Recent Advances in OMICS Technologies

1

Vikas Gupta, Satish Kumar, Disha Kamboj, Chandra Nath Mishra, Charan Singh, Gyanendra Singh, and Gyanendra Pratap Singh

Abstract

Omics is a collaborative, wide-ranging science primarily concerned with the study of the relationships of biological knowledge resulting from genome profiling, transcriptomes, proteomes, metabolomes, and a variety of other related omes. These high-throughput approaches generate huge information related to differentially expressed genes/proteins/metabolites. The interrelationship between different omics studies and high-dimensional biological data produce from these omics are analysed through statistical methods. These studies have helped to create new resistant varieties by elucidating metabolic responses of crop plants to different biotic and abiotic stresses. Interpretation of omics research has enhanced biological study resolution, which contributed to analyses of metabolic pathways and biological system research. However, it is important for the creation of strategy to work on problem-oriented and process-oriented objectives leading to crop improvement to proceed to organised disciplines such as structural genomics, transcriptomics, proteomics, and metabolomics with plant physiology, biochemistry, and plant breeding.

Keywords

Omics · Genomics · Proteomics · Metabolomics · Transcriptomics

V. Gupta · S. Kumar (✉) · D. Kamboj · C. N. Mishra · C. Singh · G. Singh · G. P. Singh
ICAR—Indian Institute of Wheat and Barley Research, Karnal, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

A. Kumar et al. (eds.), *Omics Technologies for Sustainable Agriculture and Global Food Security Volume 1*, https://doi.org/10.1007/978-981-16-0831-5_1

1.1 Introduction

Food supply is a requirement for food security, and is an important factor to reduce hunger. The agriculture industry has shown significant improvements over the past 50 years with respect to production and productivity enhancement, dominated by the cereals which form the major staples across countries. The twentieth century has seen tremendous increase in food production with the introduction of high yielding crop varieties, especially since the first 'green revolution' thus saving millions of people from starvation (Rajaram 2005). This has been achieved mainly because of development and adoption of high yielding cultivars, increased area under irrigation, and widespread use of fertilisers. However, the present crop yield is not sufficient enough to feed the 9 billion of global population by 2050 (Ray et al. 2013). In addition, global climate change has further worsened this scenario with altered rainfall patterns, extreme weather events, and changing patterns of pathogens and pests in terms of severity and distribution (Abberton et al. 2016). With population rising gradually and natural capital shrinking, the production of food must be improved that can be accomplished with the use of modern analytical methods and technology to increase the quality and quantitative improvement of crop plants (Agrawal et al. 2013; Nelson et al. 2014; Haddeland et al. 2014).

Cultural change relies on current or created genetic variation, and the genetic benefit rate in breeding systems will increase either by expanding the quantity of variation available for selection or by speeding up the selection process so that varieties can be developed more quickly. During the last few decades, the development and use of molecular markers in crop genetics led to the identification and mapping of genes and QTLs controlling key traits. The marker system initially started with the use of restriction fragment length polymorphisms (RFLPs) followed by RAPDs, SSRs, AFLPs and culminated with the SNP markers based on genome sequencing (Tanksley et al. 1989; Varshney et al. 2009). The SSR markers have been used extensively for linkage map construction, tagging of new genes/QTLs due to their abundance in the crop genomes as well as because of their high polymorphism. (Somers et al. 2004; McCouch et al. 2002; Varshney et al. 2007; Smith et al. 1997; Sharma et al. 2013). Use of SSR markers has few limitations like SSRs are not uniformly distributed in genomes, gel based assay and cannot be integrated across platforms, which limit its potential to be exploited in plant breeding programmes.

With the development of low-cost sequencing technologies and genotypic platforms, there is an exponential increase of plant sequences in databases and enhanced our understanding of molecular and physiological roles of genes. Plant genomics includes both structural and functional genomics, which covers both basic and applied aspects. Given the rapid evolution of novel technologies, especially the advent of bioinformatics, the understanding from gene to genome level has increased. The new developments in the 'omics' technologies have brought revolution in plant science research due to the next-generation sequencing (NGS) 454, Solexa Illumina, Applied Biosystems (ABI) SOLiD, and high-throughput marker genotyping technologies. Consequently, the availability of complete genome

sequences for different crop plants holds great promise for research targeted towards crop improvement.

Many different functional genomics approaches have been used to identify genes that linked to the key biological processes. The crop traits are generally complex quantitative traits, which are controlled by multiple genes involving complex genic interactions (Core et al. 2008). Omics is a collaborative, broad-based field that focuses on the study of bio informative interactions obtained from genome profiling, transcriptomes, proteomes, metabolomes, and a variety of other related omes. The three main omic technologies for the analyses of genomics, proteomics, and metabolomics are the genetic, proteome and metabolism analyses of organisms' cells and tissues (Varshney et al. 2013). In this chapter we will discuss these technologies in detail and how these can be integrated in breeding programmes for better understanding of the functioning of key genes.

1.2 Genome and Genomics

The term genome refers to the entire genome compliment contained in a haploid chromosome set of any organism and the term was coined by German botanist Hans Winkler in 1920 at the University of Hamburg (Vukmirovic and Tilghman 2000). Genomics is the study of complete DNA sequence of organisms. It has been proven as one of the most transformative impacts on biological studies. Genomics study involves recombinant DNA technology, DNA sequencing, assembly, and annotation of genome to decipher the structure, function, and evolution of genes.

The understanding of genome sequences of organisms is crucial for unravelling the roles and networks of genes, for defining evolutionary relationships and processes, and to discover previously unaware control pathways that coordinate gene operation. Genomics has become one of the most transformative forces in biological sciences, the study of the whole organism's DNA sequence. Genomics is the genetic characterisation and cloning technique of entire genomes to explain gene structure, function, and evolution. The genetic material in eukaryotic organism is deoxyribosenucleic acid (DNA), which is a double stranded molecule composed of nucleotide bases, viz., adenine (A), guanine (G), cytosine (C), and thymine (T). The sequence of nucleotides in a gene on a DNA strand determines the sequence of amino acids that make up a protein.

