants, especially in Arabidopsis. Another example can be found in the elucidation of the arogenate pathway as an alternative route for Phe biosynthesis, a similar approach in Arabidopsis, based on feeding studies and analysis of co – expression, suggested an alternative pathway to degradation of Lys in dark – induced senescent leaves (Araújo et al. 2010). However, despite the fact that these examples illustrate that combined transcriptome/metabolome studies increase our understanding of metabolic network regulation, we argue that they remain at their most powerful in gene functional annotation and the elucidation of metabolic pathway structures specific to species and/or tissue.

10.4.8 Integrating Metabolite and Proteome or Enzyme Activity Data

The combination of proteome and metabolome analyzes is less commonly used to date than combined transcriptome and metabolome analyses. In addition, they are largely used in a manner similar to the more descriptive studies discussed above hence, this gained considerable insight into the structure of metabolic networks as well as general aspects of metabolic regulation. In the first of these examples, metabolite data were studied in parallel with enzyme data (and transcriptomics data) in different wild-type diurnal cycles and an Arabidopsis starchless mutant, revealing that rapid transcript changes are integrated over time to generate substantially stable changes in many sectors of metabolism. The same group went on to apply this approach to tomato fruit development and natural variance in Arabidopsis. In tomatoes, enzyme profiles were sufficiently characteristic to distinguish developmental stages and cultivars and wild species, but the comparison of enzyme activity and metabolites revealed remarkably little connectivity between enzyme developmental changes and metabolite levels, suggesting the operation of mechanisms for post-translation modification. They documented highly coordinated changes between enzyme activities in Arabidopsis, especially within those of the Calvin-Benson cycle, as well as significant correlations between starch and growth in specific metabolite pairs. On the other hand, there were few correlations and therefore low overall connectivity. On the other hand, few correlations were observed between enzyme activity and metabolite levels(Sulpice et al. 2010), and thus low overall connectivity, but strong links were seen between starch levels and growth, which we describe below. In an alternative approach, proteomic and metabolic data were only used to extend the range of molecular entities to show that fascicular and extra fascicular phloems are isolated from each other and functionally divergent (Zhang et al. 2010).

10.4.9 Integrating Metabolite and Genome Data

Assuming that the advent of metabolomics more or less paralleled the release of the first plant genome, the integration of metabolomics and data on the entire genome sequence may be unsurprising(van der Werf et al. 2007). Suffice it to say that in such combinations there are considerable complexities; tellingly, early studies aimed at computational prediction of the size of the Escherichia coli metabolome estimated a complement of about 750 metabolites, while subsequent experimental approaches revealed many metabolites that were not computed from the genome. This discrepancy could be explained by several potential reasons (Tohge et al. 2014). The most likely reason for this is the lack of linear relationship between genes, their protein products, and metabolites, and the fact that most genomes remain incompletely annotated, including those of model organisms. Despite this serious drawback, in this section, we hope to illustrate that the integration of metabolomics and genomic data can be incredibly powerful in understanding natural metabolism variation and its regulation. Whole-Genome sequences for over 100 species of plants (including microalgae) are available. Metabolomics currently cannot match this massive acceleration provided by next-generation technologies, particularly when high-quality species are being adopted optimized approaches (Fukushima et al. 2014). The KNApSAcK database, which is one of the largest curated compendia of phytochemicals, contains over 700 compounds for early sequenced plants like Arabidopsis and rice (*Oryza sativa*) but no entries for recently sequenced species such as goatgrass (*Aegilops tauschii*) and wild tobacco (*Nicotiana tomentosiformis*). In this section, we will describe insight gained from combining metabolomic data with genome sequences in three different case studies: (1) a simple comparison of a reference genome with metabolomics data; (2) a comparison of natural allelic and metabolic variance; and (3) integrating genome sequence data into quantitative genetics approaches. The first of these has been covered in considerable detail recently (Tohge et al. 2014) so we will only briefly describe it here. The starting point is to perform genome-wide ortholog searches using functionally annotated genes; best practice is to use cross-species cluster-based BLAST searches such as those housed in the PLAZA database (Proost et al. 2009) or, in the case of photosynthetic microbes, pico-PLAZA (Vandepoele et al. 2013). Illustrations of how such analyses have been performed for central, shikimate, phenylpropanoid, terpenoid, alkaloid, and glucosinolate metabolism have been presented. Important insights into pathway evolution can be gained from such approaches, as illustrated by the recent cross-kingdom comparison of ascorbate biosynthesis (Wheeler et al. 2015). The second case study, which is similar in scope but far more targeted than genome-wide association studies to evaluate alllic and metabolic variance across natural diversity, is described below. Most recent examples of its usefulness are derived from the analysis of wild tomato species; however, it is important to note that the approach itself is essentially a modification of that adopted over decades in the cloning of natural color mutants. In recent years, this approach has significantly enhanced understanding of both primary and secondary and cuticular cell wall metabolism been enhanced considerably via this approach (Koenig et al. 2013), Although the greatest insight into the latter was ultimately clarified, as described below, through the use of the introgression line population. In essence, this approach begins with the identification of metabolic variance within a population of ecotypes, cultivars, or similarly related species and attempts to link this with alllic diversity or gene duplication, as has been achieved with acyl-sugar metabolites (Schilmiller et al. 2015), terpenes, and isoprenoids (Kang et al. 2014), or even with the presence or absence of genes, as described recently for methylated flavonoids of glandular trichomes. The previous list documents the success of this approach; however, until recently, it has been constrained by the limits of our a priori knowledge needed to select the candidate genes in which we are searching for alllic variance. The development of RNA sequencing technologies means that we are no longer limited by the amount of sequence data; however, there may still be a potential hurdle to these integrative approaches when comparing highly genetically divergent individuals, as the number of genetic polymorphisms is too large to be evaluated one by one. The quantitative trait loci approach is therefore a powerful alternative method of associating phenotypes with their underlying genetic variance. The use of such approaches in plant metabolism has been the subject of several recent comprehensive reviews (Scossa et al. 2015), however, few examples of their usefulness to advance understanding of metabolite accumulation and metabolic regulation are as-. Tomato fruit, as the model species for fleshy fruit maturation, has been the subject of combined large-scale genomic, physiological, and metabolic investigations, often using specific biparental populations or large sets of unrelated individuals, in an attempt to understand the causal variants of metabolic variations (Sauvage et al. 2014). In particular, the use of introgression lines obtained from the cross between tomato and *Solanum pennellii* (a wild tomato species) has greatly helped to identify quantitative trait loci for a large number of physiological and metabolic traits. Profiling data of primary and secondary metabolites in this population was collected over several years (along with some classical yield-related traits), revealing more than 1500 metabolic quantitative trait loci affecting levels of multiple sugars, amino acids, organic acids, vitamins, phenylpropanoids, and glycoalkaloids. In some selected metabolic quantitative trait loci, the availability of sequences of both parental genomes (Bolger et al. 2014) reduced the origin of the metabolic variation to specific genetic polymorphisms (Alseekh et al. 2015). The integration of genotypic and metabolic variance can and has been applied to large collections of unrelated individuals (metabolite-based genome-wide association studies): as in the case of biparental populations, also with this strategy, several cases of polymorphological variants of genomic sequences have been identified and related to metabolic variation. These two approaches, based either on biparental populations or on large collections of natural accessions, have been used in Arabidopsis and crop species (maize, rice, wheat (*Triticum aestivum*), and fruit trees (Luo 2015).

10.4.10 Integration of Transcriptomic and Metabolomics Level Genome-Scale Models

In the first of these studies, Arabidopsis microarray data exposed to eight different conditions of light and temperature were integrated into a model on a genome scale (Töpfer et al. 2014). We first digress to give a brief description of how genome-scale models are generated before discussing the outcome of this integration. Essentially, a model on a genome-scale match's metabolic gene with metabolic pathways in a way that generates a stoichiometrically balanced metabolic network that matches all gene functions annotated for that organism. In the turn of the century, these models were originally published for microbes, with many models for plant species subsequently available for Arabidopsis as well as crop species such as rice and maize (Simons et al. 2014). Returning to the superimposition of experimental data on the model, the addition of transcriptomic data has enabled flux capacities to be predicted and statistically assessed if these vary under the test conditions. In addition, this study introduced the concepts of metabolic sustainers and modulators, the former being metabolic functions that are differentially up-regulated with respect to the null model, while the latter are differentially down-regulated to control a certain flux and thus modulate the affected processes (Töpfer et al. 2013). In a follow-up study, predictions based on transcriptomics integration were complemented by metabolomics data from the same experiment. In doing so, the authors were able to bridge flux-centric and metabolomics-centric approaches and, in so doing, demonstrate that, under certain conditions, metabolites serving as pathway substrates in pathways defined as either modulators or sustainers display lower temporal variation with respect to all other metabolites (Töpfer et al. 2013). In addition, substantial evidence suggests that levels of specific metabolites such as Ala, pyruvate, 2-oxoglutarate, Gln, and spermidine are exceptionally stable across a wide range of cellular conditions. They also agree with observations that metabolite levels such as Ser coordinate the expression levels of genes encoding multiple steps of the pathways they themselves belong to (Timm et al. 2013). The high stability of these metabolites over a range of different stresses is in line with their requirements. It also emphasizes the fact that the most biologically relevant metabolites may be for metabolic regulation; this is an important point, since it is at odds with the manner in which the majority of the metabolomics community assesses their data. This observation additionally highlights the potential difficulties and challenges in interpreting data from a single level of the cellular hierarchy and thus provides further grounds for integrated models.

The rapid proliferation of plant and other organism genome-scale data makes it possible to study various cellular processes systematically. Because heterogeneous high-throughput data sets have been acquired from various "omics" technologies such as genomics, transcriptomics, proteomics, and metabolomics, it has become necessary to develop computational tools that can effectively integrate and analyze them. Microarrays and recently developed RNA-Seq technology have proven to be crucial tools for generating transcription data sets by simultaneously detecting thousands of gene expression. These data sets contain useful information to study gene functions in various ways including stress responses and developmental programs. Meanwhile, metabolomics, which investigates the profiles of all metabolites in an organism under specific conditions using techniques such as gas chromatography-mass spectrometry (GC-MS), has been regarded as an important research field in the postgenomic area, especially for plants due to their significant chemical diversity. New functional gene annotations have been added to various biological networks in recent years, including regulatory networks, networks of protein-protein interaction, and metabolic pathways. Despite these advances, under specific conditions, dynamic gene behaviors are still largely unexplored in specific pathways. Thus, in addition to integrating heterogeneous data sources, their analysis in the context of pathways is considered an essential step for functional studies of a complex biological system. Transcriptomic data are normally mapped into specific metabolic pathways in this type of analysis to investigate a set of genes 'coordinated behavior. It is important to develop effective tools for this type of analysis to systematically characterize and understand the dynamics of biochemical pathways by using multi-level information.

As detailed information on biological pathways has been developed, more complete and accurate pathways have been mapped, both experimentally and computationally. MetaCyc (http://metacyc.org/) and the Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.ad.jp/kegg/) are currently representative biochemical pathway databases. MetaCyc contains experimentally verified metabolic pathways and information on enzymes curated from scientific literature as well as predicted computationally predicted metabolic networks for more than 1600 different organisms (Krieger et al. 2004). KEGG is a knowledge base in terms of the network of genes and molecules resulting from their activities (Kanehisa et al. 2006). These databases are the primary resources that can be utilized to understand how genes and molecules are connected in biochemical pathways. Moreover, they can be combined with new resources or technologies for genomic and functional analysis, making it possible to expand previous databases and obtain increased depth and range of functions. For example, the database EGENES was developed to place genomic information, including ESTs of many plant species, into metabolic pathways and was integrated into the KEGG suite of databases (Kanehisa et al. 2006). Several analytical tools were developed to identify gene expression patterns that are responsible for potent biological effects by integrating large-scale transcriptomic data with various biological information such as pathways and related metabolites. Pathway Processor is a tool for visualizing metabolic pathway expression data and evaluating which transcriptional changes affect metabolic pathways. Specifically for plant species, several similar tools have been developed recently. Another plant species analysis system is KaPPA-View, a web-based tool used to display quantitative data for individual transcripts and metabolites on stored plant metabolic track maps stored in KEGG.

