MOLECULAR MECHANISMS IN PLANT ADAPTATION

Edited by Roosa A.E. Laitinen

WILEY Blackwell

Preface

This volume brings together recent findings about mechanistic understanding in diverse areas of plant adaptation. It offers the readership novel insights into contemporary results concerning the evolution, development, and stress responses. Moreover, it uniquely combines the technological and methodological perspectives. This volume aims not only to review the published results but also to introduce new concepts, discuss novel findings, and offer original views on the perspectives and challenges in this field. Therefore, it provides balanced coverage of materials suitable for both experts and newcomers to this area.

This book is organized into nine chapters. The first chapter summarizes recent advances in high-throughput technologies that are necessary to delve deeper into the molecular mechanisms of plant adaptation. Further, the use of natural variation in *Arabidopsis thaliana* in the studies of local adaptation and evolution is reviewed. The third chapter provides a specific example of how natural variation can be used in combination with candidate gene approaches to understand the mechanisms of seed dormancy and viability. The next three chapters provide unique views on mechanistic understanding of plant responses to abiotic and biotic stresses. In Chapter 4, the abiotic stress response in plants is examined from a single-cell point of view. In Chapter 5, the metabolic responses to biotic stress in plants are introduced. In Chapter 6, the latest developments in the role of small RNAs in both biotic and abiotic stress responses are presented. The next chapter deals with the evolutionary perspective in understanding adaptation. It tackles the adaptation of

flower form, with a special focus on an evo-devo approach, revealing the evolutionary history of the *SEPALLATA 3* gene. The challenge of employing the data from high-throughput technologies in understanding the mechanisms of plant adaptation may be addressed by mathematical modeling. To this end, the determination of adaptive patterns and the predictions on plant behavior are presented in Chapter 8. The final chapter highlights the importance of combining laboratory work with field experiments and is indented to help the reader formulate the guidelines on how field experiments should be performed and what factors should be taken into account. While each chapter can stand on its own, I hope that the readers will find the entire volume interesting and offering them comprehensive understanding of the current hot topics in plant adaptation.

I would like to thank all the authors for making this project thoroughly interesting and enjoyable. This book would not have been possible without their dedicated and smooth cooperation. I also thank Justin Jeffryes and Stephanie Dollan at Wiley for their support. Additionally, I appreciate the valuable comments and help by Lisa Smith, Zoran Nikoloski, Christian Schudoma, Björn Plötner, Aditya Sharma, Sebastian Proost, Alisdair Fernie, Hirofumi Ishihara, Eunyoung Chae, Vishal Kapoor, Prashant Pandey, and Matti Laitinen.

Roosa A. E. Laitinen

1
Technological Advances in Studies of Plant Adaptation

José G. Vallarino[1] and Sonia Osorio[1,2]
[1]*Department of Molecular Biology and Biochemistry, Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora" – University of Malaga-Consejo Superior de Investigaciones Científicas (IHSM-UMA-CSIC), Málaga, Spain*
[2]*Max-Planck-Institut für Molekulare Pflanzenphysiologie, Potsdam-Golm, Germany*

Introduction

In order to survive, organisms must adapt to their environment. For plants, adaptation is particularly important because they are sessile and therefore cannot move away from unfavorable conditions. Therefore, plants have evolved strategies, both on short and long time scales, to adjust their growth and development to extreme environmental conditions. One interest lies in understanding the underlying mechanisms of plant adaptation to reveal evolutionary signatures that may help to understand the selective forces of allele frequencies. The other objective is to gather information to enable us to develop crop plants that can resist the changes in environment without losing their productivity. Recent advances in high-throughput technologies have generated a wealth of data that offer new opportunities and challenges in revealing the mechanisms of plant adaptation. By combining the different high-throughput methods, it is possible to gain knowledge of the complex interactions between genotype and phenotype. The large amounts of data have also introduced new challenges in assembling, analyzing, and discovering patterns (Cronn *et al*. 2012; Kvam *et al*. 2012; Higashi and Saito 2013; Toubiana *et al*. 2013). The aim of this chapter is to bring together the recent technological advances in the studies of plant genomes, proteomes, and metabolomes and their general applications in understanding plant adaptation. In addition, we pay special attention to the way high-throughput technologies are helping us to understand nonmodel species, which would facilitate improvements in our understanding of plant breeding.

Next-Generation Sequencing Technologies

Next-generation sequencing (NGS) technologies are capable of producing billions of short nucleotide reads (50–800 bp) in parallel, at a fraction of the cost of traditional Sanger sequencing. Solexa sequencing was the first NGS technique to become commercially available, in 2005. Since the first plant genome (*Arabidopsis thaliana*; The Arabidopisis Genome Initiative, 2000) was sequenced in the year 2000, the whole-genome sequencing technologies have improved and 60 plant genomes have been sequenced till date. The reduced costs in whole-genome sequencing have led to the development of bigger sequencing projects such as the *Arabidopsis* 1001 Genomes Project (www.1001genomes.org and signal.salk.edu/atg1001) and OMAP, The Oryza Map Alignment Project, and various other large-scale projects. The Arabidopsis 1001 genomes was initiated with the goal of identifying the total genetic variation present within this species, whereas the OMAP project aims to sequence all species from the *Oryza* genus. Some of the other large-scale projects include the "1000 Plant Genomes Project" (www.onekp.org), and the "1000 Plant and Animal Genome Project" (www.1d1.genomics.cn. At present, six NGS platforms are available and a seventh one is in advanced development stage (Liu *et al*. 2012; Table 1.1). The different sequencing methods can be grouped into three main types: (i) sequencing by synthesis, (ii) sequencing by ligation, and (iii) single-molecule sequencing. The different methods are presented in more detail later.

Table 1.1 First-, second-, and third-generation DNA sequencing platforms listed in the order of commercial availability.

Platform	Current company	Sequencing method	Amplification method	Read length (bp)	Advantage	Dis
Sanger		Chain terminator	PCR	600–900	High quality, long read length	Hig thr
454 GS FLX	Roche	Synthesis (pyrosequencing)	Emulsion PCR	600–800	First next-gen sequencer, long reads	Err wit mo hig thr
Solexa GAIIx	Illumina	Synthesis	Solid-phase PCR	150	First short-read sequencer, high-throughput, current leader in advantages	Sho ass
SOLiD4	Life Technologies	Ligation	Emulsion PCR	50	Second short-read sequencer, low error rates	Sho ass
HeliScope/Helicos	Life Technologies	Synthesis	Single molecule template	25–35	Single-molecule sequencer	Sho len err
Ion Torrent	Life Technologies	Synthesis	Emulsion PCR	100–400	First post-light sequencer, first system < €100,000, high throughput	Lov thr hig bas pro hor err bet per tha
PacBio/Pacific Bioscience	Pacific Bioscience	Synthesis	Real-time single-molecule template	964 average	Real-time single-molecule sequencing, long reads, high throughput	Hig rat yie

Platform	Current company	Sequencing method	Amplification method	Read length (bp)	Advantage	Dis
Starlight[*]	Life Technologies	Synthesis	Single-molecule sequencing with quantum dots		Single-molecule sequencing with quantum dots	

Transcriptome characterization.

Targeted resequencing.

De novo BACs, plastids, microbial genomes.

Mutation detection.

De novo plant genomes.

Metagenomics.

Resequencing and transcript counting.

Others such as ChIP-Seq, μRNA-seq, Methyl-Seq, and so on.

[*]A commercial launch date for the Starlight system is not yet known but some information about its performance characteristics is known.

Sequencing by Synthesis

Sequencing by synthesis is based on determining base composition through the detection of chemiluminescence released by the incorporation of nucleotides during the synthesis of the complementary DNA strand by a DNA polymerase (Nyren *et al*. 1993). In this method, the DNA is fragmented to the appropriate size, ligated to adapter sequences, and then clonally amplified to enhance the fluorescent chemical signal. Templates are then separated and immobilized in preparation for flow cell cycles. The sequencing by synthesis method is employed by three different sequencing platforms. In Roche 454 pyrosequencing (http://www.my454.com) a single-primed DNA template is adhered to a microbead and amplified using emulsion PCR. In the beginning, this technology produced read lengths of approximately 100 bp. Nowadays, the read lengths obtained from pyrosequencing are comparable to the ones produced by Sanger sequencing (approximately 800 bp). Because of its long reads, this platform is often used for generating reads for *de novo* genome or transcriptome assembly (Strickler *et al*. 2012; Zalapa *et al*. 2012). The grape genome (published in 2007) was the first genome sequenced, based on a combination of 454 and Sanger sequencing and since then at least 18 genomes have been sequenced (Jaillon *et al*. 2007). Among them are the genomes of apple (Velasco *et al*. 2010), cocoa (Argout *et al*. 2011), potato (The Potato genome Sequencing Consortium 2011), banana (D'Hont *et al*. 2012), cotton (Li *et al*. 2014), wheat (*Aegilops tauschii*) (Jia *et al*. 2013), and bladderwort (Ibarra-laclette *et al*. 2013).

The second platform based on the sequencing by synthesis method is the Illumina Genome Analyzer (http://www.illumina.com), initially developed by Solexa. It uses solid-phase bridge amplification in which 5′- and 3′-adapters are ligated to each end of the DNA template. This method is currently the most widely used NGS platform in plant sciences (Kane *et al*. 2012; Strickler *et al*. 2012; Zalapa *et al*. 2012) because it yields the highest throughput with one of the highest raw accuracy. The Illumina HiSeq sequencer was launched in 2011. This platform is currently able to sequence up to 540–560 Gbp in a single two-flow cell in an 8.5-day run (http://illumina.com/systems/hiseq_2000.ilmn). Illumina has been used to sequence plant genomes such as cotton (Li *et al*. 2014), chickpea (Varshney *et al*. 2013), ancient lotus (Ming *et al*. 2013), pear (Wu *et al*. 2013), and watermelon (Guo *et al*. 2013).