For protein synthesis, the DNA in a gene is first transcribed to messenger RNA, followed by its translation into a sequence of amino acids through ribosomes in the cytoplasm of the cell. The proteins and products are fundamentally responsible for all cellular mechanisms. Finding out the pattern of arrangement of nucleotide bases in the entire genome is called genome sequencing. Time line for key discoveries in molecular biology has been depicted in Table 1.1 that led to the enormous data generation and formed a basis for better understanding of the key functions of the genomes. In 1972, gene sequence for bacteriophage MS2 coat protein was first determined by Walter Fiers and his team at the Laboratory of Molecular Biology of the University of Ghent, Belgium. The first DNA-based genome of bacteriophage

Table 1.1 Timeline showing key discoveries in DNA Sequencing

S. no.	Year	Discovery
1	1977	Sanger's chain termination method Chemical cleavage Maxam Gilbert's method
2	1980	Non-radioactive sequencing
3	1986	First semi-automated DNA sequencing machine
4	1987	ABI 370 fully automated sequencing machine
5	1990	Base by base sequencing
6	1995	ABI prism 310 genetic analyzer
7	1996	Pyrosequencing
8	2000	Massively parallel signature sequencing (MPSS)
9	2003	SMRT [single molecule real-time sequencing]
10	2004	ABI 3130 Genetic Analyzer
11	2005	First NGS machine Roche 45-4GS-20
12	2006	Solexa/Illumina sequence analyzer
13	2007	Second commercial NGS platform ABI SOLID sequencer
14	2008	Third commercial NGS platform 454 titanium GAII
15	2009	Illumina GAIIx SOLID 3.0
16	2010	Illumina Hi Seq2000 Ion Torrent
17	2011	Pac Bio RS sequencer CMOS non-optical sequencing MI Seq
18	2012	Ion proton
19	2014	Nanopore sequencer
20	2015	Sequel system (Pac bio)
21	2016	Oxford nanopore mini ion sequencer

Φ-X174; (5368 bp) was sequenced by Frederick Sanger (Sanger 1977). Since then genomes are being sequenced at a rapid pace ranging from viruses, human, bacteria, fungus, and many crop plants. Scientists, through a variety of functional genomics approaches, are characterising the genes that control key processes. The crop genome sequences are valuable for the discovery and insight into genetic diversity by resequencing of various genomes. The positional cloning and seed breeding promote sequence dependent markers linked to the rare elite alleles.

Affymetrix (Santa Clara, CA, USA; www.affimetrix.com) and the Illumina (San Diego, CA, USA; www.illumina.com) BeadArrayTM technology are commonly used for microarrays from GeneChipTM (GeneChipTM) and chip related technologies. Some recently built commercial platforms such as Affymetrix ® EurekaTM, and Illumina's Infinium also rely on SNP markers of high density. Furthermore, plant breeding and agricultural biotechnology have been revolutionised by the micro-array-based characterisation of plant genomes.

Ganal et al. (2011) reported a uniform large-scale SNP genotype array of over 8,00,000 SNPs, which were spread evenly throughout genome of maize. In addition, competitive PCR (KASP), TaqMan real-time PCR, and STARP experiments use low-cost PCRs to detect SNPs and provide versatility with regard to the number of

samples to be analysed (He et al. 2014; Semagn et al. 2014; Long et al. 2017). Diagnostic markers may be used without further confirmation, developed on target gene sequences, to improve application of marker-assisted selection (MAS) in breeding. The production of diagnostic markers is rapidly assisted by genome sequencing/resequencing. Such markers allow plant breeders to precisely produce new crops with targeted characteristics in future to satisfy different food requirements.

Most significant agronomic and economic characteristics in culture write traits in cultural breeding methods are several QTL regulated quantitative characteristics. Therefore, it is of considerable significance for marker-assisted breeding to explore QTL or even underlying causal genes/alleles. Linkage mapping is a standard approach for the genetic dissection of the quantitative trait loci genetic foundation (Emebiri et al. 2017; Liu et al. 2017; Zhang et al. 2017). To date, this approach has detected a significant number of QTLs. Moreover, the meta-QTL study was developed by combining QTL published to classify reliable QTL for crop enhancement with the achievement of biological informatics and large genetic knowledge (Van and McHale 2017). The extensive mapping of genomes in large populations makes it easier to map and clone target genes in crops.

Due to the available high-density SNP markers, the GWAS (Genome-Wide Association Studies) analysis, which includes natural populations, opened a path. The GWAS analysis will solve a range of traditional mapping shortcomings and deliver a powerful alternate technique for the dissection of complex features. By bringing together high-performance phenotypic and genotypic results, GWAS provides inspections of maize's complex characteristic genetic architecture, particularly in view of the rapid loss of corn imbalance (Yan et al. 2011).

Genome selection (GS) estimates the genomic estimated breeding values (GEBVs) of lines by a study of traits and high-density marker values in an artificially generated whole-genome population (Meuwissen et al. 2001; Crossa et al. 2017). GS has shown as important breeding method to develop dynamic characteristics easily. Closer relationships, including polygenic features, between the genomic estimate and true breeding values have been found (Jia and Jannink 2012). Even if also expensive, the number of QTLs that regulate a trait has proven superior to the recurring collection of markings for changes in complex characteristics in crops, so it can efficiently prevent problems.

1.3 Transcriptomics

Transcriptome is the study of total RNA, which gets edited and becomes mRNA, and finally translated into proteins. It involve the analysis of expression profiling of mRNAs in a given cell population.

