

## USING SYSTEMS BIOLOGY TO INCREASE AGRICULTURAL PRODUCTION: A CRITICAL REVIEW STUDY

<sup>1</sup>**Dr.Ashish Kumar Gupta,**

Principal and Professor, Department of Agriculture Science, SR College of Professional Studies,SRGI Ambabai Jhansi (U.P) India

Mail:drashish.research@gmail.com

<sup>2</sup>**Mayank,**

Designation – Lecturer, Department of agriculture science, SR College of Professional Studies,SRGI Ambabai Jhansi (U.P) India

**Abstract:**In recent times, the collecting of huge datasets from numerous plant species' genomes, transcriptomes, proteomes, metabolomes, and epigenomes has become routine technique. However, the vast majority of these datasets in different crop species were studied independently, making it impossible to attain comprehensive understanding of the molecular basis of complex features and biological networks. It takes a systems biology approach, which incorporates integration of various omics data, modelling, and prediction of cellular processes, to understand the flow of biological data that underlies complex traits. In this context, systems biology with integration of multiomics data is crucial and enables a complete knowledge of the complex system with the various levels of biological structure engaging with the external surroundings to generate phenotypic expression. In this article, we'll talk about the most recent developments in the field of various omics investigations, such as integrative & systems biology methodologies, with a focus on the ways in which this research might be used to crop improvement. In addition, we have analysed the challenges and opportunities associated with integrating multiomics data, modelling, and biological knowledge of complex traits that underpin yield & stress tolerance in important cereals and legumes. These subjects were covered in the preceding section.

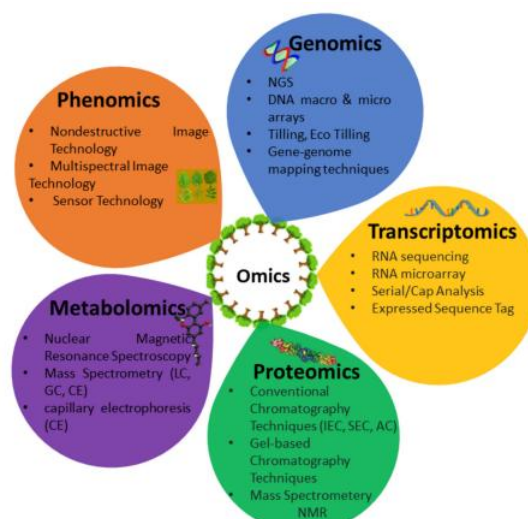
### **Introduction:**

Cereals and legumes, among other plant foods, play a significant part in the human diet since they provide the appropriate number of calories, proteins, essential amino acids, and minerals. Crop types that are capable of producing larger yields have been generated via intensive breeding efforts over the last one hundred years. Yet, due to rising populations in many parts of the world, there is an urgent need to increase production levels by 50 percent by 2050. This growth in output and productivity presents a challenge when considering the present environmental bounds and the fast-diminishing normal resources. Arabidopsis (*Arabidopsis thaliana* (L.) Heynh.) & rice (*Oryza sativa* L.) are two examples of model plants that have been subjected to significant research to gain a comprehensive grasp of plant genetics and genomics as well as describe the function of certain genes. This is necessary for harnessing genomics to breed a new cohort of climate-ready harvests, which will allow for the production of an abundance of food rich in nutrients. To this end, a strategy known as genomics-assisted breeding has the potential to speed up already established crop improvement operations significantly. Before the age of genomics, the ability to translate or transmission genetic info gathered from one species to additional was relatively restricted. This was mostly due to a lack of sufficient knowledge in genomic info and schemes biology.