The Omics Viewer package in the Pathway Tools enables scientists to visualize for any organism of interest the large-scale gene expression and metabolomics data sets on metabolic pathways predicted by the Pathway Tools. KaPPA-View and Omics Viewer, however, provide very limited statistical analysis or project management functions. A web-based system, Plant MetGenMAP, has been developed that can identify significantly altered biochemical pathways and highly affected biological processes and predict functional roles of pathway genes from transcript and metabolite profile data sets and potential pathway-related regulatory motifs. Plant MetGenMAP is a user-friendly, powerful system of analysis that supports many functions of system biology analyzes in the context of biochemical pathways and terms of gene ontology (GO). It provides an analytical platform that allows for rapid and efficient exploration of highly altered pathways through intuitive visualization and robust statistical testing. Because it allows simultaneous analysis of transcriptional and metabolic changes for each pathway, the association between gene expression and biochemical changes in specific pathways can be easily inferred under specific conditions. Functional analysis of differentially controlled pathways can help define functional roles correctly of genes within pathways. Furthermore, the system embeds a function that can putatively identify major regulators related to changing transcripts and metabolites in specific pathways. Transcript and/or metabolite profiles of the model plant species Arabidopsis (Arabidopsis thaliana) and tomato (Solanum lycopersicum) have demonstrated the functions of Plant MetGenMAP. We present comprehensive results identified with Plant MetGenMAP, including differentially regulated metabolic pathways, pathway-related gene functions, putative regulators associated with these genes, and probabilistic associations between genes, metabolites, and phenotypes. MetaCyc contains experimentally determined biochemical pathways that can be used as a metabolism reference database. MetaCyc can be used together with the Pathway Tools software to predict the metabolic pathway complement of an annotated genome computationally. More than 60 plant-specific pathways have been added or updated in MetaCyc recently to increase the breadth of pathways and enzymes. Unlike MetaCyc, which contains metabolic data for a wide range of organisms, AraCyc is a species-specific database that contains only enzymes and pathways found in the Arabidopsis (*Arabidopsis thaliana*) model plant. The first computationally AraCyc (http://arabidopsis.org/tools/aracyc/) was the first computationally predicted plant metabolism database derived from MetaCyc. AraCyc has been under ongoing curation since its initial computational construction to improve data quality and increase the breadth of pathway coverage. Recently, twenty-eight pathways were curated manually from literature. Also recently, AraCyc's pathway predictions have been updated with the latest functional annotations of Arabidopsis genes using controlled vocabulary and literature evidence. Currently, AraCyc has 1418 unique genes mapped with 1156 literature citations on 204 pathways. The Omics Viewer, a user data visualization and analysis tool, makes it possible to paint a list of genes, enzymes or metabolites with experimental values on a diagram of AraCyc full path map. Other recent improvements to both MetaCyc and AraCyc include the implementation of an ontology of evidence used to provide data quality information, the expansion of the secondary pathway ontology metabolism node to accommodate the cure of secondary metabolic pathways, and the enhancement of the ontology of the cellular component for the storage and display of enzymes and pathways within s The MetaCyc database's goal is to catalog every experimental biochemical pathway for small molecule metabolism (Krieger et al. 2004). In EcoCyc(Keseler et al. 2005) a model organism database for Escherichia coli, MetaCyc was initialized with all the manually curated pathways. MetaCyc has subsequently added pathways from more than 300 organisms, and more than 90% of its pathways are manually curated with literature quotations and species information. The other 10% of pathways originally imported from the WIT database (http://www.cme.msu.edu/WIT/) are manually curated. MetaCyc can be used as a reference database in conjunction with the Pathway Tools software to create new Pathway Genome Databases (PGDB) from annotated genomes or genes. The Pathway Tools software contains three components: (1) PathoLogic, which matches an annotated genome's gene product names against enzymes and reactions in a reference database such as MetaCyc, and predicts the organism's pathways using a scoring algorithm; (2) Pathway/Genome Editor, which allows manual updating of the derived database and supports data sharing between derived d AraCyc was PathoLogic's first plant metabolism database to predict computationally using MetaCyc as the reference database(Mueller et al. 2003). AraCyc will eventually describe a complete set of metabolic pathways for Arabidopsis (Arabidopsis thaliana) and display genes and enzymes within their metabolic context with continued manual curation. Although there are still many pathways and enzymes to be manually curated in AraCyc, AraCyc is currently the most comprehensive genome-wide metabolic database available for a single plant species. Both databases can be accessed easily through the World Wide Web (http://metacyc.org and https://www.arabidopsis.org/biocyc/). With the release of the fully sequenced plant genomes of Arabidopsis (The Arabidopsis Genome Initiative 2000) and rice (*Oryza sativa*; International Rice Genome Sequencing Project, http://rgp.dna.affrc.go.jp/IRGSP) and the initiation of sequencing projects for many other plant species, there is a fast growing desire to place the sequenced and annotated genomes in a metabolic context. Indeed, the benefits of a species-specific metabolic pathway database are substantial: (1) it depicts the biochemical components of an organism; (2) it assists comparative studies of pathways across species and facilitates metabolic engineering to improve crop metabolism and traits; (3) it can be used as a platform to integrate and analyze data from large-scale experiments, such as gene expression, protein expression, or metabolite profiling; and (4) by presenting pathway steps lacking assigned genes or having genes assigned but solely based on computational prediction, we can discern what remains to be identified and experimentally characterized. Despite these advantages, it may be labor intensive and time consuming to create a pathway database manual de novo. SoyBase (http://soybase.agron.iastate.edu/), a soybean-specific metabolic pathway database (Glycine max), is the only other manually created species-specific plant pathway database. It is also possible to predict computationally species-specific plant pathway databases as a way to jump-start manual curation. A precise and comprehensive reference database is key to the quality of the derived databases for the predictions to be useful. Examples of comprehensive pathway databases include Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/; (Kanehisa and Goto 2000) and Enzymes and Metabolic Pathways (http://www.empproject.com/). As they stand, their usefulness as reference databases for plant genomes is somewhat limited for one or more of the following reasons: (1) pathways are not linked to literature quotations and therefore it is difficult to evaluate their accuracy; (2) individual path diagrams tend to be composites taken from several different species and are therefore not accurate for any single species; and (3) they are composites taken from several different species; The approaches taken, however, have been relatively straightforward to date and have generally not been carried out at a high level of spatial resolution. There are currently several methods for obtaining data from all the methods described here at the tissue, cellular, and even subcellular levels (Aharoni and Brandizzi 2012) while still technically challenging, it seems conceivable that such methods could provide data required to better understand the cell specialization of metabolism. In addition, methods to gain accurate metabolic flux estimates following $^{13}CO_2$ labeling have recently been established (Ma et al. 2014) but are not yet fully integrated with protein or transcript data. However, it is important to note that such experiments, albeit using (^{13}C)Glc as a precursor, have already been carried out in in vitro-cultivated *Brassica napus* embryos, providing considerable insight into the systems-level regulation of this organ (Schwender et al. 2015). Moreover, it seems highly likely that future research will draw more heavily on archived genomics data than it has up to now; thus, the continued availability and quality-control curation of such data sets is imperative if their value is to be fully exploited. To facilitate our understanding of transcriptome and metabolome data, there are several metabolic pathway databases available. The Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.ad.jp/kegg/) has a pathway database (PATHWAY) containing metabolite and gene information, as well as graphical representations of metabolic pathways and complexes derived from different biological processes. The metabolic pathways for 218 organisms, including Arabidopsis and rice, have been constructed to date. Organism-specific metabolic pathway maps can be generated according to assignment information on the KEGG/GENES database. The metabolic pathway reference database named MetaCyc (Krieger et al. 2004) pathways from 302 organisms (December 2004; http://metacyc.org/). The Arabidopsis pathway database AraCyc (http://www.arabidopsis.org/tools/aracyc/) was constructed by adding plant-specific pathways and reactions to basic pathway sets in the MetaCyc pathway collection (Mueller et al. 2003). While the comprehensive database contains metabolic pathway data that is representative of the plant kingdom, the pathways and reactions involved in alkaloid and isoflavonoid biosyntheses are not well represented, as these are not found in Arabidopsis. A relatively common feature of plants is that a single enzymatic reaction is often attributed to several homologous gene products. Multigene families in plant genomes are considerably more prevalent than in animal genomes (The Arabidopsis Genome Initiative 2000). Recent research has shown that multiple genes are not simple repeats, but exhibit a variety of gene expression and therefore have a variety of functions. For example, in Arabidopsis and rice, 33 and 29 member genes are the XTH gene family, a group of genes encoding xyloglucan endotransglucosylase/hydrolase involved in xyloglucan metabolism. In dicot and monocot plants, the individual gene members exhibit tissue-specific and stage-dependent expression of growth. One of the tools on the AraCyc database is capable of painting transcript data values onto the metabolic overview diagram. However, only representative data are used for the painting when multigene families are thought to be involved in single reactions. For painting, only representative data is used. On individual metabolic pathway maps, individual transcript data is not displayed. Also, a recent AraCyc version can represent metabolite data but only on the overview diagram. A user-driven tool, MAPMAN, has recently been developed to represent transcript data on pictorial diagrams, categorizing all genes of Arabidopsis on the basis of biological function (Thimm et al. 2004). Metabolites were also categorized to represent quantitative values of each metabolite on pictorial diagrams. However, since MAPMAN provides only several metabolic pathways, users need to use the user-driven tool to prepare their own diagrams. We created a set of comprehensive metabolic pathway maps for Arabidopsis as a complementary approach to AraCyc and MAPMAN in which 1263 metabolic reactions were grouped together. We have also developed a web-based tool for analyzing plant metabolic pathways, KaPPA-View, to display quantitative data on the same set of metabolic pathway maps for individual transcripts and metabolites. We adapted Scalable Vector Graphics (SVG) to facilitate dynamic document generation with rich graphical features and pathway map editing by the user to represent data from the transcript/metabolite. We demonstrated the utility of the KaPPA – View tool by displaying the transgenic plant data set that overexpresses the PAP1gene encoding a MYB transcription factor on the metabolic pathway maps (Tohge et al. 2005).

10.5 Conclusion and Perspectives

Biological information management and computer biology are becoming more diffuse and other categories will no doubt surface in the future as this field matures. In our society, our economy and our global environment, plant life plays important and diverse roles. For modern plant biotechnology, feeding the growing world population is a challenge. Crop yields have increased during the last century and will continue to improve as agronomy re-assorting the enhanced breeding and develop new biotechnological-engineered strategies. The onset of genomics is providing massive information to improve crop phenotypes. Accumulating sequence data enables detailed genome analysis through the use of friendly access to database and retrieval of information. Genetic and molecular genome co linearity allows efficient transfer of data revealing extensive conservation of genome organization between species. Genome research's goals are to identify sequenced genes and deduct their functions through metabolic analysis and reverse genetic screens from gene knockouts.

References

Abdallah, C., Dumas-Gaudot, E., Renaut, J., & Sergeant, K. (2012). Gel-based and gel-free quantitative proteomics approaches at a glance. *International Journal of Plant Genomics, 2012*, 494572. https://doi.org/10.1155/2012/494572.

Achnine, L., Huhman, D. V., Farag, M. A., Sumner, L. W., Blount, J. W., & Dixon, R. A. (2005). Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume Medicago truncatula. *The Plant Journal, 41*, 875–887.

Adams, M. D., Soares, M. B., Kerlavage, A. R., Fields, C., & Venter, J. C. (1993). Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nature Genetics, 4*, 373–380.

Aharoni, A., & Brandizzi, F. (2012). High-resolution measurements in plant biology. *The Plant Journal, 70*, 1–4.

Alfarano, C., Andrade, C. E., Anthony, K., Bahroos, N., Bajec, M., et al. (2005). The biomolecular interaction network database and related tools 2005 update. *Nucleic Acids Research, 33*, D418–D424.

Allen, J. E., Pertea, M., & Salzberg, S. L. (2004). Computational gene prediction using multiple sources of evidence. *Genome Research, 14*, 142–148.

Alonso, R., Salavert, F., Garcia-Garcia, F., Carbonell-Caballero, J., Bleda, M., et al. (2015). Babelomics 5.0: Functional interpretation for new generations of genomic data. *Nucleic Acids Research, 43*, W1): 117–W1): 121.

Alseekh, S., Tohge, T., Wendenberg, R., Scossa, F., Omranian, N., Li, J., Kleessen, S., Giavalisco, P., Pleban, T., Mueller-Roeber, B., et al. (2015). Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. *Plant Cell, 27*, 485–512.

Al-Shahrour, F., Minguez, P., Tarraga, J., et al. (2006). BABELOMICS: A systems biology perspective in the functional annotation of genome-scale experiments. *Nucleic Acids Research, 34*, W472–W476.

Altenbach, S. B., Vensel, W. H., & DuPont, F. M. (2010). Integration of transcriptomic and proteomic data from a single wheat cultivar provides new tools for understanding the roles of individual alpha gliadin proteins in flour quality and celiac disease. *Journal of Cereal Science, 52*, 143–151.

Anderson, D. C., Campbell, E. L., & Meeks, J. C. (2006). A soluble 3D LC/MS/MS proteome of the filamentous cyanobacterium Nostoc punctiforme. *Journal of Proteome Research, 5*, 3096–3104.

Andrade, A. E., Silva, L. P., Pereira, J. L., Noronha, E. F., Reis, F. B., Jr., Bloch, C., Jr., et al. (2008). In vivo proteome analysis of Xanthomonas campestris pv. Campestris in the interaction with the host plant Brassica oleracea. *FEMS Microbiology Letters, 281*, 167–174.

Anisimov, S. V. (2008). Serial analysis of gene expression (SAGE): 13 years of application in research. *Current Pharmaceutical Biotechnology, 9*, 338–350.

Ansorge, W. J. (2009). Next-generation DNA sequencing techniques. *Nature Biotechnology, 25*, 195–203.

Aoki, K., Yano, K., Suzuki, A., Kawamura, S., Sakurai, N., Sud, K., et al. (2010). Large-scale analysis of full-length cDNAs from the tomato (Solanum lycopersicum) cultivar Micro-Tom, a reference system for the Solanaceae genomics. *BMC Genomics, 11*, 210.

Arabidopsis Interactome Mapping Consortium. (2011). Evidence for network evolution in an Arabidopsis interactome map. *Science, 333*, 601–607.

Aranda, B., Achuthan, P., Alam-Faruque, Y., Armean, I., Bridge, A., Derow, C., et al. (2010). The IntAct molecular interaction database in 2010. *Nucleic Acids Research, 38*, D525–D531.

Araújo, W. L., Ishizaki, K., Nunes-Nesi, A., Larson, T. R., Tohge, T., Krahnert, I., Witt, S., Obata, T., Schauer, N., Graham, I. A., et al. (2010). Identification of the 2-hydroxyglutarate and isovaleryl-CoA dehydrogenases as alternative electron donors linking lysine catabolism to the electron transport chain of Arabidopsis mitochondria. *Plant Cell, 22*, 1549–1563.

Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., et al. (2000). Gene ontology: Tool for the unification of biology. The gene ontology consortium. *Nature Genetics, 25*, 25–29.

Assmann, S. M., & Albert, R. (2009). Discrete dynamic modeling with asynchronous update, or how to model complex systems in the absence of quantitative information. *Methods in Molecular Biology, 553*, 207–225.

Avraham, S., Tung, C. W., Ilic, K., et al. (2008). The plant ontology database: A community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucleic Acids Research, 36*(1), D449–D454.

Babu, M. M., Luscombe, N. M., Aravind, L., et al. (2004). Structure and evolution of transcriptional regulatory networks. *Current Opinion in Structural Biology, 14*(3), 283–291.

Bagnarol, E., Popovici, J., Alloisio, N., Marechal, J., Pujic, P., Normand, P., et al. (2007). Differential Frankia protein patterns induced by phenolic extracts from Myricaceae seeds. *Physiologia Plantarum, 130*, 380–390.

Barabasi, A. L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics, 5*, 101–115.

Barakat, A., Wall, P. K., Diloreto, S., Depamphilis, C. W., & Carlson, J. E. (2007). Conservation and divergence of microRNAs in Populus. *BMC Genomics, 8*, 481.

Bard, J. B., & Rhee, S. Y. (2004). Ontologies in biology: Design, applications and future challenges. *Nature Reviews. Genetics, 5*, 213–222.

Bard, J., Rhee, S. Y., & Ashburner, M. (2005). An ontology for cell types. *Genome Biology, 6*, R21.

Barrett, T., Troup, D. B., Wilhite, S. E., Ledoux, P., Rudnev, D., Evangelista, C., et al. (2009). NCBI GEO: Archive for high-throughput functional genomic data. *Nucleic Acids Research, 37*, D885–D890.

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

Bedell, J. A., Budiman, M. A., Nunberg, A., Citek, R. W., Robbins, D., et al. (2005). Sorghum genome sequencing by methylation filtration. *PLoS Biology, 3*, e13.

Benedict, C., Geisler, M., Trygg, J., et al. (2006). Consensus by democracy. Using meta-analyses of microarray and genomic data to model the cold acclimation signaling pathway in Arabidopsis. *Plant Physiology, 141*(4), 1219–1232.

Benedito, V. A., Torres-Jerez, I., Murray, J. D., Andriankaja, A., Allen, S., Kakar, K., et al. (2008). A gene expression atlas of the model legume Medicago truncatula. *The Plant Journal, 55*, 504–513.

Bernardo, A. N., Bradbury, P. J., Ma, H., Hu, S., Bowden, R. L., Buckler, E. S., et al. (2009). Discovery and mapping of single feature polymorphisms in wheat using Affymetrix arrays. *BMC Genomics, 10*, 251.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American, 284*, 34–43.

Bevan, M. (1997). Objective: The complete sequence of a plant genome. *Plant Cell, 9*, 476–478.

Bhalla, R., Narasimhan, K., & Swarup, S. (2005). Metabolomics and its role in understanding cellular responses in plants. *Plant Cell Reports, 24*, 562–571. https://doi.org/10.1007/s00299-005-0054-9.

Bino, R. J., Hall, R. D., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B. J., Mendes, P., Roessner-Tunali, U., Beale, M. H., Trethewey, R. N., Lange, B. M., Wurtele, E. S., & Sumner, L. W. (2004). Potential of metabolomics as a functional genomics tool. *Trends in Plant Science, 9*, 418–425.

Blais, A., & Dynlacht, B. D. (2005). Constructing transcriptional regulatory networks. *Genes & Development, 19*(13), 1499–1511.

Blake-Kalff, M. M. A., Harrison, K. R., Hawkesford, M. J., Zhao, F. J., & McGrath, S. P. (1998). Distribution of sulfur within oilseed rape leaves in response to sulfur deficiency during vegetative growth. *Plant Physiology, 118*, 1337–1344.

Blakes, J., Twycross, J., Romero, F. J., et al. (2011). The Infobiotics Workbench: An integrated in silico modelling platform for systems and synthetic biology. *Bioinformatics, 27*(23), 3323–3324.

Blaschke, C., Krallinger, M., Leon, E., & Valencia, A. (2005). Evaluation of biocreative assessment of task 2. *BMC Bioinformatics, 6*, S16.

Blazej, R. G., Paegel, B. M., & Mathies, R. A. (2003). Polymorphism ratio sequencing: A new approach for single nucleotide polymorphism discovery and genotyping. *Genome Research, 13*, 287–293.

Blazejczyk, M., Miron, M., & Nadon, R. (2007). FlexArray: A statistical data analysis software for gene expression microarrays. *Genome Quebec*. *Montreal, 39*, 1208–1216.

Boguski, M. S., & Schuler, G. D. (1995). ESTablishing a human transcript map. *Nature Genetics, 10*, 369–371.

Boguski, M. S., Lowe, T. M., & Tolstoshev, C. M. (1993). dbEST—Database for 'expressed sequence tags'. *Nature Genetics, 4*, 332–333.

Bolger, A., Scossa, F., Bolger, M. E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S., Sørensen, I., Lichtenstein, G., et al. (2014). The genome of the stress-tolerant wild tomato species Solanum pennellii. *Nature Genetics, 46*, 1034–1038.

Boone, C., Bussey, H., & Andrews, B. J. (2007). Exploring genetic interactions and networks with yeast. *Nature Reviews Genetics, 8*(6), 437–449.

Brady, S. M., & Provart, N. J. (2009). Web-queryable large-scale data sets for hypothesis generation in plant biology. *Plant Cell, 21*, 1034–1051.

Brady, S. M., Orlando, D. A., Lee, J. Y., Wang, J. Y., Koch, J., Dinneny, J. R., et al. (2007). A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science, 318*, 801–806.

Breitkreutz, B. J., Stark, C., & Tyers, M. (2003). Osprey: A network visualization system. *Genome Biology, 4*(3), R22.

Brendel, V., & Zhu, W. (2002). Computational modeling of gene structure in Arabidopsis thaliana. *Plant Molecular Biology, 48*, 49–58.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., et al. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology, 18*, 630–634.

Brkljacic, J., Grotewold, E., Scholl, R., Mockler, T., Garvin, D. F., Vain, P., et al. (2011). Brachypodium as a model for the grasses: Today and the future. *Plant Physiology, 157*, 3–13.

Brown, J. R., & Sanseau, P. (2005). A computational view of microRNAs and their targets. *Drug Discovery Today, 10*, 595–601.

Buck, M. J., & Lieb, J. D. (2004). ChIP-chip: Considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics, 83*, 349–360.

Buttner, D., & Bonas, U. (2002). Getting across bacterial type III effector proteins on their way to the plant cell. *The EMBO Journal, 21*, 5313–5322.

Caicedo, A. L., Williamson, S. H., Hernandez, R. D., Boyko, A., Fledel-Alon, A., York, T. L., et al. (2007). Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genetics, 3*, 1745–1756.

Calla, B., Vuong, T., Radwan, O., Hartman, G. L., & Clough, S. J. (2009). Gene expression profiling soybean stem tissue early response to Sclerotinia sclerotiorum and in silico mapping in relation to resistance markers. *The Plant Genome Journal, 2*(2), 149–166.

Carollo, V., Matthews, D. E., Lazo, G. R., Blake, T. K., Hummel, D. D., Lui, N., et al. (2005). GrainGenes 2.0. An improved resource for the small-grains community. *Plant Physiology, 139*, 643–651.

Carrari, F., Baxter, C., Usadel, B., Urbanczyk-Wochniak, E., Zanor, M. I., NunesNesi, A., Nikiforova, V., Centero, D., Ratzka, A., Pauly, M., et al. (2006). Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. *Plant Physiology, 142*, 1380–1396.

Casati, P., Campi, M., Morrow, D. J., Fernandes, J. F., & Walbot, V. (2011). Transcriptomic, proteomic and metabolomic analysis of UV-B signaling in maize. *BMC Genomics, 12*, 321.

Caspi, R., Altman, T., Dale, J. M., et al. (2010). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research, 38*(1), D473–D479.

Chatziioannou, A., Moulos, P., & Kolisis, F. N. (2009). Gene ARMADA: An integrated multi-analysis platform for microarray data implemented in MATLAB. *BMC Bioinformatics, 10*(1), 354.

Chellappan, P., & Jin, H. (2009). Discovery of plant microRNAs and short-interfering RNAs by deep parallel sequencing. *Methods in Molecular Biology, 495*, 121–132.

Chen, T., Kao, M. Y., Tepel, M., Rush, J., & Church, G. M. (2001). A dynamic programming approach to de novo peptide sequencing via tandem mass spectrometry. *Journal of Computational Biology, 8*, 325–337.

Chisholm, S. T., Coaker, G., Day, B., & Staskawicz, B. J. (2006). Host microbe interactions: Shaping the evolution of the plant immune response. *Cell, 124*, 803–814.

Chodavarapu, R. K., Feng, S., Bernatavichute, Y. V., Chen, P. Y., Stroud, H., Yu, Y., et al. (2010). Relationship between nucleosome positioning and DNA methylation. *Nature, 466*, 388–392.

Choi, H., & Pavelka, N. (2011). When one and one gives more than two: Challenges and opportunities of integrative omics. *Frontiers in Genetics, 2*, 105.

Close, T. J., Bhat, P. R., Lonardi, S., Wu, Y., Rostoks, N., Ramsay, L., et al. (2009). Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics, 10*, 582. https://doi.org/10.1186/1471-2164-10-582.

Coen, E. S., & Meyerowitz, E. M. (1991). The war of the whorls: Genetic interactions controlling flower development. *Nature, 353*(6339), 31–37.

Cohen, A. M., & Hersh, W. R. (2005). A survey of current work in biomedical text mining. *Briefings in Bioinformatics, 6*, 57–71.

Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., et al. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature, 452*, 215–219.

Cope, L. M., Irizarry, R. A., Jaffee, H. A., Wu, Z., & Speed, T. P. (2004). A benchmark for Affymetrix GeneChip expression measures. *Bioinformatics, 20*, 323–331.

Dalby, P. A. (2003). Optimising enzyme function by directed evolution. *Current Opinion in Structural Biology, 13*, 500–505.

Dam, S., Laursen, B. S., Ornfelt, J. H., Jochimsen, B., Staerfeldt, H. H., Friis, C., et al. (2009). The proteome of seed development in the model legume Lotus japonicus. *Plant Physiology, 149*, 1325–1340.

Dancik, V., Addona, T. A., Clauser, K. R., Vath, J. E., & Pevzner, P. A. (1999). De novo peptide sequencing via tandem mass spectrometry. *Journal of Computational Biology, 6*, 327–342.

Davies, P. J. (Ed.). (2004). *Plant hormones: Biosynthesis, signal transduction, action*. Dordrecht: Kluwer Academic Publishers.

De Bodt, S., Maere, S., & Van de Peer, Y. (2005). Genome duplication and the origin of angiosperms. *Trends in Ecology & Evolution, 20*, 591–597.

de Folter, S., Immink, R. G., Kieffer, M., et al. (2005). Comprehensive interaction map of the Arabidopsis MADS box transcription factors. *Plant Cell, 17*(5), 1424–1433.

de Hoon, M., & Hayashizaki, Y. (2008). Deep cap analysis gene expression (CAGE): Genome-wide identifi cation of promoters, quantifi cation of their expression, and network inference. *BioTechniques, 44*, 627–628, 630, 632.

De Keersmaecker, S. C., Thijs, I., Vanderleyden, J., et al. (2006). Integration of omics data: How well does it work for bacteria? *Molecular Microbiology, 62*(5), 1239–1250.

Delker, C., Poschl, Y., Raschke, A., Ullrich, K., Ettingshausen, S., Hauptmann, V., et al. (2010). Natural variation of transcriptional auxin response networks in Arabidopsis thaliana. *Plant Cell, 22*, 2184–2200.

Delmotte, N., Ahrens, C. H., Knief, C., Qeli, E., Koch, M., Fischer, H.-M., et al. (2010). An integrated proteomics and transcriptomics reference data set provides new insights into the Bradyrhizobium japonicum bacteroid metabolism in soybean root nodules. *Proteomics, 10*, 1391–1400.

Depuydt, S., & Hardtke, C. S. (2011). Hormone signalling crosstalk in plant growth regulation. *Current Biology, 21*, R365–R373.

Dhar, P. K., Zhu, H., & Mishra, S. K. (2004). Computational approach to systems biology: From fraction to integration and beyond. *IEEE Transactions on NanoBioscience, 3*(3), 144–152.

Di, X., Matsuzaki, H., Webster, T. A., Hubbell, E., Liu, G., et al. (2005). Dynamic model based algorithms for screening and genotyping over 100 K SNPs on oligonucleotide microarrays. *Bioinformatics, 21*, 1958–1963.

Digman, M. A., Brown, C. M., Sengupta, P., Wiseman, P. W., Horwitz, A. R., & Gratton, E. (2005). Measuring fast dynamics in solutions and cells with a laser scanning microscope. *Biophysical Journal, 89*, 1317–1327.

Ding, J., Viswanathan, K., Berleant, D., Hughes, L., Wurtele, E. S., et al. (2005). Using the biological taxonomy to access biological literature with PathBinderH. *Bioinformatics, 21*, 2560–2562.