The transcriptome is complex, since it is mainly a representation of the genes actively expressed at any given time under different circumstances. It assesses the changes in gene expression pattern due to inner and outer factors such as biotic and abiotic stress. It is an important method for understanding responses of biological

system under different circumstances. Transcriptome research offers a crucial forum for the review of the connexions between genotype and phenotype, which provides a deeper understanding of the fundamental pathways and mechanisms involved in the regulation of the cell fate (Ruan et al. 2004).

Transcriptomic approaches such as next-generation sequencing (NGS) facilitate the comprehension of functional components of genome. From past two decades, a number of agricultural researchers have made use of advances NGS techniques that enabling them to study high-resolution linkage between gene variants and traits of interest resulting in an upsurge in the scope of transcriptomic studies, not only in number but also in the range of agricultural systems that are being studied (Imadi et al. 2015).

Microarrays, serial analysis of gene expression (SAGE), massive parallel signature sequence (MPSS), and next-generation sequencing platforms (NGSPs) are typical technologies applied for genome wide or high-throughput genetic expression analysis (Costa et al. 2010; Harbers et al. 2005; Mardis et al. 2008). High-performance, quantitative PCR is used to confirm the expression of a broad variety of genes that are elucidated by tissue omics and choice care (Czechowski et al. 2004).

1.3.1 Differential Display

Differential Display technology is a PCR based technique for analysing gene expression using an Oligo-dT (3′) and an arbitrary 5′-end oligonucleotide primers (Liang and Pardee 1992). In order to achieve good visualisation, all RNA fingerprint variants need a low annealing temperature during PCR amplification. Therefore, not only the initial concentration of the cDNA depends on the quantity of the actual amplification components, but also the consistency of a precise match between primer and prototype (McClelland et al. 1995).

Throughout the course of PCR amplification, plentiful cDNAs with bad matches to the primers used would possibly outperform rare species with perfect matching. Despite this drawback, the classification of genes expressed distinctly by plants is recorded in many ways. This include the cloning of the last (1-aminocyclopropane-1-carboxylate-oxidase) enzyme involved in the synthesis of ethylene, a hormone which is involved in tomato ripening (Barry et al. 1996); and senescence, particular genes involved in the synthesis of flavonoids (Saito et al. 1999; Yamazaki et al. 1999); genes expressed differentially during carpel growth (Yung et al. 1999) or floral transformation (Yu and Goh 2000); module such as plant/environmental interaction based nitrate conveyors (Filleur, and Daniel-Vedele 1999); light photoreceptor-regulated genes, phytochrome (Kuno et al. 2000), and abiotic stress-response proteins (Brosche and Strid 1999; Baldi et al. 1999; Kim et al. 2000). This method is also used to differentiate within a species between individuals (Lapopin et al. 1999; Ni et al. 2000).

1.3.2 cDNA-AFLP

This fingerprinting approach depends on the sub-set of DNA molecules from a diverse pool being selectively amplified. The creation of an amplified polymorphic fragment length (AFLP) gives a consistent means with which two similar individuals can distinguish the few differences (Vos et al. 1995). The method is based on highly stringent conditions that are supported by the insertion of dual strand adapters at the ends of limit fragments that act as primary sites during amplification. Selective fragment enhancement is obtained by adding one or two foundations to the PCR primers and is only successfully improved by a fragment flanked by the limiting site, minimising the amount of visualised band.

More than 260 studies presented the advantages of this technique for exposing DNA fingerprinting. As cDNA was used to amplify transcripts in plants due to the benefit of this approach in producing a good individual polymorphism. As a tool for producing differently expressed products, the use of any pair of four and six simple restricting enzymes was suggested and successfully used as a combination of Pst I/Mse I and Asen I/Taq I (Bachem et al. 1996; Money et al. 1996). This technique seems more effective than that to recognise the individual RNA fingerprint and discern it; it is possible to quickly clone and subsequently characterise a certain band, particularly if silver stain is used to establish the fingerprint pattern. There have been accounts of definitions of many ESTs (Suarez et al. 2000; Durrant et al. 2000; Bachem et al. 2000).

1.3.3 DNA Microarray

A high efficiency, most frequently used technique for gene expression analysis, and a central factor in today‘s practical genomics research (Aharoni and Vorst 2001) is the DNA microarray (also referred to as a DNA chip or microarray). In short, this procedure involves the extraction, first of all by means of a laser-capture microdissection (LCM) technology of individual cells from two or more biological samples, and then the extraction from cells captured of complete RNA. The copies of each RNA that has been expressed are multiplied. Thus generated RNAs (and hybridised into a DNA chip under machine control be radioactively labelled or labelled with fluorescent colours) are engineered to visualise which genes are transcribed into the RNA in the cell when the sample is taken.

For any colouring position of the chip, scanners are used to read signals and fluorescence measurements are made. For data retrieval and interpretation (http://www.parisdevelopment.com), advanced applications and data collection systems are then used. A variety of plant procedures including the circulation clock, plant protection, environmentally stressful responses, fruit maturing, phytochrome signalling, seed growth, or nitrate assimilation have been investigated using microarray technology (Aharoni and Vorst 2001).

