Advances in Omics and Bioinformatics Tools for Systems Analyses of Plant Functions

Keiichi Mochida1,2,3,* and Kazuo Shinozaki1,2
1RIKEN Biomass Engineering Program, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045 Japan
2RIKEN Plant Science Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045 Japan
3Kihara Institute for Biological Research, Yokohama City University, 641-12 Maioka-cho, Totsuka-ku, Yokohama, Kanagawa, 230-0045, Japan
*Corresponding author: E-mail, mochida@psc.riken.jp; Fax: +81-45-503-9591.
(Received October 2, 2011; Accepted October 27, 2011)

Omics and bioinformatics are essential to understanding the molecular systems that underlie various plant functions. Recent game-changing sequencing technologies have revitalized sequencing approaches in genomics and have produced opportunities for various emerging analytical applications. Driven by technological advances, several new omics layers such as the interactome, epigenome and hormonome have emerged. Furthermore, in several plant species, the development of omics resources has progressed to address particular biological properties of individual species. Integration of knowledge from omics-based research is an emerging issue as researchers seek to identify significance, gain biological insights and promote translational research. From these perspectives, we provide this review of the emerging aspects of plant systems research based on omics and bioinformatics analyses together with their associated resources and technological advances.

Keywords: Bioinformatics • Data integration • Genome-scale approach • Omics • Systems analysis.

Abbreviations: ChIP, chromatin immunoprecipitation; EMS, ethyl methanesulfonate; EST, expressed sequence tag; FOX, full-length cDNA overexpressor; GWAS, genome wide association study; InDel, insertion–deletion polymorphism; miRNA, microRNA; mQTL, metabolite quantitative trait locus; NGS, next-generation sequencing; RNA-seq, RNA sequencing; SNP, single nucleotide polymorphism; SRA, Sequence Read Archive; sRNA, small RNA; TILLING, targeting induced local lesions in genomes; Y2H, yeast two-hybrid.

Introduction

In the first decade of the 21st century, genomic studies in model plants promoted gene discovery and provided a breadth of gene function knowledge. In the second decade, driven by recent technological advances, omics-based approaches have allowed us to address the complex global biological systems that underlie various plant functions. These technological advances have also accelerated the development of genome-scale resources in applied and emerging model plant species, and have promoted translational research by integrating knowledge across plant species (Mochida and Shinozaki 2010).

The recent game-changing advances in DNA sequencing technology have revolutionized the sciences field, featuring unprecedented innovations in sequencing scale and throughput, and implementations of various novel applications beyond genome sequencing. In particular, while accelerating genome projects, next-generation sequencing (NGS) technology provides feasible applications such as whole-genome re-sequencing for variation analysis, RNA sequencing (RNA-seq) for transcriptome and non-coding RNAome analysis, quantitative detection of epigenomic dynamics, and Chip-seq analysis for DNA–protein interactions (Lister et al. 2009). In addition to these approaches, which focus on transcriptional regulatory networks, other approaches have been developed, including interactome analysis for networks formed by protein–protein interactions (Arabidopsis Interactome Mapping Consortium 2011), hormone analysis for phytohormone-mediated cellular signaling (Kojima et al. 2009), and metabolome analysis for metabolic systems (Saito and Matsuda 2010). These rapidly developing omics fields with associated genome-scale resources treat each molecular network as an interlocked component of the plant cellular system (Fig. 1). Bioinformatics has been crucial in every aspect of omics-based research to manage various types of genome-scale data sets effectively and extract valuable knowledge. Integration of accumulating omics outcomes will deepen and update our understanding and facilitate knowledge exchange with other model organisms (Shinozaki and Sakakibara 2009). Therefore, databases for omics outcomes derived from recent innovative analytical platforms comprise the essential infrastructure for systems analysis.

In this review, we provide an overview of several recently emerged resources derived from a systems approach to omics and bioinformatics analyses in plant science. We describe NGS-based applications and illustrative outcomes of plant genomics. We then review the status of emerging omics topics...
that have recently appeared in plant omics science, including
the interactome, epigenome, hormonome and metabolome.
We also highlight the status of genomic resource developments
in the families Solanaceae, Gramineae and Fabaceae, each of
which includes emerging plant species and/or important
applied plant species. We also discuss the integration and visu-
alization of omics data sets as key issues for effective analysis
and better understanding of biological insights. Finally, we use
omics-based systems analyses to understand plant functions.
Throughout this review, we provide examples of recent out-
comes in plant omics and bioinformatics, and present an over-
view of available resources.