The development of next-generation sequencing (NGS) technologies has ushered in a new age of 'big data' by radically altering and quickening the rate at which genomes and transcriptomics data are produced. This has led to a rebellion in the field of genomics. Genomes and transcriptomes from many different plant species have been sequenced using next-generation sequencing technologies developed by companies like Illumina (with their MiSeq/HiSeq system), Roche (with their 454/FLX system),

ABI/Life Technologies (with their SOLiD system), and Invitrogen (with their Ion Proton system). Third-generation sequencing technologies, such as the single-molecule sequence analysis developed by Helicos Biosciences (HeliScope), single-molecule real-time sequencing established by Pacific Biosciences (PacBio), as well as Nanopore sequencing developed by Oxford Nanopore Technologies, accelerated the generation of large-scale sequencing data. This has had a significant impact on the sequencing processes and contributed to the creation of high-quality genome assemblies in a variety of agricultural plants, including those with complicated and large-scale genomes. In order to get a better sympathetic of the difficulty of biological schemes, the massive amounts of data gleaned from omics studies are put through a number of sophisticated software tools and examined in various ways. Either the term "integrative biology" or "systems biology" is now often used to refer to this emerging field of study. Integrative biology is founded on the Aristotelian premise that the whole is larger than the amount of the parts. Its primary goal is to combine the many omics layers to get new insights. Schemes biology emphasizes on merging different omics layers to develop models that describe the behaviour of systems and that have the predictive capacity to suggest what would happen if certain biological pathways were altered or mutated. In addition, systems biology may be used in the analysis of complicated agronomic characteristics and the model-based prediction of phenotypes under various environmental factors. By enhancing selection tools & allowing for a more nuanced design of crossover programmes or layering of gene alterations, the creation of user-friendly pipelines & bioinformatic tools to examine the large data generated by omics techniques will further improve breeding efforts. Gene editing stacking is a viable option for achieving this goal.

Uses in plant research & breeding are highlighted throughout this evaluation of current omics resources, algorithms to detect data integration, & systems biology methodologies (Figure 1). We highlight both established and novel methods that advance our knowledge of the biology of complex characteristics and the comprehensive enhancement of yield, tolerance, and resistance to both abiotic and biotic challenges. Specifically, we focus on improving the yield of crops by increasing their tolerance and resistance. In addition, the potential and problems that are faced by multiomics data integration, modelling, and systems-level studies have been examined. This is especially relevant in light of the rapidly developing omics technologies. The ideas discussed in this overview provide perspectives on how integrated research on different omics might be used to systems biology projects pertaining to crop development.



**Figure 1: Principal Methods in the Field of Plant Omics**

### Approaches:

The investigation of a specific species' DNA, RNA, proteins, and metabolites, amongst other biological entities, has been made possible by high-throughput technologies, which have been notable for their contribution to the revolution in plant research brought about by these technologies. This high-

throughput measurement has resulted in the development of a variety of methodologies that end in the suffix omics. Genomics, pangenomics, omics, proteomics, metabolic engineering, epigenomics, and even single-cell omics, phenomics, as well as QTL-omics are all examples of such approaches. The current integration of these approaches across several omics layers provides a chance to understand the information flow underlying trait biology. Following, we will discuss the rapidly evolving omics strategies now being investigated for crop improvement.

To better comprehend how genes have been altered or modified by both artificial and natural selection in response to environmental stresses, it may be useful to get an appreciation for the structure, organization, and dynamics of the genomes of various plant species. It has been established through research that modified and adapted genes can be used for crop improvement, and not just within a species. Species can perform this action on themselves (Kawashima et al., 2016). Many plant genomes have been sequenced with the intention of accomplishing these goals (Kersey, 2019). Only a few plant species having relatively compact genomes were read in the past to get an understanding of genome architecture due to the high cost of sequencing & limited knowledge available at the time. These plant types were called "models" in the industry. The first plant genomes to have their sequence information created were for the species *Arabidopsis thaliana* (L.) Heynh., rice (*Oryza sativa*) L., black cottonwood, grapevine (*Vitis vinifera*) L., and maize (*Zea mays*) L. This was followed by the sequencing of the genomes of many legumes, including soy bean, pigeonpea, & chickpea (*Cicer arietinum* L.).

The growth of next-generation sequencing (NGS) also third-generation sequencing technologies has led to ever-increasing throughput as well as lower costs associated with sequencing. As a direct consequence of this, over 600 whole plant genome assemblies are already accessible in public repositories, and the sequencing of many more is currently underway. The knowledge about the genome that is obtained by high-throughput sequencing of germplasm collections makes it possible to find and sequence hundreds of genetic markers simultaneously across complete genomes. These new sequencing technologies are particularly helpful for validating and assessing genetic markers in populations, making them a very important resource. In addition, it is now feasible to identify all of the genes present in a plant, which, in turn, will assist in understanding the genetic features and networks that lead to the development of better crop types.