The third platform using sequencing by synthesis is Ion Torrent by Life Technologies (http://www.iontorrent.com). This platform is the only NGS technology that is not based on fluorescent dyes but rather measures the pH change as the result of the release of a H^+ ion upon nucleotide incorporation, using the semiconductor technology (Rothberg *et al*. 2011). By sequentially adding nucleotides, the sequencer is able to detect which nucleotide has been incorporated into the elongating strand (Howden *et al*. 2011; Rothberg *et al*. 2011). Life Technologies currently offers two series of NGS instruments: the large-scale 5500 series, which can yield up to >20 Gbp per day (75bp reads), and the small-scale Ion Torrent series yielding up to 10 Gbp per run in less than a day. The Ion Torrent series (PGM and Ion Proton) are smaller instruments that use a semiconductor chip technology to capture the signal emission after incorporation of a single base to the elongating strand of DNA. The Ion Torrent has the lowest throughput but also the fastest turnaround times of all commercially available NGS systems. It can yield several hundred-thousand reads in less than 2 h. Publications on research that have utilized the Ion Torrent platform currently focus on the sequencing of microbial genomes (Howden *et al*. 2011; Rothberg *et al*. 2011), but this platform has clearly made its way into programs pursuing plant-based objectives. One example is the recent study by Mascher *et al*. (2013) in which they compare Ion Torrent and Illumina HiSeq 2000 platforms by sequencing a barley recombinant line population.

Sequencing by Ligation

Sequencing by ligation methods harnesses the mismatch sensitivity of DNA ligase to determine the sequence of nucleotides in a given DNA strand (Landegren *et al*. 1988). These methods use oligonucleotide probes of varying lengths, which are labeled with fluorescent tags. Methods based on sequencing by ligation usually differ in their probe usage and read length. The SOLiD platform (http://www.appliedbiosystems.com) utilizes sequencing by ligation method to determine the sequence composition of DNA. These methods are often used in resequencing studies (Ashelford *et al*. 2011), transcriptomics, or in genomic sequencing. So far, only two genomes have been sequenced using SOLiD sequencing: strawberry (Shulaev *et al*. 2011) and tomato (The tomato Genome Consortium 2012).

Single-Molecule Sequencing

Single-molecule sequencing technology is also known as the “third-generation sequencing” technology. This technology is based on a detectable signal produced by nucleotide incorporation via chemiluminescence from a single nucleic acid molecule, thus eliminating the need for DNA template amplification. This method has been used for direct RNA sequencing, thus removing the biases created during cDNA amplification in RNA-seq studies (Ozsolak *et al*. 2009). Single-molecule sequencing has some benefits over the other NGS methods, one of them being simplified sample preparation that can use degraded or low concentrations of starting material (Orlando *et al*. 2011) and the avoidance of PCR errors and biases introduced during template amplification. As mentioned before, NGS technologies are evolving at a very rapid pace. Companies are constantly improving the performance of the technology used. These emerging technological developments may herald the fourth generation of NGS techniques. Several optical sequencing technologies are being explored that enable long DNA strands to be read and sequenced with greater efficiency. Other research is being done on nanopores as a means of reading DNA sequences based solely on the inherent electronic or chemical properties of the native nucleotides (Thompson and Milos 2011; Maitra *et al*. 2012). The different nanopore sequencing strategies that are in development enable individual base detection, based on the measurement of conductivity changes across a lipid membrane while a DNA fragment is pulled through a nanoscale pore by an electric current. Although nanopore sequencing faces several challenges, it seems to have a promising future.

Applications of Next-Generation Sequencing

NGS enables progress in studying the genetics of plant adaptation beyond what is possible with current genetic methods. Most sequencing applications can be divided into two categories: (i) *de novo* sequencing and (ii) resequencing. For *de novo* sequencing, reads are obtained from an unknown sequence and assembled to reconstruct the sequence, whereas in resequencing the unknown sequences are compared to a known reference sequence. *De novo* applications are usually slower and more computationally intensive than resequencing. Major resequencing applications include polymorphism discovery, transcriptome profiling, and epigenome analysis.

Polymorphism Detection, Genome-Wide Association Studies (GWAS), and Gene Identification

The analysis of genomic variation is an essential part of studying plant adaptation. Studies that search for a statistical association between a phenotype and a particular locus or loci by screening across the entire genome are called genome-wide association studies (GWASs). During the past decades, the use of genotyping has enabled the characterization and mapping of genes and metabolic pathways in plants, as well as the study of the genetic variation and evolutionary history, marker-assisted selection (MAS), and germplasm characterization. Single nucleotide polymorphism (SNP) markers are the most widely used genotyping markers due to their abundance in the genome and the relative ease in determining their frequency in a collection of individuals. The development of markers as well as their scoring across populations traditionally has been a high-cost process, with many labor-intensive and time-consuming steps. With the help of NGS technologies, several methods have been developed for high-throughput genetic marker discovery. All the methods involve (i) the digestion of multiple samples of genomic DNA with one or more restriction enzymes, (ii) the selection of the resulting restriction fragments, and (iii) NGS of the final set of fragments, which should be less than 1 kb in size. Polymorphisms in the resulting sequenced fragments can be used as genetic markers. All these methods can be classified into three different categories: (i) reduced representation sequencing (reduced representation libraries, RRLs) and complexity reduction of polymorphic sequences (CRoPS), (ii) restriction-site-associated (RAD-seq), and (iii) low-coverage genotyping, including multiplexed shotgun genotyping (MSG), and genotyping by sequencing (GBS).

RRLs and CRoPS are the two methods for sampling and sequencing a small set of genome-wide regions without sequencing the entire genome. The RRL approach (adapted to NGS) has been used to generate tens of thousands of candidate SNPs, for example, in maize (Gore *et al*. 2009) and soybean (Hyten *et al*. 2010). RAD-seq is a method that uses Illumina NGS for genotyping. The RAD-seq approach involves a genome-wide survey of nucleotide diversity of regions flanking restriction sites and allows the simultaneous detection and genotyping of thousands of genome-wide SNPs (Wagner *et al*. 2013). For example, RAD-seq has been used to construct linkage maps in barley (Chutimanitsakun *et al*. 2011) and ryegrass (Pfender *et al*. (2011). The high cost of multiplexing prevented the genotyping of population or pooled samples for initial iterations of the method, but emerging pipelines, such as double digest RAD-seq (ddRAD-seq), have allowed cheaper polymorphism discovery and genotyping for large samples (Peterson *et al*. 2012). Currently, ddRAD-seq offers the possibility to obtain genomic data necessary for inferences about population structure, especially when its consequences are not extreme (such as local adaptation). An advantage is that RADseq can be applied not only to species with available reference genomes, but also to study those species in which no reference genome is available.

Large GWASs require hundreds of thousands of markers to generate sufficient information and coverage, and getting such a resolution has been greatly facilitated by the emergence of NGS technologies. Recently, NGS technologies have been used to resequence recombinant inbred lines (RILs) in many plant species. A collection of 5000 maize RILs have been resequenced and a total of 1.4 million SNPs and 0.2 million indels (large insertions and deletions) were generated, which span the 5000 inbred lines (Gore *et al*. 2009). Seeds of the RILs can be used to grow and phenotype plants for any trait of interest (McMullen *et al*. 2009). Using this population, Buckler *et al*. (2009) identified 50 loci that contribute to variation in the genetic architecture of flowering time, with many loci showing small effects determining leaf architecture. Poland *et al*. (2011) identified candidate genes for resistance to northern leaf blight in 29 loci, which included quantitative trait loci (QTL) with small additive effects. In yet another study, the resequencing of 150 RILs derived from a cross between Indica and Japonica rice cultivars resulted in the discovery of 1,226,791 SNPs, separated by 40 kbp on average (Wang *et al*. 2011). Haplotypes and recombination breakpoints could be determined for each RIL, using the parental origins of SNPs in discrete regions of the genome and a recombination map of 2334 bins for the 150 RILs was constructed from the haplotypes. Using each bin as a genetic marker, various phenotypes were mapped to 49 QTLs, including five QTLs physically located at positions overlapping with known candidate genes (Wang *et al*. 2011). Another example of how NGS greatly enhances the ability to find associations between phenotypes and the underlying genetic variation is a study performed on lodgepole pines (*Pinus contorta*). Data obtained for more than 95,000 SNPs across 98 lodgepole pines could be used to identify 11 loci associated with degree of serotiny (Parchman *et al*. 2012). These results provided a first genome-wide association map of an adaptive trait (serotiny) in pines.

Transcriptome Analysis

RNA-seq is a rapidly growing application of NGS to study gene expression (transcriptomics). Short-read NGS technologies such as Illumina and SOLiD have allowed the development of transcription profiling strategies that are more sensitive and accurate than other high-throughput technologies such as microarrays. In RNA-seq, total or messenger RNA is fragmented and converted into complementary DNA (cDNA). Alternatively, RNA can first be converted into cDNA and then fragmented. Adapters are attached to one or both ends, and reads are sequenced as single or paired ends (Wang *et al*. 2009a; Marguerat and Bahler 2010). Depending on the genomic resources available for the organism of interest, the resulting sequences can be aligned to either a reference genome (or reference transcripts) or the genome can be assembled *de novo*. Therefore, RNA-seq is practical for nonmodel species as it provides information on the transcriptome, including gene structure, expression levels, presence of multiple isoforms, and sequence polymorphisms. RNA-seq can also be used for transcriptome characterization, SNP detection, and comparative gene expression (Strickler *et al*. 2012). Large-scale transcriptomic profiling can provide important insights, for example, into the response of individuals to climatic changes predicted due to global warming. Realistic heat wave conditions were applied in a common stress garden to southern and northern populations of the seagrass *Zostera marina* (Franssen *et al*. (2011). These results suggested that transcriptomic patterns could be used to predict how populations get adapted to thermal stress.