Generational shifts in plant genomics by
emerging sequencing technologies

Innovation in DNA sequencing technologies that rapidly pro-
duce huge amounts of sequence information has triggered
a paradigm shift in genomics (Lister et al. 2009). A review on
the historical changes in DNA sequencing was provided by
Hutchinson (2007). A number of so-called NGS are available
(Gupta 2008) as widespread platforms, including the 454 FLX
(Roche) (Margulies et al. 2005), the Genome Analyzer/Hiseq
(Illumina Solexa) (Bennett 2004, Bennett et al. 2005) and the
SOLiD (Life Technologies), and newer platforms such as
Heliscope (Helicos) (Milos 2008) and PacBio RS (Pacific
Biosciences) (Eid et al. 2009) for single molecular sequencing,
and Ion Torrent (Life Technologies), based on a semi-
conductor chip (Rothberg et al. 2011), are also available. For
example, a long reader of NGS, the current version of the Roche
454 platform, is capable of generating 1 million reads in
an 400 bp run, and, a short reader, the current version of
the Illumina Hiseq2000, is capable of producing ~600 Gb of
sequence data in 100 bp x 2 paired end reads in a run. A
number of reviews on NGS technologies including experimen-
tal applications and computational methods have recently
been published (Lister et al. 2009, Varshney et al. 2009,
Horner et al. 2010, Metzker 2010). We provide an overview
of recent NGS-based approaches and outcomes in plants.

Genome and RNA sequencing by NGS

Several plant genomes have recently been sequenced using
NGS technology (Table 1) (Huang et al. 2009, S. Sato et al.

![Fig. 1](https://academic.oup.com/pcp/article-abstract/52/12/2017/1821864/links)
2011, Shulaev et al. 2011, Wang et al. 2011, Xu et al. 2011). Since de novo assembly of complex plant genomes remains a challenge, combinatorial approaches using Sanger and/or Roche pyrosequencing methods with other NGS platforms provide more efficient methods of assembly than does a single NGS platform. Re-sequencing coupled with reference genome sequencing data is a pronounced application that fulfills the features of NGS technologies. Rapid acquisition of genome-scale variant data sets enables high-throughput identification of the candidate mutations and alleles associated with phenotypic diversity (DePristo et al. 2011). A methodological pipeline to identify ethyl methanesulfonate (EMS)-induced mutations in Arabidopsis rapidly was developed by taking advantage of the NGS technology (Uchida et al. 2011). DNA polymorphisms such as single nucleotide polymorphisms (SNPs) and insertion–deletion polymorphisms (InDels) have been identified using NGS-based re-sequencing, which enabled the identification of even those polymorphisms in closely related ecotypes and cultivars (Yamamoto et al. 2010, Arai–Kichise et al. 2011). High-throughput polymorphism analysis is an essential tool for facilitating any genetic map-based approach. Genome-wide association study (GWAS) is rapidly becoming an effective approach to dissect the genetic architecture of complex traits in plants (Atwell et al. 2010). The aim of the Arabidopsis 1001 Genomes Project, launched at the beginning of 2008, is to discover the whole-genome sequence variation in 1,001 Arabidopsis strains (accessions) using NGS technologies (http://1001genomes.org). In the first phase of the project, whole-genome re-sequencing of 80 Arabidopsis strains from eight geographic regions revealed 4,902,039 SNPs and 810,467 small InDels (Cao et al. 2011). According to the Arabidopsis 1001 genomes data center, nucleotide polymorphisms of >400 sequenced strains are available (http://www.1001genomes.org/datacenter/). Re-sequencing of plant genomes using NGS technologies has been made possible by decreased costs and increased throughput, improving the efficiency of our analyses of the genotype–phenotype relationship (Rounsley and Last 2010).