In addition to the information that has been made available from the sequencing of cultivar genomes, the description of the genetic variety that is present in wild crop families and landraces that have been preserved in gene banks is a valuable source of novel genes that have the potential to improve yield & resistance to biotic and abiotic stresses. In this setting, attempts to resequence large-scale germplasm collections have developed into a major area of research. For instance, the 3,000 Rice Genomes Project has sequenced the genomes of 3,010 varieties of farmed rice from Asia (Wang et al., 2018). According to the research findings, 29 million single nucleotide polymorphisms (SNPs), 2.4 million minor indels, and more than 90,000 structural variants (SVs) contribute to diversity both within and across populations. This research aimed to emphasise the genetic variety present in rice germplasm banks and how it relates to characteristics that are significant to agriculture. In addition, the research revealed how detected SNPs may be used in trait mapping analyses for highly heritable characteristics in rice such as grain length, grain breadth, and resistance to bacterial blight. Another kind of maize was sequenced, and the results showed substantial diversity in terms of SNPs (27 million), indels (287,504), and copy number variations. Future maize breeding programs may use these findings as a selection signal (Jiao et al., 2012). The study found that nucleotide diversity has decreased and the proportion of rare alleles has increased due to modern breeding's introduction of dynamic genetic alterations into the maize genome and additional artificial selection's effect on thousands of targets, such as genes and nongenetic regions (Jiao et al., 2012). For example, in the case of dryland cereals, 8 million SNPs, 1.9 million indels, and specific gene loss and gain events were identified through the resequencing of 44 sorghum lines (*Sorghum bicolor* (L.) Moench) that depicted the primary gene pool and spanned aspects of geographic origin, end use, and taxonomic group (Mace et al., 2013). This

study concludes that not only do the numerous sorghum races have an abundant, untapped pool of diversity, but so does the allopatric Asian species *S. propinquum* (Kunth) Hitchc. Strong racial structure and complex domestication processes were found among the explored accessions. Similarly, more than 29 million SNPs and 3 million indels were discovered by the homologous recombination of 994 lines of pearl millet (*Cenchrus americanus* (L.) Morrone), which will aid in the study of phenotypic diversity and hasten the rate of genetic improvement (Varshney et al., 2017b). This research aims to examine the use of commonly encountered information to better comprehend the population dynamics, genetic diversity, & domestication of pearl millet. A genome-wide associations study (GWAS) was conducted to predict the yield-associated properties of the crop in both irrigated and drought-like situations, and genomic prediction was utilized to estimate the performance of pearl millet hybrids. In the case of legumes, including such soybeans, resequencing of 302 wild and farmed soybean accessions resulted in the detection of 9 million SNPs including 876,799 indels. These discoveries identified genes linked with domestication as well as tools for genomics-enabled agricultural crops (Zhou et al., 2015). (Zhou et al., 2015). The genomes of 17 wild soybeans and 14 cultivated soybeans were resequenced in a different investigation (Lam et al., 2010), which revealed patterns of genetic variation between wild and farmed soybeans. As a result of this study, a total of 205,614 SNPs were gathered from wild soybean, increasing our ability to employ this species for QTL mapping and association investigations. In a research paper that appeared not too long ago, Valliyodan et al. (2021) evaluated the genetic diversity & organisation of 481 distinct soybean accessions by resequencing them. These accessions included 52 wild selections and 429 cultivated varieties (landraces and elites) (landraces and elites). This study's findings suggest that different geographic regions are exercising independent selection pressures, leading to the emergence of distinct lineages. Thousands of chickpea genomes from around the world are being sequenced and phenotyped as part of the 3,000 Chickpea Genome Sequencing Project. In fact, we just finished up with this project. In this study, we resequenced 429 chickpeas representing 45 countries to create a map of 4.97 million SNPs, and we conducted GWAS to identify 262 markers and several candidate genes for a total of 13 distinct drought and heat tolerance mechanisms (Varshney et al., 2019b). A similar study was undertaken on pigeonpea, and after resequencing 292 pigeonpea accessions, researchers found 15.1 million SNPs and 2.1 million indels. Genetic regions associated with domestication and markers associated with key traits like blooming duration, seed development, or pod dehiscence were discovered (Varshney et al., 2017a). Evidence for the utilization of genomes & SNPs in trait mapping analyses was uncovered through sequencing and resequencing tests performed on numerous crop species. Genetic variation discovered through these studies is expected to aid in the direction and acceleration of crop breeding, which is crucial for the improvement of a wide range of crops and the development of more sustainable agricultural practices in the future.

