

Chapter 2

Omics and System Biology Approaches in Plant Stress Research

Abstract The continuous development of analytical and experimental technologies as well as instruments resulted in the development of very specialized experimental approaches that can identify, measure and quantify particular types of cellular molecules. These technologies are known as “Omics Technologies”. Most of the omics technologies are high throughput with very fast data generation rates and humongous outputs. Thus, they are highly dependent on bioinformatics and computational tools. These technologies have made noticeable contributions to the current advancements in our understanding of plant biology in general and plant stress tolerance and response in particular. In this chapter, we will introduce the main omics technologies employed in plant biology and the bioinformatics platforms associated with them.

Keywords Plant stress • Biotic stress • Abiotic stress • Omics • Genomics • Proteomics • Proteogenomics • Transcriptomics • Metabolomics • Databases • Bioinformatics

2.1 Introduction

In the last two decades, molecular biology and systems biology experienced unprecedented advancements either in the accuracy of the analysis or at their overall scales (El-Metwally et al. 2014a). At first sight, molecular biology and systems biology look to be opposites due to the reductionistic nature of molecular biology and the holistic perspective of systems biology respectively. However, in modern research settings, both molecular and systems biology complement each other, providing new perspectives in approaching complex topics such as the study of plant stress or the improvement of plant stress responses (Duque et al. 2013). Figure 2.1 shows the common approaches utilized in plant stress research including omics-based researches.

Genome sequencing represented a landmark in the development of biological sciences and the methodologies of approaching biological problems. It enhanced our grasp of biological systems by examining the base of life (i.e. DNA) and allowed an enriched understanding of gene structures and functions (El-Metwally et al. 2014a). Next-generation sequencing (NGS) of the genome elevates the utility of genome sequencing by providing cheap, fast and easy genome sequencing platforms, though not without some challenges (El-Metwally et al. 2013, 2014a). NGS presents a great

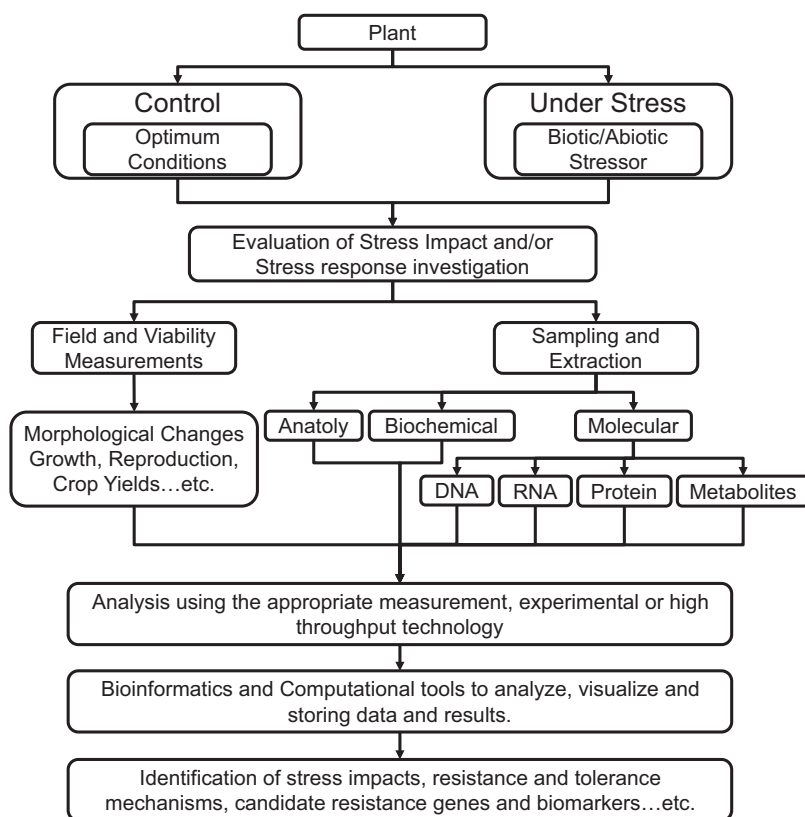


Fig. 2.1 Schematic overview of common approaches in plant stress research

foundation for several other methodologies to be developed as well as several new approaches for studying biological systems in the so-called post-genomic era (Duque et al. 2013).

Genomics, transcriptomics, proteomics, proteogenomics and metabolomics are modern methodologies and approaches that have been recently applied in the study of plant stress mechanism responses. They provide new insights and open new horizons for understanding stresses and responses as well as the improvement of plant responses and resistance to stresses (Duque et al. 2013). Due to the large-scale nature of these approaches, bioinformatics and computational approaches are highly associated with the above for either developing new data analytical methods, better visualization or storage in sustainable online resources (Helmy et al. 2011, 2012a, b, c). Figure 2.2 shows the main omics approaches employed in plant stress research, its primary technologies and expected outcomes.

Since the focus of this book is the integrated omics approaches in plant stress tolerance, we will introduce the applications of each of the omics and bioinformatics approaches in detail.