**NGS-based transcriptome analysis**

NGS technologies in RNA-seq are a popular approach to collecting and quantifying the large-scale sequences of coding and non-coding RNAs rapidly (Wang et al. 2009c, Garber et al. 2011). RNA-seq has been used with NGS technologies to identify splicing variants accurately by mapping sequenced fragments onto a reference genome sequence such as in Arabidopsis and rice (Filichkin et al. 2010, Lu et al. 2010). These approaches have yielded information about alternative splicing events in transcriptomes at single base resolution as well as quantitative profiles of each isoform. Because it permits simultaneous acquisition of sequences, expression profiles and polymorphisms, NGS-based RNA-seq has been used for the rapid development of genomic resources in applied and emerging plant species (Table 2) (González-Ballester et al. 2010, Zenoni et al. 2010, Castruita et al. 2011, Gowik et al. 2011). Even when reference genome sequences are not available, de novo assembly from RNA-seq analysis enables rapid acquisition of preliminary genome-scale sequence data to provide early characterization of new plant species (Fu et al. 2011, Su et al. 2011, Wong et al. 2011).

NGS-based sequencing applications have rapidly expanded in plant genomics (Table 3). By browsing the Sequence Read Archive (SRA) in NCBI (http://www.ncbi.nlm.nih.gov/sra), European Nucleotide Archive (http://www.ebi.ac.uk/ena/home) and DDBJ Sequence Read Archive (http://trace.ddbj.nig.ac.jp/dra/index_e.shtml), all of which store raw sequencing

---

### Table 1: Outcomes from next-generation sequencing-based genome sequencing in plants

<table>
<thead>
<tr>
<th>Approach</th>
<th>Species</th>
<th>Analytical methods</th>
<th>Sequencing outcomes</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole-genome shotgun</td>
<td><em>Cucumis sativus</em></td>
<td>De novo genome</td>
<td>26,682 protein-coding genes predicted</td>
<td>Huang et al. (2009)</td>
</tr>
<tr>
<td>sequencing</td>
<td><em>Fragaria vesca</em></td>
<td>assembly</td>
<td>34,809 protein-coding genes predicted</td>
<td>Shulaev et al. (2011)</td>
</tr>
<tr>
<td></td>
<td><em>Jatropha carcas</em></td>
<td></td>
<td>40,929 protein-coding genes predicted</td>
<td>S. Sato et al. (2011)</td>
</tr>
<tr>
<td></td>
<td><em>Solanum tuberosum</em></td>
<td></td>
<td>39,031 protein-coding genes predicted</td>
<td>Xu et al. (2011)</td>
</tr>
<tr>
<td></td>
<td><em>Brassica rapa</em></td>
<td></td>
<td>41,174 protein-coding genes predicted</td>
<td>Wang et al. (2011)</td>
</tr>
<tr>
<td>Genome re-sequencing</td>
<td><em>Oryza sativa</em> cv. Koshikari</td>
<td>Mapping to reference</td>
<td>Identification of 67,051 SNPs</td>
<td>Yamamoto et al. (2010)</td>
</tr>
<tr>
<td></td>
<td><em>O. sativa</em> cv. Omashi</td>
<td>genome</td>
<td>Identification of 132,662 SNPs, 16,448 insertions and 19,318 deletions</td>
<td>Arai-Kichise et al. (2011)</td>
</tr>
<tr>
<td></td>
<td><em>O. sativa</em> (517 rice landsraces) (pooled sample)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Arabidopsis lyrata</em></td>
<td></td>
<td>High-density haplotype map generation and GWAS for 14 genomic traits</td>
<td>Huang et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>(80 ecotypes)</td>
<td></td>
<td>Identification of numerous mutations strongly associated with serpentine adaptation</td>
<td>Turner et al. (2010)</td>
</tr>
<tr>
<td></td>
<td><em>Arabidopsis thaliana</em></td>
<td></td>
<td>Whole-genome re-sequencing of 80 Arabidopsis strains from 8 geographic regions</td>
<td>Cao et al. (2011)</td>
</tr>
<tr>
<td></td>
<td><em>Zea mays</em> (6 elite inbred lines)</td>
<td></td>
<td>More than 1,000,000 SNPs, 30,000 InDels and 101 low-sequence-diversity chromosomal intervals</td>
<td>Lai et al. (2010)</td>
</tr>
<tr>
<td></td>
<td><em>Glycine max</em> (17 wild and 14 cultivars)</td>
<td></td>
<td>Identification of 205,614 tag SNPs</td>
<td>Lam et al. (2010)</td>
</tr>
</tbody>
</table>
data from NGS platforms, users can determine how thoroughly a given species has been sequenced and retrieve the publicly available sequencing data for further use.

