

Plant systems biology: insights, advances and challenges

Bhavisha P. Sheth · Vrinda S. Thaker

Received: 23 December 2013 / Accepted: 6 March 2014 / Published online: 27 March 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract Plants dwelling at the base of biological food chain are of fundamental significance in providing solutions to some of the most daunting ecological and environmental problems faced by our planet. The reductionist views of molecular biology provide only a partial understanding to the phenotypic knowledge of plants. Systems biology offers a comprehensive view of plant systems, by employing a holistic approach integrating the molecular data at various hierarchical levels. In this review, we discuss the basics of systems biology including the various ‘omics’ approaches and their integration, the modeling aspects and the tools needed for the plant systems research. A particular emphasis is given to the recent analytical advances, updated published examples of plant systems biology studies and the future trends.

Keywords Omics · Modeling · Plant systems biology · Holistic · Integration · Network

God does not care about our mathematical difficulties. He integrates empirically.

—Albert Einstein

Introduction

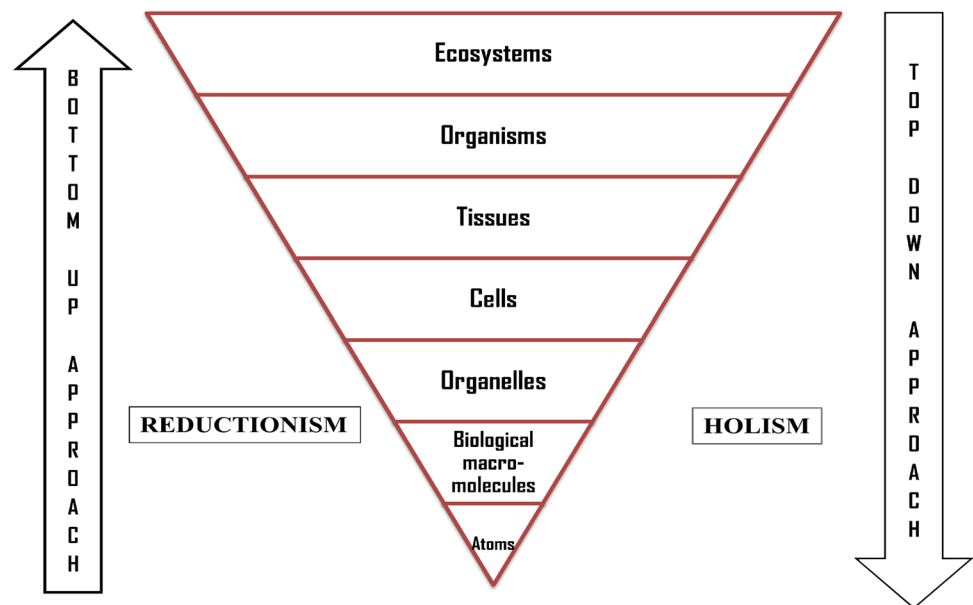
The reductionism versus holism is an age old debate in all realms of science. Reductionism explains the individual factors responsible for the systems behavior. On the contrary, holism advocates the behavior of the systems in totality and hence favors the integrative approach to understand a system. The concept of holism can be traced back, from the Aristotle’s *Metaphysica* which says “The whole is greater than the sum of its parts”.

Science finds its roots in reductionism right from its infancy. Reductionist thinking still holds a significant influence on science, including biology, especially after the rise of molecular biology which interpreted life being guided by molecular means including heredity. Molecular biology is grounded in reductionist views wherein complexity of biological systems is explicated in terms of the physicochemical properties of the individual components. The reductionist understanding, finds its roots, way back in the exemplary the one-gene one-polypeptide hypothesis of Beadle and Tatum, who actually demonstrated the direct relationship of the genotype to the phenotype at the molecular level. The technological advances in the DNA sequencing and other analytical platforms have resulted in an exponential increase in the molecular data (Ahmadian et al. 2006; Lister et al. 2009). In this context, the main challenge confronting the field is not to look back (incorporating previous findings is critical but will be comparatively easy) but to look forward to how one might plan and interpret the enormous new data that soon will be generated. Although as far as the interpretation of the data into meaningful information is concerned, bioinformatics has been crucial in every aspect of omics-based research to manage various types of genome-scale datasets efficiently and extract valuable knowledge (Shinozaki and Sakakibara 2009).

B. P. Sheth · V. S. Thaker (✉)
Department of Biosciences, Centre for Advanced Studies in Plant
Biotechnology and Genetic Engineering, Saurashtra University,
Rajkot 360005, Gujarat, India
e-mail: vsthaker@sauuni.ernet.in

B. P. Sheth
e-mail: bhavisha.sheth@gmail.com

Fig. 1 Types of modeling approaches in systems biology



However, with the passage of time, the **limitations** of the reductionist molecular approach **have become** increasingly **evident**. It became clear that biological systems cannot be explained only at the genetic level. Instead they should be understood as complex systems resulting from dynamic interactions of different components at different levels, each **individually functioning as wholes**, which eventually control the phenotype. Complex systems exist at different levels of biological organization ranging from the subatomic realm to individual organisms to whole populations and beyond. Hence a need arose for an **integrative framework**, which provides a holistic understanding of the biological systems. This new realm of science, in the post-reductionist era, is called systems biology. The concept of incorporation of systems theory into biological sciences was first proposed in 1940s by the biologist Ludwig von Bertalanffy (1973) which was further developed in the 1950s by Ashby (1956).

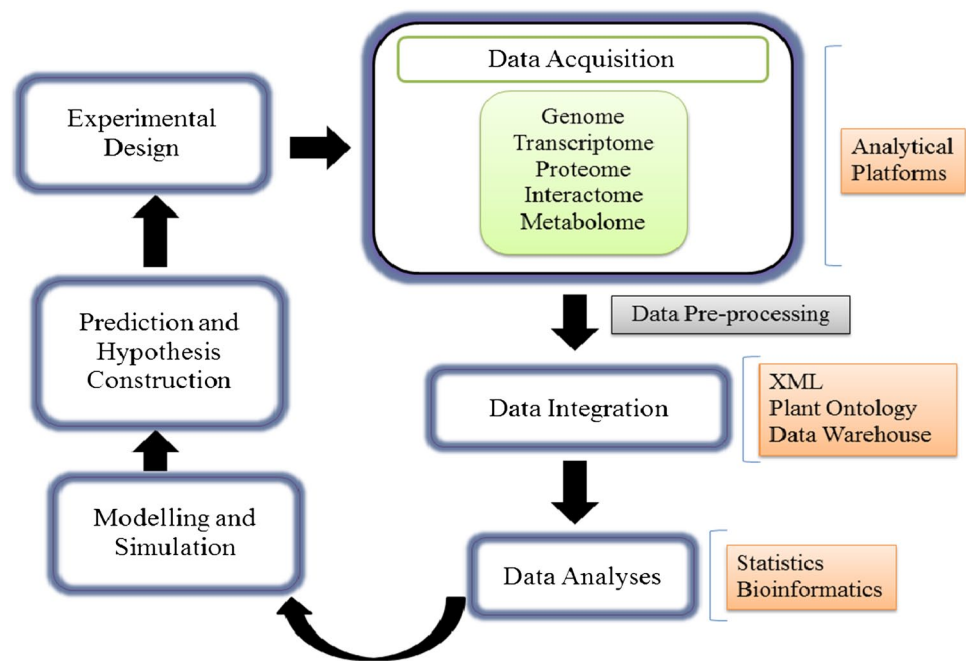
Systems biology is an approach by which a system of interacting units is analyzed as a whole rather than by analyzing its individual constituents separately. Currently three types of systems approaches exist viz. “top down” which entails the use of high-throughput ‘omics’ technologies to gain a holistic understanding to biological systems (Ideker et al. 2001); “bottom-up” which starts with molecular properties and leads to derivation of models to be tested and validated (Hartwell et al. 1999); and “middle-out” which is based on the principle to start somewhere in between the top and bottom levels, then work out toward a hierarchy of models (Noble 2002). Amongst these, the ‘top down’ and ‘bottom-up’ have been, most usually used to describe the biological systems (Fig. 1). Systems biology has actually enabled the maturation of

science from a data collection enterprise to an explanatory endeavor.

Biological systems and systems biology of plants

The biological communities, differing in their form, structure and habitat extend across the entire biosphere. Complexity and robustness are the key features of biological systems. **Robustness plays a key role in maintaining the appropriate functioning of the complex system despite internal/external perturbations (e.g., stochastic noise, physiological and developmental signals, environmental change and genetic variation)** (Stelling et al. 2004; Kitano 2007; Coruzzi et al. 2009). A thoughtful examination of the biological systems is indispensable to predict the effects of developmental programs, natural or human-induced perturbations on the composition and function of biological systems. The molecular interrogation and investigation of a complex system, along with maintenance of a global perspective is needed to understand the system function is eventually a key challenge to biological intellect. The various levels of biological systems’ organization include cells, molecules, organelles, tissues, organs, biochemical pathways, and whole plants and so on. The systems biology focuses on integration and not the dissimulation of the parts below the cellular level and hence the biological systems extend toward levels such as genome, transcriptome, proteome, metabolome, interactome and cellome (Ideker et al. 2001; Kitano 2002). **To gain a comprehensive idea, the assimilation of the networks at different levels of biological organization involving experimental and computational modeling approaches is required** (Keurentjes et al. 2011).

Fig. 2 The schematic diagram of a typical plant systems biology approach



Plant systems biology

Plant biology has the potential of providing partial solutions to several of the most daunting problems faced by our planet in the twenty-first century viz. increasing scarcity of food, the depletion of global oil reserves, and a shortage of fresh-water. Plants are greatly suited for systems analyses as they are easy to handle, quite good production of offspring for genetic analyses and many have comparatively short generation times. Hence modeling of plant biological phenomena can be accomplished in a quick, reiterative way by means of using the abundant molecular tools available (Keurentjes et al. 2008). The rationale behind increasing interest in systems biology is the prospects and progress in high-throughput genomics and proteomics technologies facilitating the researchers on building ample datasets with relevance to differing plant response (Minorsky 2003). Systems biology aids in our understanding of the plants using holistic systems approaches (Ideker et al. 2001; Kitano 2002).

Plant systems biology resides at the intersection of physiological, morphological, molecular, biochemical and genetic information applied to plant biology (Trewavas 2006). The need for the subject arose with necessity to integrate and interpret large datasets of high-throughput genomic, transcriptomic, proteomic, metabolomics technologies. Despite the enormous efforts in generation of huge experimental datasets, the work done on the side of development of computational platforms to analyze and decipher the data in form of biological model remains scanty (Kirschner 2005; Yuan et al. 2008). Thus, plant systems biology encompasses the studies on plants in response

to biological, genetic or chemical perturbations which includes supervision of gene, protein and biochemical pathways; application of omics approaches, integration of the data; and formulation of mathematical models that describe the structure and function responsive to the individual perturbations in form of biologically meaningful perturbations (Fig. 2).

The ‘parts’ of plant systems biology

Genomics

The ‘genome’ comprises of the complete genetic material of the organism. Genomics means the study of the organism’s whole genome. It lies at the base of the complex plant systems’ hierarchy and it provides an understanding toward the organisms’ behavioral explanation, and hence is a scientific discipline in its own premise. The earlier molecular perspective to resolve the ambiguities in plant identification and discrimination included various molecular techniques viz. PCR, RFLP, AFLP, RAPD and sequencing (Titanji et al. 2007; Saliba-Colombani et al. 2000). The era of single gene sequencing marked the beginning of plant genomics followed by whole genome sequencing, single nucleotide polymorphism (SNP) and medium density arrays, and eventually led to the current whole genome resequencing. As such, it is a prerequisite for understanding the role of the genes in the development of an organism, hence acting as a driving force from the genomics to the systems biology approach for gaining an insight in the totality.

The sequencing applications known today rely on one of the two applicative aspects viz. de novo sequencing, wherein the sequence reads are obtained from an unknown sequence followed by the assembly; or resequencing, wherein reads are mapped/aligned to a known reference sequence. The de novo applications are far more labor and cost intensive than the resequencing ones. Major resequencing applications include polymorphism discovery and transcription profiling. The state-of-the-art genome sequencing architecture involving several path breaking technologies has enabled us to gain an insight into the gene complement of the biosphere. The genome-scale technological advances have triggered the whole genome sequencing of the complex plant systems at lower costs. Publication of the first complete genome sequence of *Arabidopsis thaliana* in 2000 and subsequently that of first monocotyledonous plant *Oryza sativa* in 2005 revolutionized research in the realm of plant genomics. These genomes used the traditional clone-by-clone strategy which involved sequencing of overlapped bacterial artificial chromosomes (BAC) clones selected from a physical map, deploying the Sanger dideoxynucleotide chain termination method. The next approach introduced was the whole genome shotgun (WGS) which was used to produce the draft sequences of many plants like *Sorghum bicolor* (Patterson et al. 2009), *Vitis vinifera* (Jaillon et al. 2007) and so on. Since 2005, new sequencing platforms have emerged, having the high-throughput and cost-efficient capabilities, so called as ‘Next Generation Sequencing (NGS)’ technologies. A variety of NGS technologies include the 454 FLX (Roche) (Margulies et al. 2005), the Genome Analyzer/Hiseq (Illumina Solexa) (Bennett et al. 2005) and the SOLiD (Life Technologies), as well as newer platforms such as Heliscope (Helicos) (Milos 2008), PacBio RS (Pacific Biosciences) (Eid et al. 2009) for single molecular sequencing, and Ion Torrent (Life Technologies), based on a semiconductor chip (Rothberg et al. 2011), are also available. A good No. of plant genomes have recently been sequenced using NGS technology (Huang et al. 2009; Sato et al. 2011; Shulaev et al. 2010; Wang et al. 2011; Xu et al. 2011). Some examples of draft plant genome sequences include that of cucumber (*Cucumis sativus*) and cassava (*Manihot esculenta*) (Feuillet et al. 2011).