Interaction Studies

Study of transcription factors (TFs) and other chromatin-associated proteins are essential in elucidating the complex phenotype-influencing mechanisms. Determining how proteins interact with DNA to regulate gene expression is essential to fully understand the processes of plant adaptation. Traditional methods have successfully identified TF-binding

sites and specific DNA-associated protein modifications and their roles in regulating genes, but these experiments are limited in scale and resolution. The powerful Illumina whole-genome chromatin IP sequencing (ChIP-Seq) application provides a snapshot of a single protein's direct physical interactions with DNA at a particular time in a particular tissue on a genome-wide scale (Mardis 2007; Kaufmann *et al*. 2010).

Specific DNA sites in direct physical interaction with TFs can be isolated by chromatin immunoprecipitation (ChIP). ChIP produces a library of target DNA sites that a given factor was bound to, *in vivo*. NGS technology allows the determination of the sequences of ChIP-isolated DNA fragments for identification and quantification of the sites bound by a protein of interest. The big advantage of NGS technology is that a single sequencing run can scan for the protein–DNA interactions on a genome-wide basis with high resolution. For example, ChIP-Seq has been applied to elucidate the role of the MADS domain protein FLOWERING LOCUS C (FLC) as a floral repressor (Michaels and Amasino 1999; Deng *et al*. 2011). FLC was found to bind to 505 specific sites in the *A. thaliana* genome, but binding was nonrandom. As expected, FLC was found to mainly bind to promoter regions with $CC(A/T)_6GG$ motifs (Deng *et al*. 2011). In another example, ChIP-Seq was used to elucidate the binding sites for RIN, one of the main ripening TFs in tomato (Zhong *et al*. 2013). These results provided some insight into the systems regulation underlying fruit ripening, showing that the epigenome is not static during fruit development (Zhong *et al*. 2013).

Methylome Analysis and Small RNA Characterization

Traditionally, the material that is carried from one generation to the next and is responsible for the phenotypic variation is associated with genes and DNA. However, there exists the phenotypic variation that cannot be explained by differences in DNA sequences but by changes in gene expression patterns that influence the phenotype. These are called epigenetic mechanisms. Yet, it is still largely unknown how the interplay between the epigenetic modifications and genes could influence adaptation and evolution. Therefore, it is essential to identify not only genetic variation but also natural epigenetic variation. The known epigenetic mechanisms include DNA methylation, histone modification, and RNA-directed DNA methylation (Bird 2007; Becker and Weigel 2012; Schmitz and Erker 2012). Functional genomics aims to interrogate the functional elements and regulatory mechanisms in the genome, including DNA methylation and histone modifications. One important consideration is that both the epigenome and methylome are larger than the genome of an organism. As a major part of the epigenome, the methylome consists of the sum of genome and methylation states at every cytosine location. NGS can be integrated into epigenomic studies and several new and innovative sequencing-based methods have been developed together with bioinformatics and analytical tools (Horner *et al*. 2010; Huss 2010). Before NGS, epigenetic studies were mostly limited to individual genes or sets of candidate genes or regions. One exception is the work done in Arabidopsis by Zhang *et al*. (2007), which provided the first genome-wide study in plants and considerable information on methylation distribution and its effect on gene expression. The use of NGS technologies coupled with bisulfite conversion (BC), restriction digestion, or immunoprecipitation strategies facilitate genome-wide methylome analysis in plants and play an important role in further characterization of epiregulation in plants (Zhang and Jeltsch 2010). DNA methylomes in 10 *A. thaliana* lines, derived 30 generations ago from a common ancestor (Shaw *et al*. 2000), captured the formation of pure epialleles, some of which resulted in significant transcriptional variations (approximately 10- to 1000-fold changes) of the affected locus (Schmitz *et al*. 2011). Another study in the same population showed that the number of epimutations does not increase with time, indicating that many are not stably inherited over the long term (Becker *et al*. 2011). Also, they found that transposon methylation was highly conserved. They concluded that the biased distribution and frequent reversion of

epimutations determine the ability of any type of allele to be subject to Darwinian selection (Becker *et al*. 2011).

In addition to the analysis of the methylome, NGS technologies have been used to identify small RNAs. Small RNAs are short nonprotein coding RNA molecules ranging from 20 to 30 nucleotides that have a role in development, genome maintenance, and responses to environmental stresses (Simon *et al*. 2009). The role of small RNAs in plant adaptation is reviewed in detail in Chapter 7 of this book. Most small RNAs belong to two groups, microRNAs (miRNAs) and small interfering RNAs (siRNAs). miRNAs are about 21 bases long and usually play a post-transcriptional regulatory role by directing cleavage of a specific transcript. siRNAs are normally 24 bases long and influence the post-transcriptional gene silencing (Vaucheret 2006). Small noncoding RNAs in the low-molecular-weight total RNA fraction of plants were detected before the development of high-throughput sequencing techniques (Gupta *et al*. 1989). Subsequently, Illumina, SOLiD, and Roche 454 platforms manifested optimal features for short-read sequencing and small RNA detection (Zhang *et al*. 2009; Gonzalez-Ibeas *et al*. 2011). Thus, construction and sequencing of small RNA libraries, coupled with bioinformatical analysis for miRNA prediction, is currently the most powerful experimental method for miRNA identification (Kurtoglu *et al*. 2014). For example, miRNAs were discovered using this method in wheat grown under normal conditions (Kenan-Eichler *et al*. 2011; Li *et al*. 2013; Meng *et al*. 2013; Sun *et al*. 2013) as well as under biotic stress caused by pathogens, powdery mildew (Xin *et al*. 2010), and under abiotic stress due to extreme heat (Xin *et al*. 2010; Yao *et al*. 2010), cold (Yao *et al*. 2010; Tang *et al*. 2012), salinity (Yao *et al*. 2010), or dehydration (Yao *et al*. 2010). In some of these reports, wheat miRNAs were discovered via sequencing on an Illumina platform (Xin *et al*. 2010; Tang *et al*. 2012; Meng *et al*. 2013; Sun *et al*. 2013) while in others a Roche 454 platform was used (Yao *et al*. 2010; Li *et al*. 2013).

Metagenomics

Metagenomics is the study of the plant-associated microbiota based in genomic analysis. These studies provide insight into the composition and physiological potential of plant-associated microorganisms. In metagenomics, NGS technologies are used to identify organismal communities from small amounts of DNA. These could be used to characterize biogeographical patterns of diversity and functional capabilities of soil microbes in intact and reconstructed soils (Harris 2009; Fierer *et al*. 2013). As an example, Ruzicka *et al*. (2013) used high-throughput sequencing to characterize the transcriptomes of both tomato and its arbuscular mycorrhizal fungal symbiont in the field. Instead of culturing the symbiont, a metagenomic sequencing strategy was employed, where RNA from a wild-type tomato plant and a mutant for reduced mycorrhizal colonization were sequenced and separated using bioinformatics (Ruzicka *et al*. 2013). This metagenomic analysis revealed a cluster of genes for transport and cell wall remodeling, which is required for the symbiotic relationship. Metagenomic sequencing opens up the opportunity to explore additional symbiotic relationships and further functionally characterize aspects of a genome that are not innate to the genomic sequence.

Proteome Analysis in Understanding Plant Adaptation

The proteome is defined as the total set of proteins or gene products present in a biological unit. The proteomics approach aims to know “how,” “where,” and “when,” the several thousands of individual proteins are produced in a cell. Advances in proteomics have been possible due to continuous improvement in the methods of protein extraction, purification, and separation, as well as improvements in equipment, and protein identification, quantification, and characterization. With the combination of proteomics and NGS technologies, identification and annotation of proteins and their isoforms are

becoming much more straightforward. Proteogenomics can be referred to as a field largely based on improvement in genome annotation using proteomics information based on MS (Tejedor-Lorenzo *et al.* 2009). Table 1.2 summarizes the representative databases and web sites related to plant proteomics.

Table 1.2 Current plant proteomics and metabolomics databases.

Plant proteome databases and web sites		
AT-CHLORO	Database developed for the chloroplast proteome from *A. thaliana* using LC-MS/MS analysis	www.grenoble.prabi.fr/at_chloro
AtPID	Database that contains predicted protein–protein interaction for *A. thaliana*	www.megabionet.org/atpid/webfile
TAIR	A comprehensive genome database for *A. thaliana*. Regarding proteomics, it includes (i) primary protein sequences, (ii) protein domains, (iii) protein structures including, (iv) protein–protein interactions, (v) biochemical properties including enzymes and biochemical pathways	www.arabidopsis.org
DIPOS	DIPOS provide information on interacting protein in rice, where the interactions are predicted using two computational methods, interologs and	csb.shu.edu.cn/dipos

	domain-based methods	
Rice proteome database	This database contains proteome data from rice based on 2-DE techniques	gene64.dna.affrc.go.jp/RPD
GabiPD	This database provides integrated plant “Omics” data. Data from different “Omics” are integrated and interactively connected. 2D electrophoresis gel images were collected from different tissues of *A. thaliana* and *Brassica napus*	www.gabipd.org
Seed Proteome Web Portal	This database provides information both on quantitative seed proteomic data and on seed-related protocols	www.seed-proteome.com
Plant metabolome databases and web sites		
AraCyc	A tool to visualize biochemical pathways of *Arabidopsis*	www.arabidopsis.org/tools/aracyc/
ArMet	A framework for the description of plant metabolomics experiments and their results	www.armet.org
MetaCyc	A metabolic pathway database that	metacyc.org

	contains pathways from >150 different organisms. MetaCyc describes metabolic pathways, reactions, enzymes and substrate compounds	
MapMan	U A user-driven tool that displays large datasets onto diagrams of metabolic pathways. It is composed of multiple modules for hierarchical grouping of transcript and metabolite data that can be visualized using a separate user-guided module	Gabi.rzpd.de/projects/MapMan
MetNet	Contains a suite of open-source software tools for systems biology and is designed to provide a framework for the formulation of testable hypothesis regarding the function of specific genes	www.public.iastate.edu/≈mash/MetNet/homepage.html
METLIN	Metabolite database for metabolomics containing over 64,000 structures. MS/MS Data	http://metlin.scripps.edu/index.php

	was obtained on a 6510 Q-TOF(Agilent Technologies)	
MassBank	High-quality mass spectral database	http://www.massbank.jp/en/database.html
KNApSAck	A Comprehensive Species-Metabolite Relationship Database	http://kanaya.naist.jp/KNApSAcK/
GMD@CSB.DB-The Golm Metabolome Database	Public access to custom mass spectra libraries, metabolite profiling experiments and other necessary information related to the field of metabolomics.	http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html
Metabolomics Science	Scripps Center for Metabolomics and Mass Spectrometry. Compilations of the most inclusive and freely accessible databases that currently exist	http://masspec.scripps.edu/metabo_science/metadbase.ph