### Emerging layers in plant omics

#### Plant epigenome analysis

The epigenome, genome-scale properties of epigenetic modifications, has garnered attention as a new omics area that has been advanced by NGS technology-based solutions (Schmitz and Zhang 2011). Epigenetic modifications in plants can be directed and mediated by small RNAs (sRNAs) (Matzke et al. 2009). The epigenomic regulation of chromatin structure and genome stability is crucial to the interpretation of genetic information (Law and Jacobsen 2010, He et al. 2011).

NGS technology-based sequencing of the cytosine methylome (methylC-seq), transcriptome (mRNA-seq) and Small RNA transcriptome (small RNA-seq) in Arabidopsis inflorescences revealed genome-scale methylation patterns and a direct relationship between the location of sRNAs and DNA methylation (Lister et al. 2008). At the whole plant level, a genome-scale map of methylated cytosine in Arabidopsis was generated by bisulfite sequencing using NGS technologies, the so-called BS-seq (Cokus et al. 2008). These NGS-based methylome analyses allowed for holistic understanding of genome-scale methylation patterns at a single base resolution. In addition to DNA methylation, histone N-terminal tail modifications such as acetylation, methylation are crucial in plant development (He et al. 2011) and defense mechanisms (Sokol et al. 2007, J.M. Kim et al. 2008). A genome-wide analysis of the

### Table 2: Examples of recent outcomes from next-generation sequencing-based transcriptome sequencing in plants

<table>
<thead>
<tr>
<th>Approach</th>
<th>Species</th>
<th>Properties</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA-seq with reference genome</td>
<td>A. thaliana</td>
<td>Confirmation of a majority of annotated introns and identification of</td>
<td>Filichkin et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>O. sativa indica and</td>
<td>thousands of novel, alternatively spliced mRNA isoforms</td>
<td></td>
</tr>
<tr>
<td></td>
<td>japonicus subspecies</td>
<td>15,708 novel transcriptionally active regions and ~48% of the genes</td>
<td>Lu et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>Chlamydomonas reinhardtii</td>
<td>Transcription analysis on acclimation to sulfur deprivation. RNA-seq</td>
<td>González-Ballester et al.</td>
</tr>
<tr>
<td></td>
<td>Vitis vinifera</td>
<td>showed larger dynamic range of expression than microarray</td>
<td></td>
</tr>
<tr>
<td>RNA-seq with cross-species mapping</td>
<td>Flaveria species</td>
<td>Transcription analysis during berry development</td>
<td>Castruita et al. (2011)</td>
</tr>
<tr>
<td>De novo assembly for RNA-seq data</td>
<td>Orchid species</td>
<td>A large collection of ESTs using NGS and in combination with Sanger</td>
<td>Zenoni et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>Phalaenopsis aphrodite</td>
<td>sequencing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acacia species</td>
<td>A large collection of ESTs using Roche 454 pyrosequencing and Illumina</td>
<td>Gowik et al. (2011)</td>
</tr>
<tr>
<td></td>
<td>Citrusus lanatus</td>
<td>Transcriptome analysis of species ranging from C₃ to C₄–C₈ intermediate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sesamum indicum</td>
<td>Transcriptome comparison between Flaveria bidentis (C₃) and</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Flaveria pringlei (C₄)</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3: Status of next-generation sequencing technology in plant species