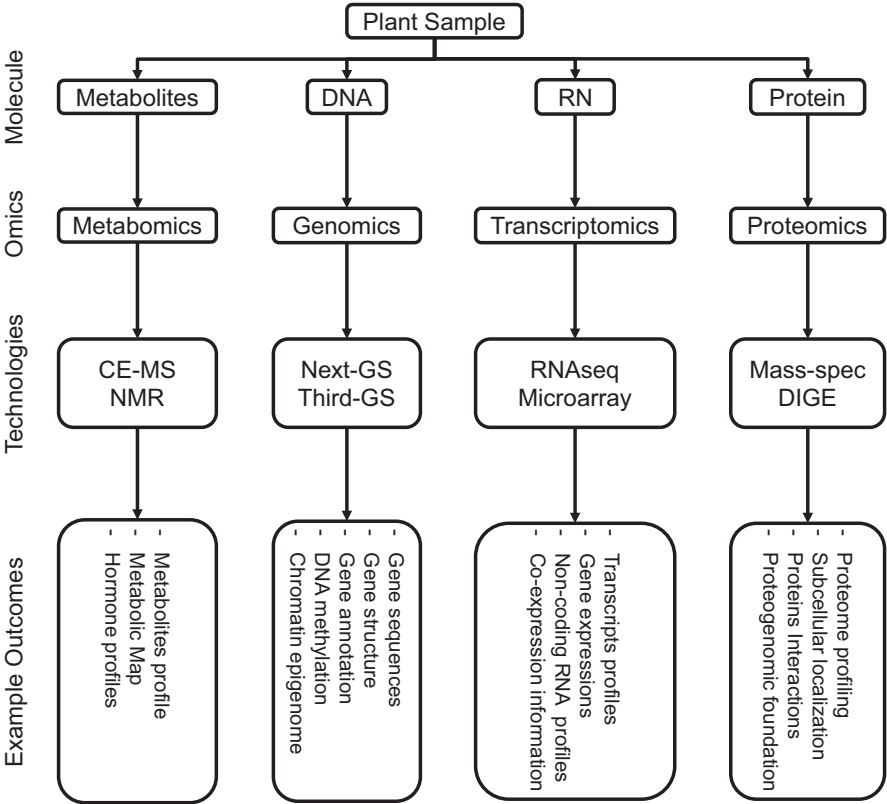


Fig. 2.2 Schematic overview of the main molecular and systems biology approaches and their technologies as well as expected outcomes in plant biology and stress research

2.2 Genomics

Genomics is the study of all the genes in a given genome including the identification of gene sequences, intragenic sequences, gene structures and annotations (Duque et al. 2013). The technology of choice for genomics is genome sequencing that began with the first generation of methods in the 1970s, followed by the next-generation sequencing (NGS) technologies in the middle 1990s as well as the more recent third-generation sequencing technologies (El-Metwally et al. 2014b, c). The process involves DNA extraction, amplification using polymerase chain reaction (PCR) techniques, DNA sequencing and sequence assembly as well as quality assessment (El-Metwally et al. 2013, 2014a, d). Following DNA sequencing and assembly, the gene structural and functional annotation takes place revealing invaluable information about the biology of the organism in question.

It would be challenging to list all the contributions of genomics to the study of the plant stress response and tolerance research. Therefore, we will list the main applications of genomics in this field. Genomics mainly helped in identifying the functional relevance of genes involved in abiotic and biotic stress responses in plants via functional genomic approaches (Cao et al. 2005; Govind et al. 2009; Ramegowda et al. 2013, 2014). Combined with other techniques, genomics helped plant breeders create new breeds that can tolerate several biotic and abiotic stresses and, consequently, have increased crop yields. This includes new breeds for drought and cold tolerance as well as pathogen resistance (Yadav et al. 2010; Yao et al. 2011; Le et al. 2012; Chen et al. 2012; Shankar et al. 2013; Wang et al. 2013; Agarwal et al. 2014). Furthermore, the huge online databases, repositories and archives of plant genomic information serve as a foundation for transcriptomics, genome engineering and proteogenomics (Matthews et al. 2009; Batley and Edwards 2009; Mochida and Shinozaki 2010; Jung and Main 2014).

2.3 Transcriptomics

The transcriptome is the RNA expression profile of an organism. Unlike the genome which remains constant despite age, organ or growth conditions, the transcriptome is highly dynamic (El-Metwally et al. 2014a). Therefore, the term transcriptomics refers to the capturing of the RNA expression profile in spatial and temporal bases in certain plant organs, tissues and cells within particular context (Duque et al. 2013; El-Metwally et al. 2014a). This particular context can be growth or environmental conditions, treatment with certain nutrients or biotic stress conditions. The RNA/gene expression profiling is mostly accomplished using microarray, RNA sequencing (RNAseq) through next-generation sequencing (NGS), serial analysis of gene expression (SAGE) and digital gene expression profiling (Kawahara et al. 2012; Duque et al. 2013; De Cremer et al. 2013).

Transcriptomics mainly helps in finding genes that are associated with alterations in the plant phenotype under different conditions. For instance, transcriptomics can be used in finding candidate genes that contribute to stress tolerance through the comparison of transcriptomes of the same plant under optimal and stress conditions (Le et al. 2012; Zhang et al. 2014b). Transcriptomics applications in plant stress response and tolerance can also include searching for abiotic stress candidate genes, predicting tentative gene functions and providing a better understanding of the plant-pathogen relationship (Kawahara et al. 2012; Jogaiah et al. 2013; De Cremer et al. 2013; Agarwal et al. 2014). The recent increase in the availability of online resources, databases and archives of transcriptome data allows for performing novel genome-wide analysis of plant stress responses and tolerances (Mochida and Shinozaki 2011; Duque et al. 2013; Jogaiah et al. 2013).

2.4 Proteomics

The proteome is the total expressed protein under certain conditions in a given organism, organ, cell, tissue or microorganism population (Tyers and Mann 2003). Similar to the transcriptome, the proteome is highly dynamic and changes based on temporal and environmental factors. Proteomics are the techniques used in capturing and measuring (or “profiling”) the expressed proteins in a specific context (Tyers and Mann 2003). There are several types of proteomes that can be measured, and each of them can reveal particular information about the expressed proteins. The most common proteomes to be measured in plant stress tolerance and other plant related studies are the whole proteome and the phosphoproteome (Helmy et al. 2011, 2012b, c). The whole proteome is the quantitative and/or qualitative profiling of all the expressed proteins in a given sample, while the phosphoproteome is the quantitative and/or qualitative profiling of the phosphorylated proteins expressed in a given sample (Nakagami et al. 2012).

The technology of choice for proteomics is mass spectrometry (MS) including several approaches such as liquid chromatography–mass spectrometry (LC-MS/MS), Ion Trap–mass spectrometry (IT-MS) and matrix-assisted laser desorption/ionization–mass spectrometry (MALDI-MS) (Helmy et al. 2011, 2012a; Komatsu et al. 2014; Shao et al. 2014). These technologies are basically used in measuring the mass and charge of small protein fragments (or “peptides”) that result from protein enzymatic digestion with special enzymes called proteases, such as trypsin (Helmy et al. 2011; Nakagami et al. 2012). The output of a standard MS-based proteomic analysis is a set of peptide fingerprints called MS spectra. MS spectra require another layer of interpretation to reveal the peptide sequences associated with each of them, the protein of each peptide and the modification occurring in each protein after being translated (Tyers and Mann 2003; Helmy et al. 2012c; Nakagami et al. 2012). Furthermore, several proteomics labs use protein electrophoresis technologies such as two-dimensional electrophoresis and Difference Gel Electrophoresis (DIGE) in plant proteomics (Cramer and Westermeier 2012; Duque et al. 2013; Komatsu et al. 2014; Arentz et al. 2014).