The main challenges in the whole genome sequencing of plants include: size and complexity (in the form of varied ploidy levels). Overall, the size of plant genomes (both number of chromosomes and total nucleotide base pairs) exhibits the greatest variation amongst other forms of life, with the average representative size of a plant genome being ~6 Gb (Gregory et al. 2007)—far larger than the average size of genomes sequenced so far from other organisms. Yet other hurdle includes the heterozygosity, observed in a variety of plants species, and to solve

Table 1 List of de novo genome assembly software tools for plant genomics

Sr. no.	De novo assembler	Sequencing platform
1	VCAKE	Solexa reads
2	SHARGS	Solexa reads
3	SSAKE	Solexa reads
4	Edena	Solexa reads
5	ALLPATHS	Solexa reads
6	ABYSS	Solexa reads
7	Euler-USR	454, Solexa reads
8	Velvet	Solexa reads
9	LOCAS	Solexa reads

this issue, the homozygous derivatives are selected for sequencing, whose sequences would then act as templates for mapping the variation observed on the heterozygotes (Jaillon et al. 2007). The other factors affecting the success rate include the presence of repetitive and transposable elements (Ming et al. 2008) as well as high-copy chloroplast and mitochondria organelles (Schnable et al. 2009), which complicate the de novo assembly of plant genome sequences and skew the coverage levels. Despite the limitations, plants do offer some advantages like ability of clonal propagation, which effectively immortalizes the genotypes of interest. The various de novo assemblers for next-generation sequencing platforms are listed in Table 1 (Imelfort and Edwards 2009; Zhang et al. 2011).

The final step in genomics is the annotation and deposition of sequence data into public consortia of databases. The vast technological attainments have pushed the needs for improvement of bioinformatics capabilities to mine complex data for valuable biological information. The current databanks [e.g., NCBI Genbank, EMBL (European Molecular Biology Laboratory), DDBJ (DNA Databank of Japan)] contain wealth of ~7.4 million plant DNA sequences, relatively few whole organellar and fewer nuclear genomes.

Genome-wide association studies (GWAS)

The interaction of the individual genes with the environment enables an organism to produce a phenotype. Many phenotypes are quantitative in nature, and complex in etiology, with multiple environmental and genetic causes. There are innumerable examples of quantitative traits in plants viz. plant yield, flowering time, sugar content, disease resistance and fruit weight, which result from the segregation of many genes and are influenced by environmental interactions (Paran and Zamir 2003). The classical approach to study the complex quantitative trait loci (QTLs) was linkage analysis. In this approach, the identification of polymorphisms is carried out in two parents followed in a

large segregating population. The linkage can be inferred from the identification of the recombinants showing phenotypic differences from that of parents. The limitation of this approach was that the resultant QTLs were restricted in allelic diversity and in having limited genomic resolution (Borevitz and Nordborg 2003). Another approach to map QTLs is association mapping, which resolves complex trait variation down to the sequence level by exploiting historical and evolutionary recombination events at the level of whole populations (Nordborg and Tavaré 2002). The latter can be investigated at the candidate gene level as well as a genome-wide study, involving multiple gene associations. The current advancement in the NGS technologies and decrease in associated costs, have enabled the paradigm shift from candidate gene-based studies to a comparatively new approach of genome-wide association studies (GWAS) to map the plant QTLs efficiently (Rafalski 2010, Schneeberger and Weigel 2011). Genome-wide association mapping, or genome scan, includes the investigation of genetic variation in the whole genome to find signals of association for various complex quantitative traits. The main advantages of GWAS over the classical linkage mapping includes: (1) increased mapping resolution, (2) reduced research time and (3) greater allele number (Yu and Buckler 2006).

The GWAS strategy comprises of single or combinatorial use of high-throughput NGS techniques, whole genome oligonucleotide arrays and genotyping by sequencing approaches for elucidation of complex phenotypic variations in plants. The GWAS approach was successfully exploited in many plant species like *Arabidopsis* (Atwell et al. 2010), rice (Huang et al. 2010b), maize (McMullen et al. 2009), barley (Pasam et al. 2012), sorghum (Morris et al. 2013), etc. In several plant species, diverse germplasm panels are being established for whole-genome association analysis (Caldwell et al. 2006; Hamblin et al. 2006; Nordborg et al. 2005; Yu and Buckler 2006).

The ideal technological platforms for GWAS include 454-GS FLX and Illumina 1 G Genome Analyzer for identifying SNPs through short resequencing reads of alleles from different individuals. Despite the technological and thoughtful innovations, GWAS have given rise to several systems level bioinformatics challenges of modeling the complex genotype–phenotype relationships using computational means, interpretation of genetic associations using biological resources and further development of powerful yet user-friendly softwares for the interaction between genome biologists and bioinformaticians.

Epigenomics

Phenotypic variation is usually attributed to genetic variation corresponding to the nucleotide sequences during

the course of evolution. The candidate gene and genome-wide analyses are imperative for mapping the genotype–phenotype relationship. Over a long span of time, numerous biological phenomena like paramutation (Brink 1958), parental imprinting (Kermicle 1970), control of transposon activity (McClintock 1984) and transgene silencing (Baulcombe 2004; Matzke et al. 2009), failed to be understood by the classical mendelian notion of inheritance of acquired traits, long before their molecular details were known. These phenomena overruled the Mendelian laws and appeared to follow the non-mendelian mode of inheritance (Grant-Downton and Dickinson 2005). The intensive efforts to find suitable reasons and mechanisms underlying, these apparent ‘exceptions’ to Mendelian suppositions have led to shaping a new field of ‘epigenetics’ (literally meaning ‘above genetics’).

The term “Epigenesis” was coined by Aristotle and later in 1942, the British developmental biologist Conrad H. Waddington coined the term “epigenetics” (Waddington 1942). In current scenario, epigenetics is the study of heritable changes in gene expression and function that cannot be explained by changes in DNA sequence (Bird 2007). The epigenome refers to the description of various epigenetic regulators across the whole genome (He et al. 2011). However, in contrast to the genome, which is identical in all cell types throughout life, the epigenome is dynamic and varies between cell types, developmental stages or in response to environmental stimuli.

Plants, being sessile autotrophs, demonstrate a high degree of developmental plasticity, to survive stressful environmental changes. Recent studies show that various epigenetic mechanisms underlie the survival strategies of plants under environmentally harsh conditions.

Epigenetic pathways are important components of plant growth, development and reproduction regulatory processes in plants. Various plant physiological processes in plant development, including flowering time, gametogenesis, stress response, light signaling and morphological change are modulated directly or indirectly by epigenetic marks. DNA methylation, histone modifications and non-coding RNA-based mechanisms are the chief modes by which ‘epigenetic’ information can be encoded in plants.

DNA methylation

Cytosine methylation, i.e., the chemical modification of the cytosine bases with the methyl group at the 5′ end is a common epigenetic mark of the plant genome. The patterns of cytosine methylation are varied and have been elaborately reviewed by Castiglione et al. (2002).

These patterns of cytosine methylation are highly dynamic and can change significantly with both the developmental state and the environmental perturbations (e.g.,

Burn et al. 1993). The DNA methylation may appear to be more complex particularly in plants with reference to the sequence sites which are methylated. DNA methylation could take place at symmetric (CpG and CpNpG, where N is any nucleotide) as well as asymmetric sites (any other cytosine) on the genome. The CpG and CpNpG, sites are called symmetric as the sequence is self complementary with methylable cytosines in pairs on opposite stands. Asymmetric methylation sites are rarely found in plants as compared to their symmetric counterparts.

The various techniques associated with the comprehensive study of DNA methylation on a genome-wide scale include bisulfite sequencing, restriction endonuclease digestion coupled to microarray technology and affinity purification (Beck and Rakyen 2008; Suzuki and Bird 2008). While sequencing the DNA methylome, bisulfite treatment converts unmethylated cytosines to uracils but does not alter methylated cytosines, thereby scoring the correct methylated cytosines. Very recently, single-molecule-based methods such as nanopore sequencing (Clarke et al. 2009) have been used to directly sequence the DNA methylome without bisulfite treatment. The activities of methylation-sensitive restriction endonucleases like *HpaII* and *SmaI* are inhibited by methylated cytosines, thereby allowing differentiation between methylated and unmethylated DNA fragments (Zilberman and Henikoff 2007). The combination of the methylation-sensitive enzymatic digestion with microarray technologies was used in early cases of plant DNA methylome profiling (Tompá et al. 2002; Tran et al. 2005a, b). The affinity purification includes immunoprecipitation of DNA with an antibody that specifically recognizes methylated cytosine (mCIP). Recently, the combination of this technique with high-end sequencing techniques has been used, i.e., methylated Chromatin Immunoprecipitation Sequencing (mCIP-Seq) to profile the DNA methylome in *Arabidopsis* (Gehring et al. 2009) and rice (Yan et al. 2010). BS-Seq is presently the most useful and widely used technology for DNA methylation analyses in plants (Cokus et al. 2008; Feng et al. 2010; Hsieh et al. 2009; Lister et al. 2008; Zemach et al. 2010).

Histone modifications

The histone proteins (H2A, H2B, H3 and H4), involved in the compaction of the chromosomes, can be covalently modified by post-translational modifications like methylation, acetylation, ubiquitination, phosphorylation, biotinylation and ADP-ribosylation. They are thought to determine the accessibility of DNA to transcriptional regulators (Bird 2001; Schones and Zhao 2008). Histone acetylation generally creates an accessible chromatin conformation while histone deacetylation, often coupled to histone methylation, initiates a compressed chromatin conformation

that promotes silencing and the formation of heterochromatin (Berger 2002). Histone methylation can confer both an active or repressed transcriptional state depending upon which lysine is methylated viz. (Histone 3 lysine 9: H3K9), H4K20 and H3K27 are silencing modifications, while H3K4 methylation produces active chromatin (Feng and Jacobsen 2011). Histone modifications and DNA methylation are often intertwined; each epigenetic mark can influence the other's recruitment to reinforce differential epigenetic states (Tariq and Paszkowski 2004; Cedar and Bergman 2009). Histone modifications can be studied via deploying the different variants of the traditional chromatin immunoprecipitation (ChIP) technology, based on the principle of probing histone–DNA interactions in vivo and to determine the genomic location of chromatin-associated proteins (Solomon et al. 1988). The enriched DNA from a ChIP assay can be examined by genomic tiling microarray hybridization (ChIP-chip) or various sequencing-based technologies (ChIP-Seq) enabling genome-wide analyses of histone modifications using instrumental sophistication (Mendenhall and Bernstein 2008; Park 2009; Schones and Zhao 2008).

Non-coding RNA

Small non-coding RNAs (miRNAs, siRNAs, snoRNAs and rasiRNAs), about 21–30 nucleotides long, play crucial epigenetic regulatory roles at different developmental stages in plants in response to environmental perturbations (Bonnet et al. 2006). In the past some years, smRNAs have been found to be involved in a plethora of roles in developmental biology, plant physiology and genome stability as adaptive responses to environmental stimuli (Mirouze 2012). They also control the movement of transposable elements at the transcriptional and post-transcriptional level (Vaucheret 2006). The analytics used for genome-wide profiling of smRNAs is the massively parallel signature sequencing (MPSS) approach (Lu et al. 2006; Nobuta et al. 2007). This technique briefly involves smRNA purification, adaptor ligation, reverse transcription and amplification of cDNA products followed by high-throughput sequencing of the same (Wang et al. 2009a).

High-resolution profiling of the epigenome has uncovered genome-wide combinatorial interactions of DNA methylation, histone modifications and siRNAs with complex relationships to chromatin accessibility and mRNA transcription in plants (Elling and Deng 2009). Recent technical advances used to decipher the epigenetic mechanisms have started to convert epigenetic research into a high-throughput enterprise, to which bioinformatics is expected to make significant contributions. The Epigenomics of Plants International Consortium web site (<https://www.plant-epigenome.org/>) provides hyperlinks to

plant epigenome data resources. Various other web-based epigenetic resources include SIGnAL (Zhang et al. 2006) and Epigara (Bouyer et al. 2011) for the Arabidopsis-based research. Various bioinformatics tools are available for explaining epigenetic phenomena (Bock and Lengauer 2008) like for ChIP on chip (Ringo, Telescope), ChIP-seq (BLAT, BLASTn, ELAND, Solexa Analysis pipeline, SXOligosearch), CpG sequence prediction (Glimmer, Grai-LEXP), Bisulfite sequencing (ESME, BiQAnalyzer).

Transcriptomics

The term ‘transcriptome’ encompasses all the genomic counterparts which are expressed as RNA transcripts, including coding (mRNA) and non-coding (e.g., tRNA, miRNA) RNAs at a given time in a cell or population of cells under a given set of environmental conditions (Wang et al. 2010). Traditional transcriptional analyses comprised of northern blots, but the advent of high-throughput technologies, have enabled us to elucidate the entire transcriptome of model and non-model plants. The modern methods of transcriptome elucidation include microarray analysis and NGS technologies. Microarray technology has been used since past many years and has matured into developing our insight toward gene regulatory networks and their behavior under varying environmental conditions. Particularly, the ATH1 Genome Array developed by Affymetrix™ has been extensively used for transcriptional studies in Arabidopsis (Busch and Lohmann 2007). Microarray technology was deployed to characterize transcriptomes for other plant species, including maize (Zhu et al. 2009) rice (Hazen et al. 2005), barley (Delp et al. 2009), soybean (O’Rourke et al. 2009) and tomato (Auge et al. 2009). The success of microarray techniques and extension toward other ‘omics’ approaches have been reviewed in Prowar (2012) and Tohge and Fernie (2012). A variety of bioinformatics tools have emanated in the microarray-based research of the plant fraternity viz. Genevestigator (Zimmermann et al. 2004), NASCArrays (Craigon et al. 2004), ArrayExpress (Parkinson et al. 2005), the Gene Expression Omnibus (Edgar et al. 2002) and Stanford Microarray Database (Ball et al. 2005).