<table>
<thead>
<tr>
<th>Taxonomic classification (common name)</th>
<th>No. of SRA experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlamydomonas reinhardtii</td>
<td>131</td>
</tr>
<tr>
<td>Physcomitrella patens</td>
<td>8</td>
</tr>
<tr>
<td>Solanum lycopersicum (tomato)</td>
<td>44</td>
</tr>
<tr>
<td>Solanum tuberosum (potato)</td>
<td>48</td>
</tr>
<tr>
<td>Glycine max (soybean)</td>
<td>160</td>
</tr>
<tr>
<td>Phaseolus vulgaris (kidney bean)</td>
<td>83</td>
</tr>
<tr>
<td>Medicago truncatula (barrel medic)</td>
<td>118</td>
</tr>
<tr>
<td>Quercus robur (truffle oak)</td>
<td>11</td>
</tr>
<tr>
<td>Manihot esculenta (cassava)</td>
<td>20</td>
</tr>
<tr>
<td>Populus (poplars)</td>
<td>20</td>
</tr>
<tr>
<td>Brassica napus (rape)</td>
<td>60</td>
</tr>
<tr>
<td>Brassica rapa (field mustard)</td>
<td>10</td>
</tr>
<tr>
<td>Arabidopsis lyrata (lyrate rockcress)</td>
<td>10</td>
</tr>
<tr>
<td>Arabidopsis thaliana (thale cress)</td>
<td>578</td>
</tr>
<tr>
<td>Vitis vinifera (wine grape)</td>
<td>22</td>
</tr>
<tr>
<td>Oryza sativa (rice)</td>
<td>760</td>
</tr>
<tr>
<td>Oryza sativa Japonica group (Japanese rice)</td>
<td>23</td>
</tr>
<tr>
<td>Hordeum vulgare (barley)</td>
<td>38</td>
</tr>
<tr>
<td>Triticum aestivum (bread wheat)</td>
<td>34</td>
</tr>
<tr>
<td>Triticum turgidum (poulard wheat)</td>
<td>22</td>
</tr>
<tr>
<td>Sorghum bicolor (sorghum)</td>
<td>5</td>
</tr>
<tr>
<td>Zea mays (maize)</td>
<td>161</td>
</tr>
<tr>
<td>Viridiplantae (green plants)</td>
<td>3,745</td>
</tr>
</tbody>
</table>

Plant species with ≥5 entries of SRA experiments are basically listed.

---

K. Mochida and K. Shinozaki

nucleosomes positioning combined with profiles of DNA methylation revealed 10 base periodicities in the DNA methylation status of nucleosome-bound DNA (Chodavarapu et al. 2010).

Chromatin immune precipitation (ChIP) plus NGS technologies or tiling array, the so-called ChIP-seq or ChIP-chip, has been used to generate a genome-scale map of histone modifications. For example, the genome-scale epigenetic map for 8-histone modification—H3K4me2, H3K4me3, H3K27me1, H3K27me2, H3K36me3, H3K56ac, H4K20me1 and H2Bub—was recently generated in Arabidopsis (Roudier et al. 2011). The Epigenomics of Plants International Consortium web site (https://www.plant-epigenome.org/) provides hyperlinks to plant epigenome data resources. In fact, numerous efforts have been made to acquire epigenome information from plant species (Table 4) (Wang et al. 2009b. He et al. 2010).

**Plant interactome analysis**

Protein–protein interactions are essential to almost all cellular processes. The interactome, a comprehensive set of all protein–protein interactions in an organism, is crucial to our understanding of the molecular networks underlying cellular systems (Cusick et al. 2005, Morsy et al. 2008). Interactome analysis has been used to characterize plant cellular functions such as the cell cycle, Ca++/calmodulin-mediated signaling, auxin signaling and membrane protein–signaling protein interactions in Arabidopsis (Boruc et al. 2010, Lalonde et al. 2010, Reddy et al. 2011, Vernoux et al. 2011). The Arabidopsis Interactome Mapping Consortium recently presented a proteome-wide binary protein–protein interaction map of Arabidopsis containing about 6,200 highly reliable interactions between about 2,700 proteins (Arabidopsis Interactome Mapping Consortium 2011).