Proteomics is a very informative approach that is used to reveal invaluable information when studying plant stress response and tolerance, either in a genome-wide or sample scale (Nakagami et al. 2012). It can be used to profile all the expressed proteins in multiple conditions (e.g. optimal, stress and prolonged stress conditions) and cross compare these different sets to pinpoint the proteins involved in stress tolerance (Evers et al. 2012; Yan et al. 2014). Quantitative proteomics reveals the proteins that are differentially expressed under the condition changes, which points towards its contribution in the stress response process as well (Liu et al. 2015). Phosphoproteomics is more associated with the identification of proteins activated and functioning under certain condition. Therefore, it is very useful in identifying pathways involved in a particular function or process through ascertaining the set of proteins that are exclusively activated under the condition that triggered this function. Through phosphoproteomics, proteins and signaling pathways involved in

response to particular stress can be identified (Sugiyama et al. 2008; Lassowskat et al. 2014; Zhang et al. 2014a). Both whole proteomics and phosphoproteomics can be combined in one comprehensive study to provide a better understanding of the stress in question (Margaria et al. 2013; Yang et al. 2013; Hopff et al. 2013).

2.5 Proteogenomics

Proteogenomics is a comprehensive approach that combines large-scale proteomic data with genomic and/or transcriptomics data in genome annotation refinement and the elucidation of novel regulatory mechanisms (Helmy et al. 2012a; Ansong et al. 2008). The proteomics data generated by means of MS-based proteomics is well known for its high throughput and accuracy. Therefore, it provides a rich source of translation-level information about the expressed proteins and can be used as a source of affordable large-scale experimental evidence for several predictions (Helmy et al. 2012b, c; Ansong et al. 2008; de Groot et al. 2009; Armengaud 2010). In a standard proteogenomics study, the naturally expressed proteins are identified using MS-based proteogenomics followed by mapping them back to the genomic or transcriptomic data (Helmy et al. 2012a; Ansong et al. 2008).

In the last decade, proteogenomics has helped in elevating our understanding of the biology of plants in general as well as plant stress research in particular. For instance, a large-scale proteogenomics study of *Arabidopsis thaliana* identified 57 new genes and corrected the annotations of hundreds of its genes using intensive sampling from the *Arabidopsis* organs under several conditions and in different life stages (Baerenfaller et al. 2008). Another study reported corrections and new identifications in about 13% of the annotated genes in *Arabidopsis* (Castellana et al. 2008). Furthermore, several major cultivated crops such as *Oryza sativa* and *Zea mays* benefited from proteogenomics studies (Helmy et al. 2011; Castellana et al. 2014).

In plant stress research, the use of proteogenomics provided a deeper understanding of the major abiotic stress factors including bacteria such as *Bradyrhizobium diazoefficiens* (Chapman and Bellgard 2014), fungi such as *Aspergillus niger* (the black mold fungus) and *Stagonospora nodorum* (Wright et al. 2009; Bringans et al. 2009), insects such as *Drosophila melanogaster* (Tress et al. 2008; Loevenich et al. 2009) and nematodes such as *Pristionchus pacificus* (Borchert et al. 2010). It also presented new insight into the investigation of the host-pathogen relationship such as understanding the relationship between the plant and the phyllosphere bacteria (Delmotte et al. 2009), identifying novel effectors in fungal diseases (Cooke et al. 2014), providing a better understanding of the host-parasite relationship (Lasonder et al. 2002; Bindschedler et al. 2009) as well as shedding light on the mechanisms of environmental adaptation and ecological diversity (de Groot et al. 2009; Denef et al. 2010).

2.6 Metabolomics

The metabolome is the complete set of metabolites that can be identified in a given organism, organ, tissue or biological fluid. Thus, metabolomics refers to techniques and methods used to study the metabolome (Duque et al. 2013). Due to differences in the chemical and physical properties of the metabolites, a combination of several analytical and separation techniques is required to obtain the metabolic profile of a plant or given sample (Jogaiah et al. 2013). Although Capillary Electrophoresis-liquid-chromatography Mass Spectrometry (CE-MS) is considered the most advanced metabolomics technology to date (Soga et al. 2002), several other analytical instruments and separation technologies are employed in metabolomics such as Gas Chromatography (GC), Mass Spectrometry (MS) and Nuclear Magnetic Resonance (NMR) (Saito and Matsuda 2010; Duque et al. 2013; Jogaiah et al. 2013).

Plants are able to synthesize a wide spectrum of chemical and biological compounds that are crucial for regulating the response to different types of biotic and abiotic stress. Therefore, identifying the metabolites produced by the plant under each stress condition provides a better understanding of the regulation processes as well as joins the genotype with the phenotype and investigates the changes in phenotype that take place under stress conditions (Badjakov et al. 2012). Metabolomics is usually used in combination with other omics analysis (e.g. transcriptomics or proteomics) to investigate the correlation between metabolite levels and the expression level of genes/proteins (Srivastava et al. 2013). A strong correlation between stress metabolites and a certain gene/protein indicates the role of this gene/protein in the response process (Urano et al. 2010; Duque et al. 2013; Jogaiah et al. 2013). Metabolomics is used to provide a better understanding of the stress response and tolerance process in model plants such as *Arabidopsis* (Cook et al. 2004) as well as production crops such as the common bean (*Phaseolus vulgaris*) (Broughton et al. 2003), poplars (*Populus x canescens*) (Behnke et al. 2010), cereals (Sicher and Barnaby 2012) and other food crops (Hernández et al. 2007; Duque et al. 2013).

2.7 Bioinformatics

The brief introduction of each of the omics approaches that we provided above shows that all of them share similar high throughput and large-scale properties. Furthermore, these approaches can be genome-wide as well as through the combination of several genomes or several species, which results in producing huge amounts of data that requires proper handling, analysis, visualization and storage (El-Metwally et al. 2014e). Therefore, all omics research is tightly bound with strong bioinformatics and computational tools that perform the various analysis tasks as well as allow integration between several types of data “multi-omics” and enable knowledge exchange between different organisms (Shinozaki and Sakakibara 2009; Mochida and Shinozaki 2011; El-Metwally et al. 2014a).