Donaldson, I., Martin, J., de Bruijn, B., Wolting, C., Lay, V., et al. (2003). PreBIND and Textomy—Mining the biomedical literature for protein-protein interactions using a support vector machine. *BMC Bioinformatics, 4*, 11.

Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science, 284*, 2124–2129.

Drăghici, S. (2011). *Statistics and data analysis for microarrays using R and bioconductor*. Boca Raton: CRC Press.

Driever, S. M., & Kromdijk, J. (2013). Will C3 crops enhanced with the C4 CO2- concentrating mechanism live up to their full potential (yield)? *Journal of Experimental Botany, 64*, 3925–3935. https://doi.org/10.1093/jxb/ert103.

Duvick, J., Fu, A., Muppirala, U., Sabharwal, M., Wilkerson, M. D., Lawrence, C. J., et al. (2008). PlantGDB: A resource for comparative plant genomics. *Nucleic Acids Research, 36*, D959–D965.

Edwards, J. S., & Palsson, B. O. (2000). The Escherichia coli MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the United States of America, 97*, 5528–5533.

Eilbeck, K., Lewis, S. E., Mungall, C. J., Yandell, M., Stein, L., et al. (2005). The sequence ontology: A tool for the unification of genome annotations. *Genome Biology, 6*, R44.

Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America, 95*, 14863–14868.

Emmert-Buck, M. R., Bonner, R. F., Smith, P. D., Chuaqui, R. F., Zhuang, Z., et al. (1996). Laser capture microdissection. *Science, 274*, 998–1001.

Enfissi, E. M., Barneche, F., Ahmed, I., Lichtle, C., Gerrish, C., McQuinn, R. P., et al. (2010). Integrative transcript and metabolite analysis of nutritionally enhanced DE-ETIOLATED1 downregulated tomato fruit. *Plant Cell, 22*, 1190–1215.

Fazzari, M. J., & Greally, J. M. (2004). Epigenomics: Beyond CpG islands. *Nature Reviews Genetics, 5*, 446–455.

Feltus, F. A., Wan, J., Schulze, S. R., Estill, J. C., Jiang, N., & Paterson, A. H. (2004). An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. *Genome Research, 14*, 1812–1819.

Fiehn, O. (2001). Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comparative and Functional Genomics, 2*, 155–168.

Forrester, J. W. (1958). Industrial dynamics: A major breakthrough for decision makers. *Harvard Business Review, 36*(4), 37–66.

Forrester, J. W. (1961). *Industrial dynamics*. Portland: Productivity Press.

Foster, I. (2002). What is the grid? A three point checklist. In *GRIDToday* (p. 4). Chicago: Argonne National Lab & University of Chicago.

Fouracre, J. P., Ando, S., & Langdale, J. A. (2014). Cracking the Kranz enigma with systems biology. *Journal of Experimental Botany, 65*(13), 3327–3339. https://doi.org/10.1093/jxb/eru015.

Fu, J., Keurentjes, J. J., Bouwmeester, H., America, T., Verstappen, F. W., Ward, J. L., et al. (2009). System-wide molecular evidence for phenotypic buffering in Arabidopsis. *Nature Genetics, 41*, 166–167.

Fujimura, Y., Kurihara, K., Ida, M., Kosaka, R., Miura, D., Wariishi, H., et al. (2011). Metabolomics-driven nutraceutical evaluation of diverse green tea cultivars. *PLoS One, 6*, e23426.

Fujita, M., Horiuchi, Y., Ueda, Y., Mizuta, Y., Kubo, T., Yano, K., et al. (2010). Rice expression atlas in reproductive development. *Plant & Cell Physiology, 51*, 2060–2081.

Fukuda, H., & Higashiyama, T. (2011). Diverse functions of plant peptides: Entering a new phase. *Plant & Cell Physiology, 52*, 1–4.

Fukuda, H., Hirakawa, Y., & Sawa, S. (2007). Peptide signaling in vascular development. *Current Opinion in Plant Biology, 10*, 477–482.

Fukushima, A., Kanaya, S., & Nishida, K. (2014). Integrated network analysis and effective tools in plant systems biology. *Frontiers in Plant Science, 5*, 598.

Galindo González, L. M., El Kayal, W., Ju, C. J. T., et al. (2012). Integrated transcriptomic and proteomic profiling of white spruce stems during the transition from active growth to dormancy. *Plant, Cell & Environment, 35*(4), 682–701.

Garcia-Hernandez, M., Berardini, T. Z., Chen, G., Crist, D., Doyle, A., et al. (2002). TAIR: A resource for integrated Arabidopsis data. *Functional & Integrative Genomics, 2*, 239–253.

Garcia-Seco, D., Chiapello, M., Bracale, M., Pesce, C., Bagnaresi, P., et al. (2017). Transcriptome and proteome analysis reveal new insight into proximal and distal responses of wheat to foliar infection by Xanthomonas translucens. *Scientific Reports, 7*, 10157.

Gechev, T. S., Benina, M., Obata, T., Tohge, T., Sujeeth, N., Minkov, I., Hille, J., Temanni, M. R., Marriott, A. S., Bergström, E., et al. (2013). Molecular mechanisms of desiccation tolerance in the resurrection glacial relic Haberlea rhodopensis. *Cellular and Molecular Life Sciences, 70*, 689–709.

Gehlenborg, N., O'Donoghue, S. I., Baliga, N. S., et al. (2010). Visualization of omics data for systems biology. *Nature Methods, 7*, S56–S68.

Gibbs, R. A., & Weinstock, G. M. (2003). Evolving methods for the assembly of large genomes. *Cold Spring Harbor Symposia on Quantitative Biology, 68*, 189–194.

Glaubitz, U., Li, X., Schaedel, S., Erban, A., Sulpice, R., Kopka, J., et al. (2017). Integrated analysis of rice transcriptomic and metabolomic responses to elevated night temperatures identifies sensitivity-and tolerance-related profiles. *Plant, Cell & Environment, 40*(1), 121–137.

Glinski, M., & Weckwerth, W. (2006). The role of mass spectrometry in plant systems biology. *Mass Spectrometry Reviews, 25*, 173–214. https://doi.org/10.1002/mas.20063.

Goda, H., Sasaki, E., Akiyama, K., Maruyama-Nakashita, A., Nakabayashi, K., Li, W., et al. (2008). The AtGenExpress hormone and chemical treatment data set: Experimental design, data evaluation, model. *The Plant Journal, 55*(3), 526–542.

Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., et al. (2002). A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science, 296*, 92–100.

Gomez-Gomez, L., Felix, G., & Boller, T. (1999). A single locus determines sensitivity to bacterial flagellin in *Arabidopsis thaliana*. *The Plant Journal, 18*, 277–284. https://doi.org/10.1046/j.1365-313X.1999.00451.x.

Gong, C. Y., & Wang, T. (2013). Proteomic evaluation of genetically modified crops: Current status and challenges. *Frontiers in Plant Science, 4*, 41. https://doi.org/10.3389/fpls.2013.00041.

Gonzalez, N., De Bodt, S., Sulpice, R., et al. (2010). Increased leaf size: Different means to an end. *Plant Physiology, 153*, 1261–1279.

Gorg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., et al. (2000). The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis, 21*, 1037–1053.

Gourion, B., Rossignol, M., & Vorholt, J. A. (2006). A proteomic study of Methylobacterium extorquens reveals a response regulator essential for epiphytic growth. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 13186–13191.

Grant, D., Nelson, R. T., Cannon, S. B., & Shoemaker, R. C. (2010). SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Research, 38*, D843–D846.

Gras, R., & Muller, M. (2001). Computational aspects of protein identification by mass spectrometry. *Current Opinion in Molecular Therapeutics, 3*, 526–532.

Grimsrud, P. A., den Os, D., Wenger, C. D., Swaney, D. L., Schwartz, D., Sussman, M. R., et al. (2010). Large-scale phosphoprotein analysis in Medicago truncatula roots provides insight into in vivo kinase activity in legumes. *Plant Physiology, 152*, 19–28.

Gygi, S. P., Rochon, Y., Franza, B. R., et al. (1999). Correlation between protein and mRNA abundance in yeast. *Molecular and Cellular Biology, 19*, 1720–1730.

Hannah, M. A., Caldana, C., Steinhauser, D., Balbo, I., Fernie, A. R., & Willmitzer, L. (2010). Combined transcript and metabolite profiling of Arabidopsis grown under widely variant growth conditions facilitates the identification of novel metabolite-mediated regulation of gene expression. *Plant Physiology, 152*, 2120–2129.

Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., et al. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research, 32*, D258–D261.

He, D., & Yang, P. (2013). Proteomics of rice seed germination. *Frontiers in Plant Science, 4*, 246. https://doi.org/10.3389/fpls.2013.00246.

He, G., Zhu, X., Elling, A. A., Chen, L., Wang, X., Guo, L., et al. (2010). Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell, 22*, 17–33.

He, G., Elling, A. A., & Deng, X. W. (2011). The epigenome and plant development. *Annual Review of Plant Biology, 62*, 411–435.

He, G., Chen, B., Wang, X., et al. (2013). Conservation and divergence of transcriptomic and epigenomic variation in maize hybrids. *Genome Biology, 14*(6), R57.

Heesacker, A., Kishore, V. K., Gao, W., Tang, S., Kolkman, J. M., Gingle, A., et al. (2008). SSRs and INDELs mined from the sunfl ower EST database: Abundance, polymorphisms, and cross-taxa utility. *Theoretical and Applied Genetics, 117*, 1021–1029.

Heinrich, R., & Schuster, S. (1996). *The regulation of cellular systems*. New York: Chapman & Hall.

Heisler, M. G., Ohno, C., Das, P., Sieber, P., Reddy, G. V., et al. (2005). Patterns of auxin transport and gene expression during primordium development revealed by live imaging of the Arabidopsis inflorescence meristem. *Current Biology, 15*, 1899–1911.

Helmy, M., Tomita, M., & Ishihama, Y. (2011). OryzaPG-DB: Rice proteome database based on shotgun proteogenomics. *BMC Plant Biology, 11*, 63.

Hirai, M. Y., Yano, M., Goodenowe, D. B., Kanaya, S., Kimura, T., Awazuhara, M., et al. (2004). Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 10205–10210.

Hobo, T., Suwabe, K., Aya, K., Suzuki, G., Yano, K., Ishimizu, T., et al. (2008). Various spatio-temporal expression profiles of anther-expressed genes in rice. *Plant & Cell Physiology, 49*, 1417–1428.

Hoffmann, R., & Valencia, A. (2004). A gene network for navigating the literature. *Nature Genetics, 36*, 664.

Hoops, S., Sahle, S., Gauges, R., et al. (2006). COPASI—A complex pathway simulator. *Bioinformatics, 22*(24), 3067–3074.

Hori, K., Sato, K., & Takeda, K. (2007). Detection of seed dormancy QTL in multiple mapping populations derived from crosses involving novel barley germplasm. *Theoretical and Applied Genetics, 115*, 869–876.

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., et al. (2009). High-throughput genotyping by whole-genome resequencing. *Genome Research, 19*, 1068–1076.

Hucka, M., Finney, A., Sauro, H. M., et al. (2003). The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics, 19*(4), 524–531.

Hucka, M., Finney, A., Bornstein, B. J., Keating, S. M., Shapiro, B. E., et al. (2004). Evolving a lingua franca and associated software infrastructure for computational systems biology: The Systems Biology Markup Language (SBML) Project. *Systematic Biology, 1*, 41–53.

Hulsen, T., de Vlieg, J., & Groenen, P. M. (2006). PhyloPat: Phylogenetic pattern analysis of eukaryotic genes. *BMC Bioinformatics, 7*, 398.

Iijima, Y., Nakamura, Y., Ogata, Y., Tanaka, K., Sakurai, N., Suda, K., et al. (2008). Metabolite annotations based on the integration of mass spectral information. *The Plant Journal, 54*, 949–962.

Ikeda, S., Okubo, T., Anda, M., Nakashita, H., Yasuda, M., Sato, S., et al. (2010). Community- and genome-based views of plant-associated bacteria: Plant–bacterial interactions in soybean and rice. *Plant & Cell Physiology, 51*, 1398–1410.

Inada, D. C., Bashir, A., Lee, C., Thomas, B. C., Ko, C., et al. (2003). Conserved noncoding sequences in the grasses. *Genome Research, 13*, 2030–2041.

International Brachypodium Initiative. (2010). Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature, 463*, 763–768.

International Rice Genome Sequencing Project. (2005). The map-based sequence of the rice genome. *Nature, 436*, 793–800.

Itoh, T., Tanaka, T., Barrero, R. A., Yamasaki, C., Fujii, Y., Hilton, P. B., et al. (2007). Curated genome annotation of Oryza sativa ssp. Japonica and comparative genome analysis with Arabidopsis thaliana. *Genome Research, 17*, 175–183.

Izawa, T., Mihara, M., Suzuki, Y., Gupta, M., Itoh, H., Nagano, A. J., et al. (2011). Os-GIGANTEA confers robust diurnal rhythms on the global transcriptome of rice in the field. *Plant Cell, 23*, 1741–1755.

Jacobs, J. M., Babujee, L., Meng, F., Milling, A., & Allen, C. (2012). The in planta transcriptome of Ralstonia solanacearum: Conserved physiological and virulence strategies during bacterial wilt of tomato. *MBio, 3*, e00114–e00112.

Janeway, C. A., & Medzhitov, R. (2002). Innate immune recognition. *Annual Review of Immunology, 20*, 197–216.