A protein can only be identified and quantified when it is extracted, solubilized, and visualized. For this reason, attention must be paid to procedures for protein extraction, especially in the case of plant recalcitrant tissue. There is no single protein extraction protocol that can be applied to the full proteome. Therefore, the protocol must be optimized for each tissue, keeping in mind the particular research objective. Frequently used protocols involve a precipitation step, which should separate proteins from interfering compounds. There are three protocols that are more commonly used in proteomics: (i) The TCA method (Sarry *et al*. 2004) in which the sample is homogenized and then extracted with trichloroacetic acid (TCA), (ii) the acetone method (Saravanan and Rose 2004) in which acetone is used for protein precipitation, and (iii) the phenol method (Chan *et al*. 2008) that uses an equal volume of Tris-HCl pH 7.8 buffered phenol and methanol for final precipitation. Different reports comparing the three protocols in different plant tissues agreed that the phenol method results in a higher protein yield and greater resolution (Saravanan and Rose; 2004; Wang *et al*. 2009b; Chan 2013).

Proteomics Technologies

Proteomics can be a holistic or targeted approach. In terms of methodology, proteomics research was originally based on two-dimensional electrophoresis (2-DE) protein separation, coupled with mass spectrometry (MS) analysis of spots (first generation) moving then to liquid chromatography (LC) based shotgun strategies (second generation), and later to quantification approaches including label and label-free variants (third generation).

MS analysis can be performed with either total protein extracts (called top-down proteomics) or peptides (called bottom-up proteomics) resulting from protease treatment (usually trypsin). Alternatively, MS analysis can be performed after the protein separation step, by either one- or two-dimensional gel electrophoresis (gel-based approach) or by LC-based techniques (gel-free analysis). In terms of proteome coverage and number of identified proteins, the combination of 1-DE and nanoLC-MS has proved to be the most powerful technique (Irar *et al*. 2010; Gammulla *et al*. 2011). For MS-based quantification, two approaches are usually followed, the label (DIGE, ICAT, iTRAQ, and SILAC) and label-free (peak area or spectral counting) approaches. Recently, the absolute quantification of proteins based on stable isotope-labeled integral standard peptides and LC coupled to selective reaction monitoring (SRM) triple quadrupole-MS (QqQ-MS) has been successfully applied to highly complex proteomics samples (Wienkoop and Weckwerth 2006). Finally, the techniques used in proteomics can be grouped according to the MS equipment, resulting in different combinations of ionizer and analyzer, with the MALDI source usually coupled to a time-of-flight (TOF) analyzer and electrospray ionization (ESI) to quadrupoles, although lately a number of hybrid techniques have been commercialized.

Applications of Proteomics

The proteomics approach can be used to understand plant response to different environmental conditions (Kawamura and Uemura 2003; Yamazaki *et al*. 2008; Komatsu *et al*. 2009; Ahsan *et al*. 2012). Stress-induced changes in gene expression modulate metabolic processes through alteration of cellular protein abundance and function. Therefore, understanding the change in the function of proteins under stressed conditions is crucial for clarifying the molecular mechanisms underlying stress tolerance (Wang *et al*. 2013). In agriculture, the combination of proteomics with other MAS approaches is especially useful in breeding programs. Damerval *et al*. (1994) used an approach that brought proteomics and MAS together. They identified protein quantity loci (PQL) that explained some of the spot intensity variation. Of the 72 proteins analyzed, 70 PQLs were identified for 42 proteins, 20 of which had more than one PQL. Therefore, plant breeders should be able to integrate the selected genes in marker-assisted breeding programs to improve the trait under study (Salekdeh and Komatsu 2007). Several studies on the soybean proteome have revealed that some common proteins involved in redox systems, carbon metabolism, photosynthesis, signaling, and amino acid metabolism are associated with various stress responses (Zhen *et al*. 2007; Aghaei *et al*. 2009; Yamaguchi *et al*. 2010; Qin *et al*. 2013). These candidate proteins can be used for the functional annotation of genes present in QTL regions or found to be differentially expressed under stress conditions.

A specific advantage of proteomics is the ability to reveal post-translational modifications, which is very important to determine the function of the protein. Till date, about 300 post-translational modifications have been identified through proteomic analyses (Grimsrud *et al*. 2010; Bindschedler and Cramer 2011; Bond *et al*. 2011).

A number of proteomic studies have identified protein phosphorylation events involved in plant–microbe interactions. Gerber *et al*. (2006, 2008) used MALDI-TOF-TOF MS/MS to identify differentially phosphorylated *Nicotiana tabacum* proteins upon elicitation with

lipopolysaccharides from *Burkholderia cepacia*. Trapphoff *et al*. (2009) combined immunoblotting and MS to identify phosphorylated proteins in defense-related pathways in *Medicago truncatula* cell suspension cultures after inoculation with *Aphanomyces euteiches*. A recent work used 2-DE to separate ^{33}P-labeled samples of *Lotus japonicas* roots treated with symbiotic and pathogen-elicited (flg22) signaling molecules, revealing phosphorylation of both unique and shared proteins during symbiotic and defense responses (Serna-Sanz *et al*. 2011).

Metabolome Analysis in Plant Adaptation

Plant species are estimated to produce 200,000 different metabolites that can be divided into primary and secondary metabolites (Pichersky and Gang 2000). Metabolomics was first established as a powerful screening approach in toxicology (Nicholson *et al*. 1999), but has found wide application after the development of high-throughput analytical techniques. Techniques used in metabolomics allow the detection of hundreds of different metabolites at the given state of a cell. Higher plants have a remarkable ability to synthesize a vast array of metabolites that differ in chemical complexity and biological functions, and play an indispensable role in plant adaptation (Saito and Matsuda 2010). Therefore, metabolites and/or metabolite patterns are effective biomarkers for phenotyping and diagnostic analysis of plants. Table 1.2 summarizes the representative databases and web sites relating to plant metabolomics.

Metabolomics Technologies

The two techniques that dominate metabolomics are MS and nuclear magnetic resonance (NMR) spectroscopy. Gas chromatography mass spectrometry (GC-MS) and liquid chromatography mass spectrometry (LC-MS) are currently the standard MS methods for metabolite profiling. GC-MS is the best-established analytical technique used in metabolomics. It enables the identification and robust quantification of a few hundred primary metabolites in a single extract (Roessner *et al*. 2001). Advanced injection systems allow split or splitless injections of small or large volume, cold or hot injections. Moreover, headspace GC-MS systems allow the detection of volatile metabolites and provide quantitative volatile profiles. Increased chromatographic resolution can be achieved by GC × GC-MS, where two columns are coupled to allow increased separation of closely eluted metabolites from the first column onto the second column (Pripdeevech *et al*. 2010). While GC-MS is mainly applicable to the analysis of volatile organic compounds and nonvolatile/polar metabolites after derivatization, LC-MS allows detection of a wider range of metabolites without a prior derivatization step. The resolution of LC-MS chromatographic signals is highly dependent on both the LC separation and the scan speed of the mass analyzer. Ultrahigh-performance LC-MS systems allow the detection of up to several hundred metabolites from a single analysis of a plant sample (de Vos *et al*. 2007). However, LC-MS-based metabolite databases are not as rich as GC-MS libraries, making LC-MS peak annotation difficult. Nevertheless, the choice of ionization and ion separation methods of MS is much greater in LC-MS than in GC-MS. Additionally, the application of tandem mass spectrometry (MS/MS) has further assisted the identification of unknown metabolites (Xu *et al*. 2007). In addition to these machines, capillary electrophoresis-mass spectrometry (CE-MS) is one of the most versatile analytical techniques that is used in proteomics and metabolomics (Oikawa *et al*. 2008). CE-MS is a highly sensitive methodology that can detect low-abundance metabolites and provides good analyte separation. NMR approaches, which rely on the detection of magnetic nuclei of atoms after application of constant magnetic field, are the main alternative to MS-based approaches for metabolic profiling. The intramolecular magnetic field around an atom in a molecule changes the resonance frequency, thus providing information on the electronic structure of a molecule. NMR spectroscopy is probably the most widely used

analytical technique in metabolomics. In comparison with GC-MS, LC-MS, and CE-MS, NMR is less biased and less destructive to the sample. The main advantage of NMR over MS-based techniques is that it can quantitatively detect all metabolites present in complex mixtures, irrespective of their volatility, polarity, molecular weight, size, chemical structure, and the sample matrix. NMR can provide subcellular information and it is easier to derive atomic information for flux modeling from NMR than from MS-based approaches.