2.7.1 Data Handling and Analysis

The primary reason for including informatics analysis and computational tools as well as associated methods and algorithms in biology is to allow biological data analysis in an accurate, fast, human-error free and easily reproducible manner (Orozco et al. 2013). Hence, several bioinformatics tasks became indispensable in biological research in general and plant stress multi-omics research in particular. This includes the standard tasks involved in genome sequence assembly (El-Metwally et al. 2013, 2014f, g), sequence alignment (Altschul et al. 1990; Tatusova and Madden 1999), gene prediction (Stanke and Morgenstern 2005), peptides and proteins sequence identification (MS-spectra interpretation tools) (Eng et al. 1994; Perkins et al. 1999), gene and protein function prediction (Falda et al. 2012; Yachdav et al. 2014), DNA-protein and protein-protein interaction prediction (McDowall et al. 2009; Franceschini et al. 2013), interaction and regulatory networks analysis (Chaouiya 2012) and several other essential tasks (Polpitiya et al. 2008; Henry et al. 2014).

2.7.2 Data and Results Visualization

The large amount of data generated by modern analytical and experimental instruments such as genome sequencers and mass spectrometers as well as the information resulting from the analysis and processing of this data requires special types of visualization. Thus, several tools were developed to help visualize the biological data and results in a manner that would maximize the utility of the data. These genomic data visualizing tools include Gbrowse, UCSC Genome Browser and Integrated Genome Viewer (IGV) (Stein et al. 2002; Karolchik et al. 2003; Pang et al. 2014), proteomics data visualization tools such as PRIDE Inspector and ConPath (Kim et al. 2008; Wang et al. 2012), proteogenomics data or multi-omics data visualization tools such as PGFeval, 3Omics, Peppy (Helmy et al. 2011, 2012b; Kuo et al. 2013; Risk et al. 2013), metabolomics visualization tools such as MultiExperiment Viewer (MeV) (Saeed et al. 2003) and network visualization tools such as Cytoscape and its associated web versions Cytoscape.js and Cytoscape web (Shannon et al. 2003; Lopes et al. 2010; Ono et al. 2014).

2.7.3 Data and Results Storage and Maintenance

High throughput data is very fruitful in that we can gain more knowledge from it by applying different types of analyses or by combining several datasets into one large-scale comparative analysis. However, this requires the data and results to be sustainably available and accessible to the scientific community (Smalter Hall et al. 2013; Helmy et al. 2016). Therefore, several types of databases are available online for depositing and storing the biological data and results. The databases range from

those that store plant information as classification, growth, production, geographical distribution (Wilkinson et al. 2012), plant genomic information (Yu et al. 2013; Zhao et al. 2014), plant transcriptomic information (Priya and Jain 2013), plant proteomic information (Komatsu and Tanaka 2005; Cheng et al. 2014), plant proteogenomic information (Helmy et al. 2011, 2012b) and plant metabolomic information (Deborde and Jacob 2014). Furthermore, some databases are specialized in storing and maintaining plant stress resistance and tolerance information such as STIFDB2, the Arabidopsis stress responsive gene database, QlicRice and the fungal stress response database (FSRD) (Smita et al. 2011; Borkotoky et al. 2013; Karányi et al. 2013; Naika et al. 2013).

In general, several tools provide more than one of the above-mentioned features such as data analysis and visualization (Eng et al. 1994; Perkins et al. 1999; Cargile et al. 2004; Helmy et al. 2011) or data visualization and storage (Helmy et al. 2012b). Furthermore, several tools provide these services for multi-omics data and results (Kuo et al. 2013).

References

- Agarwal P, Parida SK, Mahto A, et al (2014) Expanding frontiers in plant transcriptomics in aid of functional genomics and molecular breeding. *Biotechnol J* 9:1480–1492. doi:[10.1002/biot.201400063](https://doi.org/10.1002/biot.201400063)
- Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
- Ansong C, Purvine SO, Adkins JN, et al (2008) Proteogenomics: needs and roles to be filled by proteomics in genome annotation. *Br Funct Genomic Proteomic* 7:50–62.
- Arentz G, Weiland F, Oehler MK, Hoffmann P (2014) State of the art of 2D DIGE. *Proteomics Clin Appl* 9:277–288. doi:[10.1002/prca.201400119](https://doi.org/10.1002/prca.201400119)
- Armengaud J (2010) Proteogenomics and systems biology: quest for the ultimate missing parts. *Expert Rev Proteomics* 7:65–77. doi:[10.1586/epr.09.104](https://doi.org/10.1586/epr.09.104)
- Baerenfaller K, Grossmann J, Grobei MA, et al (2008) Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. *Science* 320:938–941
- Batley J, Edwards D (2009) Genome sequence data: management, storage, and visualization. *Biotechniques* 46:333–334., 336. doi:[10.2144/000113134](https://doi.org/10.2144/000113134)
- Behnke K, Kaiser A, Zimmer I, et al (2010) RNAi-mediated suppression of isoprene emission in poplar transiently impacts phenolic metabolism under high temperature and high light intensities: a transcriptomic and metabolomic analysis. *Plant Mol Biol* 74:61–75. doi:[10.1007/s11103-010-9654-z](https://doi.org/10.1007/s11103-010-9654-z)
- Bindschedler L V, Burgis TA, Mills DJS, et al (2009) In planta proteomics and proteogenomics of the biotrophic barley fungal pathogen *Blumeria graminis* f. sp. *hordei*. *Mol Cell Proteomics MCP* 8:2368–2381. doi:[10.1074/mcp.M900188-MCP00](https://doi.org/10.1074/mcp.M900188-MCP00)
- Borchert N, Dieterich C, Krug K, et al (2010) Proteogenomics of *Pristionchus pacificus* reveals distinct proteome structure of nematode models. *Genome Res* 20:837–846. doi:[10.1101/gr.103119.109](https://doi.org/10.1101/gr.103119.109)
- Borkotoky S, Saravanan V, Jaiswal A, et al (2013) The Arabidopsis stress responsive gene database. *Int J Plant Genomics* 2013:949564. doi:[10.1155/2013/949564](https://doi.org/10.1155/2013/949564)
- Bringans S, Hane JK, Casey T, et al (2009) Deep proteogenomics; high throughput gene validation by multidimensional liquid chromatography and mass spectrometry of proteins