Jiang, N., Bao, Z., Zhang, X., Eddy, S. R., & Wessler, S. R. (2004). Pack-MULE transposable elements mediate gene evolution in plants. *Nature, 431*, 569–573.

Jiao, Y., Lau, O. S., & Deng, X. W. (2007). Light-regulated transcriptional networks in higher plants. *Nature Reviews Genetics, 8*(3), 217–230.

Jones, J. D., & Dangl, J. L. (2006). The plant immune system. *Nature, 444*, 323–329.

Jorrín-Novo, J. V., Pascual, J., Sánchez-Lucas, R., Romero-Rodríguez, M. C., Rodríguez-Ortega, M. J., Lenz, C., et al. (2015). Fourteen years of plant proteomics reflected in proteomics: Moving from model species and 2DE−based approaches to orphan species and gel-free platforms. *Proteomics, 15*, 1089–1112. https://doi.org/10.1002/pmic.201400349.

Joshi-Tope, G., Gillespie, M., Vastrik, I., et al. (2005). Reactome: A knowledgebase of biological pathways. *Nucleic Acids Research, 33*(1), D428–D432.

Joyce, A. R., & Palsson, B. O. (2006). The model organism as a system: Integrating 'omics' data sets. *Nature Reviews. Molecular Cell Biology, 7*, 198–210.

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research, 28*, 27–30.

Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., & Hirakawa, M. (2006). From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Research, 34*, D354–D357.

Kanehisa, M., Goto, S., Sato, Y., et al. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research, 40*(D1), D109–D114.

Kang, J. H., Gonzales-Vigil, E., Matsuba, Y., Pichersky, E., & Barry, C. S. (2014). Determination of residues responsible for substrate and product specificity of Solanum habrochaites short-chain cis-prenyltransferases. *Plant Physiology, 164*, 80–91.

Kanno, Y., Jikumaru, Y., Hanada, A., Nambara, E., Abrams, S. R., Kamiya, Y., et al. (2010). Comprehensive hormone profiling in developing Arabidopsis seeds: Examination of the site of ABA biosynthesis, ABA transport and hormone interactions. *Plant & Cell Physiology, 51*, 1988–2001.

Karlin, S., & Altschul, S. F. (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proceedings of the National Academy of Sciences of the United States of America, 87*, 2264–2268.

Katari, M. S., Nowicki, S. D., Aceituno, F. F., et al. (2010). VirtualPlant: A software platform to support systems biology research. *Plant Physiology, 152*(2), 500–515.

Kawaguchi, M., & Minamisawa, K. (2010). Plant–microbe communications for symbiosis. *Plant & Cell Physiology, 51*(9), 1377–1380.

Kell, D. B., Brown, M., Davey, H. M., Dunn, W. B., Spasic, I., & Oliver, S. G. (2005). Metabolic footprinting and systems biology: The medium is the message. *Nature Reviews. Microbiology, 3*, 557–565.

Keseler, I. M., Collado-vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I. T., Peralta-Gil, M., & Karp, P. D. (2005). EcoCyc: A comprehensive database resource for Escherichia coli. *Nucleic Acids Research, 33*, D334–D337.

Khatri, P., & Draghici, S. (2005). Ontological analysis of gene expression data: Current tools, limitations, and open problems. *Bioinformatics, 21*, 3587–3595.

Khojasteh, M., Khahani, B., Taghavi, M., & Tvakol, E. (2017). Identification and characterization of responsive genes in rice during compatible interactions with pathogenic pathovars of Xanthomonas oryzae. *European Journal of Plant Pathology, 151*(1), 141–153.

Kim, H. J., Baek, K. H., Lee, S. W., Kim, J., Lee, B. W., Cho, H. S., et al. (2008). Pepper EST database: Comprehensive in silico tool for analyzing the chili pepper (Capsicum annuum) transcriptome. *BMC Plant Biology, 8*, 101.

Kitano, H. (2002). Systems biology: A brief overview. *Science, 295*(5560), 1662–1664.

Klamt, S., Stelling, J., Ginkel, M., & Gilles, E. D. (2003). FluxAnalyzer: Exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics, 19*, 261–269.

Koenig, D., Jiménez-Gómez, J. M., Kimura, S., Fulop, D., Chitwood, D. H., Headland, L. R., Kumar, R., Covington, M. F., Devisetty, U. K., Tat, A. V., et al. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proceedings of the National Academy of Sciences of the United States of America, 110*, E2655–E2662.

Kogel, K. H., Voll, L. M., Schäfer, P., et al. (2010). Transcriptome and metabolome profiling of field-grown transgenic barley lack induced differences but show cultivar-specific variances. *Proceedings of the National Academy of Sciences of the United States of America, 107*(14), 6198–6203.

Kojima, M., Kamada-Nobusada, T., Komatsu, H., Takei, K., Kuroha, T., Mizutani, M., et al. (2009). Highly sensitive and high-throughput analysis of plant hormones using MS-probe modification and liquid chromatography-tandem mass spectrometry: An application for hormone profiling in Oryza sativa. *Plant & Cell Physiology, 50*(7), 1201–1214.

Komatsu, S., Mock, H. P., Yang, P., & Svensson, B. (2013). Application of proteomics for improving crop protection/artificial regulation. *Frontiers in Plant Science, 4*, 522. https://doi.org/10.3389/fpls.2013.00522. Published 2013 Dec 19.

Kondou, Y., Higuchi, M., Takahashi, S., Sakurai, T., Ichikawa, T., Kuroda, H., et al. (2009). Systematic approaches to using the FOX hunting system to identify useful rice genes. *The Plant Journal, 57*, 883–894.

Kosová, K., Vítámvás, P., Prášil, I. T., & Renaut, J. (2011). Plant proteome changes under abiotic stress–contribution of proteomics studies to understanding plant stress response. *Journal of Proteomics, 74*, 1301–1322. https://doi.org/10.1016/j.jprot.2011.02.006.

Kouchi, H., Imaizumi-Anraku, H., Hayashi, M., Hakoyama, T., Nakagawa, T., Umehara, Y., et al. (2010). How many peas in a pod? Legume genes responsible for mutualistic symbioses underground. *Plant & Cell Physiology, 51*, 1381–1397.

Krallinger, M., Rodriguez-Penagos, C., Tendulkar, A., et al. (2009). PLAN2L: A web tool for integrated text mining and literature-based bioentity relation extraction. *Nucleic Acids Research, 37*(2), W160–W165.

Krieger, C. J., Zhang, P., Mu¨ller, L. A., Wang, A., Paley, S., Arnaud, M., Pick, J., Rhee, S. Y., & Karp, P. D. (2004). MetaCyc: A multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Research, 32*, D438–D442.

Kusano, M., Tohge, T., Fukushima, A., Kobayashi, M., Hayashi, N., Otsuki, H., et al. (2011). Metabolomics reveals comprehensive reprogramming involving two independent metabolic responses of Arabidopsis to UV-B light. *The Plant Journal, 67*, 354–369.

Laakso, M., & Hautaniemi, S. (2010). Integrative platform to translate gene sets to networks. *Bioinformatics, 26*(14), 1802–1803.

Langridge, P., & Fleury, D. (2011). Making the most of 'omics' for crop breeding. *Trends in Biotechnology, 29*, 33–40. https://doi.org/10.1016/j.tibtech.2010.09.006.

Le Novere, N., Bornstein, B., Broicher, A., et al. (2006). BioModels database: A free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research, 34*(1), D689–D691.

Lee, S. W., Jeong, K. S., Han, S. W., Lee, S. E., Phee, B. K., Hahn, T. R., et al. (2008). The Xanthomonas oryzae pv. oryzae PhoPQ twocomponent system is required for AvrXA21 activity, hrpG expression, and virulence. *Journal of Bacteriology, 190*, 2183–2197.

Lelandais-Briere, C., Naya, L., Sallet, E., Calenge, F., Frugier, F., Hartmann, C., et al. (2009). Genome-wide Medicago truncatula small RNA analysis revealed novel microRNAs and isoformsdifferentially regulated in roots and nodules. *Plant Cell, 21*, 780–2796.

Lewin, B. (2003). *Genes VIII*. Upper Saddle River: Prentice Hall.

Li, F., Kitashiba, H., Inaba, K., & Nishio, T. (2009). A Brassica rapa linkage map of EST-based SNP markers for identification of candidategenes controlling flowering time and leaf morphological traits. *DNA Research, 16*, 311–323.

Li, P., Zang, W., Li, Y., Xu, F., Wang, J., & Shi, T. (2011). AtPID: The overall hierarchical functional protein interaction network interface and analytic platform for Arabidopsis. *Nucleic Acids Research, 39*, D1130–D1133.

Liang, C., Jaiswal, P., Hebbard, C., Avraham, S., Buckler, E. S., Casstevens, T., et al. (2008). Gramene: A growing plant comparative genomics resource. *Nucleic Acids Research, 36*, D947–D953.

Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R. J., Franklin, L. D., et al. (2010). An integrated transcriptome atlas of the crop model Glycine max, and its use in comparative analyses in plants. *The Plant Journal, 63*, 86–99.

Lin, Q., Wang, C., Dong, W., Jiang, Q., Wang, D., Li, S., Chen, M., Liu, C., Sun, C., & Chen, K. (2015). Transcriptome and metabolome analyses of sugar and organic acid metabolism in ponkan (Citrus reticulata) fruit during fruit maturation. *Gene, 554*, 64–74.

Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., et al. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell, 133*, 523–536.

Liu, X., Noll, D. M., Lieb, J. D., & Clarke, N. D. (2005). DIP-chip: Rapid and accurate determination of DNA-binding specificity. *Genome Research, 15*, 421–427.

Lobell, D. B., Schlenker, W., & Costa-Roberts, J. (2011). Climate trends and global crop production since 1980. *Science, 333*, 616–620.

Loew, L. M., & Schaff, J. C. (2001). The virtual cell: A software environment for computational cell biology. *Trends in Biotechnology, 19*(10), 401–406.

Long, T. A., Brady, S. M., & Benfey, P. N. (2008). Systems approaches to identifying gene regulatory networks in plants. *Annual Review of Cell and Developmental Biology, 24*, 81–103.

Looger, L. L., Dwyer, M. A., Smith, J. J., & Hellinga, H. W. (2003). Computational design of receptor and sensor proteins with novel functions. *Nature, 423*, 185–190.

Lord, P. W., Stevens, R. D., Brass, A., & Goble, C. A. (2003). Investigating semantic similarity measures across the Gene Ontology: The relationship between sequence and annotation. *Bioinformatics, 19*, 1275–1283.

Luo, J. (2015). Metabolite-based genome-wide association studies in plants. *Current Opinion in Plant Biology, 24*, 31–38.

Ma, J. F., Yamaji, N., Mitani, N., Tamai, K., Konishi, S., Fujiwara, T., et al. (2007). An efflux transporter of silicon in rice. *Nature, 448*, 209–212.

Ma, F., Jazmin, L. J., Young, J. D., & Allen, D. K. (2014). Isotopically nonstationary 13C flux analysis of changes in Arabidopsis thaliana leaf metabolism due to high light acclimation. *Proceedings of the National Academy of Sciences of the United States of America, 111*, 16967–16972.

Mace, E. S., Rami, J. F., Bouchet, S., Klein, P. E., Klein, R. R., Kilian, A., et al. (2009). A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. *BMC Plant Biology, 9*, 13.

Macho, A. P., Boutrot, F., Rathjen, J. P., & Zipfel, C. (2012). Asparate oxidase plays an important role in Arabidopsis stomatal immunity. *Plant Physiology, 159*, 1845–1856.

Makita, Y., Kobayashi, N., Mochizuki, Y., et al. (2009). PosMed-plus: An intelligent search engine that inferentially integrates crossspecies information resources for molecular breeding of plants. *Plant & Cell Physiology, 50*(7), 1249–1259.

Manandhar-Shrestha, K., Tamot, B., Pratt, E. P. S., Saitie, S., Bräutigam, A., Weber, A. P. M., et al. (2013). Comparative proteomics of chloroplasts envelopes from bundle sheath and mesophyll chloroplasts reveals novel membrane proteins with a possible role in C4-related metabolite fluxes and development. *Frontiers in Plant Science, 4*, 65. https://doi.org/10.3389/fpls.2013.00065.

Manavalan, L. P., Guttikonda, S. K., Tran, L. S., & Nguyen, H. T. (2009). Physiological and molecular approaches to improve drought resistance in soybean. *Plant & Cell Physiology, 50*, 1260–1276.

Mao, X., Cai, T., Olyarchuk, J. G., & Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics, 21*, 3787–3793.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature, 437*, 376–380.

Maruyama, K., Takeda, M., Kidokoro, S., Yamada, K., Sakuma, Y., Urano, K., et al. (2009). Metabolic pathways involved in cold acclimation identified by integrated analysis of metabolites and transcripts regulated by DREB1A and DREB2A. *Plant Physiology, 150*, 1972–1980.

Masoudi-Nejad, A., Tonomura, K., Kawashima, S., Moriya, Y., Suzuki, M., Itoh, M., et al. (2006). EGassembler: Online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. *Nucleic Acids Research, 34*, W459–W462.

Matros, A., & Mock, H.-P. (2013). Mass spectrometry based imaging techniques for spatially resolved analysis of molecules. *Frontiers in Plant Science, 4*, 89. https://doi.org/10.3389/fpls.2013.00089.

Matsumura, H., Reich, S., Ito, A., Saitoh, H., Kamoun, S., Winter, P., et al. (2003). Gene expression analysis of plant host–pathogen interactions by SuperSAGE. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 15718–15723.