### **High-throughput phenotyping and phenomic analysis:**

The advantages from genetics resources can only be fully realised via phenomics. The field of agricultural phenomics has advanced significantly in recent years, allowing researchers to collect and analyse data on a wide range of phenotypes, from the single cell to the whole plant to the entire population. Quantitative, automated, and high-throughput measurements of plant phenotypes including size, shape, and composition are essential components. Important phenotyping data gathering and processing, which had previously been a key impediment for functional genomics investigations and crop breeding, is now feasible because to recent technical advancements. High-throughput phenotyping systems might be used to assess a wide variety of phenotypes, including but not limited to high-throughput shoot phenotyping, root phenotyping, canopy, and leaf attributes (Jin et al., 2020). New sensor technologies allow for comprehensive documentation of plant environmental history, and hence the real-time reaction of crops to their surroundings. UAVs, or unmanned aerial vehicles, then pocketPlant3D (a smartphone app) with a variety of sensors including hyperspectral imaging and

computed tomography imaging as well as specific metabolic sensors are used to estimate plant attributes like leaf area index, find weeds & diseases, and forecast crop yields.

High-throughput phenotyping in carefully regulated settings has made significant strides in recent years (Pratap et al., 2019).

Vision-guided robots is one example of a field-applied technology that is advancing quickly at the moment (Pieruschka & Schurr, 2019). Many model plants and agricultural species were studied by collecting high-throughput shoot and root phenotypic data under greenhouse settings (Yang et al., 2020). During its growth and development, a plant or crop's performance is influenced by a large number of genes that interact with a wide range of environmental factors. The dynamic response of crops may now be recorded using sophisticated sensor, machine vision, and automation technologies, and this data can then be combined with sequence data (Jin et al., 2020). As phenomics employs many different kinds of sensors all at once, it is essential to collect data methodically from the time of experimental design through data analysis and interpretation. Now that the genomes of both model and nonmodel crop species have been sequenced, a comprehensive description of the phenotype of the crop as a whole is urgently needed. This is crucial for determining essential adaptation features in crops and linking genes and QTL to phenotypes (Yang et al., 2020).

### **Bio Systems:**

Maximizing production while decreasing losses due to stress is a primary focus of crop biology studies. The issue's complexity necessitates resolution and an equally complex and comprehensive approach. Generating massive amounts of data from the genome, transcriptome, proteome, metabolome, & high-throughput phenome has become normal, making it feasible to understand the biological response at each level of the organisation. Nonetheless, most of the times, these data have been investigated separately till lately. The discovery and dissection of complex plant regulatory networks would be substantially aided by the integration of transcriptomics, proteomics, and metabolomics (Urano et al., 2010). Systems biology is a new area of study that combines the analysis of massive omics data sets with well-crafted mathematical models to test hypotheses and make predictions about living organisms and their environments. Important for the future breeding of climate-ready crops, this gives a more comprehensive knowledge of the system-wide response throughout growth, growth, and pressure adaptation. Yet, developing hypotheses based on existing information is the first stage in a systems biology strategy. Most importantly, this is the first step in developing a systems biology experimental design (Pinu et al., 2019). The process of using a systems biology approach in agricultural studies has been summarised.

### **Systems Biology Networks:**

Many transcriptional and translational levels of regulation allow cells to adapt to changing genetics and environments. Various methods, including chromatin remodelling, epigenetic modifications, mutations, transcription factors, and intergenic interactions, control gene expression. Plants' ability to respond to and adapt to stress is largely governed by regulatory networks. Understanding these regulatory networks and the useful interaction amongst the biological entities may be aided by pathway and network reconstruction employing transcriptome, proteome, and metabolome data. Genes, proteins, and metabolites all interact with one another in complex ways, and these interactions may be mapped out using networks and graphs to shed light on their interdependencies (Weckwerth, 2011; Weckwerth et al., 2004). In a nutshell, normalising omics data yields a similarity matrix, and the adjacency matrix that is obtained from that is converted into an undirected graph, or a network concept.