1.3.4 Serial Analysis of Gene Expression (SAGE)

The key difference between DNA hybridisation and SAGE approaches is that the latter does not need previous sequence knowledge to be studied, as SAGE is a gene expression profiling technique focused on sequencing (Velculescu et al. 2000). SAGE can be used to achieve full transcriptional profiles of expressed genes, although unknown, for species with poorly defined and expressed sequences. A new modification of SAGE known as Long SAGE makes it possible to use the derived transcriptome in the annotation of genetic expression in the genome (Saha et al. 2002). SAGE is a genuinely global and neutral technique for gene expression. The SAGE procedure uses several enzyme steps, amplifying and cleaning PCR and cloning. The SAGE protocol begins with mRNA cleaning connected with solid-phase magnetic oligo(dT) beads. The cDNA is synthesised directly on the oligo (dT) bead and then digested with the anchoring enzyme NlaIII (AE) to reveal the 3′ most restriction site on the oligo(dT) bead. The 4-bp site-detection enzyme NlaIII is expected to take place per 256 bp and thus to be present on most mRNA organisms in most SAGE experiments. However, a second SAGE library can be used for the identification of transcripts without a NlaIII site with a separate anchoring enzyme and also for the reconfirmation of a transcript identity in both sites. This could decrease the work relevant to research substantially, but the marginal importance of this method remains to be seen. Next, the sample is also divided into two separate tubes and attached to two separate connectors, A and B. The BsmFI, a type IIS restrictive enzyme that reduces 10 bp 3′ from the position of anchoring enzyme recognition, is recognised by both linked instruments. A special oligonucleotide known as the SAGE tag, hence known as the Tagging Enzyme (TE) is produced by BsmFI. The SAGE tags released from the beads of oligo (dT) are then removed and binding to each other for the purpose of making the ditag was amplified by the help of an electronic sequencer (Patino et al. 2002) and released from the ligators. The gel is purified, serially ligated, clone and sequenced.

1.4 Proteomics

Proteomics includes the recognition and characterisation of the full collection of genome proteins (Wilkins et al. 1996) sufficient for the preservation of the structure and maintenance of important regulatory functions (Whitelegge 2002). Proteins are huge amino acid molecules, each with a special genetically-defined series that specify the basic shape and role of the proteins in various ways, to form tens of thousands of proteins. Proteomics analyses the sequence of amino acids and various post-translation modifications to determine their relative abundance (Brygos and Joyards 2004). A cell, tissue, or organism's protein is the complete complement of proteins expressed by its genome at any given time (Renaut et al. 2006).

It includes systematic study of proteins expressed by a genome, from their amino acid sequence to relative quantities assessment, post-translocation transition state

and the interplay with other proteins or molecules of various forms to protein function and structure classification (Brygoo and Joyard 2004; Rhee et al. 2006). It starts with functionally modified proteins and the corresponding gene sequences, as opposed to genomics.

Proteomics is simply a method for plant functional genomics. The proteomics research leads to understanding diverse biological mechanisms and cellular response to environmental stress. It has emerged as an effective method to understand the role of proteins that establish homeostasis in cells, participate in cell signals, and are needed to maintain the structure. Protein-level studies are thus important to demonstrate molecular mechanisms that underlie plant growth, production, and environmental interactions (Chen and Harmon 2006). These are particularly compatible with crops because they can provide insights, not only in terms of nutritional worth, but also in terms of yield and the impact of adverse conditions on these factors (Solekdeh and Komatsu 2007).

Many emerging proteome approaches are available, each seeking to increase isolation, resolution, and automation in conjunction with various experimental objectives and with the chemical and physical properties of the target proteins. The significant technology used during research is the identifying and characterisation of 2-D polyacrylamide gel electrophoresis of separated proteins (Klose 1975; Fenn et al. 1989; Kara et al. 1989) and Mass Spectroscopy (Fenn et al. 1989; Kara et al. 1989) for the identification and characterisation of separated proteins. 2-DE is based on isoelectric concentration, where the proteins are segregated in polyacrylamide gels pH gradient (first dimension) and in the SDS (sulphate sodium dodecyl) PAGE according to their pI (second dimension) (Klose 1975; Klose and Kobalz 1995) and are separated according to their molecular weight.

The visualisation is accomplished with the use of staining methods of isolated protein spots. Electron-spray-ionisation (ESI) is carried out to transform peptides into ions by transmitting them through high-voltage columns. Proteins within the points of interest are first detected by digesting them into peptides usually with trypsin and then analysed by the mass spectrometry (MS). Flight Time (TOF) is an operation that analyses the mass of peptide ions in mass spectrometry. The MALDI TOF (Matrix-assisted laser desorption and ionisation time-of-flight) technique is the most widely used MS technique. These methods are used to trace peptides to classify protein and its complex interaction (Kersten et al. 2002; Henzel et al. 1993).

An alternative approach for analysis of protein products directly by MS, without isolation of gel, has been developed that enables the automatic analyses of peptide mixtures that are formed from complex protein samples, to be carried out using a capillary, high-performance liquid chromatography pair (HPLC), MS or MS/MS (Appella et al. 1995; Washburn et al. 2001, in French). In addition, the ground breaking reagent known as isotope encoded affinity tag (ICAT) in the LC-MS/MS method is made possible for quantitative proteomics. Yeast two-hybrid (Y2H) assay helps in detection of the protein–protein interactions (Fields and Song 1989). The large-scale protein-interaction map was developed using Y2H systems, which were used in the study of signalling complexes (Chen and Harmon 2006). These maps are

also available. This genetic technique enables the speedup detection and separation of the relationship between in vivo protein–protein interactions of the related nucleic acid sequences. In holding the intracellular balance, the relationship between protein and protein is important (Ozbabacan et al. 2011). The study of proteomics is therefore important for understanding the dynamics of complex biological processes. Proteomic analytics are becoming a good method for tracking growth and for obtaining insight into the molecular workings of plants. Attempts were made to improve the photosynthesis and abiotic stress resistance in crops. C4 plants were found to possess two forms of chloroplasts, making them more energy-efficient. A comparative proteomic analysis of C3 and C4 plant chloroplast was carried out to determine the proteins that cause more effective light fastening (Zhao et al. 2013).

In Arabidopsis the application of proteome analysis resulted in better understanding of the complex cellular events while analysing the role of gibberellic acid (GA) during the initial stages of seed germination and the effect of scarification on seed germination. During the production of food, after proteome review, a number of previously unknown, novel enzyme encoding genes were found in maize. The genetic diversity at the level of expressed proteins has been extensively tested with proteomics (Canovas et al. 2004). Genetically heterogeneity at the level of the expressed proteins has been commonly applied to proteomics (Canovas et al. 2004). The male sterility proteins fundamentally required for the selection of hybrids have been identified by Wang et al. (2013). The conditions of heat stress influence crop plant growth and yield. The organ-specific proteomic trial was conducted in various crop plants (Hossain and Komatsu 2013) to determine proteins directly acquired during abiotic stress. The proteomic information has been extended to the development of desired transgenic plants (Gong and Wong 2013).