- from the fungal wheat pathogen *Stagonospora nodorum*. BMC Bioinformatics 10:301. doi:[10.1186/1471-2105-10-301](https://doi.org/10.1186/1471-2105-10-301)
- Broughton WJ, Hernández G, Blair M, et al (2003) Beans (*Phaseolus* spp.)—model food legumes. Plant and Soil 252:55–128. doi:[10.1023/A:1024146710611](https://doi.org/10.1023/A:1024146710611)
- Cao X, Zhou P, Zhang X, et al (2005) Identification of an RNA silencing suppressor from a plant double-strandedRNA virus. J Virol 79:13018–13027. doi:[10.1128/JVI.79.20.13018-13027.2005](https://doi.org/10.1128/JVI.79.20.13018-13027.2005)
- Cargile BJ, Bundy JL, Freeman TW, Stephenson Jr. JL (2004) Gel based isoelectric focusing of peptides and the utility of isoelectric point in protein identification. J Proteome Res 3:112–119.
- Castellana NE, Payne SH, Shen Z, et al (2008) Discovery and revision of Arabidopsis genes by proteogenomics. Proc Natl Acad Sci U S A 105:21034–21038.
- Castellana NE, Shen Z, He Y, et al (2014) An automated proteogenomic method uses mass spectrometry to reveal novel genes in *Zea mays*. Mol Cell Proteomics MCP 13:157–167. doi:[10.1074/mcp.M113.031260](https://doi.org/10.1074/mcp.M113.031260)
- Chapman B, Bellgard M (2014) High-throughput parallel proteogenomics: a bacterial case study. Proteomics 14:2780–2789. doi:[10.1002/pmic.201400185](https://doi.org/10.1002/pmic.201400185)
- Chen S, Jiang J, Li H, Liu G (2012) The salt-responsive transcriptome of *Populus simonii* × *Populus nigra* via DGE. Gene 504:203–212. doi:[10.1016/j.gene.2012.05.023](https://doi.org/10.1016/j.gene.2012.05.023)
- Cheng H, Deng W, Wang Y, et al (2014) dbPPT: a comprehensive database of protein phosphorylation in plants. Database 2014:bau121. doi:[10.1093/database/bau121](https://doi.org/10.1093/database/bau121)
- Claudine Chaouiya (2012). Logical Modelling of Gene Regulatory Networks with GINsim in Methods in molecular biology Edited by N.J. Clifton, Humana Press, Print ISBN 1940-6029
- Cook D, Fowler S, Fiehn O, Thomashow MF (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of Arabidopsis. Proc Natl Acad Sci U S A 101:15243–15248. doi:[10.1073/pnas.0406069101](https://doi.org/10.1073/pnas.0406069101)
- Cooke IR, Jones D, Bowen JK, et al (2014) Proteogenomic analysis of the *Venturia pirina* (Pear Scab Fungus) secretome reveals potential effectors. J Proteome Res 13:3635–3644. doi:[10.1021/pr500176c](https://doi.org/10.1021/pr500176c)
- Cramer, Rainer, Westermeier R (2012) Difference Gel Electrophoresis (DIGE) - Methods and Protocols. Humana Press, Print ISBN: 9781617795732
- De Cremer K, Mathys J, Vos C, et al (2013) RNAseq-based transcriptome analysis of *Lactuca sativa* infected by the fungal necrotroph *Botrytis cinerea*. Plant Cell Environ 36:1992–2007. doi:[10.1111/pce.12106](https://doi.org/10.1111/pce.12106)
- de Groot A, Dulerio R, Ortet P, et al (2009) Alliance of proteomics and genomics to unravel the specificities of Sahara bacterium *Deinococcus deserti*. PLoS Genet 5:e1000434. doi:[10.1371/journal.pgen.1000434](https://doi.org/10.1371/journal.pgen.1000434)
- Deborde C, Jacob D (2014) MeRy-B, a metabolomic database and knowledge base for exploring plant primary metabolism. Methods Mol Biol 1083:3–16. doi:[10.1007/978-1-62703-661-0_1](https://doi.org/10.1007/978-1-62703-661-0_1)
- Delmotte N, Knief C, Chaffron S, et al (2009) Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. Proc Natl Acad Sci U S A 106:16428–16433. doi:[10.1073/pnas.0905240106](https://doi.org/10.1073/pnas.0905240106)
- Denef VJ, Kalnejais LH, Mueller RS, et al (2010) Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. Proc Natl Acad Sci U S A 107:2383–2390. doi:[10.1073/pnas.0907041107](https://doi.org/10.1073/pnas.0907041107)
- Duque AS, de Almeida AM, da Silva AB, da Silva JM, et al (2013) Abiotic stress—plant responses and applications in agriculture. doi:[10.5772/45842](https://doi.org/10.5772/45842)
- El-Metwally S, Hamza T, Zakaria M, Helmy M (2013) Next-generation sequence assembly: four stages of data processing and computational challenges. PLoS Comput Biol 9:e1003345. doi:[10.1371/journal.pcbi.1003345](https://doi.org/10.1371/journal.pcbi.1003345)
- El-Metwally S, Ouda OM, Helmy M (2014a) Next generation sequencing technologies and challenges in sequence assembly.
- El-Metwally S, Ouda OM, Helmy M (2014b) First- and next-generations sequencing methods. Next Gener Seq Technol Challenges Seq Assem. doi:[10.1007/978-1-4939-0715-1_3](https://doi.org/10.1007/978-1-4939-0715-1_3)
- El-Metwally S, Ouda OM, Helmy M (2014c) New horizons in next-generation sequencing. Next Gener Seq Technol Challenges Seq Assem. doi:[10.1007/978-1-4939-0715-1_6](https://doi.org/10.1007/978-1-4939-0715-1_6)