A better alternative to microarray technology is sequence-based transcriptome profiling. The significant advantages of sequence-based transcriptome analyses are potential to quantify the abundance of transcript and second, indifference to the availability of sequenced genome. The high-throughput NGS technologies in RNA-seq are a popular approach to collecting and quantifying the large-scale sequences of coding and non-coding RNA pools at a low cost (Wang et al. 2009b; Garber et al. 2011). The two main NGS platforms that have been used successfully for transcriptomic studies of non-model plants include

Roche/454 (Margulies et al. 2005) and Solexa/Illumina (San Diego, California, USA). This approach has been used for the swift development of genomic resources in applied and emerging plant species (Zenoni et al. 2010; Gowik et al. 2011). Majority of RNA-seq projects have utilized a reference genome for identification of splice variants like in Arabidopsis and rice (Filichkin et al. 2010; Lu et al. 2010). Although, in cases of unavailability of reference genomes, the de novo assembly from RNA-seq data is used to provide transcriptional profiling in new species (Fu et al. 2011; Su et al. 2011; Wong et al. 2011). Apart from mRNA-based transcriptome studies, NGS technologies were also exploited to studies focused on non-coding smRNA studies in plants like *Solanum lycopersicum* (Moxon et al. 2008), *Medicago truncatula* (Szittyá et al. 2008; Lelandais-Brière et al. 2009) and rice (Zhou et al. 2010).

Various bioinformatics tools aid the *de novo* assembly in RNA-seq based applications viz. CAP3 (Huang and Madan 1999), CLCbio Genomics Workbench (CLC Bio), gsAssembler (454 Life Sciences), MIRA (Chevreux et al. 2004), TGICL (Pertea et al. 2003) and Trinity (Grabherr et al. 2011) and Velvet (Zerbino and Birney 2008) etc. Similarly, there are others available for reference-guided assembly in RNA-seq like BLAT (Kent 2002), Bowtie (Langmead et al. 2009), BWA (Li and Durbin 2009), GSNAP (Wu and Nacu 2010) and an RNA sequence analysis tool is RX (Hong et al. 2012) etc. Moreover, programs like TopHat (Trapnell et al. 2009) and SpliceMap (Au et al. 2010) have been used to map splice junctions onto the sequence.

Proteomics

Proteome, termed by Wilkins et al. (1996) is the entire protein complement of the system, expressed at a given time and under-defined environmental conditions. The systemic analysis of the proteome is called proteomics (Pandey and Mann 2000; Patterson and Aebersold 2003; Phizicky et al. 2003). The complexity of the proteome is much greater than that of the transcriptome due to the huge amount of possible post-translational modifications, making the former highly dynamic (Glinski and Weckwerth 2006). The general workflow of proteomics analyses involves separation of proteins, digestion with proteases followed by peptide mass fingerprinting, determination of mass of the fragments and peptide sequencing by MS/MS method followed by database searching and identification of protein (Thelen and Peck 2007). Proteomics basically deals with analyzing changes in protein expression, study of protein structure, function and post-translational modifications that majorly include phosphorylation and ubiquitination (You-ping et al. 2010). A wide variety of analytical platforms for elucidation of the proteomic aspects are currently available which include, the classical 2D Gel electrophoresis, Edman

sequencing and mass spectrometric methods. Amongst these, mass spectrometric variants like Matrix-Assisted Laser Desorption Ionization Mass Spectrometry (MALDI-TOF MS), Electron spray ionization (ESI-MS), MS–MS, etc. have contributed immensely to the development of the field (Park 2004; Domon and Aebersold 2006). Recently, a new approach to analyze proteins directly by MS, without gel separation, has been developed, which is referred to as multidimensional protein identification technology (MudPIT) (Appella et al. 1995; Washburn et al. 2001; Wolters et al. 2001). The isotope-coded affinity tag (ICAT)-MS has been used for quantitative proteomics. Recently a new method using fluorescence resonance energy transfer (FRET) between fluorescent tags on interacting proteins, using green, cyan and yellow fluorescent protein (Phizicky et al. 2003) has been reported, which involves the *in vivo* analysis by microscopy.

Bioinformatics has immensely helped in the detailed analysis of the proteomics data. Various 2D PAGE-related databases are available on the web like ECO-2DBASE (VanBogelen et al. 1999), SWISS-2DPAGE (Hoogland et al. 2000) and WORLD-2DPAGE (specifically for plant proteins). Various other programs are available for the image analysis of the 2D gels like Melanie (Appel et al. 1997), PDQuest (BioRad), Progenesis (Rosengren et al. 2003) and Delta 2D (Decodon GmbH). The protein mass fingerprinting analysis tools include Mascot, Sequest, Aldente, Popitam, Phenyx, FindMod, Profound, PepFrag, MS-Fit, OMSSA, TagIdent, etc. The protein databases available over the web include, SwissProt/UniProt knowledgebase, Tr-EMBL, Genbank, Ensembl, PIR and EST database (Vihinen 2001). The Arabidopsis and rice proteome databases are used by various proteomics labs today. The subcellular proteomic database (SUBA) hosts information on subcellular localization of plant proteins, based on GFP tagging and proteomics methods. The various other types of proteomics databases have been reviewed by Komatsu (2006).

An attempt to map proteomes of different plant tissues from rice and Arabidopsis was seen in the publications by Tsugita et al. (1996). The works of Vener et al. (2001) and Sherrier et al. (1999) were emphasized on the post-translational modifications of proteins. Also, the subcellular proteomes and protein complexes in plants, e.g., proteins in the plasma membranes, chloroplasts, mitochondria and nuclei have been worked out (Rouquie et al. 1997; Peltier et al. 2000, 2001; Prime et al. 2000; Kruff et al. 2001; Miliar et al. 2001; Bae et al. 2003). A genome-scale proteomics study was carried out in *A. thaliana* with the identification of 13,000 proteins and almost half of predicted gene models (Baerenfaller et al. 2008). More recently, the International Plant Proteomics Organization (www.inppo.com), which is a recent global initiative to develop and improve

connections between plant proteomics researchers and related fields, has been established (Agrawal et al. 2011). The field of ‘proteomics’ is still in its infancy; and for the prosperity of the same, the challenges have to be overcome with further logical and technological advancement (van Wijk 2001).

Metabolomics

The term “metabolome”, suggested by Oliver et al. (1998), includes the entire set of small molecule metabolites, produced by any organism. Metabolomics, coined by, Oliver Fiehn (2002), is the comprehensive analysis of all the metabolites under given set of conditions, in an organism. The set of metabolites in an organism, represents more heterogeneity compared to genes and proteins in terms of their physical and chemical properties, varying widely with respect to size, polarity, quantity and stability. This is responsible for the dynamism of the metabolome having both temporal and spatial constraints (Fiehn 2002). Moreover, the metabolomes spatially defined as organs, tissues, cells have different metabolite profiles (Ebert et al. 2010; Schadt et al. 2005; Sumner et al. 2011). The metabolic wealth is attributed not only to the amount of genes (20,000–50,000) but also to the multiple substrate specificities for many enzymes (Aharoni et al. 2000), subcellular compartmentation, and the occurrence of nonenzymatic reactions. An estimated 200,000 metabolites exist in plant (Pichersky and Gang 2000), although only ~50,000 have been elucidated (De Luca and St. Pierre 2000). Metabolic profiles provide a biochemical phenotypic assessment of the plants and hence are the most valuable in systems biology studies, so regarded as a cornerstone of systems biology (Hall 2006; Saito and Matsuda 2010).

The general workflow of a metabolomic analysis comprises of four main stages viz. preparation of the sample, data acquisition using analytical methods, data mining and compound identification plus quantification using the statistical and bioinformatics analyses. The final task is to draw meaningful biological interpretations from the analyzed data (Fiehn 2002). Various analytical platforms may be used in metabolomics like, nuclear magnetic resonance (NMR), LC-NMR, mass spectrometry (MS): gas chromatography–mass spectrometry (GC-MS), capillary electrophoresis–mass spectrometry (CE-MS), liquid chromatography–mass spectrometry (LC-MS), liquid chromatography–electrochemistry–mass spectrometry (LC-EC-MS), direct infusion mass spectrometry (DIMS), fourier transform ion cyclotron mass spectrometry (FTMS); infrared spectroscopy (IR), thin layer chromatography (TLC), high-performance liquid chromatography (HPLC) equipped with different kinds of detectors: UV or photodiode array (PDA), fluorescent, electrochemical, etc., Fourier transform infrared (FT-IR)- and Raman

spectroscopies (Fiehn et al. 2000; Verpoorte et al. 2008; Moco et al. 2008; Allwood and Goodacre 2010; Weckwerth and Morgenthal 2005). Amongst these, NMR and MS are the chiefly applied (Fiehn 2002; Allwood and Goodacre 2010). Statistics and bioinformatics are indispensable tools for the processing of the large metabolome datasets generated by high-performance instrumentation mentioned above. Statistical analyses include unsupervised and supervised algorithm-based methods. Unsupervised methods include principal component analysis (PCA), hierarchical clustering (HCA), K-means clustering, Soft Independent Modeling of Class Analogy (SIMCA) (Wold and Sjostrom 1977) and self-organizing maps (SOMs) while the supervised methods include discriminant function analysis (DFA) (Lachenbruch 1975), partial least squares (PLS), ANOVA, feed-forward neural networks (Cowan and Sharp 1988), support vector machines (Cristianini and Shawe-Taylor 2000), orthogonal partial least squares discriminant analysis (OPLS-DA) (Bylesjö et al. 2006), genetic algorithms (Goldberg 1989) and genetic programming (Koza 1992). Various bioinformatics tools and databases are available for handling, processing and analyzing metabolomics data. The diverse resources used for plant metabolomics-based studies have been excellently reviewed by Fukushima and Kusano (2013).

Several attempts are made to explore the plant metabolome in various aspects like mutations (Bino et al. 2005; Yonekura-Sakakibara et al. 2008), identification of novel metabolite compounds (Xie et al. 2008), environmental disturbances (Ahuja et al. 2010; Allwood et al. 2006, 2008, 2010; Choi et al. 2006; Jansen et al. 2008; Sanchez et al. 2010; van Dam and van der Meijden 2011; Ward et al. 2010), genotyping (Sumner et al. 2003; Fiehn et al. 2000) and genetic introgression (Fernie and Schauer 2009; Keurentjes et al. 2006; Schauer et al. 2006). Systems analyses have also been applied to secondary metabolite in plants (Huang et al. 2010a; Breitling et al. 2013; Spiering et al. 2014). An integrated systems approach has also been applied to the investigation of tomato fruit development, where combined transcript, protein and metabolite analyses are carried out (Osorio et al. 2011). Recent studies pertaining to secondary metabolites have been carried out in Jasmonates (De Geyter et al. 2012), Flavonoids (Groenenboom et al. 2013), polyphenolics (Bovy et al. 2010), alkaloids analysis (Kim and Verpoorte 2010) and glucosinolate analysis (Hall et al. 2010). The current challenges in the field include, the heterogeneity and complexity of the plant metabolome and corresponding massive analytical needs to elucidate the same.

Interactomics

The functioning of a cell or any system is attributed to the dynamic and harmonized interactions of its macromolecular constituents, like DNA, RNA, lipids, proteins, and

other small molecules, with varied biochemical properties. Amongst these, the protein–protein interactions are most abundantly reported followed by the DNA/RNA–protein interactions in plants. DNA–protein interactions include the histone proteins bound to DNA to form the chromatin structure, which function in the epigenetic regulation of various physiological processes, discussed earlier. The RNA-binding proteins (RBPs), influence the processing, synthesis, nuclear export, stability and translation of RNA subsets contribute to the coordination of gene expression, but the reports are scanty (Fedoroff 2002; Belostotsky and Rose 2005) owing to the genetic compartmentation (nuclear and organellar) and eventual larger repertoire of the RBPs than in other well-explored biological systems (Barkan 2009). Thus, interactomics, the comprehensive analysis of the interactions between different macromolecules, predominantly protein–protein interactions in an organism, is critical to our understanding of the cellular systems (Cusick et al. 2005; Morsy et al. 2008). Protein–protein interaction (PPI) studies aid in elucidation of the biophysical basis for all the cellular processes and building a framework for functional characterization of individual proteins. Profiling protein–protein interactions has been the major focus of interactomics in the past few years (Charbonnier et al. 2008) largely due to the technological advances and developing insight in the field. Plants are expected to have 75,000–15,000 protein interaction pairs (Morsy et al. 2008), from proteomes of 30,000–40,000 estimated proteins (Sterck et al. 2007).

Plant protein interactomes can be mapped using in vivo, in vitro and in silico methods. The various in vivo technological platforms used for the plant interactome mapping studies include, yeast two-hybrid (Y2H), split-ubiquitin system (SUS), bimolecular fluorescence complementation (BiFC), Split-luciferase system, fluorescence resonance energy transfer (FRET) and bioluminescence resonance energy transfer (BRET) (Morsy et al. 2008). The various in vitro analytics include affinity purification mass spectrometry (AP-MS), protein microarrays and surface plasmon resonance. The in silico prediction methods often integrate multiple types of data from various analyses like co-expression, co-localisation, co-evolution, functional comparison, occurrence of orthologs or interologs (i.e., interactions conserved across species), times of occurrence and type of interaction (e.g., direct, physical association, genetic interaction, etc.) and so on (Sharan and Ideker 2006; Sharan et al. 2007).