Table 4  Epigenome analysis and resources in plant

<table>
<thead>
<tr>
<th>Resource</th>
<th>Species</th>
<th>URL</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiGNaL</td>
<td>Arabidopsis</td>
<td><a href="http://signal.salk.edu/cgi-bin/methylome">http://signal.salk.edu/cgi-bin/methylome</a></td>
<td>Zhang et al. (2006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://neomorph.salk.edu/epigenome/">http://neomorph.salk.edu/epigenome/</a></td>
<td>Lister et al. (2008)</td>
</tr>
<tr>
<td>Yale Plant Genomics</td>
<td>Rice and maize</td>
<td><a href="http://plantgenomics.biology.yale.edu/">http://plantgenomics.biology.yale.edu/</a></td>
<td>Li et al. (2008); Wang et al. (2009); He et al. 2010</td>
</tr>
<tr>
<td>Epigara</td>
<td>Arabidopsis</td>
<td><a href="http://epigara.biologie.ens.fr/cgi-bin/browse/a2e/">http://epigara.biologie.ens.fr/cgi-bin/browse/a2e/</a></td>
<td>Bouyer et al. (2011); Moghaddam et al. (2011); Roudier et al. (2011)</td>
</tr>
</tbody>
</table>

To generate the large-scale Arabidopsis interactome map, the consortium prepared about 8,000 open reading frames of Arabidopsis protein-coding genes, and then analyzed all pairwise combinations of the proteins encoded by these constructs using an improved high-throughput binary interactome mapping pipeline based on the yeast two-hybrid (Y2H) system. Using the Y2H pipeline method, a large-scale plant pathogen effector interactome network was generated (Mukhtar et al. 2011). In rice, a focused interactome analysis addressed biotic and abiotic stress responses (Seo et al. 2011). A number of databases for protein–protein interaction data sets have been available on the web (Table 5) (Stark et al. 2006, Swarbreck et al. 2008, Aranda et al. 2010). In addition to the curated data sets, predicted protein–protein interaction data sets are a valuable complement to experimental approaches (Cui et al. 2008, Lin et al. 2009, De Bodt et al. 2010, Li et al. 2011, Lin et al. 2011a, Lin et al. 2011b).

**Plant hormonome analysis**

Plant hormones play a crucial role as signaling molecules in the regulation of plant development and environmental responses. To date, a number of low molecular weight plant hormones have been identified, including auxin, ABA, cytokinin, gibberellins, ethylene, brassinosteroids, jasmonates and salicylic acid (Davies 2004). In addition, a novel plant hormone, strigolactone, was recently identified as a shoot branching inhibitor (Gomez-Roldan et al. 2008, Umehara et al. 2008). A special issue of *Plant and Cell Physiology* on strigolactone was published to collate current knowledge on the subject (Yamaguchi and Kyozuka 2010). Small peptides (peptide hormones) also function as signaling molecules in the regulation of plant growth and development (Matsubayashi and Sakagami 2006, Fukuda et al. 2007). A special issue on peptide hormones was also published to describe recent advances in the area of plant peptide research (Fukuda and Higashiyama 2011).

During the last decade, there have been many remarkable advances in our understanding of the molecular basis of plant hormones, including biosynthesis, transport, perception and response (Santner and Estelle 2009). Umezawa et al. in 2010 provided a review of recent exciting advances in understanding the molecular basis of regulatory networks in the ABA response. One recent remarkable advance was the discovery of the receptors for several important phytohormones, including auxin, gibberellins, ABA and jasmonates (Santner and Estelle 2009). Structural analysis of each complex revealed the structural basis of the interaction between each receptor and phytohormone and their signaling mechanisms (Tan et al. 2007, Shimada et al. 2008, Miyazono et al. 2009, Sheard et al. 2010). According to a number of recent mutant analyses, it is almost certain that all plant hormones cross-talk with one or more other hormones depending on the tissue, developmental stage and environmental changes (Santner and Estelle 2009, Depuydt and Hardtke 2011).
Because of the interplay between multiple plant hormones, a comprehensive analysis known as hormonome analysis, which is based on high-throughput, high-sensitivity and simultaneous profiling of plant hormones, is a key approach to a holistic understanding of the plant hormone network and its association with biological phenomena. A recently developed analytical platform for high-sensitivity, high-throughput measurements of plant hormones enables simultaneous measurement of 43 molecular species of cytokinin, auxin, ABA and gibberellin (Kojima et al. 2009). The platform was used to acquire hormonome profiles of the organ distribution patterns of plant hormones in rice. The hormonome analysis of endogenous levels of cytokinin, gibberellin, ABA and auxin in the wild type and gibberellin signaling mutants indicated that the metabolism of cytokinin, ABA and auxin cross-talk with the gibberellin signaling system. Comprehensive hormone profiling in Arabidopsis was used to analyze the accumulation of ABA, gibberellins, IAA, cytokinins, jasmonates and salicylic acid in the developing seeds of wild-type Arabidopsis and an ABA-deficient mutant. The results of the hormonome approach suggested that ABA interacts with other hormones to regulate seed development (Kanno et al. 2010).