1.5 Metabolomics

The quality and derivative products of crop plants depend directly on their metabolites. The taste, fragrance, colour and texture of crops, their storage properties, and their field efficiency are determined by metabolite concentration (Memelink 2005). A plant's metabolite material is its metabolome. It is a technology for comprehensive analysis which identifies and quantifies all the metabolites of an organism at a particular time (Fiehn 2002). Building on their existence at the developmental stage, metabolites are categorised as primary and secondary.

The metabolome is very diverse: lipid soluble chemical usually present in membranes, acidic and simple ions, acidic and basic chemicals for aqueous cell elements, stable oxidisation systems and structures (Maloney 2004). This metabolome is very diverse. The plant kingdom contains an estimated 100,000–200,000 metabolites (Oksman-Caldentey and Inze 2004). A broad view of the biochemical status, which can then be used to track and determine gene expression (Fiehn et al. 2000), is provided by quantitative and qualitative measurements of this vast number of cellular metabolites. In several plant resistance and stress reactions the metabolites are stated to be working (Bino et al. 2004).

The organism's biochemical reaction to conditional disruption is characterised by the impact it has on differential aggregation of each metabolite (Raamsdonk et al. 2001). As metabolites reflect catabolic and anabolic behaviour by proteins at a given time (Maloney 2004), macro-molecular processes such as feedback inhibition and signalising molecules (Dixon et al. 2006) are increasingly understood as modulating (metabolite). It originated as a practical genome approach to help us understand the diverse molecular dynamics in biological processes (Hall et al. 2002).

In contrast with genes and proteins, metabolites are much more variable in the order for atoms and subgroups. The invention of new methods such as mass spectrometry (MS), nuclear resonance magnetism (NMR), and four-transform infrared spectroscopy (FTIR) makes the study of metabolites much simpler. The study of metabolomics relies on the sample being extracted at a certain point in time (Hall 2006). The techniques of metabolomics give us an insight into the diverse metabolism and diversity of the organic reactions involved in plant growth, differentiation, stress tolerance, and pathogens defence mechanisms. The performance of the metabolome analysis depends on a few important factors, such as biological material/sample production and the processing of samples/metabolite detection (Hall 2006).

Bino et al. (2004) proposed a metabolomics research (MIAMET) protocol that includes experimental design; sampling; planning, extraction, and deriving of metabolites; design and design of metabolite profiles; and calculation of metabolites and parameters of metabolites. Because of complicated metabolic data for complex analytics, various analytical methods have been developed. These methods include target analysis, profiling of metabolites, metabolomics, and metabolism. Metabolite target analysis requires a variety of approaches for the preparation and analysis of samples for one or a few compounds from diverse mixtures. Samples based on genotype, phenotype, or biologic significance are tested by metabolic fingerprinting on a variety of metabolites (Shanks 2005).

Metabolomics in the strict sense is the calculation of all metabolites in a given organism. It is not yet theoretically feasible, and would possibly require platform of complementary technologies, since no single methodology is systematic, selective, and efficient enough to quantify them all (Weckwerth 2003). Technology for the study of target compounds and metabolic profiling is such as gas chromatography (GC), high-performance liquid chromatography (HPLC), and NMR. These methods focus on chromatographic isolation, often paired with sophisticated analytical calibrations. Samples are studied using NMR, direct injection mass spectrometry (MS), or FTIR spectroscopy for metabolic fingerprinting as crude extracts without separation. Often paired with multivariate analyses, these fingerprinting methods get the best out of the results.

In order to produce metabolic profile for study of different stress, economically significant characteristics, and heredity results, studies in metabolomics were carried out in the agricultural sector. Seasonal variations, geographical area, and natural variation have all been successfully used to determine the effect of metabolic profiles. The metabolic profiles of transgenic plants have been studied in order to characterize them. Metabolomics allows researchers to classify and distinguish

genotypes and phenotypes based on metabolic levels in a non-biased way. This information will be important for tracking crop quality characteristics (Hall et al. 2005). The detection and enhancement of trait production in farm and biological refining products (Reid et al. 2004) will identify possible biological markers for commodity contaminations and adulteration (Dixon et al. 2006).

Plant properties can be enhanced in different ways, for example, through raising metabolic fluxes by metabolic engineering to useful biochemical routes (e.g. increasing food's nutritional value, reducing pesticidal or fertiliser requirements, etc.) or through utilising the required routes for pharmaceutical development in plants (Giddings et al. 2000). In the same way, the incorporation of a foreign group of enzymes that contribute to the creation of desired finished products from other or more upstream precursors can lead to metabolistic shortcuts, and the synthesis of new metabolites may also generate foreign enzymes (Bino et al. 2004; Memelink 2005). Metabolic fluid analyses helped to define the similar catalytic properties of orthologous enzymes. Metabolistic experiments to classify the growth profile, production, and chemical taxis research were also conducted. Biochemical network construction has been carried out by creating relative metabolite profiles (Mendes 2002). The incorporation of metabolome and transcriptome evidence indicates that regulatory networks are being recognised and genetic material is correlated with phenotypic characteristics (Urano et al. 2009).