Matsumura, H., Kruger, D. H., Kahl, G., & Terauchi, R. (2008). SuperSAGE: A modern platform for genome-wide quantitative transcript profiling. *Current Pharmaceutical Biotechnology, 9*, 368–374.

Matzke, M., Kanno, T., Daxinger, L., Huettel, B., & Matzke, A. J. (2009). RNA-mediated chromatin-based silencing in plants. *Current Opinion in Cell Biology, 21*, 367–376.

Mayer, K. F., Martis, M., Hedley, P. E., Simkova, H., Liu, H., Morris, J. A., et al. (2011). Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell, 23*, 1249–1263.

McCann, H. C., & Guttman, D. S. (2008). Evolution of the type III secretion system and its effectors in plant–microbe interactions. *The New Phytologist, 177*, 33–47. https://doi.org/10.1111/j.1469-8137.2007.02293.x.

Mehta, R. A., Cassol, T., Li, N., Ali, N., Handa, A. K., & Mattoo, A. K. (2002). Engineered polyamine accumulation in tomato enhances phytonutrient content, juice quality, and vine life. *Nature Biotechnology, 20*, 613–618.

Meihls, L. N., Handrick, V., Glauser, G., Barbier, H., Kaur, H., Haribal, M. M., Lipka, A. E., Gershenzon, J., Buckler, E. S., Erb, M., et al. (2013). Natural variation in maize aphid resistance is associated with 2,4-dihydroxy-7- methoxy-1,4-benzoxazin-3-one glucoside methyltransferase activity. *Plant Cell, 25*, 2341–2355.

Mendes, P. (1997). Biochemistry by numbers: Simulation of biochemical pathways with Gepasi 3. *Trends in Biochemical Sciences, 22*, 361–363.

Meng, Y., Shao, C., Wang, H., et al. (2011). The regulatory activities of plant microRNAs: A more dynamic perspective. *Plant Physiology, 157*(4), 1583–1595.

Meyers, B. C., Galbraith, D. W., Nelson, T., & Agrawal, V. (2004). Methods for transcriptional profiling in plants. Be fruitful and replicate. *Plant Physiology, 135*, 637–652.

Miyagi, A., Takahara, K., Takahashi, H., Kawai-Yamada, M., & Uchimiya, H. (2010). *Metabolomics, 6*, 497–510. https://doi.org/10.1007/s11306-010-0220-0.

Mochida, K., Saisho, D., Yoshida, T., Sakurai, T., & Shinozaki, K. (2008). TriMEDB: A database to integrate transcribed markers and facilitate genetic studies of the tribe Triticeae. *BMC Plant Biology, 8*, 72.

Mochida, K., Furuta, T., Ebana, K., Shinozaki, K., & Kikuchi, J. (2009). Correlation exploration of metabolic and genomic diversities in rice. *BMC Genomics, 10*, 568.

Mochida, K., Yoshida, T., Sakurai, T., Yamaguchi-Shinozaki, K., Shinozaki, K., & Tran, L. S. (2010). LegumeTFDB: An integrative database of Glycine max, Lotus japonicus and Medicago truncatula transcription factors. *Bioinformatics, 26*, 290–291.

Mochida, K., Uehara-Yamaguchi, Y., Yoshida, T., Sakurai, T., & Shinozaki, K. (2011). Global landscape of a co-expressed gene network in barley and its application to gene discovery in Triticeae crops. *Plant & Cell Physiology, 52*, 785–803.

Mockler, T. C., & Ecker, J. R. (2005). Applications of DNA tiling arrays for whole-genome analysis. *Genomics, 85*, 1–15.

Moco, S., Bino, R. J., Vorst, O., Verhoeven, H. A., de Groot, J., van Beek, T. A., et al. (2006). A liquid chromatography–mass spectrometry-based metabolome database for tomato. *Plant Physiology, 141*, 1205–1218.

Moran, N. A., McLaughlin, H. J., & Sorek, R. (2009). The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science, 323*, 379–382.

Morsy, M., Gouthu, S., Orchard, S., Thorneycroft, D., Harper, J. F., Mittler, R., et al. (2008). Charting plant interactomes: Possibilities and challenges. *Trends in Plant Science, 13*, 183–191.

Mostafavi, S., Ray, D., Warde-Farley, D., et al. (2008). GeneMANIA: A real-time multiple association network integration algorithm for predicting gene function. *Genome Biology, 9*(1), S4.

Mueller, L. A., Zhang, P., & Rhee, S. Y. (2003). AraCyc: A biochemical pathway database for Arabidopsis. *Plant Physiology, 132*, 453–460.

Mukhtar, M. S., Carvunis, A. R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., et al. (2011). Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science, 333*, 596–601.

Nagasaki, M., Saito, A., Jeong, E., et al. (2010). Cell illustrator 4.0: A computational platform for systems biology. *In Silico Biology, 10*(1), 5–26.

Nakabayashi, R., & Saito, K. (2015). Integrated metabolomics for abiotic stress responses in plants. *Current Opinion in Plant Biology, 24*, 10–16.

Nakamura, Y., Teo, N. Z., Shui, G., Chua, C. H., Cheong, W. F., Parameswaran, S., Koizumi, R., Ohta, H., Wenk, M. R., & Ito, T. (2014). Transcriptomic and lipidomic profiles of glycerolipids during Arabidopsis flower development. *The New Phytologist, 203*, 310–322.

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

Nashilevitz, S., Melamed-Bessudo, C., Izkovich, Y., Rogachev, I., Osorio, S., Itkin, M., et al. (2010). An orange ripening mutant links plastid NAD(P)H dehydrogenase complex activity to central and specialized metabolism during tomato fruit maturation. *Plant Cell, 22*, 1977–1997.

Neumann, E. (2005). A life science semantic web: Are we there yet? *Science STKE, 283*, pe22.

Newton, A. C., Fitt, B. D. L., Atkins, S. D., Walters, D. R., & Daniell, T. J. (2010). Pathogenesis, parasitism and mutualism in the trophic space of microbe–plant interactions. *Trends in Microbiology, 18*, 365–373.

Nishimura, D. (2001). BioCarta. *Biotech Software & Internet Report, 2*, 117–120.

Nobuta, K., Venu, R. C., Lu, C., Belo, A., Vemaraju, K., Kulkarni, K., et al. (2007). An expression atlas of rice mRNAs and small RNAs. *Nature Biotechnology, 25*, 473–477.

Nobuta, K., Lu, C., Shrivastava, R., Pillay, M., De Paoli, E., Accerbi, M., et al. (2008). Distinct size distribution of endogeneous siRNAs in maize: Evidence from deep sequencing in the mop1-1 mutant. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 14958–14963.

Noel, J. P., Austin, M. B., & Bomati, E. K. (2005). Structure-function relationships in plant phenylpropanoid biosynthesis. *Current Opinion in Plant Biology, 8*, 249–253.

Nomura, M., Arunothayanan, H., Dao, T. V., Le, H. T. P., Takakazu Kaneko, T., Sato, S., et al. (2010). Differential protein profiles of Bradyrhizobium japonicum USDA110 bacteroid during soybean nodule development. *Soil Science & Plant Nutrition, 56*, 579–590.

Obayashi, T., Hayashi, S., Saeki, M., Ohta, H., & Kinoshita, K. (2009). ATTED-II provides coexpressed gene networks for Arabidopsis. *Nucleic Acids Research, 37*, D987–D991.

Ogasawara, O., Otsuji, M., Watanabe, K., Iizuka, T., Tamura, T., Hishiki, T., et al. (2006). BodyMap-Xs: Anatomical breakdown of 17 million animal ESTs for cross-species comparison of gene expression. *Nucleic Acids Research, 34*, D628–D631.

Okazaki, Y., Shimojima, M., Sawada, Y., Toyooka, K., Narisawa, T., Mochida, K., et al. (2009). A chloroplastic UDP-glucose pyrophosphorylase from Arabidopsis is the committed enzyme for the first step of sulfolipid biosynthesis. *Plant Cell, 21*, 892–909.

Olivier, B. G., & Snoep, J. L. (2004). Web-based kinetic modelling using JWS online. *Bioinformatics, 20*, 2143–2144.

Ozaki, S., Ogata, Y., Suda, K., Kurabayashi, A., Suzuki, T., Yamamoto, N., et al. (2010). Coexpression analysis of tomato genes and experimental verification of coordinated expression of genes found in a functionally enriched coexpression module. *DNA Research, 17*, 105–116.

Ozinsky, A., Underhill, D. M., Fontenot, J. D., Hajjar, A. M., Smith, K. D., Wilson, C. B., Schroeder, L., & Aderem, A. (2000). The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors. *Proceedings of the National Academy of Sciences of the United States of America, 97*, 13766–13771.

Pabinger, S., Rader, R., Agren, R., et al. (2011). MEMOSys: Bioinformatics platform for genome-scale metabolic models. *BMC Systems Biology, 5*(1), 20.

Paine, J. A., Shipton, C. A., Chaggar, S., Howells, R. M., Kennedy, M. J., Vernon, G., Wright, S. Y., Hinchliffe, E., Adams, J. L., Silverstone, A. L., & Drake, R. (2005). Improving the nutritional value of Golden Rice through increased pro-vitamin a content. *Nature Biotechnology, 23*, 482–487.

Papin, J. A., Reed, J. L., & Palsson, B. O. (2004). Hierarchical thinking in network biology: The unbiased modularization of biochemical networks. *Trends in Biochemical Sciences, 29*, 641–647.

Pappin, D. J., Hojrup, P., & Bleasby, A. J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology, 3*, 327–332.

Park, P. J. (2009). ChIP-seq: Advantages and challenges of a maturing technology. *Nature Reviews Genetics, 10*, 669–680.

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., et al. (2009). The Sorghum bicolor genome and the diversifi cation of grasses. *Nature, 457*, 551–556.

Patil, N., Berno, A. J., Hinds, D. A., Barrett, W. A., Doshi, J. M., et al. (2001). Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science, 294*, 1719–1723. 102.

Peña, P. A., Quach, T., Sato, S., Ge, Z., Nersesian, N., et al. (2017). Expression of the maize Dof 1 transcription factor in wheat and sorghum. *Frontiers in Plant Science, 8*, 434.

Pérez-Delgado, C. M., Moyano, T. C., García-Calderón, M., Canales, J., Gutiérrez, R. A., et al. (2016). Use of transcriptomics and co-expression networks to analyze the interconnections between nitrogen assimilation and photorespiratory metabolism. *Journal of Experimental Botany, 67*(10), 3095–3108.

Pichersky, E., & Gang, D. R. (2000). Genetics and biochemistry of secondary metabolites: An evolutionary perspective. *Trends in Plant Science, 5*, 439–445.

Pires, N. D., Yi, K., Breuninger, H., et al. (2013). Recruitment and remodeling of an ancient gene regulatory network during land plant evolution. *Proceedings of the National Academy of Sciences of the United States of America, 110*(23), 9571–9576.

Pop, M., Phillippy, A., Delcher, A. L., & Salzberg, S. L. (2004). Comparative genome assembly. *Briefings in Bioinformatics, 5*, 237–248.

Poultney, C. S., Gutiérrez, R. A., Katari, M. S., et al. (2007). Sungear: Interactive visualization and functional analysis of genomic datasets. *Bioinformatics, 23*(2), 259–261.

Proietti, S., Bertini, L., Timperio, A. M., et al. (2013). Crosstalk between salicylic acid and jasmonate in Arabidopsis investigated by an integrated proteomic and transcriptomic approach. *Molecular BioSystems, 9*(6), 1169–1187.

Proost, S., Van Bel, M., Sterck, L., Billiau, K., Van Parys, T., Van de Peer, Y., & Vandepoele, K. (2009). PLAZA: A comparative genomics resource to study gene and genome evolution in plants. *Plant Cell, 21*, 3718–3731.

Rhodes, D., Yu, J., Shanker, K., Deshpande, N., Varambally, R., et al. (2004). Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 9309–9314.

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., et al. (2000). Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. *Science, 290*, 2105–2110.

Rischer, H., Orešič, M., Seppänen-Laakso, T., et al. (2006). Gene-tometabolite networks for terpenoid indole alkaloid biosynthesis in Catharanthus roseus cells. *Proceedings of the National Academy of Sciences, 103*(14), 5614–5619.

Roberts, C., Nelson, B., Marton, M., Stoughton, R., Meyer, M., et al. (2000). Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science, 287*, 873–880.

Roth, F. P., Hughes, J. D., Estep, P. W., & Church, G. M. (1998). Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nature Biotechnology, 16*, 939–945.

Roux, M., Schwessinger, B., Albrecht, C., Chinchilla, D., Jones, A., Holton, N., et al. (2011). The Arabidopsis leucine-rich repeat receptor-like kinases BAK1/SERK3 and BKK1/SERK4 are required for innate immunity to hemibiotrophic and biotrophic pathogens. *Plant Cell, 23*, 2440–2455.

Ruiz-Ferrer, V., & Voinnet, O. (2009). Roles of plant small RNAs in biotic stress responses. *Annual Review of Plant Biology, 60*, 485–510.

Saal, L. H., Troein, C., Vallon-Christersson, J., Gruvberger, S., Borg, A., & Peterson, C. (2002). BioArray Software Environment: A platform for comprehensive management and analysis of microarray data. *Genome Biology, 3*, software000.

Saisho, D., & Takeda, K. (2011). Barley: Emergence as a new research material of crop science. *Plant & Cell Physiology, 52*, 724–727.

