MOLECULAR MECHANISMS IN PLANT ADAPTATION

Edited by Roosa A.E. Laitinen

WILEY Blackwell

Molecular Mechanisms in Plant Adaptation

Molecular Mechanisms in Plant Adaptation

Edited by

Roosa A. E. Laitinen

WILEY Blackwell

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Molecular mechanisms in plant adaptation / edited by Roosa A.E. Laitinen.
p. ; cm.
Includes bibliographical references index.
ISBN 978-1-118-86017-5 (cloth)
1. Plants–Adaptation. 2. Plant molecular biology. I. Laitinen, Roosa A. E., editor.
[DNLM: 1. Adaptation, Physiological. 2. Plants–genetics. 3. Computational Biology. 4. Plant Physiological Phenomena. QK 912]
SB112.5.M67 2015
581.4–dc23

2015005263

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Contents

List of Contributors

Ana M.R. Almeida
Departments of Plant and Microbial Biology and Integrative Biology and the University and Jepson Herbaria, The University of California, Berkeley, CA, USA
Programa de Pós-Graduação em Genética e Biodiversidade, Universidade Federal da Bahia (UFBA), Salvador, BA, Brazil

Leónie Bentsink
Wageningen Seedlab, Laboratory of Plant Physiology, Wageningen University, Wageningen, The Netherlands

William R. Chezem
Department of Molecular, Cellular & Developmental Biology, Yale University, New Haven, CT, USA

Nicole K. Clay
Department of Molecular, Cellular & Developmental Biology, Yale University, New Haven, CT, USA

Matthew Gilliham
ARC Centre of Excellence in Plant Energy Biology, School of Agriculture, Food and Wine, University of Adelaide, Glen Osmond, SA, Australia

Sam W. Henderson
ARC Centre of Excellence in Plant Energy Biology, School of Agriculture, Food and Wine, University of Adelaide, Glen Osmond, SA, Australia

Sabrina Kleessen
Systems Biology and Mathematical Modeling, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany
Targenomix GmbH, Potsdam-Golm, Germany

Karin I. Köhl
Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

Roosa A.E. Laitinen
Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

Sascha Laubinger
Center for Plant Molecular Biology, University of Tübingen, Tübingen, Germany
Chemical Genomics Centre of the Max Planck Society, Dortmund, Germany
Max Planck Institute for Developmental Biology, Tübingen, Germany

Thu-Phuong Nguyen
Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, Cologne, Germany
Wageningen Seedlab, Laboratory of Plant Physiology, Wageningen University, Wageningen, The Netherlands

Zoran Nikoloski
Systems Biology and Mathematical Modeling, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

Sonia Osorio
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
Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

Alma Piñeyro-Nelson
Departments of Plant and Microbial Biology and Integrative Biology and the University and Jepson Herbaria, The University of California, Berkeley, CA, USA

Lisa M. Smith
Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

Chelsea D. Specht
Departments of Plant and Microbial Biology and Integrative Biology and the University and Jepson Herbaria, The University of California, Berkeley, CA, USA

Luis A.A. Toledo-Filho
Center for Plant Molecular Biology, University of Tübingen, Tübingen, Germany
Chemical Genomics Centre of the Max Planck Society, Dortmund, Germany
Max Planck Institute for Developmental Biology, Tübingen, Germany

José G. Vallarino
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

Roxana Yockteng
Departments of Plant and Microbial Biology and Integrative Biology and the University and Jepson Herbaria, The University of California, Berkeley, CA, USA
Institut de Systématique, Évolution, Biodiversité, UMR 7205 CNRS, Muséum National d'Histoire Naturelle, Paris Cedex, France

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

Molecular Mechanisms in Plant Adaptation, First Edition. Edited by Roosa Laitinen.

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.

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

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	Disadvantage	Primary applications
Sanger		Chain terminator	PCR	600–900	High quality, long read length	High cost, low throughput	
454 GS FLX	Roche	Synthesis (pyrosequencing)	Emulsion PCR	600–800	First next-gen sequencer, long reads	Error rate with polybase more than six, high cost, low throughput	1, 2, 3, 5, 6, 8
Solexa GAIIx	Illumina	Synthesis	Solid-phase PCR	150	First short-read sequencer, high-throughput, current leader in advantages	Short-read assembly	1, 2, 3, 4, 5, 6, 7, 8
SOLiD4	Life Technologies	Ligation	Emulsion PCR	50	Second short-read sequencer, low error rates	Short-read assembly	2, 4, 7, 8
HeliScope/ Helicos	Life Technologies	Synthesis	Single molecule template	25–35	Single-molecule sequencer	Short read lengths, high error rates	7, 8

(*continued*)

Table 1.1 (*Continued*)

Platform	Current company	Sequencing method	Amplification method	Read length (bp)	Advantage	Disadvantage	Primary applications
Ion Torrent	Life Technologies	Synthesis	Emulsion PCR	100–400	First post-light sequencer, first system < €100,000, high throughput	Low throughput, high cost per base, still prone to homopolymer error, though better performance than 454	1, 2, 3, 5, 8
PacBio/Pacific Bioscience	Pacific Bioscience	Synthesis	Real-time single-molecule template	964 average	Real-time single-molecule sequencing, long reads, high throughput	High error rates, low yield	1, 2, 3, 6, 8
Starlight*	Life Technologies	Synthesis	Single-molecule sequencing with quantum dots		Single-molecule sequencing with quantum dots		1, 3, 6, 8

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.

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