- El-Metwally S, Ouda OM, Helmy M (2014d) Assessment of next-generation sequence assembly. *Next Gener Seq Technol Challenges Seq Assem*. doi:10.1007/978-1-4939-0715-1_10
- El-Metwally S, Ouda OM, Helmy M (2014e) Novel next-generation sequencing applications. *Next Gener Seq Technol Challenges Seq Assem*. doi:10.1007/978-1-4939-0715-1_7
- El-Metwally S, Ouda OM, Helmy M (2014f) Next-generation sequence assembly overview. *Next Gener Seq Technol Challenges Seq Assem*. doi:10.1007/978-1-4939-0715-1_8
- El-Metwally S, Ouda OM, Helmy M (2014g) Next-generation sequence assemblers. *Next Gener Seq Technol Challenges Seq Assem*. doi:10.1007/978-1-4939-0715-1_11
- Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5:976–989. doi:10.1016/1044-0305(94)80016-2
- Evers D, Legay S, Lamoureux D, et al (2012) Towards a synthetic view of potato cold and salt stress response by transcriptomic and proteomic analyses. *Plant Mol Biol* 78:503–514. doi:10.1007/s11103-012-9879-0
- Falda M, Toppo S, Pescarolo A, et al (2012) Argot2: a large scale function prediction tool relying on semantic similarity of weighted Gene Ontology terms. *BMC Bioinformatics* 13(Suppl 4): S14. doi:10.1186/1471-2105-13-S4-S14
- Franceschini A, Szklarczyk D, Frankild S, et al (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 41:D808–15. doi:10.1093/nar/gks1094
- Govind G, Harshavardhan VT, Patricia JK, et al (2009) Identification and functional validation of a unique set of drought induced genes preferentially expressed in response to gradual water stress in peanut. *Mol Genet Genomics* 281:607. doi:10.1007/s00438-009-0441-y
- Helmy M, Tomita M, Ishihama Y (2011) OryzaPG-DB: rice proteome database based on shotgun proteogenomics. *BMC Plant Biol* 11:63. doi:10.1186/1471-2229-11-63
- Helmy M, Sugiyama N, Tomita M, Ishihama Y (2012a) Mass spectrum sequential subtraction speeds up searching large peptide MS/MS spectra datasets against large nucleotide databases for proteogenomics. *Cell Mech* 17:633–644. doi:10.1111/j.1365-2443.2012.01615.x
- Helmy M, Sugiyama N, Tomita M, Ishihama Y (2012b) The rice proteogenomics database oryza PG-DB: development, expansion, and new features. *Front Plant Sci* 3:65. doi:10.3389/fpls.2012.00065
- Helmy M, Tomita M, Ishihama Y (2012c) Peptide identification by searching large-scale tandem mass spectra against large databases: bioinformatics methods in proteogenomics. *Gene Genome Genomics* 6:76–85.
- Helmy M, Crits-Christoph A, Bader GD, et al (2016) Ten simple rules for developing public biological databases. *PLoS Comput Biol* 12:e1005128. doi:10.1371/journal.pcbi.1005128
- Henry VJ, Bandrowski AE, Pepin A-S, et al (2014) OMICtools: an informative directory for multi-omic data analysis. *Database (Oxford)* 2014:bau069. doi:10.1093/database/bau069
- Hernández G, Ramírez M, Valdés-López O, et al (2007) Phosphorus stress in common bean: root transcript and metabolic responses. *Plant Physiol* 144:752–767. doi:10.1104/pp.107.096958
- Hopff D, Wienkoop S, Luthje S (2013) The plasma membrane proteome of maize roots grown under low and high iron conditions. *J Proteomics* 91:605–618. doi:10.1016/j.jprot.2013.01.006
- Ilian Badjakov, Violeta Kondakova and Atanas Atanassov (2012). *Current View on Fruit Quality in Relation to Human Health in Sustainable Agriculture and New Biotechnologies*, Edited by Nouredine Benkeblia, CRC Press, Boca Raton, Pages 303–319, Print ISBN: 978-1-4398-2504-4, eBook ISBN: 978-1-4398-2505-1. doi: 10.1201/b10977-14
- Jogaiah S, Govind SR, Tran L-SP (2013) Systems biology-based approaches toward understanding drought tolerance in food crops. *Crit Rev Biotechnol* 33:23–39. doi:10.3109/07388551.2012.659174
- Jung S, Main D (2014) Genomics and bioinformatics resources for translational science in Rosaceae. *Plant Biotechnol Rep* 8:49–64. doi:10.1007/s11816-013-0282-3
- Karányi Z, Holb I, Hornok L, et al (2013) FSRD: fungal stress response database. *Database (Oxford)* 2013:bat037. doi:10.1093/database/bat037