For the analysis of the different metabolites, there are two main approaches: (i) metabolite fingerprinting, which is a semiquantitative technique with largely putative metabolite identification and (ii) metabolite profiling, which involves accurate chemical identification and quantification of a large number of metabolites (Dunn and Ellis 2005). The metabolite fingerprinting approach has been a powerful technique for identifying general patterns in the metabolic response to stress or in population studies (Kunin *et al*. 2009; Simpson *et al*. 2012). Metabolite fingerprinting aims to analyze metabolites in a complex mixture by either injecting the crude extract into a mass spectrometer (Dunn and Ellis 2005) or by obtaining NMR spectra of the whole mixture (Forseth and Schroeder 2011). The resulting metabolite fingerprints should be verified by quantitative metabolic profiling techniques. For metabolite profiling, the metabolites have to be separated prior to identification by chromatographic techniques such as GC, GC × GC or LC. Whichever separation approach is used, the metabolites within the extract will have to be identified either by their molecular mass, fragmentation pattern using MS, or chemical shift using NMR.

Applications of Metabolic Profiling

Metabolic approaches provide the opportunity to understand the functional phenotype of individuals and at the same time to get insight into plant plasticity to different environments and/or stresses. One approach to study this is by metabolome analysis of species within a genus spread across different populations and locations. There are recent examples showing the use of metabolite fingerprinting in the study of population dynamics in identifying the origin of coffee beans (Choi *et al*. 2010), categorizing populations of tobacco plants from China and Zimbabwe (Li *et al*. 2010), and identifying populations of the arctic-alpine plant *Arabidopsis lyrata* ssp. *petraea* from isolated regions across Europe (Davey *et al*. 2011). Furthermore, recent research done on a number of herbs obtained from a grassland biodiversity experiment suggested that variation in the metabolome is a functional indicator of the competitive ability and subsequent biodiversity of plants within experimental plant communities (Scherling *et al*. 2010). They found greater metabolic diversity in small herbaceous species compared to taller, more dominant plants grown in the experimental fields. Metabolomic analysis provided evidence for negative effects of resource competition on the investigated small-statured herbs that might explain their decreasing performance in terms of physical characteristics with increasing plant diversity. In contrast, taller species that often become dominant in mixed plant communities did not show modified metabolite profiles in response to altered resource availability with increasing plant diversity.

Metabolomics can also contribute significantly in finding answers to ecological questions on gross carbon, nitrogen, and phosphate allocation in plants (Peñuelas and Sardans 2009) (Rivas-Ubach *et al*. 2012) and also under environmental stresses (Shulaev *et al*. 2008). Metabolic profiling has been used to characterize stress responses to abiotic stresses such as water deficiency, dehydration, and high salinity (Urano *et al*. 2009), extreme temperature, cold and heat shock (Cook *et al*. 2004; Hannah *et al*. 2006), and ozone (Cho *et al*. 2008) among other stresses. Metabolic comparison between Arabidopsis and *Thellungiella halopita* showed that *Thellungiella* contained higher levels of various osmolytes after these results were corroborated with transcriptome analysis in which

several stress-related genes were expressed at high levels in *Thellungiella*, even in the absence of salt stress (Gong *et al*. 2005). These results suggest that a constant state of stress anticipation exists in *Thellungiella*. Recent analysis using GC-MS approach on GMO (Bt-maize) and non-GMO maize varieties grown under different conditions, including several growing locations and seasons, showed that a total of 3% of metabolome difference was related to genotype and up to 42% difference was caused by different growing conditions (Frank *et al*. 2012). A number of studies performed within the HEALTHGRAIN diversity screen program also demonstrate the effects of environment and genotype on the metabolome of wheat cultivars (Andersson *et al*. 2010; Fernandez-Orozco *et al*. 2010; Lampi *et al*. 2010).

Marker-Assisted Breeding and Association Mapping

Metabolomics is especially useful in genomics-assisted breeding. For example, a combination of QTL map-based cloning, transgenic approach, and association mapping has been used to reveal the amino acid sequence of the enzyme acyl-CoA:diacylglycerol acyltransferase responsible for determining oil content and composition in maize (Zheng *et al*. 2008). In a similar approach, screening of a tomato introgression line population harboring introgression of the wild species *Solanum pennellii* resulted in the identification of multiple QTL for total soluble solid content. One of these introgression lines (IL9-2-5) was delimited to a single base-pair change in *LIN5*, a cell wall invertase, and the line containing the allele from wild tomato had greater ability to bind sucrose, and, as a result, an increased sugar yield (Fridman *et al*. 2004). A broad profiling in the same *S*. *pennellii* ILs described for volatiles yielded 100 QTL (Tieman *et al*. 2005). Physiological studies of two of these volatiles, 2-phenyletanol and 2-phenylacetaldehyde confirmed the biological pathway of these important aromatic compounds in tomato (Tieman *et al*. 2006). A similar study on the same IL population using MS for primary metabolites analysis, resulted in the identification of 889 QTL governing the accumulation of 74 metabolites, including sugars, organic acids, essential amino acids, and vitamins. Although in many cases the metabolite content was increased, this was often associated with a yield penalty (Schauer *et al*. 2006). The GC-MS metabolic analysis of 289 diverse inbred maize lines showed the power of metabolomics for linking genotype and phenotype (Riedelsheimer *et al*. 2012). This study dealt with genome-wide association mapping of maize leaf metabolome including 118 distinct metabolites, 56,110 SNPs, and several agronomic traits of mature maize plants. This approach demonstrated that 26 distinct metabolites were highly associated with 26 SNPs and showed that lignin precursors, *p*-coumaric, and caffeic acids were strongly associated with a region of chromosome 9 encoding a key enzyme in monolignol synthesis. Recently, a core collection of 168 *Brassica rapa* accessions of different morphotypes and origins was explored to find genetic association between markers and metabolites (Pino del Carpio *et al*. 2011). In this study, eight myb and AFLP markers and two microsatellite markers associated with important metabolites that are considered as healthy compounds were identified (named tocopherols, carotenoids, chlorophylls, and folate). Nevertheless, further validation of these markers is needed to be used for MAS in *B*. *rapa* (Pino del Carpio *et al*. 2011). The success of the metabolite approaches indicates that metabolomics studies could greatly benefit from the advantages afforded by a multiparallel approach because this would encompass the use of a higher mapping resolution, a greater allele number, and a reduced time span to establish association as opposed to linkage analysis (Yu and Buckler 2006). It seems likely to be only a matter of time before the efficacy of such strategies can be effectively assessed.

Concluding Remarks and Future Prospects

The application of a suite of advanced technologies provides a rich comparative database of the molecular changes underlying plant adaptation. There is a super complex system of

biological processes that links genotype and phenotype, with the latter reflecting the end product of a complex transduction from genotype through the transcriptomic, proteomic, and metabolomics networks that lead to the phenotype (Fig. 1.1). Combining the different methods shows how changes at the genetic level propagate through various levels of organization to confer the targeted phenotypes. However, further challenges lie ahead in the integration of genomics, transcriptomics, proteomics, and metabolomics research to elucidate the mechanisms that underlie the behavior of these networks. Nevertheless, current advances indicate a potential paradigm shift from tissue-scale "omics" studies to single cell "omics," which will complement other "omics" methods on the way toward the development of unified systems biology of a single cell.

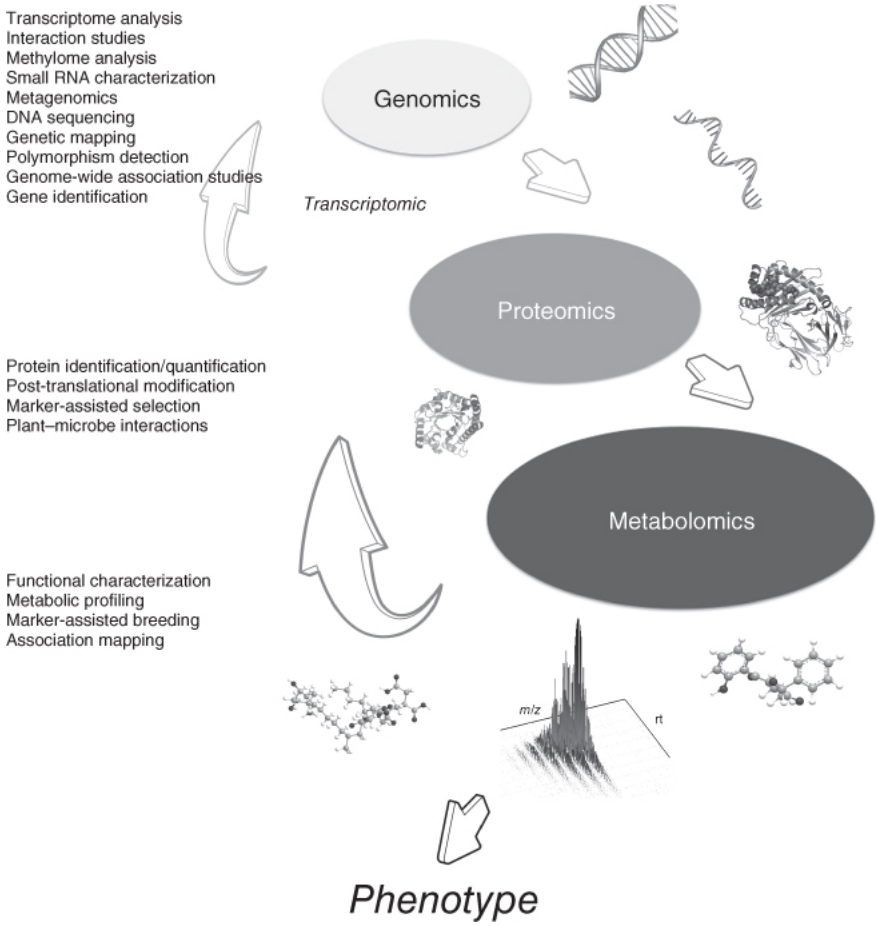

Figure 1.1 General scheme of "Omics" organization with their major applications in different integrated approaches.