1.6 Bioinformatics

Over the past 15 years, bioinformatics has steadily developed and has become a modern major field in biology. A vast range of genetically and genome dependent results have been created using next-generation sequence technology to produce massive genetic and genome data, too detailed to be processed by the human brain. Bioinformatics includes one discipline: genetics, computer science, and IT (Rhee et al. 2006; Vassilev et al. 2005). It uses specialised tools and computers to search and interpret datasets from gene sequences and other sources. The National Science Foundation (NSF) began a project following the genome sequence of *Arabidopsis thaliana* to determine the function of 25,000 genes of Arabidopsis (Ruperao et al. 2014). Rice was the first genome-sequenced crop (International Rice Genome Sequencing Project 2005), and subsequently a variety of large crop genomes were sequenced. A vast number of data were published from all these sequencing initiatives. A series of bioinformatics instruments have been developed to organise and interpret these data that have significantly contributed to the drawing of significant biological conclusions, gene function predictions, etc. In addition, it allows researchers to define quantitative trait loci characteristics (QTLs) in the production of unorthodox mapping communities and online marker tools.

For the DNA sequencing research, transcriptomic, proteomics and metabolomics, multiple bioinformatics instruments and databases were developed and different researchers have given a detailed list of the same for separate purposes (Bino et al. 2004; Varshney et al. 2005; Rhee et al. 2006). These databases formed the

foundations for generating theories, developing tests, and inferring knowledge of a single organism. Furthermore, various species have used datasets and omic tools to evaluate the 'omic' properties of species which allow for further analysis of preserved genes and evolutionary relationships. The role of bioinformatics in gluing fundamental science to applied research will continue to be important, and biotechnology is crucial in solving the pressing issues of society such as sustainable energy production, reduction of world malnutrition, hunger and protection of the environment (Rhee et al. 2006).

1.7 Conclusion

The genomics, proteomics, and metabolomics technology platform is extremely successful. They significantly increase the number of proteins/genes which can be concurrently identified and are able to connect complex mixtures in gene/protein expression profiles with complex results. Constitute infrastructure for the implementation of genomics information and techniques (and other 'omics') was considerably advanced to allow alterations in crop plant characteristics to boost real and potential production, increase resource use efficiency, and improve health in the crop system. The need for developing new crop varieties with higher yield, heat resistance, and lower use of pesticides is to meet the demands of the growing population, improvements in climate trends, and environmental stress.

High-dimensional biological data have been produced by advanced technologies, analysed using statistical methods to detect interrelationships within omics studies. In order to grow new resistant varieties, the study of metabolic responses of cultivars against different biotic and abiotic stresses will be helpful. However, over time, stage, and environmental factors the transcriptome, proteome, and metabolomic data have become extremely variable and thus the processing and interpretation of samples needs to be conducted with great caution. The understanding of omics also enhanced the resolution of biological experiments, which led to analyses of metabolic pathways and biological system research. Disciplines including structural genomics, transcriptomics, proteomics, and metabolomics are also required for the production of strategy for problem-oriented and process-oriented processes in plant physiology, bio-methods, and plant breeding.

References

Abberton M, Batley J, Bentley A, Bryant J, Cai H et al (2016) Global agricultural intensification during climate change: a role for genomics. Plant Biotechnol J 14:1095–1098

Agrawal GK, Sarkar A, Righetti PG, Pedreschi R, Carpentier S et al (2013) A decade of plant proteomics and mass spectrometry: translation of technical advancements to food security and safety issues. Mass Spectrom Rev 32:335–365

Aharoni A, Vorst O (2001) DNA microarrays for functional plant genomics. Plant Mol Biol 48:99–118

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
Bachem CW, Van Der Hoeven RS, De Bruijn SM, Vreugdenhil D, Zabeau M et al (1996) Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. Plant J 9:745–753
Bachem CW, Oomen RJF, Kuyt S, Horvath BM, Claassens MM et al (2000) Antisense suppression of a potato alpha-SNAP homologue leads to alterations in cellular development and assimilate distribution. Plant Mol Biol 43:473–482
Baldi P, Grossi M, Pecchioni N, Vale G, Cattivelli L (1999) High expression level of a gene coding for a chloroplastic amino acid selective channel protein is correlated to cold acclimation in cereals. Plant Mol Biol 41:233–243
Barry CS, Blume B, Bouzayen M, Cooper W, Hamilton AJ et al (1996) Differential expression of the 1-aminocyclopropane-1-carboxylate oxidase gene family of tomato. Plant J 9:525–535
Bino RJ, Hall RD, Fiehn O et al (2004) Potential of metabolomics as a functional genomics tool. Trends Pl Sci 9:418–425
Brosche M, Strid A (1999) Cloning, expression, and molecular characterization of a small pea gene family regulated by low levels of ultraviolet B radiation and other stresses. Plant Physiol 121:479–487
Brygoo H, Joyard J (2004) Introduction – focus on plant proteomics. Plant Physiol Biochem 42:913–917
Canovas FM, Dumas-Gaudot E, Recorbet G et al (2004) Plant proteome analysis. Proteomics 4:285–298
Chen S, Harmon AC (2006) Advances in plant proteomics. Proteomics 6:5504–5516
Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals wide spread pausing and divergent initiation at human promoters. Science 322:1845–1848
Costa V, Gallo MA, Letizia F et al (2010) The molecular basis of PPAR function. PPAR Res 2010:409168. https://doi.org/10.1155/2010/409168
Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D et al (2017) Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci 22:961–975
Czechowski T, Bari RP, Stitt M et al (2004) Real-time RT-PCR profiling of over 1400 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. Plant J 38:366–379
Dixon RA, Gang DR, Charlton AD et al (2006) Applications of metabolomics in agriculture. J Agric Food Chem 54:8984–8994
Durrant WE, Rowland O, Piedras P, Hammond-Kosack KE, Jones JD (2000) cDNA-AFLP reveals a striking overlap in race specific resistance and wound response gene expression profiles. Plant Cell 12:963–977
Emebiri LC, Tan MK, El-Bouhssini M, Wildman O, Jighly A et al (2017) QTL mapping identifies a major locus for resistance in wheat to Sunn pest (Eurygaster integriceps) feeding at the vegetative growth stage. Theor Appl Genet 130:309–318
Fenn JB, Mann M, Meng CK et al (1989) Electrospray ionization for mass spectrometry of large biomolecules. Science 246:64–71
Fiehn O (2002) Metabolomics – the link between genotypes and phenotypes. Plant Mol Biol 48:151–171
Fiehn O, Kopka DP et al (2000) Metabolic profiling for plant functional genomics. Nat Biotechnol 18:1157–1161
Fields S, Song O (1989) A novel genetic system to detect protein-protein interactions. Nature 340:245–246
Filleur S, Daniel-Vedele F (1999) Expression analysis of a high affinity nitrate transporter isolated from *Arabidopsis thaliana* by differential display. Planta 207:461–469