Saito, K., & Matsuda, F. (2010). Metabolomics for functional genomics, systems biology, and biotechnology. *Annual Review of Plant Biology, 61*, 463–489.

Saito, T., Ariizumi, T., Okabe, Y., Asamizu, E., Hiwasa-Tanase, K., Fukuda, N., et al. (2011). TOMATOMA: A novel tomato mutant database distributing micro-tom mutant collections. *Plant & Cell Physiology, 52*, 283–296.

Sakurai, N., Ara, T., Ogata, Y., Sano, R., Ohno, T., Sugiyama, K., et al. (2011). KaPPA-View4: A metabolic pathway database for representation and analysis of correlation networks of gene co-expression and metabolite co-accumulation and omics data. *Nucleic Acids Research, 39*, D677–D684.

Sanchez, L., Courteaux, B., Hubert, J., Kauffmann, S., Renault, J.-H., Clement, C., et al. (2012). Rhamnolipids elicit defense responses and induce disease resistance against biotrophic, hemibiotrophic, and necrotrophic pathogens that require different signaling pathways in Arabidopsis and highlight a central role for salicylic acid. *Plant Physiology, 160*, 1630–1641.

Santner, A., & Estelle, M. (2009). Recent advances and emerging trends in plant hormone signalling. *Nature, 459*, 1071.

Sauro, H. M., Hucka, M., Finney, A., et al. (2003). Next generation simulation tools: The systems biology workbench and BioSPICE integration. *OMICS, 7*(4), 355–372.

Sauvage, C., Segura, V., Bauchet, G., Stevens, R., Do, P. T., Nikoloski, Z., Fernie, A. R., & Causse, M. (2014). Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. *Plant Physiology, 165*, 1120–1132.

Sawada, Y., Akiyama, K., Sakata, A., Kuwahara, A., Otsuki, H., Sakurai, T., et al. (2009a). Widely targeted metabolomics based on large-scale MS/MS data for elucidating metabolite accumulation patterns in plants. *Plant & Cell Physiology, 50*, 37–47.

Sawada, Y., Kuwahara, A., Nagano, M., Narisawa, T., Sakata, A., Saito, K., et al. (2009b). Omics-based approaches to methionine side chain elongation in Arabidopsis: Characterization of the genes encoding methylthioalkylmalate isomerase and methylthioalkylmalate dehydrogenase. *Plant & Cell Physiology, 50*, 1181–1190.

Schaefer, C. F., Anthony, K., Krupa, S., et al. (2009). PID: The pathway interaction database. *Nucleic Acids Research, 37*(1), D674–D679.

Schauer, N., Semel, Y., Balbo, I., Steinfath, M., Repsilber, D., Selbig, J., et al. (2008). Mode of inheritance of primary metabolic traits in tomato. *The Plant Cell, 20*, 509–523.

Scheible, W. R., Morcuende, R., Czechowski, T., et al. (2004). Genomewide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of Arabidopsis in response to nitrogen. *Plant Physiology, 136*(1), 2483–2499.

Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science, 270*, 467–470.

Schilmiller, A. L., Moghe, G. D., Fan, P., Ghosh, B., Ning, J., Jones, A. D., & Last, R. L. (2015). Functionally divergent alleles and duplicated loci encoding an acyltransferase contribute to acylsugar metabolite diversity in Solanum trichomes. *Plant Cell, 27*, 1002–1017.

Schlueter, S. D., Dong, Q., & Brendel, V. (2003). GeneSeqer@PlantGDB: Gene structure prediction in plant genomes. *Nucleic Acids Research, 31*, 3597–3600.

Schmitz, R. J., & Zhang, X. (2011). High-throughput approaches for plant epigenomic studies. *Current Opinion in Plant Biology, 14*, 130–136.

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature, 463*, 178–183.

Schwender, J., Hebbelmann, I., Heinzel, N., Hildebrandt, T., Rogers, A., Naik, D., Klapperstück, M., Braun, H. P., Schreiber, F., Denolf, P., et al. (2015). Quantitative multilevel analysis of central metabolism in developing oilseeds of oilseed rape during in vitro culture. *Plant Physiology, 168*, 828–848.

Scossa, F., Brotman, Y., de Abreu e Lima, F., Willmitzer, L., Nikoloski, Z., Tohge, T., & Fernie, A. R. (2015). Genomics-based strategies for the use of natural variation in the improvement of crop metabolism. *Plant Science*. https://doi.org/10.1016/j.plantsci.2015.05.0213.

Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., et al. (2002). Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a fulllength cDNA microarray. *The Plant Journal, 31*, 279–292.

Seo, Y. S., Chern, M., Bartley, L. E., Han, M., Jung, K. H., Lee, I., et al. (2011). Towards establishment of a rice stress response interactome. *PLoS Genetics, 7*, e1002020.

Shanks, J. V. (2005). Phytochemical engineering: Combining chemical reaction engineering with plant science. *AICHE Journal, 51*, 2–7.

Shannon, P., Markiel, A., Ozier, O., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research, 13*(11), 2498–2504.

Sheard, L. B., Tan, X., Mao, H., Withers, J., Ben-Nissan, G., Hinds, T. R., et al. (2010). Jasmonate perception by inositol-phosphatepotentiated COI1–JAZ co-receptor. *Nature, 468*, 400–405.

Shen, Y. J., Jiang, H., Jin, J. P., Zhang, Z. B., Xi, B., He, Y. Y., et al. (2004). Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant Physiology, 135*, 1198–1205.

Shen, L., Gong, J., Caldo, R. A., Nettleton, D., Cook, D., et al. (2005). Barley base—An expression profiling database for plant genomics. *Nucleic Acids Research, 33*, D614–D618.

Shingaki-Wells, R. N., Huang, S., Taylor, N. L., Carroll, A. J., Zhou, W., & Millar, A. H. (2011). Differential molecular responses of rice and wheat coleoptiles to anoxia reveal novel metabolic adaptations in amino acid metabolism for tissue tolerance. *Plant Physiology, 156*, 1706–1724.

Shoemaker, R., Deng, J., Wang, W., & Zhang, K. (2010). Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Research, 20*, 883–889.

Simons, M., Misra, A., & Sriram, G. (2014). Genome-scale models of plant metabolism. *Methods in Molecular Biology, 1083*, 213–230.

Sinha, U., Bui, A., Taira, R., Dionisio, J., Morioka, C., et al. (2002). A review of medical imaging informatics. *Annals of the New York Academy of Sciences, 980*, 168–197.

SMRS Working Group. (2005). Summary recommendations for standardization and reporting of metabolic analyses. *Nature Biotechnology, 23*, 833–838.

Song, Q. X., Liu, Y. F., Hu, X. Y., Zhang, W. K., Ma, B., Chen, S. Y., & Zhang, J. S. (2011). Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. *BMC Plant Biology, 11*, 5.

Sriram, G., Fulton, D. B., Iyer, V. V., Peterson, J. M., Zhou, R., et al. (2004). Quantification of compartmented metabolic fluxes in developing soybean embryos by employing biosynthetically directed fractional 13C labeling, two-dimensional (13C, 1H) nuclear magnetic resonance, and comprehensive isotopomer balancing. *Plant Physiology, 136*, 3043–3057.

Staab, P. R., Walossek, J., Nellessen, D., et al. (2010). SynBioWave—A realtime communication platform for molecular and synthetic biology. *Bioinformatics, 26*(21), 2782–2783.

Stacey, G., Libault, M., Brechenmacher, L., Wan, J., & May, G. D. (2006). Genetics and functional genomics of legume nodulation. *Current Opinion in Plant Biology, 9*, 110–121.

Steinfath, M., Repsilber, D., Scholz, M., et al. (2007). Integrated data analysis for genome-wide research. *EXS, 97*, 309–329.

sterck, L., Rombauts, S., Vandepoele, K., Rouze, P., & Van de Peer, Y. (2007). How many genes are there in plants (… and why are they there)? *Current Opinion in Plant Biology, 10*, 199–203.

Steuer, R., Kurths, J., Fiehn, O., & Weckwerth, W. (2003). Interpreting correlations in metabolomic networks. *Biochemical Society Transactions, 31*, 1476–1478.

Stoeckert, C. J., Jr., Causton, H. C., & Ball, C. A. (2002). Microarray databases: Standards and ontologies. *Nature Genetics, 32*(Suppl), 469–473.

Stolc, V., Samanta, M. P., Tongprasit, W., Sethi, H., Liang, S., et al. (2005). Identification of transcribed sequences in Arabidopsis thaliana by using high-resolution genome tiling arrays. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 4453–4458.

Sucaet, Y., Wang, Y., Li, J., et al. (2012). MetNet online: A novel integrated resource for plant systems biology. *BMC Bioinformatics, 13*(1), 267.

Sulpice, R., Trenkamp, S., Steinfath, M., Usadel, B., Gibon, Y., Witucka-Wall, H., Pyl, E. T., Tschoep, H., Steinhauser, M. C., Guenther, M., et al. (2010). Network analysis of enzyme activities and metabolite levels and their relationship to biomass in a large panel of Arabidopsis accessions. *Plant Cell, 22*, 2872–2893.

Sumner, L. W. (2010). Recent advances in plant metabolomics and greener pastures. *F1000 Biology Reports, 2*, 7.

Sun, W., Xu, X., Zhu, H., Liu, A., Liu, L., Li, J., et al. (2010). Comparative transcriptomic profiling of a salt-tolerant wild tomato species and a salt-sensitive tomato cultivar. *Plant & Cell Physiology, 51*, 997–1006.

Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., et al. (2008). Large-scale insertional mutagenesis using the Tnt1 retrotransposonin the model legume Medicago truncatula. *The Plant Journal, 54*, 335–347.

Taji, T., Sakurai, T., Mochida, K., Ishiwata, A., Kurotani, A., Totoki, Y., et al. (2008). Large-scale collection and annotation of full-length enriched cDNAs from a model halophyte, Thellungiella halophila. *BMC Plant Biology, 8*, 115.

Tanabe, L., Scherf, U., Smith, L. H., Lee, J. K., Hunter, L., & Weinstein, J. N. (1999). MedMiner: An internet text-mining tool for biomedical information, with application to gene expression profiling. *BioTechniques, 27*, 1210–1217.

Tanaka, T., Antonio, B. A., Kikuchi, S., Matsumoto, T., Nagamura, Y., Numa, Y., et al. (2008). The Rice Annotation Project Database (RAP-DB): 2008 update. *Nucleic Acids Research, 36*, D1028–D1033.

Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., & Paterson, A. H. (2008a). Synteny and collinearity in plant genomes. *Science, 320*, 486–488.

Tang, H., Wang, X., Bowers, J. E., Ming, R., Alam, M., & Paterson, A. H. (2008b). Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Research, 18*, 1944–1195.

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature, 408*, 796–815.

Thijs, G., Lescot, M., Marchal, K., Rombauts, S., De Moor, B., Rouze, P., & Moreau, Y. (2001). A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling. *Bioinformatics, 17*, 1113–1122.

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

Timm, S., Florian, A., Wittmiß, M., Jahnke, K., Hagemann, M., Fernie, A. R., & Bauwe, H. (2013). Serine acts as a metabolic signal for the transcriptional control of photorespiration-related genes in Arabidopsis. *Plant Physiology, 162*, 379–389.

Todaka, D., Nakashima, K., Shinozaki, K., et al. (2012). Towards understanding transcriptional regulatory networks in abiotic stress responses and tolerance in rice. *Rice, 5*(1), 1–9.

Tohge, T., & Fernie, A. R. (2010). Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. *Nature Protocols, 5*, 1210–1227.

Tohge, T., Nishiyama, Y., Hirai, M. Y., Yano, M., Nakajima, J., Awazuhara, M., et al. (2005). Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. *The Plant Journal, 42*, 218–235.

Tohge, T., de Souza, L. P., & Fernie, A. R. (2014). Genome-enabled plant metabolomics. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences, 966*, 7–20.

Tomita, M., Hashimoto, K., Takahashi, K., et al. (1999). E-CELL: Software environment for whole-cell simulation. *Bioinformatics, 15*(1), 72–84.

Töpfer, N., Caldana, C., Grimbs, S., Willmitzer, L., Fernie, A. R., & Nikoloski, Z. (2013). Integration of genome-scale modeling and transcript profiling reveals metabolic pathways underlying light and temperature acclimation in Arabidopsis. *Plant Cell, 25*, 1197–1211.

Töpfer, N., Scossa, F., Fernie, A., & Nikoloski, Z. (2014). Variability of metabolite levels is linked to differential metabolic pathways in Arabidopsis's responses to abiotic stresses. *PLoS Computational Biology, 10*, e1003656.

Torres, D., Barrier, M., Bihl, F., Quesniaux, V. J. F., Maillet, I., Akira, S., Ryffel, B., & Erard, F. (2004). Toll-like receptor 2 is required for optimal control of Listeria monocytogenes infection. *Infection and Immunity, 72*, 2131–2139.

Toyoda, T., & Shinozaki, K. (2005). Tiling array-driven elucidation of transcriptional structures based on maximum-likelihood and Markov models. *The Plant Journal, 43*, 611–621.

Turenne, N. (2011). Role of a web-based software platform for systems biology. *Journal of Computer Science & Systems Biology, 4*, 035–041.

Ulitsky, I., Maron-Katz, A., Shavit, S., Sagir, D., Linhart, C., et al. (2010). Expander: From expression microarrays to networks and functions. *Nature Protocols, 5*(2), 303–322.

Umehara, M., Hanada, A., Yoshida, S., Akiyama, K., Arite, T., Takeda- Kamiya, N., et al. (2008). Inhibition of shoot branching by new terpenoid plant hormones. *Nature, 455*, 195–200.