Acknowledgments

This work was supported by the Spanish Ministerio de Ciencia e Innovación, Ministerio de Economía y Competitividad-FEDER, and University of Málaga through the Grants no. AGL2012-40066-CO2-02 and Ramón y Cajal program.

References

Ahsan, N., Nakamura, T. & Komatsu, S. (2012) Differential responses of microsomal proteins and metabolites in two contrasting cadmium (Cd)-accumulating soybean cultivars

under Cd stress. *Amino Acids*, 42, 317–327.
Andersson, A.A.M., Kamal-Eldin, A. & Aman, P. (2010) Effects of environment and variety on alkylresorcinols in wheat in the HEALTHGRAIN diversity screen. *Journal of Agriculture of Food Chemistry*, 58, 9299–9305.
Aghaei, K., Ehsanpour, A., Shah, A. & Komatsu, S. (2009) Proteome analysis of soybean hypocotyl and root under salt stress. *Amino Acids*, 36, 91–98.
Argout, X., Salse, J., Aury, J.M., Guiltinan, M.J. *et al.* (2011) The genome of *Theobroma cacao*. *Nature Genetics*, 43, 101–108.
Ashelford, K., Eriksson, M.E., Allen, C.M. *et al.* (2011) Full genome re-sequencing reveals a novel circadian clock mutation in Arabidopsis. *Genome Biology*, 12, R28.
Becker, C. & Weigel, D. (2012) Epigenetic variation: origin and transgenerational inheritance. *Current Opinion in Plant Biology*, 15, 562–567.
Becker, C., Hagmann, J., Müller, J. *et al.* (2011) Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature*, 480, 245–249.
Bindschedler, L.V. & Cramer, R. (2011) Quantitative plant proteomics. *Proteomics*, 11, 756–775.
Bird, A. (2007) Perceptions of epigenetics. *Nature*, 447, 396–398.
Bond, A.E., Row, P.E. & Dudley, E. (2011) Post-translation modification of proteins; methodologies and applications in plant sciences. *Phytochemistry*, 72, 975–996.
Buckler, E.S., Holland, J.B., Bradbury, P.J., Acharya, C.B. *et al.* (2009) The genetic architecture of maize flowering time. *Science*, 325, 714–718.
Chan, Z.L., Wang, Q., Xu, X.B. *et al.* (2008) Functions of defense-related proteins and dehydrogenases in resistance response induced by salicylic acid in sweet cherry fruits at different maturity stages. *Proteomics*, 8, 4791–4807.
Chan, Z. (2013) Proteomic responses of fruits to environmental stresses. *Frontier in Plant Science*, 3, 311.
Cho, K., Shibato, J., Agrawal, G.K. & Jung, Y.H. (2008) Integrated transcriptomics, proteomics, and metabolomics analyses to survey ozone responses in the leaves of rice seedling. *Journal of Proteome Research*, 7, 2980–2998.
Choi, M.Y., Choi, W., Park, J.H., Lim, J. & Kwon, S.W. (2010) Determination of coffee origins by integrated metabolomic approach of combining multiple analytical data. *Food Chemistry*, 121, 1260–1268.
Chutimanitsakun, Y., Nipper, R.W., Cuesta-Marcos, A. *et al.* (2011) Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. *BMC Genomics*, 12, 4.
Cook, D.D., Fowler, S.S., Fiehn, O.O. & Thomashow, M.F.M. (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 15243–15248.
Grimsrud, P.A., Swaney, D.L., Wenger, C.D., Beauchene, N.A. & Coon, J.J. (2010) Phosphoproteomics for the masses. *ACS Chemical Biology*, 5, 105–119.
Cronn, R., Knaus, B.J. & Liston, A. (2012) Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany*, 99, 291–311.
Damerval, C., Maurice, A., Josse, J.M. & de Vienne, D. (1994) Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics*, 137, 289–301.
Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. & Blaxter, M.L. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetic*, 12, 499–510.
Deng, W., Ying, H., Helliwell, C., Peacock, W. & Dennis, E. (2011) FLOWERING LOCUS C (FLC) regulates development pathways throughout the life cycle of Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 6680–6685.

de Vos, R.C.H., Moco, S., Lommen, A., Keurentjes, J.J.B., Bino, R.J. & Hall, R.D. (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature Protocols*, 2, 778–791.
D'Hont, A., Denoeud, F., Aury, J.M., Baurens, F.C. *et al.* (2012) Thee banana (*Musa acuminata*) genome and the evolution of monocotyle-donous plants. *Nature*, 488, 213–217.
Dunn, W.B. & Ellis, D.I. (2005) Metabolomics: current analytical platforms and methodologies. *TRAC Trends in Analytical Chemistry*, 24, 285–294.
Fernandez-Orozco, R., Li, L., Harflett, C., Shewry, P.R. & Ward, J.L. (2010) Effects of environment and genotype on phenolic acids in wheat in the HEALTHGRAIN diversity screen. *Journal of Agriculture and Food Chemistry*, 58, 9341–9352.
Fierer, N., Ladau, J., Clemente, J.C. *et al.* (2013) Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. *Science*, 342, 621–624.
Forseth, R.R. & Schroeder, F.C. (2011) NMR-spectroscopic analysis of mixtures: from structure to function. *Current Opinion in Chemistry Biology*, 15, 38–47.
Frank, T., Röhlig, R.M., Davies, H.V., Barros, E. & Engel, K.H. (2012) Metabolite profiling of maize kernels--genetic modification versus environmental influence. *Journal of Agriculture and Food Chemistry*, 60, 3005–3012.
Franssen, S.U., Gu, J., Bergmann, N. *et al.* (2011) Transcriptomic resilience to global warming in the seagrass *Zostera marina*, a marine foundation species. *Proceeding of the National Academy of Sciences of the United States of America*, 108, 19276–19281.
Fridman, E., Carrari, F., Liu, Y.S., Fernie, A.R. & Zamir, D. (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science*, 305, 1786–1789.
Gammulla, C.G., Pascovici, D., Atwell, B.J. & Haynes, P.A. (2011) Differential proteomic response of rice (Oryza sativa) leaves exposed to high- and low-temperature stress. *Proteomics*, 11, 2839–2850.
Gerber, I.B., Laukens, K., De Vijlder, T., Witters, E. & Dubery, I.A. (2008) Proteomic profiling of cellular tar-gets of lipopolysaccharide-induced signalling in Nicotiana tabacum BY- 2 cells. *Biochimica Biophysica. Acta*, 1784, 1750–1762.
Gerber, I.B., Laukens, K., Witters, E. & Dubery, I.A. (2006) Lipopolysaccharide-responsive phosphoproteins in Nicotiana tabacum cells. *Plant Physiology and Biochemistry*, 44, 369–379.
Gong, Q., Li, P., Ma, S., Indu-Rupassara, S. & Bohnert, H.J. (2005) Salinity stress adaptation competence in the extremophile Thellungiella halophila in comparison with its relative Arabidopsis thaliana. *Plant Journal*, 44, 826–839.
Gore, M.A., Chia, J.M., Elshire, R.J. *et al.* (2009) A first-generation haplotype map in maize. *Science*, 32, 1115–1117.
Gonzalez-Ibeas, D., Blanca, J., Donaire, L. *et al.* (2011) Analysis of the melon (*Cucumis melo*) small RNAome by high-throughput pyrosequencing. *BMC Genomics*, 12, 393.
Guo, S., Zhang, J., Sun, H., Salse, J. *et al.* (2013) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nature Genetics*, 45, 51–58.
Gupta, R.B., Singh, N.K. & Shepherd, K.W. (1989) The cumulative effect of allelic variation in LMW and HMW glutenin subunits on dough properties in the progeny of two bread wheats. *Theoretical and Applied Genetics*, 77, 57–64.
Hannah, M.A.M., Wiese, D.D., Freund, S.S., Fiehn, O.O., Heyer, A.G.A. & Hincha, D.K.D. (2006) Natural genetic variation of freezing tolerance in Arabidopsis. *Plant Physiology*, 142, 98–112.
Harris, J. (2009) Soil microbial communities and restoration ecology: facilitators or followers? *Science*, 325, 573–574.
Higashi, Y. & Saito, K. (2013) Network analysis for gene discovery in plant-specialized metabolism. *Plant and Cell Environment*, 36, 1597–1606.
Horner, D.S., Pavesi, G., Castrignano, T. *et al.* (2010) Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Briefing in*