### Plant metabolome analysis

Plant metabolomics is now playing a significant role in systems approaches to plant functional analysis and applied plant biotechnology. Driven by advances in related technologies including instruments for metabolite measurement, analytical methodologies and information resources, there are many applications for functional genomics, systems biology and molecular breeding. A number of excellent metabolomics reviews have been published that describe emerging methodologies and attractive applications (Last et al. 2007, Saito and Matsuda 2010, Sumner 2010). Herein, we will briefly review recent advances in analytical platforms and then describe examples of practical applications to understanding plant metabolic systems.

Metabolome analysis deals with chemically diverse compounds. The plant metabolome consists of extremely large varieties of metabolites with various dynamic concentration ranges. Therefore, combined analytical techniques and data set integration from heterogeneous instruments have been key to a comprehensive understanding of diverse metabolites. Streamlined analytical platforms that integrate analytical steps such as sample preparation, data acquisition and data analysis enable us to address the complex plant metabolome (Saito and Matsuda 2010). Improved coverage and throughput for the simultaneous detection of large numbers of metabolites has significantly expanded the practical application of metabolome analysis. A widely targeted metabolomics platform that provides both coverage and throughput utilizes ultra-performance liquid chromatography–tandem quadrupole mass spectrometry (Sawada et al. 2009a). The approach enables us simultaneously to acquire accumulation patterns of hundreds or more metabolites for large numbers of samples. The platform enables us to address complex plant metabolic systems and develop practical approaches in genetics and breeding. The widely targeted metabolomics approach was used for metabolic profiling of the Arabidopsis knockout mutants for methionine chain elongation enzymes. The results suggest that these enzymes are involved in metabolism from methionine to primary and related secondary metabolites (Sawada et al. 2009b).

In addition to targeted metabolomics, non-targeted metabolome analysis with known and unknown metabolites has been important in the generation of comprehensive resources.
for the plant metabolome. Tandem mass spectrometry (MS/MS) spectral tag (MS2T) analysis, as a non-target metabolome analysis, was used to profile secondary plant metabolites and identify many previously unknown tissue-specific metabolites (Matsuda et al. 2009). The profile of accumulation of known and unknown metabolites was acquired in plant tissues throughout the Arabidopsis life cycle; the data set is available at AtMetExpress (Matsuda et al. 2010).

Metabolome profiling provides a snapshot of the accumulation patterns of metabolites in response to various kinds of biological conditions such as treatments, tissues and genotypes. For example, metabolome profiling approaches have been applied to monitor changing metabolite accumulation in response to stress conditions (Shikawa et al. 2009, Kusano et al. 2011b). Metabolome profiling has also been applied to evaluate genetic resources, not only in the model plants Arabidopsis and rice but also in various crop species for metabolic phenotyping (Akihiro et al. 2008, Mochida et al. 2009a, Yin et al. 2010, Fujimura et al. 2011, Kusano et al. 2011a). Metabolome profiling has also been used to evaluate the metabolic phenotypes of natural and/or segregation populations. A number of approaches in metabolite quantitative trait locus (mQTL) analysis have been performed in various plant species in recent years (Rowe et al. 2008, Lisec et al. 2009, Do et al. 2010).

A number of information resources have become available in recent years to provide tools for metabolome analysis, such as PRIME (http://prime.psc.riken.jp/), MeRy-B (http://www.cbib.u-bordeaux2.fr/mer yab/) and MetabolomeExpress (https://www.metabolome-express.org/) (Akiyama et al. 2008, Carroll et al. 2010, Ferry-Dumazet et al. 2011). The review by Saito and Matsuda (2010) provides a well summarized list of web resources for plant metabolomics. Synergistic application of comprehensive, high-throughput experimental procedures and computational approaches as well as genome-scale outcomes from other omics should enable the reconstruction of metabolic systems by metabolic modeling, simulation and theoretical network engineering (Stitt et al. 2010).