Redestig and Costa (2011); Langfelder and Horvath (2008); Langfelder et al. Recent publications outline the principles, methodologies, and tools of network implication for investigating biological specifics, tracing evolutionary roots, and comprehending network structure in order to anticipate biological functions. Both the existing state of knowledge and the projected interactions between

biological variables, as in Bayesian networks, place constraints on network-based techniques. Several tools exist for combining large-scale omics data sets (Misra et al., 2019). Additionally, several approaches, programmes, websites, and databases have been documented for integrating multidimensional omics data.

So, a system may be represented exactly as a collection of nodes connected by edges, where nodes are the biological entity & edges show their interaction or connection, and the nodes with the highest degree of connectivity are called hub nodes (Langfelder & Horvath, 2008). Connectivity is enhanced more so at hub nodes than at the module's outside nodes. As regulators, these nodes may affect the routes farther down the food web (McCormack et al., 2016). The ability to pinpoint hub nodes has far-reaching implications for isolating causal factors that shape both basic and complex characteristics. Hence, networks may foretell the importance of important components and their regulatory effect and reveal novel biological insights (Albert, 2007). To this end, we developed RiceNet, a genome-scale gene network for rice that has been shown to predict gene functions in other monocotyledonous species reliably (Lee et al., 2011). Moreover, a coexpressed gene network for barley (*Hordeum vulgare* L.) was created using transcriptome data. It found gene clusters related to response to drought stress and cellulose synthesis. Using a flowering gene network, researchers have been able to determine that GmCOL1a and GmCOL1b play regulatory roles in blooming in soybean. Seed size and set, nodulation and nitrogen fixation, and pollen fertility and set in pigeonpea, and the acquisition of desiccation tolerance in *Boea hygrometrica* are just a few examples of processes that have been studied using gene coexpression networks. The protein-protein interaction network, on the other hand, has proven to be an invaluable tool for studying intricate biological processes and learning how plants respond to the environment. An *Arabidopsis thaliana* Protein Interactome Database and a protein-protein interaction network of the plant tricarboxylic acid cycle have been constructed. In addition, a protein interaction network related to rice's salt tolerance was revealed. In addition, metabolite- metabolite association networks built using correlation algorithms can fully characterise the biological system's reaction to environmental disturbance. Metabolic network analysis convincingly confirmed the impact of PGR concentration and agroecosystem setting on the tomato metabolome. All of the aforementioned network inference research were conducted inside a single omics type; however, when combining data from many omics, many new and interesting insights were discovered. Using integrated omics techniques, scientists have reported on the existence of complex network linkages in agricultural plants' nitrogen metabolism and signalling (Fukushima & Kusano, 2014). In order to anticipate proteins implicated in abiotic stress resistance, namely XA21-mediated immunity, RiceNet was quantitatively merged with a proteomics dataset (Lee et al., 2011). Researchers combined data on coexpression gene modules and metabolite modules to determine the nature of the gene-metabolite interaction in tobacco (*Nicotiana tabacum* L.).

The research found new genes and potential regulators of major regulatory networks, such as the carotenoid metabolism pathway. Similar work was done by Mounet et al. (2009), who used a combination of transcriptome & metabolome data to pinpoint groups of genes essential for tomato fruit growth and metabolism.

In conclusion, global gene expression networks have shown great promise as a tool for the investigation and high-throughput prediction of specialised metabolite trails. Hence, network biology may revolutionise our comprehension of the genetic underpinnings of plant responses and adaptations to environmental change (Wisecaver et al., 2017). To achieve this goal, it is necessary to acquire omics data in a timely fashion so that it may be examined using a network biology method. Network biology may be an invaluable resource to represent the biological response to noxious environmental perturbations.