Bioinformatics, 11, 181e97.
Howden, B.P., McEvoy, C.R.E., Allen, D.L. *et al.* (2011) Evolution of multidrug resistance during Staphylococcus aureus infection involves mutation of the essential two component regulator WalKR. *PLoS Pathology*, 7, e1002359.
Huss, M. (2010) Introduction into the analysis of high-throughput-sequencing based epigenome data. *Briefing in Bioinformatics*, 11, 512e23.
Hyten, D.L., Cannon, S.B., Song, Q. *et al.* (2010) High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMC Genomics*, 11, 38.
Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G., Perez-Torres, C.A. *et al.* (2013) Architecture and evolution of a minute plant genome. *Nature*, 498, 94–98.
Irar, S., Brini, F., Goday, A., Masmoudi, K. & Pagès, M. (2010) Proteomic analysis of wheat embryos with 2-DE and liquid-phase chromatography (ProteomeLab PF-2D) – a wider perspective of the proteome. *Journal of Proteomics*, 73, 1707–1721.
Jaillon, O., Aury, J.M., Noel, B. *et al.* (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 449, 463–467.
Jia, J., Zhao, S., Kong, X., Li, Y. *et al.* (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*, 496, 91–95.
Kane, N., Sveinsson, S., Dempewolf, H. *et al.* (2012) Ultra-barcoding in cacao (Theobroma spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *American Journal of Botany*, 99, 320–329.
Kaufmann, K., Muiño, J., Osteras, M., Farinelli, L., Krajewski, P. & Angenent, G. (2010) Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). *Nature Protocols*, 5, 457–472.
Kawamura, Y. & Uemura, M. (2003) Mass spectrometric approach for identifying putative plasma membrane proteins of *Arabidopsis* leaves associated with cold acclimation. *The Plant Journal*, 36, 141–154.
Kenan-Eichler, M., Leshkowitz, D., Tal, L. *et al.* (2011) Wheat hybridization and polyploidization results in deregulation of small RNAs. *Genetics*, 188, 263–272.
Komatsu, S., Wada, T., Abaléa, Y. *et al.* (2009) Analysis of plasma membrane proteome in soybean and application to flooding stress response. *Journal of Proteome Research*, 8, 4487–4499.
Kunin, W.E., Vergeer, P., Kenta, T. *et al.* (2009) Variation at range margins across multiple spatial scales: environmental temperature, population genetics and metabolomic phenotype. *Proceeding of the Royal Society B: Biological Science*, 276, 1495–1506.
Kurtoglu, K.Y., Kantar, M. & Budak, H. (2014) New wheat microRNA using whole-genome sequence. *Functional and Integrative Genomics*, 14, 363–379.
Kvam, V.M., Liu, P. & Si, Y. (2012) A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *American Journal of Botany*, 99, 248–256.
Lampi, A.M., Nurmi, T. & Piironen, V. (2010) Effects of the environment and genotype on tocopherols and tocotrienols in wheat in the HEALTHGRAIN diversity screen. *Journal of Agriculture and Food Chemistry*, 58, 9306–9313.
Landegren, U., Kaiser, R., Sanders, J. & Hood, L. (1988) A ligase-mediated gene detection technique. *Science*, 241, 1077–1080.
Li, Q.Q., Zhao, C.C., Li, Y.Y. *et al.* (2010) Liquid chromatography/mass spectrometry-based metabolic profiling to elucidate chemical differences of tobacco leaves between Zimbabwe and China. *Journal of Separation Science*, 34, 119–126.
Li, Y.F., Zheng, Y., Jagadeeswaran, G. & Sunkar, R. (2013) Characterization of small RNAs and their target genes in wheat seedlings using sequencing-based approaches. *Plant Science*, 203, 17–24.
Li, F., Fan, G., Wang, K., Sun, F. *et al.* (2014) Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nature Genetics*, 46, 567–572.

Liu, L., Yinhu, L., Siliang, L. *et al.* (2012) Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*.: Article ID 251364
Maitra, R.D., Kim, J. & Dunbar, W.B. (2012) Recent advances in nanopore sequencing. *Electrophoresis*, 33 (23), 3418–3428.
Mardis, E.R. (2007) ChIP-seq: welcome to a new frontier. *Nature Method*, 4, 613–614.
Marguerat, S. & Bahler, J. (2010) RNA-seq: from technology to biology. *Cellular and Molecular Life Science*, 67 (4), 569–579.
Mascher, M., Wu, S., Amand, P., Stein, N. & Poland, J. (2013) Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. *Plos ONE*, 8 (10), e76925.
McMullen, M.D., Kresovich, S., Villeda, H.S., Bradbury, F. *et al.* (2009) Genetic properties of the maize nested association mapping population. *Science*, 325 (5941), 737–740.
Meng, F., Liu, H., Wang, K. *et al.* (2013) Development-associated microRNAs in grains of wheat (*Triticum aestivum* L.). *BMC Plant Biology*, 13, 140.
Michaels, S. & Amasino, R. (1999) FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell*, 11, 949–956.
Ming, R., VanBuren, R., Liu, Y., Yang, M. *et al.* (2013) Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). *Genome Biology*, 14, R41.
Nicholson, J.K., Lindon, J.C. & Holmes, E. (1999) “Metabolomics”: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, 29, 1181–1189.
Nyren, P., Pettersson, B. & Uhlén, M. (1993) Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Analytical Biochemistry*, 208 (1), 171–175.
Oikawa, A., Matsuda, F., Kusano, M., Okazaki, Y. & Saito, K. (2008) Rice metabolomics. *Rice*, 1, 63–71.
Orlando, L., Ginolhac, A., Raghavan, M. *et al.* (2011) True single-molecule DNA sequencing of a pleistocene horse bone. *Genome Research*, 21, 1705–1719.
Ozsolak, F., Platt, A.R., Jones, D.R. *et al.* (2009) Direct RNA sequencing. *Nature*, 461, 814–818.
Parchman, T.L., Gompert, Z., Mudge, J., Schilkey, F.D., Benkman, C.W. & Buerkle, C.A. (2012) Genome-wide association genetics of an adaptative trait in lodgepole pine. *Molecular Ecology*, 21 (12), 2991–3005.
Peñuelas, J. & Sardans, J. (2009) Ecological metabolomics. *Chemistry and Ecology*, 25, 305–309.
Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. & Hoekstra, H.E. (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *Plos ONE*, 7 (5), e371135.
Pichersky, E. & Gang, D. (2000) Genetics and biochemistry of secondary metabolites: an evolutionary perspective. *Trends in Plant Science*, 5, 439–445.
Pino del Carpio, D., Basnet, R.K., de Vos, R.C.H., Maliepaard, C., Paulo, M.J. & Bonnema, G. (2011) Comparative methods for association studies: a case study on metabolite variation in a *Brassica rapa* core collection. *Plos ONE*, 6 (5), e19624.
Pfender, W.F., Saha, M.C., Johnson, E.A. & Slabaugh, M.B. (2011) Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. *Theoretical and Applied Genetics*, 122, 1467–1480.
Poland, J.A., Bradbury, P.J., Buckler, E.S. & Nelson, R.J. (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (17), 6893–6898.
Pripdeevech, P., Wongpornchai, S. & Marriott, P.J. (2010) Comprehensive two-dimensional gas chromatography-mass spectrometry analysis of volatile constituents in Thai vetiver root oils obtained by using different extraction methods. *Phytochemical Analysis*, 21, 163–173.

Qin, J., Gu, F., Liu, D. *et al.* (2013) Proteomic analysis of elite soybean Jidou17 and its parents using iTRAQ-based quantitative approaches. *Proteome Science*, 11, 12.
Riedelsheimer, C., Lisec, J., Czedik-Eysenberg, A. *et al.* (2012) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 8872–8877.
Rivas-Ubach, A., Sardans, J., Pérez-Trujillo, M., Estiarte, M. & Peñuelas, J. (2012) Strong relationship between elemental stoichiometry and metabolome in plants. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 4181–4186.
Roessner, U., Luedemann, A., Brust, D. *et al.* (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell*, 13, 11–29.
Rothberg, J.M., Hinz, W., Rearick, T.M. *et al.* (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475, 348–352.
Ruzicka, D., Chamala, S., Barrios-Masias, F.H. *et al.* (2013) Inside arbuscular mycorrhizal roots-molecular probes to understand the symbiosis. *The Plant Genome*, 6 (2), 1–13.
Saito, K. & Matsuda, F. (2010) Metabolomics for Functional Genomics, Systems Biology, and Biotechnology. *Annual Review of Plant Biology*, 61, 463–489.
Salekdeh, G.H. & Komatsu, S. (2007) Crop proteomics: aim at sustainable agriculture of tomorrow. *Proteomics*, 7, 2976–2996.
Saravanan, R.S. & Rose, J.K.C. (2004) A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. *Proteomics*, 4 (9), 2522–2532.
Sarry, J.E., Sommerer, N., Sauvage, F.X. *et al.* (2004) Grape berry biochemistry revisited upon proteomic analysis of the mesocarp. *Proteomics*, 4, 201–215.
Schauer, N., Semel, Y., Roessner, U. *et al.* (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nature Biotechnology*, 24, 447–454.
Scherling, C., Roscher, C., Giavalisco, P., Schulze, E.D. & Weckwerth, W. (2010) Metabolomics unravel contrasting effects of biodiversity on the performance of individual plant species. *PLoS ONE*, 5, e12569.
Schmitz, R.J. & Erker, J.R. (2012) Epigenetic and epigenomic variation in *Arabidopsis thaliana*. *Trends in Plant Science*, 17, 149–154.
Schmitz, R.J., Schultz, M.D., Lewsey, M.G. *et al.* (2011) Transgenerational epigenetic instability is a source of novel methylation variants. *Science*, 334 (6054), 369–373.
Shaw, R.G., Byers, D.L. & Darmo, E. (2000) Spontaneous mutational effects on reproductive traits of *Arabidopsis thaliana*. *Genetics*, 155, 369–378.
Shulaev, V., Cortes, D., Miller, G. & Mittler, R. (2008) Metabolomics for plant stress response. *Physiologia Plantarum*, 132, 199–208.
Shulaev, V., Sargent, D.J., Crowhurst, R.N. *et al.* (2011) The genome of woodland strawberry (*Fragaria vesca*). *Nature Genetics*, 43, 109–116.
Serna-Sanz, A., Parniske, M. & Peck, S.C. (2011) Phosphoproteome analysis of Lotus japonicus roots reveals shared and distinct compo-nents of symbiosis and defense. *Molecular Plant-Microbe Interaction*, 24, 932–937.
Simon, S.A., Zhai, J., Nandety, R.S. *et al.* (2009) Short-read sequencing technologies for transcriptional analyses. *Annual Review of Plant Biology*, 60, 305–333.
Simpson, A.J., Simpson, M.J. & Soong, R. (2012) Nuclear magnetic resonance spectroscopy and its key role in environmental research. *Environmental Science and Technology*, 46, 11488–11496.
Strickler, S.R., Bombarely, A. & Mueller, L.A. (2012) Designing a transcriptome next-generation sequencing project for a nonmodel plant species. *American Journal of Botany*, 99, 257–266.
Sun, F., Guo, W., Du, J., Ni, Z., Su, Q. & Yao, Y. (2013) Widespread, abundant, and diverse TE-associated siRNAs in developing wheat grain. *Gene*, 522, 1–7.