The various protein–protein interaction web resources in plants include, Arabidopsis Interactome 1 (Arabidopsis Interactome Mapping Consortium 2011), Arabidopsis membrane interactome (Mukhtar et al. 2011), Rice kinase–protein interaction map (Ding et al. 2009), auxin-signaling network (Vernoux et al. 2011), InAct (Aranda

et al. 2010), TAIR Protein Interaction data (Swarbreck et al. 2008), AtPID (Cui et al. 2008), CORNET (De Bodt et al. 2012), PAIR (Lin et al. 2011), PRIN (Gu et al. 2011), MINT (Licata et al. 2012), DIP (Xenarios et al. 2002), MatrixDB (Chautard et al. 2009), BIND (Alfarano et al. 2005), BioGRID (Stark et al. 2006), APID (Prieto and De Las 2006), iPfam (Finn et al. 2005), BAR (Geisler-Lee et al. 2007), STRING (Szklarczyk et al. 2011) and AtPIN (Brandão et al. 2009), PINA (Cowley et al. 2012).

The first computationally predicted interactome for Arabidopsis was given by Geisler-Lee et al. (2007). However, the first organized experimental proteome-scale interactome map for plants—Arabidopsis Interactome 1 (AI-1), containing ~6,200 interaction among ~2,700 proteins, was published in 2011 (Arabidopsis Interactome Mapping Consortium. 2011). Mukhtar et al. (2011) utilized AI-1 to understand the pathogen–host interaction. The PPI resources containing >50 interaction have been recently reviewed by Braun et al. (2013). Protein–protein interaction networks are found to be key players in various biological phenomena like signal transduction (Schoonheim et al. 2007), response to abiotic stresses (Tardif et al. 2007), cell cycle (Van Leene et al. 2007), protein ubiquitination (Maor et al. 2007), RNA processing (Hunt et al. 2008), etc. As far as protein–protein interaction networks are concerned, network dynamics and comprehensiveness are the issues requiring great concern, to obtain better systems level understanding of plants.

Other ‘omics’ approaches

Apart from the chief omics approaches discussed earlier, some recent approaches include lipidomics (comprehensive study of the lipid entities of the organism) (Welti et al. 2007) and hormonomics (the entire set of endogenous hormones in a plant). The low molecular weight plant hormones include auxin, ABA, cytokinin, gibberellins, ethylene, brassinosteroids, jasmonates, salicylic acid (Davies 2004) and a newly identified one—strigolactone (acting as a shoot branching inhibitor) (Gomez-Roldan et al. 2008; Umehara et al. 2008); lectinomics (bioinformatics studies of carbohydrate-binding proteins–lectins) and various others. Also, a new concept which has gained much attention in this era is that of phenomics’, the high-throughput systemic analysis of phenotypes, which has probably the biggest application in plant biotechnology (Edwards and Batley 2004).

Integration of multiple ‘omics’ data

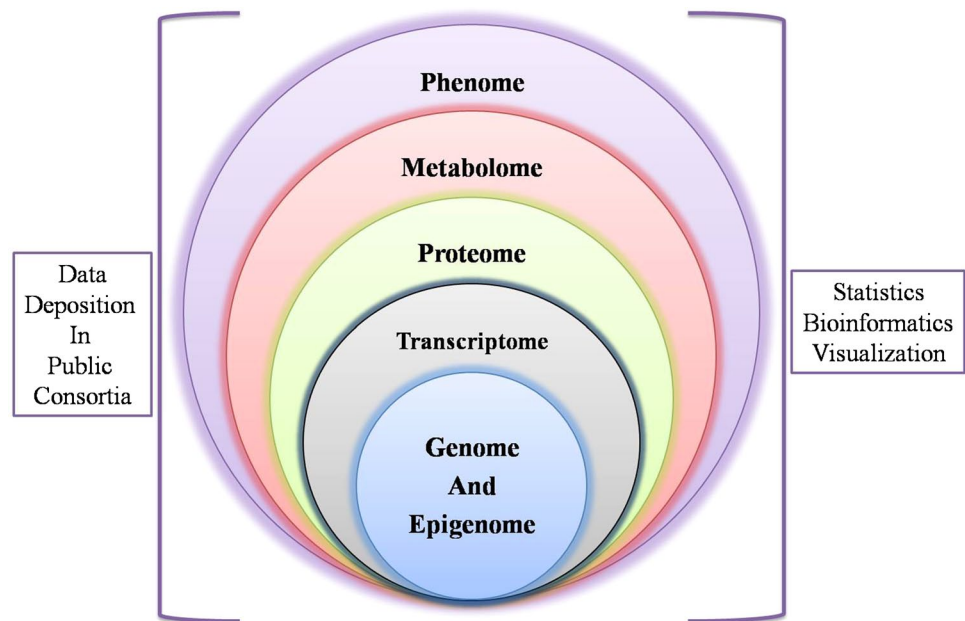
The advances in high-throughput analytics have enabled us to gain insights of individual biomolecules with the help of various ‘omics technologies’ discussed in the previous

section. However, any single ‘omics’ approach may be inadequate to characterize the complexity and behavior of biological systems as a ‘whole’ (Gygi et al. 1999). Hence the molecular research is gradually shifting toward the holistic perceptions of systems biology, by integration of the individual ‘omics’ datasets, to gain biologically meaningful interpretation of the plant systems.

Therefore, integration of multiple layers of biological information will provide a precise ‘picture’ of the ‘whole’ plant systems (Fig. 3). The integration of the multiple omics datasets must be carried out after they are preprocessed (normalization, missing value attribution and feature selection). Data integration is a key to the successful development of the systems philosophy by building comprehensive models of plant systems. Given the enormous promise of integration of the multiple omics data, logical input to the designing of various experiments and analysis of the heterogeneous data is gaining interest (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 (Fig. 3). These aspects have been elaborately discussed in Joyce and Palsson (2006).

Various instances of omics data integration are available in the literature. There are a number of reports on the elucidation of gene function by combining the metabolomic analysis with genomic and transcriptomic data (Tohge et al. 2005; Hirai et al. 2004, 2007; Hirai and Saito 2008; Saito et al. 2008; Watanabe et al. 2008; Yonekura-Sakakibara et al. 2008; Okazaki et al. 2009). Very recently, an integrated transcriptomics and epigenomics approach was used in maize hybrids (He et al. 2013). The integrated use of transcriptomic and proteomic data has been reported in various recent studies involving whole plant nitrogen economics of maize (Amiour et al. 2012), growth to dormancy transition 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 (Wang et al. 2009c), 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; Izawa et al. 2011). An integrated metabolome and proteome analysis was

Fig. 3 Integration of heterogeneous multiple ‘omics’ data



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 (Krallinger et al. 2008; 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 (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.

There are several challenges to integration of the multiple omics data (Joyce and Palsson 2006; De Keersmaecker et al. 2006; Steinfath et al. 2007). One of the problems in

complex annotation and integration is the lack of agreed formats across different omics datasets due to the heterogeneous repositories of the primary data sources. The solutions to this problem include creation of ‘data warehouses’, use of extensible markup language (XML), hypertext navigation, Unmediated MultiDB queries, creation of federated database and using controlled vocabularies. A Data Warehouse retrieves data from multiple resources, translates the formats and puts them in one 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. Plant Ontology Consortium is a collaborative effort among model plant genome databases and plant researchers that aim to create maintain and facilitate the use of a controlled vocabulary (ontology) for plants (Avraham et al. 2008). The other problem includes is of statistical analysis, i.e., evaluation of the complexity of integration differing from that of individual omics analysis and applying a suitable method thereafter. The integration of omics data is thus, far more than just ‘joining the pieces’; it is actually a journey of exploring uncharted territories and transforming information to more useful biological knowledge.

Modeling and simulation in plant system dynamics

The systems interest to biological sciences dates back to the days of von Bertalanffy (1933, 1969), 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). The successful modeling of the plants is the ultimate goal of plant systems biology. A model (*modus* in Latin, meaning manner/measure) usually represents the causal relationships in a system, in mathematics. In systems biology, cells or higher units of biological organization are understood as systems of interacting elements. For a systems level explanation, one needs to know the identity of the constituents, dynamic behavior and interactions among the constituents, of the biological system, under study (Kitano 2002). This information can eventually be combined into a model, which is not only consistent with the current knowledge but also can predict the system behavior under new unexplored perturbations. Modeling and simulation are central to bridge the gaps between theory and experiment (Dhar et al. 2004). Usually, experimental results require correct mathematical/statistical input, and model hypotheses require experimental evidences, to provide biologically meaningful interpretations. Modeling usually starts with construction of biological networks from the available molecular datasets. Network construction and analysis are the crucial components of systems biology.

A network/graph, in systems biology, has two basic parts: the elements of the system are represented as *graph nodes* (also called vertices) and the interactions are represented as *edges*, that is, lines connecting pairs of nodes. Edges may be directed (originating from a source (starting node) to a sink (ending 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 (or vertices) 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). Signs representing activation or inhibition can be shown on edges to augment the information content of the network. The important characteristics of biological networks are scale-free structure (the number of nodes that make a large number of connections with other nodes (referred to as “hubs”) is much lower than the number of nodes with few connections) and relative scarcity of hubs that connect directly to one another (Barabasi and Oltvai 2004). The nodes of the interaction network represent population of biomolecules, whose abundance varies in time and in response to the internal and environmental perturbations. To visualize the changes and create a model, the interaction network needs to be augmented by variables (expression, concentration, activity)

which indicate the state of each node and set of equations, indicating the how the state changes corresponding to the stimuli. Models may be static or dynamic depending on their behavior in the system with time. The four common types of networks in plant systems biology include, gene-to-metabolite networks, protein–protein interaction networks, transcriptional regulatory networks and gene regulatory networks, wherein the first three types are often static, while the gene regulatory network frequently is dynamic (Yuan et al. 2008) (Fig. 4).

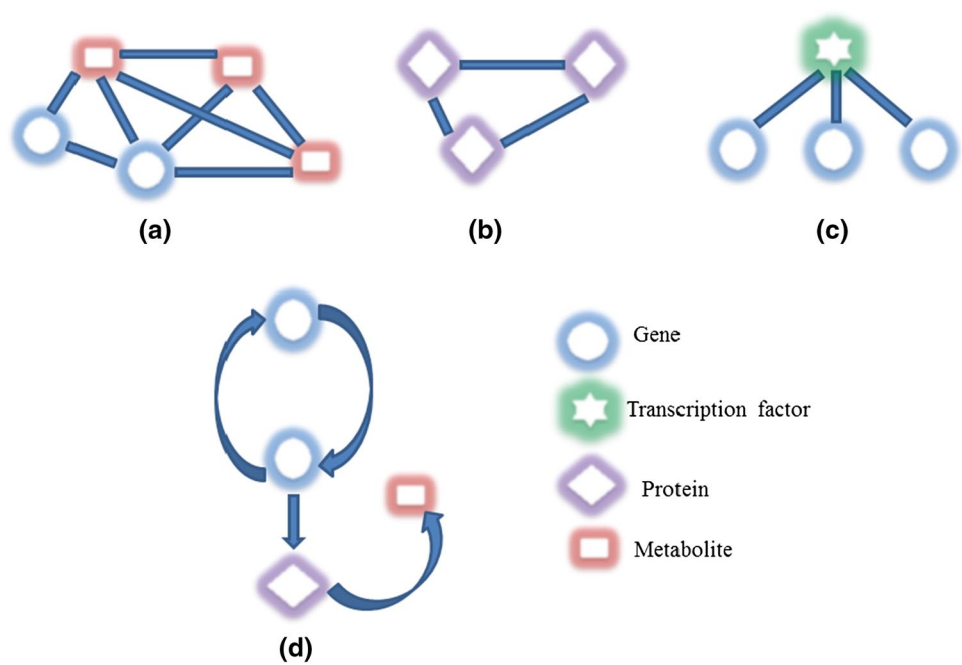
Gene-to-metabolite networks

Gene-to-metabolite networks are derived from the correlation analysis of genes and metabolites under a given set of conditions. Here the genes and metabolites act as nodes and the edges represent the regulatory interactions. The interactions are interpreted depending on the distance between the genes and the metabolites. These types of networks are highly complex and difficult to study in plants, owing to the enormous diversity and number of metabolites being produced in the cells corresponding to their sessile lifestyle. Various new research dimensions like interrelation among biological processes, gene functional annotation, discovery of new genes in biosynthesis, regulation and transport 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 (Goossens et al. 2003; Zulak et al. 2007; Carrari et al. 2006), 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 (Bläsing et al. 2005; Scheible et al. 2004) and so on.

Protein–protein interaction networks

In protein–protein interaction networks, the nodes are proteins which are connected by direct edges if the direction of information flow during their interaction is known, or non-directed edges if there is strong evidence of their physical interaction or association without an evidence for directionality of interaction (Assmann and Albert 2009). **Two types of interactions might be possible: genetic or physical.** A genetic protein–protein interaction is a network of genes characterized on the basis of genetic interactions to explain gene function within physiological processes (Boone et al. 2007). However, this approach is difficult to implement owing to 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

Fig. 4 Plant biological networks. **a** Gene-to-metabolite network, **b** protein–protein interaction network, **c** transcriptional regulatory network and **d** gene regulatory network



factors: the MADS (*MCM1*, *Agamous*, *Deficiens*, *SRF*) box transcription factors (Immink et al. 2003; de Folter et al. 2005) and the MYB (myeloblastosis) transcription factor family (Zimmermann et al. 2004). The further details regarding interactome are furnished in a preceding section in the current review namely ‘interactomics’.