### **Applications for Crop Development:**

By bridging the gap between genotype and phenotype, systems biology has great potential to increase agricultural sustainability. Agricultural output factors like plant design, efficiency of nitrogen use,

water use efficiency, or biotic and abiotic stress tolerance may all be modeled and analyzed using this method. Improvements in plant molecular biology and genetics have had a significant impact on plant breeding efforts. However, the current developments in high-throughput experimental analysis & computer capacity offer a game-changing chance to merge several disciplines into a clear answer for every given complex trait. Abiotic stress-responsive cellular, biochemical, and molecular mechanisms and signaling processes may now be studied because to the availability of whole-genome sequencing information and the creation of omics datasets using fast developing technologies, analytical tools, including software. For instance, a systems biology-based breeding strategy can be built by using well-designed mathematical models based on time series records to identify crucial candidate genes for possible implementation in the breeding programs. They also presented a convincing strategy for implementing systems biology in subsequent breeding efforts. They reasoned that if the complex linkages regulating the multigenic trait for crop development could be identified through the application of systems biology, it could be optimized. Successful breeding programs to improve agronomically important and complex traits will require full models developed through the integration of multiomics data, high throughput, and accurate phenotyping. This is important for creating crops that will thrive in the future despite biotic and abiotic obstacles, especially in marginal locations. In the near future, systems biology will likely be used to develop and implement cutting-edge breeding strategies for crop development.

### **Demands and Opportunities:**

As the price of sequencing drops and expertise increases, multiple omics methodologies (such as genomics, omics, proteomics, metabolic engineering, epigenomics, and single-cell omics) are being used increasingly frequently in plant sciences. Thus, even without the genome sequence, novel alleles from multiple sources may be found using the sequencing & resequencing data gathered for several crops. Rapid advancements in omics technology have also made it possible to generate fresh and useful data on a wide variety of plant species. Discovering the pathways and genes responsible for crucial agronomic traits would be considerably aided by the efficient integration of these genomic & functional omics data with phenotypic and genetic information. However, even in a model crop like rice, the massive amount of data generated on the functional parts of a cell is still underutilized. In a 2019 study (Muthuramalingam et al. In addition, numerous biotic and abiotic stresses under field conditions have a cumulative effect, limiting crop development and yield. The complex cellular and molecular activities that contribute to enhanced agronomic traits in crops necessitate an immediate focus of future research on identifying the essential players, molecular networks, including models that might decode these processes.

This is why systems biology has so much promise in the field of agriculture; it has the potential to radically alter our understanding of how plants respond to changes in their growth and environment.

Systems biology and the integration of omics methods paint a more complete molecular picture of plant life than do individual approaches. However, the integrative system of various omics platforms continues to be a challenge because to their inherent data discrepancies. While accurate multiomics data is essential for system modeling and prediction, the scientific community as a whole has yet to develop any uniform methods for collecting and analyzing this information. Model-based integration, however, is typically restricted to research using well-characterized model organisms.

Ever-improving databases & data processing technology would aid in the meaningful biological analysis of multiomics data.

Extensive descriptions of the various technological obstacles that may arise during experimentation, data collecting, integrations, management, sharing, & analysis are provided by both Misra et al. (2019) & Macklin (2019). (2019). Therefore, the time has come to establish and share resources, databases, and software tools, and to forge a cohesive community for plant systems biology research as opposed to separate endeavors. Given the urgency of the requirement to quadruple agricultural output by 2050 in the context of an evolving climate, this is of crucial relevance.