- Karolchik D, Baertsch R, Diekhans M, et al (2003) The UCSC genome browser database. *Nucleic Acids Res* 31:51–54.
- Kawahara Y, Oono Y, Kanamori H, et al (2012) Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. *PLoS One* 7:e49423. doi:[10.1371/journal.pone.0049423](https://doi.org/10.1371/journal.pone.0049423)
- Kim P-G, Cho H-G, Park K (2008) A scaffold analysis tool using mate-pair information in genome sequencing. *J Biomed Biotechnol* 2008:675741. doi:[10.1155/2008/675741](https://doi.org/10.1155/2008/675741)
- Komatsu S, Tanaka N (2005) Rice proteome analysis: a step toward functional analysis of the rice genome. *Proteomics* 5:938–949.
- Komatsu S, Kamal AHM, Hossain Z (2014) Wheat proteomics: proteome modulation and abiotic stress acclimation. *Front Plant Sci* 5:684. doi:[10.3389/fpls.2014.00684](https://doi.org/10.3389/fpls.2014.00684)
- Kuo T-C, Tian T-F, Tseng YJ (2013) 3Omics: a web-based systems biology tool for analysis, integration and visualization of human transcriptomic, proteomic and metabolomic data. *BMC Syst Biol* 7:64. doi:[10.1186/1752-0509-7-64](https://doi.org/10.1186/1752-0509-7-64)
- Lasonder E, Ishihama Y, Andersen JS, et al (2002) Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* 419:537–542.
- Lassowskat I, Böttcher C, Eschen-Lippold L, et al (2014) Sustained mitogen-activated protein kinase activation reprograms defense metabolism and phosphoprotein profile in *Arabidopsis thaliana*. *Front Plant Sci* 5:554. doi:[10.3389/fpls.2014.00554](https://doi.org/10.3389/fpls.2014.00554)
- Le DT, Nishiyama R, Watanabe Y, et al (2012) Differential gene expression in soybean leaf tissues at late developmental stages under drought stress revealed by genome-wide transcriptome analysis. *PLoS One* 7:e49522. doi:[10.1371/journal.pone.0049522](https://doi.org/10.1371/journal.pone.0049522)
- Liu B, Zhang N, Zhao S, et al (2015) Proteomic changes during tuber dormancy release process revealed by iTRAQ quantitative proteomics in potato. *Plant Physiol Biochem* 86:181–190. doi:[10.1016/j.plaphy.2014.12.003](https://doi.org/10.1016/j.plaphy.2014.12.003)
- Loeuenich SN, Brunner E, King NL, et al (2009) The *Drosophila melanogaster* PeptideAtlas facilitates the use of peptide data for improved fly proteomics and genome annotation. *BMC Bioinformatics* 10:59. doi:[10.1186/1471-2105-10-59](https://doi.org/10.1186/1471-2105-10-59)
- Lopes CT, Franz M, Kazi F, et al (2010) Cytoscape web: an interactive web-based network browser. *Bioinformatics* 26:2347–2348. doi:[10.1093/bioinformatics/btq430](https://doi.org/10.1093/bioinformatics/btq430)
- Margaria P, Abbà S, Palmano S (2013) Novel aspects of grapevine response to phytoplasma infection investigated by a proteomic and phospho-proteomic approach with data integration into functional networks. *BMC Genomics* 14:38. doi:[10.1186/1471-2164-14-38](https://doi.org/10.1186/1471-2164-14-38)
- Matthews DE, Lazo GR, Anderson OD (2009) Plant and crop databases. *Methods Mol Biol* 513:243–262. doi:[10.1007/978-1-59745-427-8_13](https://doi.org/10.1007/978-1-59745-427-8_13)
- McDowall MD, Scott MS, Barton GJ (2009) PIPs: human protein-protein interaction prediction database. *Nucleic Acids Res* 37:D651–6. doi:[10.1093/nar/gkn870](https://doi.org/10.1093/nar/gkn870)
- Mochida K, Shinozaki K (2010) Genomics and bioinformatics resources for crop improvement. *Plant Cell Physiol* 51:497–523. doi:[10.1093/pcp/pcq027](https://doi.org/10.1093/pcp/pcq027)
- Mochida K, Shinozaki K (2011) Advances in omics and bioinformatics tools for systems analyses of plant functions. *Plant Cell Physiol* 52:2017–2038. doi:[10.1093/pcp/pcr153](https://doi.org/10.1093/pcp/pcr153)
- Naika M, Shameer K, Mathew OK, et al (2013) STIFDB2: an updated version of plant stress-responsive transcription factor database with additional stress signals, stress-responsive transcription factor binding sites and stress-responsive genes in *Arabidopsis* and rice. *Plant Cell Physiol* 54:e8. doi:[10.1093/pcp/pcs185](https://doi.org/10.1093/pcp/pcs185)
- Nakagami H, Sugiyama N, Ishihama Y, Shirasu K (2012) Shotguns in the front line: phosphoproteomics in plants. *Plant Cell Physiol* 53:118–124. doi:[10.1093/pcp/pcr148](https://doi.org/10.1093/pcp/pcr148)
- Ono K, Demchak B, Ideker T (2014) Cytoscape tools for the web age: D3.js and cytoscape.js exporters. *F1000Research* 3:143. doi:[10.12688/f1000research.4510.2](https://doi.org/10.12688/f1000research.4510.2)
- Orozco J, Morera J, Jiménez S, Boza R (2013) A review of bioinformatics training applied to research in molecular medicine, agriculture and biodiversity in Costa Rica and Central America. *Brief Bioinform* 14:661–670. doi:[10.1093/bib/bbt033](https://doi.org/10.1093/bib/bbt033)

- Pang CNI, Tay AP, Aya C, et al (2014) Tools to covisualize and coanalyze proteomic data with genomes and transcriptomes: validation of genes and alternative mRNA splicing. *J Proteome Res* 13:84–98. doi:[10.1021/pr400820p](https://doi.org/10.1021/pr400820p)
- Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20:3551–3567.
- Polpitiya AD, Qian W-J, Jaitly N, et al (2008) DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* 24:1556–1558. doi:[10.1093/bioinformatics/btn217](https://doi.org/10.1093/bioinformatics/btn217)
- Priya P, Jain M (2013) RiceSRTFDB: a database of rice transcription factors containing comprehensive expression, cis-regulatory element and mutant information to facilitate gene function analysis. *Database (Oxford)* 2013:bat027. doi:[10.1093/database/bat027](https://doi.org/10.1093/database/bat027)
- Ramegowda V, Senthil-kumar M, Udayakumar M, Mysore KS (2013) A high-throughput virus-induced gene silencing protocol identifies genes involved in multi-stress tolerance. *BMC Plant Biol* 13:193. doi:[10.1186/1471-2229-13-193](https://doi.org/10.1186/1471-2229-13-193)
- Ramegowda V, Mysore KS, Senthil-Kumar M (2014) Virus-induced gene silencing is a versatile tool for unraveling the functional relevance of multiple abiotic-stress-responsive genes in crop plants. *Front Plant Sci* 5:323. doi:[10.3389/fpls.2014.00323](https://doi.org/10.3389/fpls.2014.00323)
- Risk BA, Spitzer WJ, Giddings MC (2013) Peppy: proteogenomic search software. *J Proteome Res* 12:3019–3025. doi:[10.1021/pr400208w](https://doi.org/10.1021/pr400208w)
- Saeed AI, Sharov V, White J, et al (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34:374–378.
- Saito K, Matsuda F (2010) Metabolomics for functional genomics, systems biology, and biotechnology. *Annu Rev Plant Biol* 61:463–489. doi:[10.1146/annurev.arplant.043008.092035](https://doi.org/10.1146/annurev.arplant.043008.092035)
- Shankar A, Singh A, Kanwar P, et al (2013) Gene expression analysis of rice seedling under potassium deprivation reveals major changes in metabolism and signaling components. *PLoS One* 8:e70321. doi:[10.1371/journal.pone.0070321](https://doi.org/10.1371/journal.pone.0070321)
- Shannon P, Markiel A, Ozier O, et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504. doi:[10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303)
- Shao S, Guo T, Aebersold R (2014) Mass spectrometry-based proteomic quest for diabetes biomarkers. *Biochim Biophys Acta* doi:[10.1016/j.bbapap.2014.12.012](https://doi.org/10.1016/j.bbapap.2014.12.012)
- Shinozaki K, Sakakibara H (2009) Omics and bioinformatics: an essential toolbox for systems analyses of plant functions beyond 2010. *Plant Cell Physiol* 50:1177–1180. doi:[10.1093/pcp/pcp085](https://doi.org/10.1093/pcp/pcp085)
- Sicher RC, Barnaby JY (2012) Impact of carbon dioxide enrichment on the responses of maize leaf transcripts and metabolites to water stress. *Physiol Plant* 144:238–253. doi:[10.1111/j.1399-3054.2011.01555.x](https://doi.org/10.1111/j.1399-3054.2011.01555.x)
- Smalter Hall A, Shan Y, Lushington G, Visvanathan M (2013) An overview of computational life science databases & exchange formats of relevance to chemical biology research. *Comb Chem High Throughput Screen* 16:189–198
- Smita S, Lenka SK, Katiyar A, et al (2011) QlicRice: a web interface for abiotic stress responsive QTL and loci interaction channels in rice. *Database (Oxford)* 2011:bar037. doi:[10.1093/database/bar037](https://doi.org/10.1093/database/bar037)
- Soga T, Ueno Y, Naraoka H, et al (2002) Simultaneous determination of anionic intermediates for *Bacillus subtilis* metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. *Anal Chem* 74:2233–2239.
- Srivastava V, Obudulu O, Bygdell J, et al (2013) OnPLS integration of transcriptomic, proteomic and metabolomic data shows multi-level oxidative stress responses in the cambium of transgenic hiPL- superoxide dismutase *Populus* plants. *BMC Genomics* 14:893. doi:[10.1186/1471-2164-14-893](https://doi.org/10.1186/1471-2164-14-893)
- Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* 33:W465–7. doi:[10.1093/nar/gki458](https://doi.org/10.1093/nar/gki458)
- Stein LD, Mungall C, Shu S, et al (2002) The generic genome browser: a building block for a model organism system database. *Genome Res* 12:1599–1610.