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 edges 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 decipher transcriptional regulatory networks include, genome-wide expression profiling, genome-wide RNA interference (RNAi) screens (Baum and Craig 2004), transcription rate assessment by measurement of mRNA decay rates (Holstege et al. 1998; Nachman et al. 2004), the evaluation of promoter co-occupancy by pairs of transcription factors (Geisberg and Struhl 2004) and computational prediction of *cis*-elements (Beer and

Tavazoie 2004). In *Arabidopsis*, a transcriptional regulatory map was created for cold signaling mediated by the ICE1 transcription factor (Benedict et al. 2006). Recent reports on transcriptional regulatory networks include that in rice to understand the role of oxidative signals in chilling stress (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.

Gene regulatory networks

In a gene regulatory network, the nodes correspond to genes, messengers or proteins and the edges represent the regulatory interactions (activation, inhibition, repression or other functional interactions) among the components of the network. Complex gene regulatory networks are composed of genes, non-coding RNAs, proteins, metabolites and signaling components (Long et al. 2008). This type of network incorporates all the stages of regulation of gene expression including regulation of DNA transcription, RNA translation, post-transcriptional RNA processing as well as the post-translational modifications like protein targeting and covalent protein modification. These networks are often used to display the dynamics of the plant systems, unlike other networks which are static in nature (Yuan et al. 2008). The ABC model, which was one of the first plant gene regulatory networks modeled, explained the interactions among transcription factors that regulate floral pattern formation across plant species (Coen and Meyerowitz

1991). Gene regulatory networks have been reported in several studies to study developmental and physiological processes in plants. The studies include the attempt to model the essential components controlling guard cell size in stomatal closure (Li et al. 2006), cell fate determination during flower development in *A. thaliana* (Espinosa-Soto et al. 2004), microRNA (miRNA)-mediated gene regulatory networks (Meng et al. 2011) and recently in explaining land plant evolution (Pires et al. 2013).

Hence, biological network construction and analyses has been an important approach in plant systems biology to explain the organism or a biological process as a whole. The high-throughput technologies in modern science provide huge amount of quantitative data. However, the use of quantitative data is obstructed in systems wherein the knowledge of mechanistic details and kinetic parameters is scarce. In such cases, a wealth of molecular data on individual constituents as well as interactions can be helpful in modeling the system (Assmann and Albert 2009). The individual key components of the systems biology viz. genomic, 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.

Softwares and algorithms for plant systems biology

The use of bioinformatics' softwares is inevitable for the comprehensive study of plant systems biology. In addition to the tools and resources used in the analyses of the individual 'omics' platforms, several resources are required for the elucidation of the 'complete picture'. The detailed discussion of various algorithms and softwares used for systems biology are listed in Joyce and Palsson (2006) and Turenne (2011). These include the tools for network visualization, modeling environments, pathway construction and visualization tools, systems biology platforms and repositories of the models.

Visualization is a means of investigative data analysis and a key method for network analysis. The purpose of omics data visualization should be to create clear, meaningful and integrated resources without being besieged by the inherent complexity of the data (Gehlenborg et al. 2010). Several tools are available which help in visualization of 'omics' data on a systems scale like Sungear (Poultney et al. 2007), MapMan (Thimm et al. 2004), Genevestigator (Zimmermann et al. 2004), Cytoscape (Shannon et al. 2003), VirtualPlant (Katari et al. 2010), REACTOME (Joshi-Tope et al. 2005). Pathway databases are used for modeling systems, since they offer a clear-cut way of

building 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 Novère 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.

Conclusion and future perspectives

Plants are a solution to many environmental problems like food and water scarcity in the twenty-first century. Despite the difficulty to find a single perfect resolution, systems biology can be a medium to understand plants through mechanistic efforts, and eventually act as a base for partial solutions. The productivity of the individual plants could be enhanced by genetic modifications. However, the comprehensive understanding of the whole plants is a prerequisite to engineer those using molecular approaches. Ample knowledge about the response of the plants to internal and external stimuli is a must, which can be gained by systems biology (Ideker et al. 2001). Investigation of the individual hierarchical biological components and their integration is a key to systems biology. Although, the reductionist molecular approaches are inevitable for expansion of our insight, holistic systems biology approaches provide a complete sense of the plant systems. In addition to the promises, there are some biological and computational challenges to the application of systems' approaches to plants (Fernie 2012). The difficulty in deciphering the highly complex architecture of plants is one of the major challenges to success of the field. This problem could be dealt by improvement of the current experimental platforms through better technical innovations. The diversity of the data formats of the experimentally derived datasets is a major computational challenge to the integration of the massive data

(Heath and Kavraki 2009). Also due to the complexity of the plant biological networks, the experimental testing is not always possible. Still, they can be helpful in assisting and predicting the most promising experimental strategies thereby reducing the labor and time, otherwise used for the ‘trial and error’-based approach. But the above challenges can be met by the integrative efforts of diverse branches of plant sciences, rather than working in isolation. Hence systems biology will give rise to immense opportunities in decoding the complexity of plants by the fruitful collaboration of the classical plant biologists and computational modelers. The ‘unity in diversity’ approach of systems biology, is growing and will continue to impact the remarkable future of the plant science, thereby getting the most of the plants’ worth to benefit humanity in a pragmatic time frame.

Acknowledgements The authors would like to thank Centre for Advanced Studies in Plant Biotechnology and Genetic engineering (CPBGE), Govt. of Gujarat and Department of Biosciences, Saurashtra University, Rajkot for providing the necessary facilities. Also, financial assistance to Sheth B.P. in form of RFSMS (Research Fellowship in Science for Meritorious Students) from University Grants Commission, New Delhi is gratefully acknowledged.

References

- Agrawal GK, Job D, Zivy M et al (2011) Time to articulate a vision for the future of plant proteomics—a global perspective: an initiative for establishing the International Plant Proteomics Organization (INPPO). *Proteomics* 11:1559–1568
- Aharoni A, Keizer LC, Bouwmeester HJ et al (2000) Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *Plant Cell* 12:647–662
- Ahmadian A, Ehn M, Hober S (2006) Pyrosequencing: history, biochemistry and future. *Clin Chim Acta* 363:83–94
- Ahuja I, De Vos RC, Bones AM et al (2010) Plant molecular stress responses face climate change. *Trends Plant Sci* 15:664–674
- Alfarano C, Andrade CE, Anthony K et al (2005) The biomolecular interaction network database and related tools 2005 update. *Nucleic acids Res* 33(1):D418–D424
- Allwood JW, Goodacre R (2010) An introduction to liquid chromatography–mass spectrometry instrumentation applied in plant metabolomic analyses. *Phytochem Anal* 21:33–47
- Allwood JW, Ellis DI, Heald JK et al (2006) Metabolomic approaches reveal that phosphatidic and phosphatidyl glycerol phospholipids are major discriminatory metabolites in responses by *Brachypodium distachyon* to challenge by *Magnaporthe grisea*. *Plant J* 46:351–368
- Allwood JW, Ellis DI, Goodacre R (2008) Metabolomic technologies and their application to the study of plants and plant-host interactions. *Physiol Plant* 132:117–135
- Allwood JW, Clarke A, Goodacre R et al (2010) Dual metabolomics: a novel approach to understanding plant-pathogen interactions. *Phytochemistry* 71:590–597
- 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 Res* 34:W472–W476
- Altenbach SB, Vensel WH, DuPont FM (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. *J Cereal Sci* 52:143–151
- Amiour N, Imbaud S, Clément G et al (2012) The use of metabolomics integrated with transcriptomic and proteomic studies for identifying key steps involved in the control of nitrogen metabolism in crops such as maize. *J Exp Bot* 63(14):5017–5033
- Appel RD, Vargas JR, Palagi PM et al (1997) Melanie II—a third-generation software package for analysis of two-dimensional electrophoresis images: II. Algorithms. *Electrophoresis* 18(15):2735–2748
- Appella E, Padlan EA, Hunt DF (1995) Analysis of the structure of naturally processed peptides bound by class I and class II major histocompatibility complex molecules. *EXS* 73:105–119
- 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 et al (2010) The IntAct molecular interaction database in 2010. *Nucleic Acids Res* 38:D525–D531
- Ashby WR (1956) An introduction to cybernetics. Chapman & Hall, London
- Assmann SM, Albert R (2009) Discrete dynamic modeling with asynchronous update, or how to model complex systems in the absence of quantitative information. *Methods Mol Biol* 553:207–225
- Atwell S, Huang YS, Vilhjálmsson BJ et al (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465(7298):627–631
- Au KF, Jiang H, Lin L et al (2010) Detection of splice junctions from paired-end RNA-seq data by SpliceMap. *Nucleic Acids Res* 38(14):4570–4578
- Auge GA, Perelman S, Crocco CD et al (2009) Gene expression analysis of light-modulated germination in tomato seeds. *New Phytol* 183(2):301–314
- Avraham S, Tung CW, Ilic K et al (2008) The Plant Ontology Database: a community resource for plant structure and developmental stages controlled vocabulary and annotations. *Nucleic Acids Res* 36(1):D449–D454
- Babu MM, Luscombe NM, Aravind L et al (2004) Structure and evolution of transcriptional regulatory networks. *Curr Opin Struc Biol* 14(3):283–291
- Bae MS, Cho EJ, Choi EY et al (2003) Analysis of the Arabidopsis nuclear proteome and its response to cold stress. *Plant J* 36:652–663
- Baerenfeller K, Grossmann J, Grobei MA et al (2008) Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. *Science* 320(5878):938–941
- Ball CA, Awad IA, Demeter J et al (2005) The Stanford Microarray Database accommodates additional microarray platforms and data formats. *Nucleic Acids Res* 33(1):D580–D582
- Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell’s functional organization. *Nat Rev Genet* 5:101–115
- Barkan A (2009) Genome-wide analysis of RNA-protein interactions in plants. In: Belostotsky D (ed) *Plant systems biology*. Humana Press, New York, pp 13–37
- Baulcombe D (2004) RNA silencing in plants. *Nature* 431(7006):356–363
- Baum B, Craig G (2004) RNAi in a postmodern, postgenomic era. *Oncogene* 23(51):8336–8339
- Beck S, Rakyan VK (2008) The methylome: approaches for global DNA methylation profiling. *Trends Genet* 24:231–237
- Beer MA, Tavazoie S (2004) Predicting gene expression from sequence. *Cell* 117(2):185–198
- Belostotsky DA, Rose AB (2005) Plant gene expression in the age of systems biology: integrating transcriptional and post-transcriptional events. *Trends Plant Sci* 10(7):347–353

- 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 Physiol* 141(4):1219–1232
- Bennett S, Barnes C, Cox A et al (2005) Toward the 1,000 dollars human genome. *Pharmacogenomics* 6:373–382
- Berger SL (2002) Histone modifications in transcriptional regulation. *Curr Opin Genet Dev* 12(2):142–148
- Bino RJ, De Vos CH, Lieberman M et al (2005) The light-hyperresponsive high pigment-2dg mutation of tomato: alterations in the fruit metabolome. *New Phytol* 166(2):427–438
- Bird A (2007) Perceptions of epigenetics. *Nature* 447(7143):396–398
- Bird AP (2001) Molecular biology-Methylation talk between histones and DNA. *Science* 294:2113–2115
- Blais A, Dynlacht BD (2005) Constructing transcriptional regulatory networks. *Gene Dev* 19(13):1499–1511
- Blakes J, Twycross J, Romero FJ et al (2011) The Infobiotics Workbench: an integrated in silico modelling platform for systems and synthetic biology. *Bioinformatics* 27(23):3323–3324
- Bläsing OE, Gibon Y, Günther M et al (2005) Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in Arabidopsis. *Plant Cell* 17(12):3257–3281
- Bock C, Lengauer T (2008) Computational epigenetics. *Bioinformatics* 24(1):1–10
- Bonnet E, Van de Peer Y, Rouzé P (2006) The small RNA world of plants. *New Phytol* 171(3):451–468
- Boone C, Bussey H, Andrews BJ (2007) Exploring genetic interactions and networks with yeast. *Nature Rev Genet* 8(6):437–449
- Borevitz JO, Nordborg M (2003) The impact of genomics on the study of natural variation in *Arabidopsis*. *Plant Physiol* 132(2):718–725
- Bouyer D, Roudier F, Heese M et al (2011) Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. *PLoS Genet* 7:e1002014
- Bovy AG, Gómez-Roldán V, Hall RD (2010) Strategies to optimize the flavonoid content of tomato fruit. *Rec Adv Polyphen Res* 2:138–162
- Brandão MM, Dantas LL, Silva-Filho MC (2009) AtPIN: *Arabidopsis thaliana* protein interaction network. *BMC Bioinform* 10(1):454
- Braun P, Aubourg S, Van Leene J et al (2013) Plant protein interactomes. *Annu Rev Plant Biol* 64:161–187
- Breitkreutz BJ, Stark C, Tyers M (2003) Osprey: a network visualization system. *Genome Biol* 4(3):R22
- Breitling R, Cenicerós A, Jankevics A, Takano E (2013) Metabolomics for secondary metabolite research. *Metabolites* 3(4):1076–1083
- Brink RA (1958) Paramutation at the R locus in maize. In: Cold Spring Harbor Symposium Quant Biol 23, Cold Spring Harbor Laboratory Press, New York, pp 379–391
- Burn JE, Bagnall DJ, Metzger JD et al (1993) DNA methylation, vernalization, and the initiation of flowering. *Proc Natl Acad Sci* 90(1):287–291
- Busch W, Lohmann JU (2007) Profiling a plant: expression analysis in *Arabidopsis*. *Current Opin Plant Biol* 10(2):136–141
- Bylesjö M, Rantalainen M, Cloarec O et al (2006) OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J Chemom* 20(8–10):341–351
- Caldwell KS, Russell J, Langridge P et al (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics* 172(1):557–567
- Carrari F, Baxter C, Usadel B et al (2006) Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. *Plant Physiol* 142(4):1380–1396
- Casati P, Campi M, Morrow DJ et al (2011) Transcriptomic, proteomic and metabolomic analysis of UV-B signaling in maize. *BMC Genom* 12(1):321
- Caspi R, Altman T, Dale JM et al (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 38(1):D473–D479
- Castiglione MR, Cremonini R, Frediani M (2002) DNA methylation patterns on plant chromosomes. *Caryologia* 55:275–282
- Cedar H, Bergman Y (2009) Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* 10(5):295–304
- Charbonnier S, Gallego O, Gavin AC (2008) The social network of a cell: recent advances in interactome mapping. *Biotech Ann Rev* 14:1–28
- Chautard E, Ballut L, Thierry-Mieg N et al (2009) MatrixDB, a database focused on extracellular protein–protein and protein–carbohydrate interactions. *Bioinformatics* 25(5):690–691
- Chevreur B, Pfisterer T, Drescher B et al (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res* 14(6):1147–1159
- Choi H, Pavelka N (2011) When one and one gives more than two: challenges and opportunities of integrative omics. *Front Genet* 2:105
- Choi YH, Kim HK, Linthorst HJM et al (2006) NMR metabolomics to revisit the tobacco mosaic virus infection in *Nicotiana tabacum* leaves. *J Nat Prod* 69:742–748
- Clarke J, Wu HC, Jayasinghe L et al (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol* 4:265–270
- Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. *Nature* 353(6339):31–37
- Cokus SJ, Feng S, Zhang X et al (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452:215–219
- Coruzzi GM, Burga AR, Katari MS et al (2009) Systems biology: principles and applications in plant research. *Ann Plant Reviews* 35:3–40
- Cowan JD, Sharp DH (1988) Neural nets and artificial intelligence. *Daedalus* 117(1):85–121
- Cowley MJ, Pinese M, Kassahn KS et al (2012) PINA v2. 0: mining interactome modules. *Nucleic Acids Res* 40(D1):D862–D865
- Craigon DJ, James N, Okyere J et al (2004) NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. *Nucleic Acids Res* 32(1):D575–D577
- Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, London
- Cui J, Li P, Li G et al (2008) AtPID: *Arabidopsis thaliana* protein interactome database—an integrative platform for plant systems biology. *Nucleic Acids Res* 36(1):D999–D1008
- Cusick ME, Klitgord N, Vidal M et al (2005) Interactome: gateway into systems biology. *Hum Mol Genet* 14(2):R171–R181
- Davies PJ (ed) (2004) Plant hormones: biosynthesis, signal transduction, action!. Springer-Verlag, Berlin
- De Bodt S, Hollunder J, Nelissen H et al (2012) CORNET 2.0: integrating plant coexpression, protein–protein interactions, regulatory interactions, gene associations and functional annotations. *New Phytol* 195(3):707–720
- de Folter S, Immink RG, Kieffer M et al (2005) Comprehensive interaction map of the Arabidopsis MADS box transcription factors. *Plant Cell* 17(5):1424–1433
- De Geyter N, Gholami A, Goormachtig S, Goossens A (2012) Transcriptional machineries in jasmonate-elicited plant secondary metabolism. *Trends Plant Sci* 17(6):349–359