**References:**

- Till, B. J., Reynolds, S. H., Weil, C., Springer, N., Burtner, C., Young, K., et al. (2004). Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biol.* 4:12. doi: 10.1186/1471-2229-4-12
- Todaka, D., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2015). Recent advances in the dissection of drought-stress regulatory networks and strategies for development of drought-tolerant transgenic rice plants. *Front. Plant Sci.* 6:84. doi: 10.3389/fpls.2015.00084
- Tokimatsu, T., Sakurai, N., Suzuki, H., Ohta, H., Nishitani, K., Koyama, T., et al. (2005). KaPPA-view: a web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. *Plant Physiol.* 138, 1289–1300. doi: 10.1104/pp.105.060525
- Tomlekova, N. B. (2010). Induced mutagenesis for crop improvement in Bulgaria. *Plant Mutat. Rep.* 2, 4–27.
- Toorchi, M., Yukawa, K., Nouri, M. Z., and Komatsu, S. (2009). Proteomics approach for identifying osmotic-stress-related proteins in soybean roots. *Peptides* 30, 2108–2117. doi: 10.1016/j.peptides.2009.09.006
- Twyman, R. M. (2013). *Principles of Proteomics*. Abingdon: Garland Science Press.
- Urano, K., Kurihara, Y., Seki, M., and Shinozaki, K. (2010). “Omics” analyses of regulatory networks in plant abiotic stress responses. *Curr. Opin. Plant Biol.* 13, 132–113. doi: 10.1016/j.pbi.2009.12.006
- Urbanczyk-Wochniak, E., Luedemann, A., Kopka, J., Selbig, J., Roessner-Tunali, U., Willmitzer, L., et al. (2003). Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Rep.* 4, 989–993. doi: 10.1038/sj.embor.embor944
- van Dijk, K., Ding, Y., Malkaram, S., Riethoven, J. J. M., Liu, R., Yang, J., et al. (2010). Dynamic changes in genome-wide histone H3 lysine 4 methylation patterns in response to dehydration stress in *Arabidopsis thaliana*. *BMC Plant Biol.* 10:238. doi: 10.1186/1471-2229-10-238
- Varshney, R. K., Thudi, M., May, G. D., and Jackson, S. A. (2010). Legume genomics and breeding. *Plant Breed. Rev.* 33, 257–304. doi: 10.1002/9780470535486.ch6
- Viant, M. R., and Sommer, U. (2013). Mass spectrometry based environmental metabolomics: a primer and review. *Metabolomics* 9, 144–158. doi: 10.1007/s11306-012-0412-x
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T., Hornes, M., et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23, 4407–4414. doi: 10.1093/nar/23.21.4407
- Wang, M., Wang, P., Tu, L., Zhu, S., Zhang, L., Li, Z., et al. (2016). Multi-omics maps of cotton fibre reveal epigenetic basis for staged single-cell differentiation. *Nucleic Acids Res.* 44, 4067–4079. doi: 10.1093/nar/gkw238
- Wang, W. S., Mauleon, R., Hu, Z. Q., Chebotarov, D., Tai, S. S., Wu, Z. C., et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557, 43–49. doi: 10.1038/s41586-018-0063-9
- Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., et al. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* 32, 947–951. doi: 10.1038/nbt.2969
- Weckwerth, W. (2003). Metabolomics in systems biology. *Annu. Rev. Plant Biol.* 54, 669–689. doi: 10.1146/annurev.arplant.54.031902.135014
- Weckwerth, W. (2010). Metabolomics: an integral technique in systems biology. *Bioanalysis* 2, 829–836. doi: 10.4155/bio.09.192
- Weckwerth, W. (2011a). Green systems biology – from single genomes, proteomes and metabolomes to ecosystems research and biotechnology. *J. Proteome* 75, 284–305. doi: 10.1016/j.jprot.2011.07.010
- Weckwerth, W. (2011b). Unpredictability of metabolism—the key role of metabolomics science in combination with next-generation genome sequencing. *Anal. Bioanal. Chem.* 400:1967. doi: 10.1007/s00216-011-4948-9



Weckwerth, W. (2019). Toward a unification of system-theoretical principles in biology and ecology—the stochastic lyapunov matrix equation and its inverse application. *Front. Appl. Math. Stat.* 5:29. doi: 10.3389/fams.2019.00029

Weckwerth, W., and Fiehn, O. (2002). Can we discover novel pathways using metabolomic analysis? *Curr. Opin. Biotechnol.* 13, 156–160. doi: 10.1016/s0958-1669(02)00299-9

Weckwerth, W., Ghatak, A., Bellaire, A., Chaturvedi, P., and Varshney, R. K. (2020). PANOMICS meets germplasm. *Plant Biotechnol. J.* 10:1111. doi: 10.1111/pbi.13372

Weckwerth, W., and Morgenthal, K. (2005). Metabolomics: from pattern recognition to biological interpretation. *Drug Discov Today* 10, 1551–1155. doi: 10.1016/S1359-6446(05)03609-3

Wen, W., Li, D., Li, X., Gao, Y. Q., Li, W. Q., Li, H. H., et al. (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat. Commun.* 5:3438. doi: 10.1038/ncomms4438