- Sugiyama N, Nakagami H, Mochida K, et al (2008) Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in Arabidopsis. *Mol Syst Biol* 4:193.
- Tatusova TA, Madden TL (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* 174:247–250.
- Tress ML, Bodenmiller B, Aebersold R, Valencia A (2008) Proteomics studies confirm the presence of alternative protein isoforms on a large scale. *Genome Biol* 9:R162. doi:[10.1186/gb-2008-9-11-r162](https://doi.org/10.1186/gb-2008-9-11-r162)
- Tyers M, Mann M (2003) From genomics to proteomics. *Nature* 422:193–197.
- Urano K, Kurihara Y, Seki M, Shinozaki K (2010) “Omics” analyses of regulatory networks in plant abiotic stress responses. *Curr Opin Plant Biol* 13:132–138. doi:[10.1016/j.pbi.2009.12.006](https://doi.org/10.1016/j.pbi.2009.12.006)
- Wang R, Fabregat A, Ríos D, et al (2012) PRIDE Inspector: a tool to visualize and validate MS proteomics data. *Nat Biotechnol* 30:135–137. doi:[10.1038/nbt.2112](https://doi.org/10.1038/nbt.2112)
- Wang M, Wang Q, Zhang B (2013) Evaluation and selection of reliable reference genes for gene expression under abiotic stress in cotton (*Gossypium hirsutum* L.). *Gene* 530:44–50. doi:[10.1016/j.gene.2013.07.084](https://doi.org/10.1016/j.gene.2013.07.084)
- Wilkinson PA, Winfield MO, Barker GLA, et al (2012) CerealsDB 2.0: an integrated resource for plant breeders and scientists. *BMC Bioinformatics* 13:219. doi:[10.1186/1471-2105-13-219](https://doi.org/10.1186/1471-2105-13-219)
- Wright JC, Sugden D, Francis-McIntyre S, et al (2009) Exploiting proteomic data for genome annotation and gene model validation in *Aspergillus niger*. *BMC Genomics* 10:61.
- Yachdav G, Kloppmann E, Kajan L, et al (2014) PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res* 42:W337–43. doi:[10.1093/nar/gku366](https://doi.org/10.1093/nar/gku366)
- Yadav R, Arora P, Kumar S, Chaudhury A (2010) Perspectives for genetic engineering of poplars for enhanced phytoremediation abilities. *Ecotoxicology* 19:1574–1588. doi:[10.1007/s10646-010-0543-7](https://doi.org/10.1007/s10646-010-0543-7)
- Yan S, Du X, Wu F, et al (2014) Proteomics insights into the basis of interspecific facilitation for maize (*Zea mays*) in faba bean (*Vicia faba*)/maize intercropping. *J Proteomics* 109:111–124. doi:[10.1016/j.jprot.2014.06.027](https://doi.org/10.1016/j.jprot.2014.06.027)
- Yang F, Melo-Braga MN, Larsen MR, et al (2013) Battle through signaling between wheat and the fungal pathogen *Septoria tritici* revealed by proteomics and phosphoproteomics. *Mol Cell Proteomics MCP* 12:2497–2508. doi:[10.1074/mcp.M113.027532](https://doi.org/10.1074/mcp.M113.027532)
- Yao D, Zhang X, Zhao X, et al (2011) Transcriptome analysis reveals salt-stress-regulated biological processes and key pathways in roots of cotton (*Gossypium hirsutum* L.). *Genomics* 98:47–55. doi:[10.1016/j.ygeno.2011.04.007](https://doi.org/10.1016/j.ygeno.2011.04.007)
- Yu J, Zhao M, Wang X, et al (2013) Bolbase: a comprehensive genomics database for Brassica oleracea. *BMC Genomics* 14:664. doi:[10.1186/1471-2164-14-664](https://doi.org/10.1186/1471-2164-14-664)
- Zhang M, Lv D, Ge P, et al (2014a) Phosphoproteome analysis reveals new drought response and defense mechanisms of seedling leaves in bread wheat (*Triticum aestivum* L.). *J Proteomics* 109:290–308. doi:[10.1016/j.jprot.2014.07.010](https://doi.org/10.1016/j.jprot.2014.07.010)
- Zhang Y, Cheng Y, Guo J, et al (2014b) Comparative transcriptome analysis to reveal genes involved in wheat hybrid necrosis. *Int J Mol Sci* 15:23332–23344. doi:[10.3390/ijms151223332](https://doi.org/10.3390/ijms151223332)
- Zhao H, Peng Z, Fei B, et al (2014) BambooGDB: a bamboo genome database with functional annotation and an analysis platform. *Database (Oxford)* 2014:bau006. doi:[10.1093/database/bau006](https://doi.org/10.1093/database/bau006)

Plant Stress Tolerance

An Integrated Omics Approach

Mosa, K.A.; Ismail, A.; Helmy, M.

2017, X, 73 p. 14 illus., 2 illus. in color., Softcover

ISBN: 978-3-319-59377-7