- De Keersmaecker SC, Thijs I, Vanderleyden J et al (2006) Integration of omics data: how well does it work for bacteria? *Mol Microbiol* 62(5):1239–1250
- De Luca V, St Pierre B (2000) The cell and developmental biology of alkaloid biosynthesis. *Trends Plant Sci* 5(4):168–173
- Delp G, Gradin T, Åhman I (2009) Microarray analysis of the interaction between the aphid *Rhopalosiphum padi* and host plants reveals both differences and similarities between susceptible and partially resistant barley lines. *Mol Genet Genomics* 281(3):233–248
- Dhar PK, Zhu H, Mishra SK (2004) Computational approach to systems biology: from fraction to integration and beyond. *IEEE Trans Nanobiosci* 3(3):144–152
- Ding X, Richter T, Chen M et al (2009) A rice kinase-protein interaction map. *Plant Physiol* 149(3):1478–1492
- Domon B, Aebersold R (2006) Mass spectrometry and protein analysis. *Science* 312(5771):212–217
- Ebert B, Zöllner D, Erban A et al (2010) Metabolic profiling of *Arabidopsis thaliana* epidermal cells. *J Exp Bot* 61(5):1321–1335
- Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30(1):207–210
- Edwards D, Batley J (2004) Plant bioinformatics: from genome to phenome. *Trends Biotechnol* 22(5):232–237
- Eid J, Fehr A, Gray J et al (2009) Real-time DNA sequencing from single polymerase molecules. *Science* 323(5910):133–138
- Elling AA, Deng XW (2009) Next-generation sequencing reveals complex relationships between the epigenome and transcriptome in maize. *Plant Signal Behav* 4(8):760–762
- Enfissi EM, Barneche F, Ahmed I et al (2010) Integrative transcript and metabolite analysis of nutritionally enhanced DE-ETIOLATED1 downregulated tomato fruit. *Plant Cell* 22(4):1190–1215
- Espinosa-Soto C, Padilla-Longoria P, Alvarez-Buylla ER (2004) A gene regulatory network model for cell-fate determination during *Arabidopsis thaliana* flower development that is robust and recovers experimental gene expression profiles. *Plant Cell* 16(11):2923–2939
- Fedoroff NV (2002) Cross-talk in abscisic acid signaling. *Sci Signal* 2002(140):e10
- Feng S, Jacobsen SE (2011) Epigenetic modifications in plants: an evolutionary perspective. *Curr Opin Plant Biol* 14(2):179–186
- Feng S, Cokus SJ, Zhang X et al (2010) Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci USA* 107:8689–8694
- Fernie AR (2012) Grand challenges in plant systems biology: closing the circle (s). *Front Plant Sci* 3:35
- Fernie AR, Schauer N (2009) Metabolomics-assisted breeding: a viable option for crop improvement? *Trends Genet* 25(1):39–48
- Feuillet C, Leach JE, Rogers J et al (2011) Crop genome sequencing: lessons and rationales. *Trends Plant Sci* 16(2):77–88
- Fiehn O (2002) Metabolomics—the link between genotypes and phenotypes. *Plant Mol Biol* 48(1–2):155–171
- Fiehn O, Kopka J, Dörmann P et al (2000) Metabolite profiling for plant functional genomics. *Nat Biotechnol* 18(11):1157–1161
- Filichkin SA, Priest HD, Givan SA et al (2010) Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res* 20(1):45–58
- Finn RD, Marshall M, Bateman A (2005) iPfam: visualization of protein–protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics* 21(3):410–412
- Forrester JW (1958) Industrial Dynamics: A Major Breakthrough for Decision Makers. Harvard Business Rev 36(4):37–66
- Forrester JW (1961) Industrial Dynamics. Productivity Press, Portland
- Fouracre JP, Ando S, Langdale JA (2014) Cracking the Kranz enigma with systems biology. *J Exp Bot*. doi:10.1093/jxb/eru015
- Fu CH, Chen YW, Hsiao YY et al (2011) OrchidBase: a collection of sequences of transcriptome derived from orchids. *Plant Cell Physiol* 52:238–243
- Fukushima A, Kusano M (2013) Recent progress in the development of metabolome databases for plant systems biology. *Front Plant Sci* 4:73
- Galindo González LM, El Kayal W, Ju CJT et al (2012) Integrated transcriptomic and proteomic profiling of white spruce stems during the transition from active growth to dormancy. *Plant Cell Environ* 35(4):682–701
- Garber M, Grabherr MG, Guttman M et al (2011) Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods* 8(6):469–477
- Gehlenborg N, O'Donoghue SI, Baliga NS et al (2010) Visualization of omics data for systems biology. *Nat Methods* 7: S56–S68
- Gehring M, Bubba KL, Henikoff S (2009) Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* 324:1447–1451
- Geisberg JV, Struhl K (2004) Quantitative sequential chromatin immunoprecipitation, a method for analyzing co-occupancy of proteins at genomic regions in vivo. *Nucleic Acids Res* 32(19):e151–e151
- Geisler-Lee J, O'Toole N, Ammar R et al (2007) A predicted interactome for Arabidopsis. *Plant Physiol* 145(2):317–329
- Glinski M, Weckwerth W (2006) The role of mass spectrometry in plant systems biology. *Mass Spectrom Rev* 25(2):173–214
- Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. Addison-Wesley, Reading
- Gomez-Roldan V, Fermas S, Brewer PB et al (2008) Strigolactone inhibition of shoot branching. *Nature* 455(7210):189–194
- Gonzalez N, De Bodt S, Sulpice R et al (2010) Increased leaf size: different means to an end. *Plant Physiol* 153:1261–1279
- Goossens A, Häkkinen ST, Laakso I et al (2003) Secretion of secondary metabolites by ATP-binding cassette transporters in plant cell suspension cultures. *Plant Physiol* 131(3):1161–1164
- Gowik U, Bräutigam A, Weber KL et al (2011) Evolution of C4 photosynthesis in the genus Flaveria: how many and which genes does it take to make C4? *Plant Cell* 23(6):2087–2105
- Grabherr MG, Haas BJ, Yassour M et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29(7):644–652
- Grant-Downton RT, Dickinson HG (2005) Epigenetics and its implications for plant biology. 1. The epigenetic network in plants. *Ann Bot* 96(7):1143–1164
- Gregory TR, Nicol JA, Tamm H et al (2007) Eukaryotic genome size databases. *Nucleic Acids Res* 35(1):D332–D338
- Groenenboom M, Gomez-Roldan V, Stigter H, Astola L, van Daelen R, Beekwilder J et al (2013) The flavonoid pathway in tomato seedlings: transcript abundance and the modeling of metabolite dynamics. *PLoS One* 8(7):e68960
- Gu H, Zhu P, Jiao Y et al (2011) PRIN: a predicted rice interactome network. *BMC Bioinform* 12(1):161
- Gygi SP, Rochon Y, Franza BR et al (1999) Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* 19:1720–1730
- Hall RD (2006) Plant metabolomics: from holistic hope, to hype, to hot topic. *New Phytol* 169(3):453–468
- Hall RD, De Vos CHR, Ward J (2010) Plant metabolomics applications in the Brassicaceae: added value for science and industry. *Acta Hort* 867:191–206
- Hamblin MT, Casa AM, Sun H et al (2006) Challenges of detecting directional selection after a bottleneck: lessons from *Sorghum bicolor*. *Genetics* 173:953–964
- Hartwell LH, Hopfield JJ, Leibler S et al (1999) From molecular to modular cell biology. *Nature* 402:47–51