Tang, Z., Zhan, L., Yuan, S., Zheng, Y. & Zhao, C. (2012) Uncovering small RNA-mediated responses to cold stress in a wheat thermosensitive genic male-sterile line by deep sequencing. *Plant Physiology*, 159 (2), 721–738.
Tejedor-Lorenzo, T.M., Mizuno, H. & Tsuyama, N. (2009) Direct single-cell molecular analysis of plant tissues by video mass spectrometry. *Analytical Science*, 25, 1053–1055.
The Arabidopsis Genome Initiative (2000) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science*, 345 (6194), 1251788.
The Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. *Nature*, 475, 189–195.
The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, 485 (7400), 635–641.
Thompson, J.F. & Milos, P.M. (2011) The properties and applications of single-molecule DNA sequencing. *Genome Biology*, 12, 217.
Tieman, D., Taylor, M., Schauer, N., Fernie, A.R., Hanson, A.D. & Klee, H.J. (2006) Tomato aromatic amino acid decarboxylases participate in synthesis of the flavor volatiles 2-phenylethanol and 2-phenylacetaldehyde. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 8287–8292.
Tieman, D.M., Zeigler, M., Schmelz, E.A. *et al.* (2005) Identification of loci affecting flavour volatile emissions in tomato fruits. *Journal of Experimental Botany*, 57, 887–896.
Toubiana, D., Fernie, A.R., Nikoloski, Z. & Fait, A. (2013) Network analysis: tackling complex data to study plant metabolism. *Trends in Biotechnology*, 31, 29–36.
Trapphoff, T., Beutner, C., Niehaus, K. & Colditz, F. (2009) Induction of distinct defense-associated protein patterns in Aphanomyces euteiches (oomycota)-elicited and -inoculated Medicago truncatula cell-suspension cultures: a proteome and phosphoproteome approach. *Molecular Plant-Microbe Interaction*, 22, 421–436.
Urano, K., Maruyama, K., Ogata, Y. & Morishita, Y. (2009) Characterization of the ABA-regulated global responses to dehydration in Arabidopsis by metabolomics. *The Plant Journal*, 57, 1065–1078.
Varshney, R.K., Song, C., Saxena, R.K., Azam, S. *et al.* (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nature Biotechnology*, 31, 240–246.
Vaucheret, H. (2006) Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Gene and Development*, 20 (7), 759–771.
Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A. *et al.* (2010) The genome of domesticated apple (*Malus x domestica* Borkh). *Nature Genetics*, 42, 833–839.
Wagner, C.E., Keller, I., Wittwer, S. *et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, 22 (3), 787–798.
Wang, K., Peng, X., Ji, Y., Yang, P., Zhu, Y. & Li, S. (2013) Gene, protein, and network of male sterility in rice. *Frontiers in Plant Science*, 4, 92.
Wang, L., Wang, A., Huang, X., Zhao, Q., Dong, G. *et al.* (2011) Mapping 49 quantitative trait loci at high resolution through sequencing-based genotyping of rice recombinant inbred lines. *Theoretical and Applied Genetics*, 122, 327–340.
Wang, Z., Gerstein, M. & Snyder, M. (2009a) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10, 57–63.
Wang, Q., Chan, Z.L., Qin, G.Z. & Tian, S.P. (2009b) Comparison of 2-DE techniques for improved proteomic analysis of fruit tissues. *Chinese Bulletin of Botany*, 44, 107–116.
Wienkoop, S. & Weckwerth, W. (2006) Relative and absolute quantitative shotgun proteomics: targeting low-abundance proteins in Arabidopsis thaliana. *Journal of Experimental Botany*, 57, 1529–1535.
Wu, J., Wang, Z., Shi, Z., Zhang, S. *et al.* (2013) The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Research*, 23 (2), 296–408.
Xin, M., Wang, Y., Yao, Y. *et al.* (2010) Diverse set of microRNAs are responsive to powdery mildew infection and heat stress in wheat (*Triticum aestivum* L.). *BMC Plant*

Biology, 10, 123.
Xu, R.N., Fan, L., Rieser, M.J. & El-Shourbagy, T.A. (2007) Recent advances in high-throughput quantitative bioanalysis by LC-MS/MS. *Journal of Pharmaceutical and Biomedical Analysis*, 44, 342–355.
Yamaguchi, M., Valliyodan, B., Zhang, J. *et al.* (2010) Regulation of growth response to water stress in the soybean primary root. I. Proteomic analysis reveals region-specific regulation of phenylpropanoid metabolism and control of free iron in the elongation zone. *Plant and Cell Environment*, 33, 223–243.
Yamazaki, T., Kawamura, Y., Minami, A. & Uemura, M. (2008) Calcium-dependent freezing tolerance in *Arabidopsis* involves membrane resealing via synaptotagmin SYT1. *Plant Cell*, 20, 3389–3404.
Yao, Y., Ni, Z., Peng, H. *et al.* (2010) Non-coding small RNAs responsive to abiotic stress in wheat (*Triticum aestivum* L.). *Functional and Integrative Genomics*, 10 (2), 187–190.
Yu, J. & Buckler, E.S. (2006) Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology*, 17, 155–160.
Zalapa, J.E., Cuevas, H., Zhu, H. *et al.* (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *American Journal of Botany*, 99, 193–208.
Zhang, Y. & Jeltsch, A. (2010) The application of next generation sequencing in DNA methylation analysis. *Genes*, 1 (1), 85–101.
Zhang, L., Chia, J.M., Kumari, S. *et al.* (2009) A genome-wide characterization of microRNA genes in maize. *PLoS Genetics*, 5, e1000716.
Zhang, P., Allen, W.B., Roesler, K. *et al.* (2007) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell*, 126 (6), 1189–1201.
Zhen, Y., Qi, J.L., Wang, S.S. *et al.* (2007) Comparative proteome analysis of differentially expressed proteins induced by Al toxicity in soybean. *Physiology Plantarum*, 131, 542–554.
Zheng, P., Allen, W.B., Poesler, K. *et al.* (2008) A phenylalanine in DGAT is a key determinant of oil content and composition in maize. *Nature Genetics*, 40 (3), 367–372.
Zhong, S., Fei, Z., Chen, Y.R. *et al.* (2013) Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nature Biotechnology*, 31, 154–159.

2
Use of Natural Variation in *Arabidopsis thaliana* to Study Adaptation

Lisa M. Smith[1] and Roosa A. E. Laitinen[2]
[1] *Department of Animal and Plant Sciences, The University of Sheffield, Sheffield, UK*
[2] *Max-Planck-Institut für Molekulare Pflanzenphysiologie, Wissenschaftspark Golm, Potsdam, Germany*

Introduction

While crops have traditionally formed the foundation of plant research, *Arabidopsis thaliana* was selected as a model plant approximately 30 years ago. This choice was based on a number of key factors that favored Arabidopsis for molecular biology and genetics research; it is a small plant ideal for growth under laboratory conditions, it has a relatively short generation time, it is diploid and so facilitates genetics studies, and it is self-compatible, producing a high number of progeny (Koornneef and Meinke 2010). In addition, Arabidopsis has a broad geographical range, so has adapted to a number of different growth habitats. While native to Northern Africa and Eurasia, its range has extended to the New World through human migration. Arabidopsis is mostly selfing; therefore, accessions collected in the wild are generally mostly isogenic. The natural variation among diploid homogenous accessions is of significant advantage in addressing correlations between genotype and phenotype.

In recent years, the use of natural variation in Arabidopsis to understand adaptation has been accelerated through the advent of next-generation sequencing. Since the first Arabidopsis genome (accession Col-0) was published in 2000, Illumina sequencing has been used extensively to sequence further Arabidopsis accessions (Lin *et al*. ; Mayer *et al*. 1999; Arabidopsis Genome Initiative 2000; Salanoubat *et al*. 2000; Tabata *et al*. 2000; Theologis *et al*. 2000). The genomes of Bur-0 and Tsu-1 were released in 2008 (Ossowski *et al*. 2008), followed by reference-guided assemblies of the next four resequenced genomes which were made available in 2011 and cover most of the ~120 Mb genome (Schneeberger *et al*. 2011). Since then, many more Arabidopsis genomes have been published: 19 accessions from the MAGIC lines (see Genetic Natural Variation section for further details; Gan *et al*. 2011); 80 Eurasian accessions (Cao *et al*. 2011); 180 Swedish accessions (Long *et al*. 2013); 217 diverse accessions (Schmitz *et al*. 2013); and a high-quality Ler-0 genome from sequencing company PacBio (http://blog.pacificbiosciences.com/2013/08/new-data-release-arabidopsis-assembly.html). Some of these resequenced genomes are part of the 1001 genomes project, which has sought to expand our knowledge and use of natural genetic variation in Arabidopsis (http://1001genomes.org).

The genome information of the different accessions has been advantageous in understanding plant adaptation, not only in developing new and faster methods to analyze the genetic basis of adaptive traits, but also in determining how much natural variation has arisen in Arabidopsis over an evolutionary timescale. Whole genome resequencing of 80 European and Asian strains of Arabidopsis found variation (single nucleotide polymorphisms (SNPs) or small insertions/deletions of up to 20 bp) at approximately 5% of