Ganal MW, Durstewitz G, Polley A, Berard A, Buckler ES et al (2011) A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One 6:e28334

Giddings G, Allison G, Brooks D, Carter A (2000) Transgenic plants as factories for pharmaceuticals. Nat Biotechnol 18:1151–1155

Gong CY, Wang T (2013) Proteomic evaluation of genetically modified crops current status and challenges. Front Plant Sci 4:41

Haddeland I, Heinke J, Biemans H, Eisner S, Flörke M et al (2014) Global water resources affected by human interventions and climate change. Proc Natl Acad Sci U S A 111:3251–3256

Hall RD (2006) Plant metabolomics: from holistic hope to hype, to hot topic. New Phytol 169:453–468

Hall R, Beale M, Fiehn O et al (2002) Plant metabolomics—the missing link in functional genomics strategies. Plant Cell 14:1437–1440

Hall RD, Vos CHR, Varhoeven HA, Bino RJ (2005) Metabolomics for the assessment of functional diversity and quality traits in plants. In: Vaidyanathan S, Harrigan GG, Goodacre R (eds) Metabolome analyses: strategies for systems biology. Springer, New York

Harbers M, Carninci P (2005) Tag-based approaches for transcriptome research and genome annotation. Nat Methods 2:495–502

He C, Holme J, Anthony J (2014) SNP genotyping: the KASP assay. Crop Breed: Meth Protocol 1145:75–86

Henzel WJ, Billeci TM, Stults JT et al (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. Proc Natl Acad Sci U S A 90:5011–50115

Hossain Z, Komatsu S (2013) Contribution of proteomic studies towards understanding plant heavy metal stress responses front. Plant Sci 3:310

Imadi SR, Kazi AG, Ahanger MA, Gucel S, Ahmad P (2015) Plant transcriptomics and responses to environmental stress: an overview. J Genet 94:525–537

International Rice Genome Sequencing Project., Sasaki, T. The map-based sequence of the rice genome. Nature 436, 793–800 (2005). https://doi.org/10.1038/nature03895

Jia Y, Jannink JL (2012) Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192:1513–1522

Kara M, Bahr U, Ingendoh AI, Hillenkamp F (1989) Laser description-ionization mass spectrometry of proteins with masses 100,000 to 250,000 Dalton. Angew Chem Int 28:760–761

Kersten B, Bürkle L, Kuhn EJ et al (2002) Large-scale plant proteomics. Plant Mol Biol 48:133–141

Kim C, Lee S, Park H, Bae C, Cheong Y et al (2000) Identification of rice blast fungal elicitor responsive genes by differential display analysis. Mol Plant Microbe Interact 13:470–474

Klose J (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik 26:231–243

Klose J, Kobalz U (1995) Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. Electrophoresis 16:1034–1059

Kuno N, Muramatsu T, Hamazato F, Furuya M (2000) Identification by large scale screening of phytochrome regulated genes in etiolated seedlings of *Arabidopsis* using a fluorescent differential display technique. Plant Physiol 122:15–24

Lapopin L, Gianinazzi-Pearson V, Franken P (1999) Comparative differential RNA display analysis of arbuscular mycorrhiza in *Pisum sativum* wild type and a mutant defective in late stage development. Plant Mol Biol 41:669–677

Liang P, Pardee AB (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257:967–971

Liu M, Lei L, Miao F, Powers C, Zhang X et al (2017) The STENOFOLIA gene from Medicago alters leaf width, flowering time and chlorophyll content in transgenic wheat. Plant Biotechnol J 16:186–196. https://doi.org/10.1111/pbi.12759

Long YM, Chao WS, Ma GJ, Xu SS, Qi LL (2017) An innovative SNP genotyping method adapting to multiple platforms and throughputs. Theor Appl Genet 130:597–607

Maloney V (2004) Plant metabolomics. Biotechnol J 2:92–99

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. Trends Genet 24:133–141

Mcclelland M, Mathieu-Daude F, Welsh J (1995) RNA fingerprinting and differential display using arbitrary primed PCR. Trends Genet 11:242–246

McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K et al (2002) Development and mapping of 2240 new SSR markers for rice (Oryza sativa L.). DNA Res 9:199–207

Memelink J (2005) Tailoring the plant metabolome without a loose stitch. Trends Pl Sci 10:305–307

Mendes P (2002) Emerging bioinformatics for the metabolome. Brief Bioinform 3:34–45

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Money T, Reader S, Qu L, Dunford R, Moore G (1996) AFLP-based mRNA fingerprinting. Nucleic Acids Res 24:2616–2617

Nelson GC, Valin H, Sands RD, Havlík P, Ahammad H et al (2014) Climate change effects on agriculture: economic responses to biophysical shocks. Proc Natl Acad Sci U S A 111:3274–3279

Ni Z, Sun Q, Liu Z, Wu L, Wang X (2000) Identification of a hybrid specific expressed gene encoding novel RNA binding protein in wheat seedling leaves using differential display of mRNA. Mol Gen Genet 263:934–938

Oksman-Caldentey KM, Inz D (2004) Plant cell factories in the post genomic era: new ways to produce designer secondary metabolites. Trends Pl Sci 9:433–440

Ozbabacan SE, Engin HB, Gursoy A, Keskin O (2011) Transient protein protein interactions. Protein Eng Des Sel 24:635–648