- Hazen SP, Pathan MS, Sanchez A et al (2005) Expression profiling of rice segregating for drought tolerance QTLs using a rice genome array. *Funct Integr Genomics* 5(2):104–116
- He G, Elling AA, Deng XW (2011) The epigenome and plant development. *Ann Rev Plant Biol* 62:411–435
- He G, Chen B, Wang X et al (2013) Conservation and divergence of transcriptomic and epigenomic variation in maize hybrids. *Genome Biol* 14(6):R57
- Heath AP, Kavraki LE (2009) Computational challenges in systems biology. *Comput Sci Review* 3(1):1–17
- Heinrich R, Schuster S (1996) The regulation of cellular systems. Chapman & Hall, New York
- Hirai MY, Saito K (2008) Analysis of systemic sulfur metabolism in plants using integrated ‘-omics’ strategies. *Mol Biosyst* 4:967–973
- Hirai MY, Yano M, Goodenowe DB et al (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 101:10205–10210
- Hirai MY, Sugiyama K, Sawada Y et al (2007) Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc Natl Acad Sci USA* 104:6478–6483
- Holstege FC, Jennings EG, Wyrick JJ et al (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95:717–728
- Hong D, Rhie A, Park SS et al (2012) FX: an RNA-Seq analysis tool on the cloud. *Bioinformatics* 28(5):721–723
- Hoogland C, Sanchez JC, Tonella L et al (2000) The 1999 SWISS-2DPAGE database update. *Nucleic Acids Res* 28:286–288
- Hoops S, Sahle S, Gauges R et al (2006) COPASI—a complex pathway simulator. *Bioinformatics* 22(24):3067–3074
- Hsieh TF, Ibarra CA, Silva P et al (2009) Genome-wide demethylation of *Arabidopsis* endosperm. *Science* 324:1451–1454
- Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. *Genome Res* 9(9):868–877
- Huang S, Li R, Zhang Z et al (2009) The genome of the cucumber, *Cucumis sativus* L. *Nat Genet* 41:1275–1281
- Huang L, Gao W, Zhou J, Wang R (2010a) Systems biology applications to explore secondary metabolites in medicinal plants. *China J Chin Materia Medica* 35(1):8–12
- Huang X, Wei X, Sang T et al (2010b) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* 42:961–967
- Hucka M, Finney A, Sauro HM et al (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19(4):524–531
- Hunt AG, Xu R, Addepalli B et al (2008) *Arabidopsis* mRNA polyadenylation machinery: comprehensive analysis of protein–protein interactions and gene expression profiling. *BMC Genom* 9(1):220
- Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: systems biology. *Ann Rev Genom Hum G* 2(1):343–372
- Imelfort M, Edwards D (2009) De novo sequencing of plant genomes using second-generation technologies. *Briefings Bioinform* 10(6):609–618
- Immink RG, Ferrario S, Busscher-Lange J et al (2003) Analysis of the petunia MADS-box transcription factor family. *Mol Genet Genomics* 268(5):598–606
- Izawa T, Mihara M, Suzuki Y et al (2011) Os-GIGANTEA confers robust diurnal rhythms on the global transcriptome of rice in the field. *Plant Cell* 23(5):1741–1755
- Jaillon O, Aury JM, Noel B et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449(7161):463–467
- Jansen JJ, Allwood JW, Marsden-Edwards E et al (2008) Metabolomic analysis of the interaction between plants and herbivores. *Metabolomics* 5:150–161
- Jiao Y, Lau OS, Deng XW (2007) Light-regulated transcriptional networks in higher plants. *Nature Rev Genet* 8(3):217–230
- Joshi-Tope G, Gillespie M, Vastrik I et al (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res* 33(1):D428–D432
- Joyce AR, Palsson BO (2006) The model organism as a system: integrating ‘omics’ data sets. *Nat Rev Mol Cell Biol* 7:198–210
- Kanehisa M, Goto S, Sato Y et al (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40(D1):D109–D114
- Katari MS, Nowicki SD, Aceituno FF et al (2010) VirtualPlant: a software platform to support systems biology research. *Plant Physiol* 152(2):500–515
- Kent WJ (2002) BLAT—the BLAST-like alignment tool. *Genome Res* 12(4):656–664
- Kermicle JL (1970) Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. *Genetics* 66(1):69
- Keurentjes JJ, Fu J, De Vos CR et al (2006) The genetics of plant metabolism. *Nature Genet* 38(7):842–849
- Keurentjes JJ, Koornneef M, Vreugdenhil D (2008) Quantitative genetics in the age of omics. *Curr Opin Plant Biol* 11(2):123–128
- Keurentjes JJ, Angenent GC, Dicke M et al (2011) Redefining plant systems biology: from cell to ecosystem. *Trends Plant Sci* 16(4):183–190
- Kim HK, Verpoorte R (2010) Sample preparation for plant metabolomics. *Phytochem Anal* 21(1):4–13
- Kirschner MW (2005) The meaning of systems biology. *Cell* 121:503–504
- Kitano H (2002) Systems biology: a brief overview. *Science* 295(5560):1662–1664
- Kitano H (2007) Towards a theory of biological robustness. *Mol Syst Biol* 3(1):137
- Kogel KH, Voll LM, Schäfer P et al (2010) Transcriptome and metabolome profiling of field-grown transgenic barley lack induced differences but show cultivar-specific variances. *Proc Natl Acad Sci USA* 107(14):6198–6203
- Komatsu S (2006) Plant proteomics databases: their status in 2005. *Curr Bioinform* 1(1):33–36
- Koza JR (1992) Genetic programming: On the programming of computers by means of natural selection, vol. 1, MIT press, Cambridge
- Krallinger M, Morgan A, Smith L et al (2008) Evaluation of text-mining systems for biology: overview of the second BioCreative community challenge. *Genome Biol* 9(2):S1
- 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 Res* 37(2):W160–W165
- Kruft V, Eubel H, Jansch L et al (2001) Proteomic approach to identify novel mitochondrial proteins in *Arabidopsis*. *Plant Physiol* 127(4):1694–1710
- Laakso M, Hautaniemi S (2010) Integrative platform to translate gene sets to networks. *Bioinformatics* 26(14):1802–1803
- Lachenbruch PA (1975) Zero-mean difference discrimination and the absolute linear discriminant function. *Biometrika* 62(2):397–401
- Langmead B, Trapnell C, Pop M et al (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25
- Le Novère 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 Res* 34(1):D689–D691
- Lelandais-Brière C, Naya L, Sallet E (2009) Genome-wide *Medicago truncatula* small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. *Plant Cell* 21(9):2780–2796
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760
- Li S, Assmann SM, Albert R (2006) Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. *PLoS Biol* 4(10):e312
- Licata L, Briganti L, Peluso D et al (2012) MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* 40(D1):D857–D861
- Lin M, Shen X, Chen X (2011) PAIR: the predicted Arabidopsis interaction resource. *Nucleic Acids Res* 39(1):D1134–D1140
- Lister R, O'Malley RC, Tonti-Filippini J et al (2008) Highly integrated singlebase resolution maps of the epigenome in *Arabidopsis*. *Cell* 133:523–536
- Lister R, Gregory BD, Ecker JR (2009) Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Curr Opin Plant Biol* 12:107–118
- Loew LM, Schaff JC (2001) The virtual cell: a software environment for computational cell biology. *Trends Biotechnol* 19(10):401–406
- Long TA, Brady SM, Benfey PN (2008) Systems approaches to identifying gene regulatory networks in plants. *Annu Rev Cell Dev Biol* 24:81–103
- Lu C, Kulkarni K, Souret FF et al (2006) MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. *Genome Res* 16(10):1276–1288
- Lu T, Lu G, Fan D et al (2010) Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome Res* 20(9):1238–1249
- Makita Y, Kobayashi N, Mochizuki Y et al (2009) PosMed-plus: an intelligent search engine that inferentially integrates cross-species information resources for molecular breeding of plants. *Plant Cell Physiol* 50(7):1249–1259
- Maor R, Jones A, Nühse TS et al (2007) Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. *Mol Cell Proteom* 6(4):601–610
- Margulies M, Egholm M, Altman WE et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057):376–380
- Matzke M, Kanno T, Daxinger L et al (2009) RNA-mediated chromatin-based silencing in plants. *Curr Opin Cell Biol* 21(3):367–376
- McCintock B (1984) The significance of responses of the genome to challenge. *Science* 226:792–801
- McMullen MD, Kresovich S, Villeda HS et al (2009) Genetic properties of the maize nested association mapping population. *Science* 325(5941):737–740
- Mendenhall EM, Bernstein BE (2008) Chromatin state maps: new technologies, new insights. *Curr Opin Genet Dev* 18:109–115
- Mendes P (1997) Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3. *Trends Biochem Sci* 22:361–363
- Meng Y, Shao C, Wang H et al (2011) The regulatory activities of plant microRNAs: a more dynamic perspective. *Plant Physiol* 157(4):1583–1595
- Millar AH, Sweetlove LJ, Giegé P et al (2001) Analysis of the Arabidopsis mitochondrial proteome. *Plant Physiol* 127(4):1711–1727
- Milos P (2008) Helicos BioSciences. *Pharmacogenomics* 9:477–480
- Ming R, Hou S, Feng Y et al (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452(7190):991–996
- Minorsky PV (2003) Achieving the in silico plant. *Systems biology and the future of plant biological research. Plant Physiol* 132(2):404–409
- Mirouze M (2012) The small RNA-based odyssey of epigenetic information in plants: from cells to species. *DNA Cell Biol* 31(12):1650–1656
- Moco S, Forshed J, De Vos RC et al (2008) Intra- and inter-metabolite correlation spectroscopy of tomato metabolomics data obtained by liquid chromatography-mass spectrometry and nuclear magnetic resonance. *Metabolomics* 4(3):202–215
- Morris GP, Ramu P, Deshpande SP et al (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc Natl Acad Sci USA* 110(2):453–458
- Morsy M, Gouthu S, Orchard S et al (2008) Charting plant interactomes: possibilities and challenges. *Trends Plant Sci* 13(4):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 Biol* 9(1):S4
- Moxon S, Jing R, Szittyá G et al (2008) Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Res* 18(10):1602–1609
- Mueller LA, Zhang P, Rhee SY (2003) AraCyc: a biochemical pathway database for Arabidopsis. *Plant Physiol* 132(2):453–460
- Mukhtar MS, Carvunis AR, Dreze M et al (2011) Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science* 333:596–601
- Nachman I, Regev A, Friedman N (2004) Inferring quantitative models of regulatory networks from expression data. *Bioinformatics* 20(1):i248–i256
- Nagasaki M, Saito A, Jeong E et al (2010) Cell Illustrator 4.0: a computational platform for systems biology. *In Silico Biol* 10(1):5–26
- Nakashima K, Ito Y, Yamaguchi-Shinozaki K (2009) Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. *Plant Physiol* 149(1):88–95
- Nishimura D (2001) BioCarta. *Biotech Softw Internet Rep* 2:117–120
- Noble D (2002) Modeling the heart—from genes to cells to the whole organ. *Science* 295(5560):1678–1682
- Nobuta K, Venu RC, Lu C et al (2007) An expression atlas of rice mRNAs and small RNAs. *Nat Biotechnol* 25:473–477
- Nordborg M, Tavaré S (2002) Linkage disequilibrium: what history has to tell us. *Trends Genet* 18(2):83–90
- Nordborg M, Hu T, Ishino Y et al (2005) The genomic pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol* 3:e196
- Okazaki Y, Shimajima M, Sawada Y 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
- Oliver SG, Winson MK, Kell DB, Baganz F (1998) Systematic functional analysis of the yeast genome. *TIPTECH* 16:373–378
- Olivier BG, Snoep JL (2004) Web-based kinetic modelling using JWS online. *Bioinformatics* 20:2143–2144
- O'Rourke JA, Nelson RT, Grant D et al (2009) Integrating microarray analysis and the soybean genome to understand the soybeans iron deficiency response. *BMC Genom* 10(1):376
- Osorio S, Alba R, Damasceno CM, Lopez-Casado G, Lohse M, Zanor MI et al (2011) Systems biology of tomato fruit development: combined transcript, protein, and metabolite analysis of tomato transcription factor (nor, rin) and ethylene receptor (Nr) mutants reveals novel regulatory interactions. *Plant Physiol* 157(1):405–425
- Pabinger S, Rader R, Agren R et al (2011) MEMOSys: bioinformatics platform for genome-scale metabolic models. *BMC Syst Biol* 5(1):20
- Pandey A, Mann M (2000) Proteomics to study genes and genomes. *Nature* 405(6788):837–846

- Paran I, Zamir D (2003) Quantitative traits in plants: beyond the QTL. *Trends Genet* 19(6):303–306
- Park OK (2004) Proteomic studies in plants. *J Biochem Mol Biol* 37(1):133–138
- Park PJ (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 10:669–680
- Parkinson H, Sarkans U, Shojatalab M et al (2005) ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* 33(1):D553–D555
- Pasam RK, Sharma R, Malosett M et al (2012) Genome-wide association studies for agronomical traits in a worldwide spring barley collection. *BMC Plant Biol* 12(1):16
- Paterson AH, Bowers JE, Bruggmann R et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457(7229):551–556
- Patterson SD, Aebersold RH (2003) Proteomics: the first decade and beyond. *Nature Genet* 33:311–323
- Peltier JB, Friso G, Kalume DE et al (2000) Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. *Plant Cell* 12:319–342
- Peltier JB, Ytterberg J, Liberles DA et al (2001) Identification of a 350 kDa ClpP protease complex with 10 different Clp isoforms in chloroplasts of *Arabidopsis thaliana*. *J Biol Chem* 276:16318–16327
- Pertea G, Huang X, Liang F et al (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* 19(5):651–652
- Phizicky E, Bastiaens PI, Zhu H et al (2003) Protein analysis on a proteomic scale. *Nature* 422:208–215
- Pichersky E, Gang DR (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci* 5(10):439–445
- Pires ND, Yi K, Breuninger H et al (2013) Recruitment and remodeling of an ancient gene regulatory network during land plant evolution. *Proc Natl Acad Sci USA* 110(23):9571–9576
- Poultney CS, Gutiérrez RA, Katari MS et al (2007) Sungear: interactive visualization and functional analysis of genomic datasets. *Bioinformatics* 23(2):259–261
- Prieto C, De Las Rivas J (2006) APID: agile protein interaction data analyzer. *Nucleic Acids Res* 34(2):W298–W302
- Prime T, Sherrier D, Mahon P et al (2000) A proteomic analysis of organelles from *Arabidopsis thaliana*. *Electrophoresis* 21:3488–3499
- Proietti S, Bertini L, Timperio AM et al (2013) Crosstalk between salicylic acid and jasmonate in *Arabidopsis* investigated by an integrated proteomic and transcriptomic approach. *Mol Bio Sys* 9(6):1169–1187
- Provart NJ (2012) Correlation networks visualization. *Front Plant Sci* 3:240
- Rafalski JA (2010) Association genetics in crop improvement. *Curr Opin Plant Biol* 13(2):174–180
- Rischer H, Orešič M, Seppänen-Laakso T et al (2006) Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells. *Proc Natl Acad Sci* 103(14):5614–5619
- Rosengren AT, Salmi JM, Aittokallio T et al (2003) Comparison of PDQuest and Progenesis software packages in the analysis of two-dimensional electrophoresis gels. *Proteomics* 3(10):1936–1946
- Rothberg JM, Hinz W, Rearick TM et al (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475(7356):348–352
- Rouquie D, Peltier JB, Marquis Mansion M et al (1997) Construction of a directory of tobacco plasma membrane proteins by combined two-dimensional gel electrophoresis and protein sequencing. *Electrophoresis* 18:654–660
- Saito K, Matsuda F (2010) Metabolomics for functional genomics, systems biology, and biotechnology. *Annu Rev Plant Biol* 61:463–489
- Saito K, Hirai MY, Yonekura-Sakakibara K (2008) Decoding genes with coexpression networks and metabolomics—‘majority report by precogs’. *Trends Plant Sci* 13:36–43
- Saliba-Colombani V, Causse M, Gervais L et al (2000) Efficiency of RFLP, RAPD, and AFLP markers for the construction of an intraspecific map of the tomato genome. *Genome* 43(1):29–40
- Sanchez DH, Szymanski J, Erban A et al (2010) Mining for robust transcriptional and metabolic responses to long-term salt stress: a case study on the model legume *Lotus japonicus*. *Plant Cell Environ* 33:468–480
- Sato S, Hirakawa H, Isobe S et al (2011) Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. *DNA Res* 18:65–76
- Sauro HM, Hucka M, Finney A et al (2003) Next generation simulation tools: the systems biology workbench and BioSPICE integration. *OMICS* 7(4):355–372
- Schad M, Mungur R, Fiehn O et al (2005) Metabolic profiling of laser microdissected vascular bundles of *Arabidopsis thaliana*. *Plant Methods* 1(1):2
- Schaefer CF, Anthony K, Krupa S et al (2009) PID: the pathway interaction database. *Nucleic Acids Res* 37(1):D674–D679
- Schauer N, Semel Y, Roessner U et al (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat Biotechnol* 24:447–454
- Scheible WR, Morcuende R, Czechowski T et al (2004) Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of *Arabidopsis* in response to nitrogen. *Plant Physiol* 136(1):2483–2499
- Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326(5956):1112–1115
- Schneeberger K, Weigel D (2011) Fast-forward genetics enabled by new sequencing technologies. *Trends Plant Sci* 16(5):282–288
- Schones DE, Zhao K (2008) Genome-wide approaches to studying chromatin modifications. *Nat Rev Genet* 9:179–191
- Schoonheim PJ, Veiga H, da Costa Pereira D et al (2007) A comprehensive analysis of the 14-3-3 interactome in barley leaves using a complementary proteomics and two-hybrid approach. *Plant Physiol* 143(2):670–683
- Shannon P, Markiel A, Ozier O et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–2504
- Sharan R, Ideker T (2006) Modeling cellular machinery through biological network comparison. *Nat Biotechnol* 24(4):427–433
- Sharan R, Ulitsky I, Shamir R (2007) Network-based prediction of protein function. *Mol Sys Biol* 3(88)
- Sherrier DJ, Prime TA, Dupree P (1999) Glycosylphosphatidylinositol-anchored cell-surface proteins from *Arabidopsis*. *Electrophoresis* 20:2027–2035
- Shingaki-Wells RN, Huang S, Taylor NL et al (2011) Differential molecular responses of rice and wheat coleoptiles to anoxia reveal novel metabolic adaptations in amino acid metabolism for tissue tolerance. *Plant Physiol* 156(4):1706–1724
- Shinozaki K, Sakakibara H (2009) Omics and bioinformatics: an essential toolbox for systems analyses of plant functions beyond 2010. *Plant Cell Physiol* 50(7):1177–1180
- Shulaev V, Sargent DJ, Crowhurst RN et al (2010) The genome of woodland strawberry (*Fragaria vesca*). *Nat Genet* 43(2):109–116
- Solomon MJ, Larsen PL, Varshavsky A (1988) Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell* 53:937–947