Umezawa, T., Sakurai, T., Totoki, Y., Toyoda, A., Seki, M., Ishiwata, A., et al. (2008). Sequencing and analysis of approximately 40 000 soybean cDNA clones from a full-length-enriched cDNA library. *DNA Research, 15*, 333–346.

Umezawa, T., Nakashima, K., Miyakawa, T., Kuromori, T., Tanokura, M., Shinozaki, K., et al. (2010). Molecular basis of the core regulatory network in ABA responses: Sensing, signaling and transport. *Plant & Cell Physiology, 51*, 1821–1839.

Urano, K., Maruyama, K., Ogata, Y., Morishita, Y., Takeda, M., Sakurai, N., et al. (2009). Characterization of the ABA-regulated global responses to dehydration in Arabidopsis by metabolomics. *The Plant Journal, 57*, 1065–1078.

Urbanczyk-Wochniak, E., Luedemann, A., Kopka, J., Selbig, J., RoessnerTunali, U., Willmitzer, L., & Fernie, A. R. (2003). Parallel analysis of transcript and metabolic profiles: A new approach in systems biology. *EMBO Reports, 4*, 989–993.

van der Werf, M. J., Overkamp, K. M., Muilwijk, B., Coulier, L., & Hankemeier, T. (2007). Microbial metabolomics: Toward a platform with full metabolome coverage. *Analytical Biochemistry, 370*, 17–25.

van Helden, J. (2003). Regulatory sequence analysis tools. *Nucleic Acids Research, 31*, 3593–3596.

Van Helden, J., Rios, A. F., & Collado-Vides, J. (2000). Discovering regulatory elements in non-coding sequences by analysis of spaced dyads. *Nucleic Acids Research, 28*, 1808–1818.

Vandepoele, K., Van Bel, M., Richard, G., Van Landeghem, S., Verhelst, B., Moreau, H., Van de Peer, Y., Grimsley, N., & Piganeau, G. (2013). picoPLAZA, a genome database of microbial photosynthetic eukaryotes. *Environmental Microbiology, 15*, 2147–2153.

Varshney, R. K., Nayak, S. N., May, G. D., & Jackson, S. A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology, 27*, 522–530.

Velculescu, V. E., Zhang, L., Vogelstein, B., & Kinzler, K. W. (1995). Serial analysis of gene expression. *Science, 270*, 484–487.

Vernoux, T., Brunoud, G., Farcot, E., Morin, V., Van den Daele, H., Legrand, J., et al. (2011). The auxin signalling network translates dynamic input into robust patterning at the shoot apex. *Molecular Systems Biology, 7*, 508.

Voit, E. O. (2000). *Computational analysis of biochemical systems: A practical guide for biochemists and molecular biologists*. Cambridge: Cambridge University Press.

von Bertalanffy, L. (1933). *Modern theories of development*. London: Oxford University Press.

von Bertalanffy, L. (1968). General systems theory. In G. Braziller (Ed.), *Foundations, development, applications*. New York: George Braziller.

Walbot, V. (2009). 10 reasons to be tantalized by the B73 maize genome. *PLoS Genetics, 5*, e1000723.

Wall, P. K., Leebens-Mack, J., Muller, K. F., Field, D., Altman, N. S., & dePamphilis, C. W. (2008). PlantTribes: A gene and gene family resource for comparative genomics in plants. *Nucleic Acids Research, 36*, D970–D976.

Wan, X., & Xu, D. (2005). Computational methods for remote homolog identification. *Current Protein & Peptide Science, 6*, 527–546.

Wang, H., Schauer, N., Usadel, B., Frasse, P., Zouine, M., Hernould, M., et al. (2009). Regulatory features underlying pollination-dependent and -independent tomato fruit set revealed by transcript and primary metabolite profiling. *Plant Cell, 21*, 1428–1452.

Wang, K., Peng, X., Ji, Y., Yang, P., Zhu, Y., & Li, S. (2013). Gene, protein, and network of male sterility in rice. *Frontiers in Plant Science, 4*, 92. https://doi.org/10.3389/fpls.2013.00092.

Ware, D. H., Jaiswal, P., Ni, J., Yap, I. V., Pan, X., et al. (2002). Gramene, a tool for grass genomics. *Plant Physiology, 130*, 1606–1613.

Weckwerth, W. (2003). Metabolomics in systems biology. *Annual Review of Plant Biology, 54*, 669–689. https://doi.org/10.1146/annurev.arplant.54.031902.135014.

Wei, C.-F., Hsu, S.-T., Deng, W.-L., Wen, Y.-D., & Huang, H.-C. (2012). Plant innate immunity induced by flagellin suppresses the hypersensitive response in non-host plants elicited by Pseudomonas syringae pv. Averrhoi. *PLoS One, 7*, e41056. https://doi.org/10.1371/journal.pone.0041056.

Weigel, D., & Mott, R. (2009). The 1001 genomes project for Arabidopsis thaliana. *Genome Biology, 10*, 107.

Wenzl, P., Raman, H., Wang, J., Zhou, M., Huttner, E., & Kilian, A. (2007). A DArT platform for quantitative bulked segregant analysis. *BMC Genomics, 8*, 196.

Weston, D. J., Karve, A. A., Gunter, L. E., Jawdy, S. S., Yang, X., Allen, S. M., et al. (2011). Comparative physiology and transcriptional networks underlying the heat shock response in Populus trichocarpa, Arabidopsis thaliana and Glycine max. *Plant, Cell & Environment, 34*, 1488–1506.

Weston, D. J., Hanson, P. J., Norby, R. J., Tuskan, G. A., & Wullschleger, S. D. (2012). From systems biology to photosynthesis and wholeplant physiology. *Plant Signaling & Behavior, 7*(2), 260–262.

Wheeler, G., Ishikawa, T., Pornsaksit, V., & Smirnoff, N. (2015). Evolution of alternative biosynthetic pathways for vitamin C following plastid acquisition in photosynthetic eukaryotes. *eLife, 4*, e06369.

Wiechert, W., Mollney, M., Petersen, S., & de Graaf, A. A. (2001). A universal framework for 13C metabolic flux analysis. *Metabolic Engineering, 3*, 265–283.

Wiener, N. (1948). *Cybernetics* (p. 112). New York: Wiley.

Wienkoop, S., Morgenthal, K., Wolschin, F., Scholz, M., Selbig, J., & Weckwerth, W. (2008). Integration of metabolomic and proteomic phenotypes analysis of data covariance dissects starch and RFO metabolism from low and high temperature compensation response in Arabidopsis thaliana. *Molecular & Cellular Proteomics, 7*, 1725–1736. https://doi.org/10.1074/mcp.M700273-MCP200.

Windram, O., Madhou, P., McHattie, S., Hill, C., Hickman, R., et al. (2012). Arabidopsis defense against Botrytis cinerea: Chronology and regulation deciphered by high-resolution temporal transcriptomic analysis. *The Plant Cell, 24*(9), 3530–3557.

Winnenburg, R., Wächter, T., Plake, C., et al. (2008). Facts from text: Can text mining help to scale-up high-quality manual curation of gene products with ontologies? *Briefings in Bioinformatics, 9*(6), 466–478.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., & Provart, N. J. (2007). An 'electronic fluorescent pictograph' browser for exploring and analyzing large-scale biological data sets. *PLoS One, 2*, e718.

Witte, C. E., Archer, K. A., Rae, C. S., Sauer, J. D., Woodward, J. J., & Portnoy, D. A. (2012). Innate immune pathways triggered by Listeria monocytogenes and their role in the induction of cell-mediated immunity. *Advances in Immunology, 113*, 135–156.

Woo, Y., Affourtit, J., Daigle, S., Viale, A., Johnson, K., et al. (2004). A comparison of cDNA, oligonucleotide, and affymetrix GeneChip gene expression microarray platforms. *Journal of Biomolecular Techniques, 15*, 276–284.

Woodward, J. J., Iavarone, A. T., & Portnoy, D. A. (2010). C-di-AMP secreted by intracellular Listeria monocytogenes activates a host type I interferon response. *Science, 328*, 1703–1705.

Wu H, Yang H, Churchill GA (2011) *R/MAANOVA: An extensive R environment for the analysis of microarray experiments*.

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., et al. (2011). Genome sequence and analysis of the tuber crop potato. *Nature, 475*, 189–195.

Yamada, K., Lim, J., Dale, J. M., Chen, H., Shinn, P., et al. (2003). Empirical analysis of transcriptional activity in the Arabidopsis genome. *Science, 302*, 842–846.

Yamaguchi, S., & Kyozuka, J. (2010). Branching hormone is busy both underground and overground. *Plant & Cell Physiology, 51*, 1091–1094.

Yamakawa, H., & Hakata, M. (2010). Atlas of rice grain filling-related metabolism under high temperature: Joint analysis of metabolome and transcriptome demonstrated inhibition of starch accumulation and induction of amino acid accumulation. *Plant & Cell Physiology, 51*(5), 795–809.

Yamamoto, Y. Y., & Obokata, J. (2008). ppdb: A plant promoter database. *Nucleic Acids Research, 36*, D977–D981.

Yamamoto, Y. Y., Yoshitsugu, T., Sakurai, T., Seki, M., Shinozaki, K., & Obokata, J. (2009). Heterogeneity of Arabidopsis core promoters revealed by high-density TSS analysis. *The Plant Journal, 60*, 350–362.

Yang, F., Jacobsen, S., Jørgensen, H. J. L., Collinge, D. B., Svensson, B., & Finnie, C. (2013). *Fusarium graminearum* and its interactions with cereal heads: Studies in the proteomics era. *Frontiers in Plant Science, 4*, 37. https://doi.org/10.3389/fpls.2013.00037.

Yates, J. R., 3rd, Eng, J. K., McCormack, A. L., & Schieltz, D. (1995). Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Analytical Chemistry, 67*, 1426–1436.

Ye, X., Al-Babili, S., Klöti, A., Zhang, J., Lucca, P., Beyer, P., & Potrykus, I. (2000). Engineering the provitamin A (-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science, 287*, 303–305.

Yeager, A. F. (1927). Determinate growth in the tomato. *The Journal of Heredity, 18*, 263–265.

Yona, G., & Levitt, M. (2002). Within the twilight zone: A sensitive profile-profile comparison tool based on information theory. *Journal of Molecular Biology, 315*, 1257–1275.

Yonekura-Sakakibara, K., Tohge, T., Matsuda, F., Nakabayashi, R., Takayama, H., Niida, R., et al. (2008). Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene–metabolite correlations in Arabidopsis. *The Plant Cell, 20*, 2160–2176.

Young, N. D., & Udvardi, M. (2009). Translating Medicagotruncatula genomics to crop legumes. *Current Opinion in Plant Biology, 12*, 193–201.

Yuan, J. S., Galbraith, D. W., Dai, S. Y., et al. (2008). Plant systems biology comes of age. *Trends in Plant Science, 13*(4), 165–171.

Yun, K. Y., Park, M. R., Mohanty, B., et al. (2010). Transcriptional regulatory network triggered by oxidative signals configures the early response mechanisms of japonica rice to chilling stress. *BMC Plant Biology, 10*(1), 16.

Zeller, G., Henz, S. R., Widmer, C. K., Sachsenberg, T., Ratsch, G., Weigel, D., et al. (2009). Stress-induced changes in the Arabidopsis thaliana transcriptome analyzed using whole-genome tiling arrays. *The Plant Journal, 58*, 1068–1082.

Zhang, M. Q. (2002). Computational prediction of eukaryotic protein-coding genes. *Nature Reviews. Genetics, 3*, 698–709.

Zhang, H., Sreenivasulu, N., Weschke, W., Stein, N., Rudd, S., Radchuk, V., et al. (2004). Large-scale analysis of the barley transcriptome based on expressed sequence tags. *The Plant Journal, 40*, 276–290.

Zhang, J., Leiderman, K., Pfeiffer, J. R., Wilson, B. S., Oliver, J. M., & Steinberg, S. L. (2006a). Characterizing the topography of membrane receptors and signaling molecules from spatial patterns obtained using nanometer-scale electron-dense probes and electron microscopy. *Micron, 37*, 14–34.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W., Chen, H., et al. (2006b). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell, 126*, 1189–1201.

Zhang, B., Tolstikov, V., Turnbull, C., Hicks, L. M., & Fiehn, O. (2010). Divergent metabolome and proteome suggest functional independence of dual phloem transport systems in cucurbits. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 13532–13537.

Zhang, Z., Wu, Y., Gao, M., Zhang, J., Kong, Q., Liu, Y., et al. (2012). Disruption of PAMP-induced MAP kinase cascade by a Pseudomonas syringae effector activates plant immunity mediated by the NB-LRR protein SUMM2. *Cell Host & Microbe, 11*, 253–263.

Zheng, Y., Ren, N., Wang, H., Stromberg, A. J., & Perry, S. E. (2009). Global identification of targets of the Arabidopsis MADS domain protein AGAMOUS-Like15. *Plant Cell, 21*, 2563–2577.

Zhu, T., & Wang, X. (2000). Large-scale profiling of the Arabidopsis transcriptome. *Plant Physiology, 124*, 1472–1476.

Zhu, H., Bilgin, M., & Snyder, M. (2003). Proteomics. *Annual Review of Biochemistry, 72*, 783–812.

Zimmermann, I. M., Heim, M. A., Weisshaar, B., et al. (2004a). Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins. *The Plant Journal, 40*(1), 22–34.

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