Patino WD, Mian OY, Hwang PM (2002) Serial analysis of gene expression. Circ Res 91:565–569

Raamsdonk LM et al (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat Biotechnol 19:45–50

Rajaram S (2005) Role of conventional plant breeding and biotechnology in future wheat production. Turk J Agric For 29:105–111

Ray DK, Mueller ND, West PC, Foley JA (2013) Yield trends are insufficient to double global crop production by 2050. PLoS One 8:e66428

Reid LM, Donell CPO, Downey G (2004) Potential of SPME-GC and chemometrics to detect adulteration of soft fruit purees. J Agric Food Chem 52:421–427

Renaut J, Hausman JF, Wisniewski ME (2006) Proteomics and low temperature studies: bridging the gap between gene expression and metabolism. Physiol Plant 126:97–109

Rhee SY, Dickerson J, Xu D (2006) Bioinformatics and its applications in plant biology. Annu Rev Plant Biol 57:335–360

Ruan Y, Le Ber P, Ng HH, Liu ET (2004) Interrogating the transcriptome. Trends Biotechnol 22:23–30

Ruperao P, Edwards D (2014) Identification of markers from next-generation sequence data plant genotyping. Meth Mole Biol Bioinform 1245:29–47

Saha S, Sparks AB, Rago C (2002) Using the transcriptome to annotate the genome. Nat Biotechnol 20:508–512

Saito K, Kobayashi M, Gong Z, Tanaka Y, Yamazaki M (1999) Direct evidence for anthocyanidin synthase as a 2-oxoglutarate dependent oxygenase: molecular cloning and functional expression of cDNA from some red forma of *Perilla frutescens*. Plant J 17:181–189

Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR et al (1977) Nucleotide sequence of bacteriophage phi X174 DNA. Nature 265:687–695

Semagn K, Babu R, Hearne S, Olsen M (2014) Single nucleotide polymorphism genotyping using Kompetitive allele specific PCR (KASP): overview of the technology and its application in crop improvement. Mol Breed 33:1–14

Shanks JV (2005) Phytochemical engineering: combining chemical reaction engineering with plant science. AICh EJ 51:2–7

Sharma SK, Bolser D, de Boer J, Sønderkær M, Amoros W et al (2013) Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. G3 (Bethesda) 3:2031–2047

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ et al (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPS and pedigree. Theor Appl Genet 95:163–170

Solekdeh GH, Komatsu S (2007) Crop proteomics: aim at sustainable agriculture of tomorrow. Proteomics 7:2976–2996

Somers DJ, Isaac P, Edwards K (2004) A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor Appl Genet 109:1105–1114

Suarez MC, Bernal A, Gutierrez J, Tohme J, Fregene M (2000) Developing expressed sequence tags (ESTs) from polymorphic transcript-derived fragments (TDFs) in cassava (*Manihot esculenta* Crantz). Genome 43:62–67

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Nat Biotechnol 7:257–264

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

Van K, McHale LK (2017) Meta-analyses of QTLs associated with protein and oil contents and compositions in soybean [Glycine max (L.) Merr.] seed. Int J Mol Sci 18:1180

Varshney RK, Graner A, Sorrels ME (2005) Genomics-assisted breeding for crop improvement. Trends Pl Sci 10:621–630

Varshney RK, Marcel TC, Ramsay L, Russell J, Roder MS et al (2007) A high density barley microsatellite consensus map with 775 SSR loci. Theor Appl Genet 114:1091–1103

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522–530

Varshney RK, Song C, Saxena RK, Azam S, Yu S et al (2013) Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. Nat Biotechnol 31:240–248

Vassilev D, Leunissen J, Atanassov A et al (2005) Applications of bioinformatics in plant breeding. Australas Biotechnol 19:139–152

Velculescu VE, Vogelstein B, Kinzler KW (2000) Analysing uncharted transcriptomes with SAGE. Trends Genet 16:423–425

Vos P, Hogers R, Bleeker M, Reijans M, Van De Lee T et al (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Vukmirovic OG, Tilghman S (2000) Exploring the genome space. Nature 405:820–822

Wang K, Peng X, Ji Y, Yang P, Zhu Y et al (2013) Gene, protein and network of male sterility in rice. Front Plant Sci 92:1–10

Washburn MP, Wolters D, Yates JR (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nature Biotech 19:242–247

Weckwerth W (2003) Metabolomics in systems biology. Annu Rev Plant Biol 54:669–689

Whitelegge JP (2002) Plant proteomics: blasting out of a mud PIT. PNAS 99:11564–11566

Wilkins Marc R, Jean-Charles S, Gooley Andrew A, Appel Ron D, Ian H-S 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 Rev 13:19–50

Yamazaki M, Gong Z, Fukuchi-Mizatani M, Fukui Y, Tanaka Y et al (1999) Molecular cloning and biochemical characterization of a novel anthocyanin 5-O-glucosyltransferase by mRNA differential display for plant forms regarding anthocyanin. J Biol Chem 274:7405–7411

Yan J, Warburton M, Crouch J (2011) Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. Crop Sci 51:433–449

Yu H, Goh C (2000) Identification and characterization of three orchid MADS-box genes of the AP1/AGL9 subfamily during floral transition. Plant Physiol 123:458–464

Yung M, Schaffer R, Putterill J (1999) Identification of genes expressed during early Arabidopsis carpel development by mRNA differential display: characterisation of ATCEL2, a novel endo-1,4-beta-D-glucanase gene. Plant J 17:203–208

Zhang X, Huang C, Wu D, Qiao F, Li W et al (2017) High-throughput phenotyping and QTL mapping reveals the genetic architecture of maize plant growth. Plant Physiol 173:1554–1564

Zhao Q, Chen S, Dai S (2013) C4 photosynthetic machinery: insights from maize chloroplast proteomics. Front Plant Sci 4:85