- Spiering MJ, Kaur B, Parsons JF et al (2014) Systems approaches to unraveling plant metabolism: identifying biosynthetic genes of secondary metabolic pathways. In: Plant Metabolism. Humana Press, pp 253–273
- Staab PR, Walossek J, Nellessen D et al (2010) SynBioWave—a real-time communication platform for molecular and synthetic biology. *Bioinformatics* 26(21):2782–2783
- Stark C, Breitkreutz BJ, Reguly T et al (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* 34(1):D535–D539
- Steinfath M, Reipsilber D, Scholz M et al (2007) Integrated data analysis for genome-wide research. *EXS* 97:309–329
- Stelling J, Sauer U, Szallasi Z et al (2004) Robustness of cellular functions. *Cell* 118(6):675–685
- Sterck L, Rombauts S, Vandepoele K et al (2007) How many genes are there in plants (... and why are they there)? *Curr Opin Plant Biol* 10(2):199–203
- Su CL, Chao YT, Alex Chang YC et al (2011) De novo assembly of expressed transcripts and global analysis of the *Phalaenopsis aphrodite* transcriptome. *Plant Cell Physiol* 52:1501–1514
- Sucaet Y, Wang Y, Li J et al (2012) MetNet online: a novel integrated resource for plant systems biology. *BMC Bioinform* 13(1):267
- Sumner L, Mendes P, Dixon R (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* 62:817–836
- Sumner LW, Yang DS, Bench BJ et al (2011) Spatially—resolved metabolomics—challenges for the future. In: Hall RD (ed) The biology of plant metabolomics. Blackwell-Wiley, London
- Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 9:465–476
- Swarbreck D, Wilks C, Lamesch P et al (2008) The Arabidopsis information resource (TAIR): gene structure and function annotation. *Nucleic Acids Res* 36(1):D1009–D1014
- Szitty G, Moxon S, Santos DM et al (2008) High-throughput sequencing of *Medicago truncatula* short RNAs identifies eight new miRNA families. *BMC Genom* 9:593
- Szklarczyk D, Franceschini A, Kuhn M et al (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 39(1):D561–D568
- Tardif G, Kane NA, Adam H et al (2007) Interaction network of proteins associated with abiotic stress response and development in wheat. *Plant Mol Biol* 63:703–718
- Tariq M, Paszkowski J (2004) DNA and histone methylation in plants. *Trends Genet* 20(6):244–251
- Thelen JJ, Peck SC (2007) Quantitative proteomics in plants: choices in abundance. *Plant Cell* 19(11):3339–3346
- Thimm O, Bläsing O, Gibon Y et al (2004) Mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J* 37(6):914–939
- Titanji VP, Ngwa AA, Ngemenya M (2007) Applications of biotechnology techniques to the study of medicinal plants. *Afr J Med Medical Sci* 36:23–29
- 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 AR (2012) Co-expression and co-responses: within and beyond transcription. *Front Plant Sci* 3:248
- Tohge T, Nishiyama Y, Hirai MY et al (2005) Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. *Plant J* 42:218–235
- Tomita M, Hashimoto K, Takahashi K et al (1999) E-CELL: software environment for whole-cell simulation. *Bioinformatics* 15(1):72–84
- Tompa R, McCallum CM, Delrow J et al (2002) Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. *Curr Biol* 12:65–68
- Tran RK, Henikoff JG, Zilberman D et al (2005a) DNA methylation profiling identifies CG methylation clusters in *Arabidopsis* genes. *Curr Biol* 15:154–159
- Tran RK, Zilberman D, de Bustos C et al (2005b) Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in Arabidopsis. *Genome Biol* 6:R90
- Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105–1111
- Trewavas A (2006) A brief history of systems biology “Every object that biology studies is a system of systems”. *Francois Jacob* (1974). *Plant Cell* 18(10):2420–2430
- Tsugita A, Kamo M, Kawakami T et al (1996) Two-dimensional electrophoresis of plant proteins and standardization of gel patterns. *Electrophoresis* 17:855–865
- Turenne N (2011) Role of a web-based software platform for systems biology. *J Comput Sci Syst Biol* 4:035–041
- Umehara M, Hanada A, Yoshida S et al (2008) Inhibition of shoot branching by new terpenoid plant hormones. *Nature* 455(7210):195–200
- van Dam NM, van der Meijden E (2011) A role for metabolomics in plant ecology. In: Hall RD (ed) Biology of Plant Metabolomics. Wiley-Blackwell, Chichester
- Van Leene J, Stals H, Eeckhout D et al (2007) Tandem affinity purification-based technology platform to study the cell cycle interactome in *Arabidopsis thaliana*. *Mol Cell Proteomics* 6:1226–1238
- van wijk KJ (2001) Challenges and prospects of plant proteomics. *Plant Physiol* 126(2):501–508
- VanBogelen RA, Schiller EE, Thomas JD et al (1999) Diagnosis of cellular states of microbial organisms using proteomics. *Electrophoresis* 20:2149–2159
- Vaucheret H (2006) Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Gene Dev* 20(7):759–771
- Vener AV, Harms A, Sussman MR et al (2001) Mass spectrometric resolution of reversible protein phosphorylation in photosynthetic membranes of *Arabidopsis thaliana*. *J Biol Chem* 276:6959–6966
- Vernoux T, Brunoud G, Farcot E, et al (2011) The auxin signalling network translates dynamic input into robust patterning at the shoot apex. *Mol Syst Biol* 7(508)
- Verpoorte R, Choi YH, Mustafa NR et al (2008) Metabolomics: back to basics. *Phytochem Rev* 7(3):525–537
- Vihinen M (2001) Bioinformatics in proteomics. *Biomol Eng* 18(5):241–248
- Voit EO (2000) Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists. Cambridge University Press, Cambridge
- von Bertalanffy, L (1933) Modern theories of development. Oxford University Press, London
- von Bertalanffy, L (1968) General systems theory. In: Braziller G (ed) foundations, development, applications. New York
- von Bertalanffy, L (1973) General systems theory. In: Braziller G (ed) foundations, development, applications (revised edition)
- Waddington CH (1942) Canalization of development and the inheritance of acquired characters. *Nature* 150(3811):563–565
- Wang X, Elling AA, Li X et al (2009a) Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. *Plant Cell* 21:1053–1069
- Wang H, Schauer N, Usadel B et al (2009b) Regulatory features underlying pollination-dependent and-independent tomato fruit

- set revealed by transcript and primary metabolite profiling. *Plant Cell* 21(5):1428–1452
- Wang Z, Gerstein M, Snyder M (2009c) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10(1):57–63
- Wang L, Li P, Brutnell TP (2010) Exploring plant transcriptomes using ultra high-throughput sequencing. *Brief Funct Genomics* 9(2):118–128
- Wang X, Wang H, Wang J et al (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* 43(10):1035–1039
- Ward JL, Forcat S, Beckmann M et al (2010) The metabolic transition during disease following infection of *Arabidopsis thaliana* by *Pseudomonas syringae* pv. *tomato*. *Plant J* 63:443–457
- Washburn MP, Wolters D, Yates JR (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19(3):242–247
- Watanabe M, Mochida K, Kato T et al (2008) Comparative genomics and reverse genetics analysis reveal indispensable functions of the serine acetyltransferase gene family in *Arabidopsis*. *Plant Cell* 20:2484–2496
- Weckwerth W, Morgenthal K (2005) Metabolomics: from pattern recognition to biological interpretation. *Drug Discov Today* 10(22):1551–1558
- Welti R, Shah J, Li W et al (2007) Plant lipidomics: discerning biological function by profiling plant complex lipids using mass spectrometry. *Frontiers Biosci* 12:2494–2506
- Weston DJ, Karve AA, Gunter LE, Jawdy SS, Yang X, Allen SM et al (2011) Comparative physiology and transcriptional networks underlying the heat shock response in *Populus trichocarpa*, *Arabidopsis thaliana* and *Glycine max*. *Plant Cell Environ* 34:1488–1506
- Weston DJ, Hanson PJ, Norby RJ, Tuskan GA, Wulfschleger SD (2012) From systems biology to photosynthesis and whole-plant physiology. *Plant Signal Behav* 7(2):260–262
- Wiener N (1948) *Cybernetics*. Wiley, New York, p 112
- Wilkins MR, Sanchez JC, Gooley AA et al (1996) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng* 13(1):19–50
- 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? *Brief Bioinform* 9(6):466–478
- Winter D, Vinegar B, Nahal H et al (2007) An “Electronic Fluorescent Pictograph” browser for exploring and analyzing large-scale biological data sets. *PLoS One* 2(8):e718
- Wold S, Sjostrom M (1977) SIMCA: a method for analyzing chemical data in terms of similarity and analogy. *Chemom Theory Appl* 52:243–282
- Wolters DA, Washburn MP, Yates JR (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* 73(23):5683–5690
- Wong MM, Cannon CH, Wickneswari R (2011) Identification of lignin genes and regulatory sequences involved in secondary cell wall formation in *Acacia auriculiformis* and *Acacia mangium* via de novo transcriptome sequencing. *BMC Genom* 12:342
- Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26:873–881
- Xenarios I, Salwinski L, Duan XJ et al (2002) DIP, the Database of interacting proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res* 30(1):303–305
- Xie G, Plumb R, Su M et al (2008) Ultra-performance LC/TOF MS analysis of medicinal Panax herbs for metabolomic research. *J Sep Sci* 31(6–7):1015–1026
- Xu X, Pan S, Cheng S et al (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475:189–195
- 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 Physiol* 51(5):795–809
- Yan H, Kikuchi S, Neumann P et al (2010) Genome-wide mapping of cytosine methylation revealed dynamic DNA methylation patterns associated with genes and centromeres in rice. *Plant J* 63:353–365
- Yonekura-Sakakibara K, Tohge T, Matsuda F et al (2008) Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene–metabolite correlations in *Arabidopsis*. *Plant Cell* 20(8):2160–2176
- You-ping D, Jun-mei AI, Pei-gen X (2010) Application of bioinformatics and systems biology in medicinal plant studies. *Chin Herbal Med* 2(3):170–179
- Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. *Curr Opin Biotech* 17(2):155–160
- Yuan JS, Galbraith DW, Dai SY et al (2008) Plant systems biology comes of age. *Trends Plant Sci* 13(4):165–171
- Yun KY, Park MR, 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 Biol* 10(1):16
- Zemach A, McDaniel IE, Silva P et al (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328:916–919
- Zenoni S, Ferrarini A, Giacomelli E et al (2010) Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. *Plant Physiol* 152(4):1787–1795
- Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821–829
- Zhang X, Yazaki J, Sundaresan A et al (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* 126:1189–1201
- Zhang W, Chen J, Yang Y et al (2011) A practical comparison of de novo genome assembly software tools for next-generation sequencing technologies. *PLoS One* 6(3):e17915
- Zhou M, Gu L, Li P et al (2010) Degradome sequencing reveals endogenous small RNA targets in rice (*Oryza sativa* L. ssp. *indica*). *Front Biol* 5(1):67–90
- Zhu Y, Fu J, Zhang J et al (2009) Genome-wide analysis of gene expression profiles during ear development of maize. *Plant Mol Biol* 70(1–2):63–77
- Zilberman D, Henikoff S (2007) Genome-wide analysis of DNA methylation patterns. *Development* 134:3959–3965
- Zimmermann IM, Heim MA, Weisshaar B et al (2004) Comprehensive identification of *Arabidopsis thaliana* MYB transcription factors interacting with R/B-like BHLH proteins. *Plant J* 40(1):22–34
- Zulak KG, Cornish A, Daskalchuk TE et al (2007) Gene transcript and metabolite profiling of elicitor-induced opium poppy cell cultures reveals the coordinate regulation of primary and secondary metabolism. *Planta* 225(5):1085–1106