Omics resources in emerging plant species

The development of genomic resources has progressed in a number of plant species. As a typical example, genome sequence data sets of 25 plant species are available at Phytozome (ver. 7.0, http://www.phytozome.net/). Recently sequenced plant species were typically nominated for genomic resource development because: (i) they have particular systems not covered by ‘conventional’ model plants; (ii) they are evolutionarily important; or (iii) they provide a commodity resource such as food and energy. Here, we briefly review the current status of available genomic resources for emerging plant species with a focus on examples of the Solanaceae, Poaceae (Gramineae) and Fabaceae (Leguminosae) families.

Solanaceae species

The Solanaceae family includes a number of important agricultural crops, such as tomato, potato, pepper, paprika, petunia and tobacco. The tomato (Solanum lycopersicum) is a representative crop species for which there has been significant progress in genomic resources. The tomato is an important crop that is sold fresh and used in processed foods. In addition to its agricultural importance, the tomato is a model plant for studying Solanaceae species due to its small genome size and shared conserved synten with other Solanaceae genomes. The tomato has also become a model plant for the study of fruit development, ripening, maturation and metabolic systems. The International Tomato Genome Sequencing Project was initiated in 2004. Following the initial BAC-by-BAC approach for the euchromatic regions, a whole-genome shotgun approach was initiated in 2009. The International Tomato Annotation Group provided the official annotation of the tomato genome assembly (http://solgenomics.net/organism/Solanum_lycopersicum/genome). A full-length cDNA resource from the tomato cultivar Micro-Tom was recently launched (Aoki et al. 2010; http://www.pgb.kazusa.or.jp/kafom/). Transcriptome data from 296 samples of 16 series using the Affymetrix GeneChip tomato genome array can be found in NCBI GEO (September 8, 2011). The tomato GeneChip data deposited in NCBI GEO includes, for example, data sets acquired for co-expression analysis using cultivar Micro-Tom (Ozaki et al. 2010), for comparative transcriptome analysis between salt-tolerant and salt-sensitive wild tomato species (Sun et al. 2010) and for examining the transcriptome of the ripening process of an orange ripening mutant (Nashilevitz et al. 2010).

Significant progress has been made with metabolome analyses such as metabolome profiling and mQTL analysis (Schauer et al. 2008, Do et al. 2010, Enfissi et al. 2010, Schilmiller et al. 2010, Tieman et al. 2010). Information resources related to metabolome analysis are available and updated, and provide data archives for tomato metabolome data sets and analytical platforms such as Plant MetGenMAP (Joung et al. 2009), The Metabolome Tomato Database (MotoDB) (Moco et al. 2006), KaPPA-View4 SOL (Sakurai et al. 2011a) and KOMICS (Iijima et al. 2008). As an integrative information resource, the Tomato Functional Genomics Database provides data on gene expression, metabolites and microRNAs (miRNAs) through a web interface (Fei et al. 2011). As a tomato mutant resource, TOMATOMA was launched as a web-based database for a Micro-Tom EMS mutant collection with phenotypic classifications and a Targeting Induced Local Lesions IN Genomes (TILLING) resource was established (Okabe et al. 2011, Saito et al. 2011). The potato genome was recently sequenced using a homozygous doubled-monoploid potato clone, and assembly of 86% of the 844 Mb genome revealed 39,031 predicted protein-coding genes (Xu et al. 2011; http://www.potatogenome.net/index.php/Main_Page).
A large-scale expressed sequence tag (EST) analysis of the chili pepper yielded 116,412 ESTs with related annotations that are available at the pepper EST database (http://genepool.kribb.re.kr/pepper/) (Kim et al. 2008). To integrate collected ESTs from Solanaceae species, the SoLEST database (D’Agostino et al. 2009) includes ESTs and related annotations from the tomato, potato and pepper (http://biosrv.cab.unina.it/solestdb/index.php). The Sol Genomics Network (SGN; http://solgenomics.net/) is an information portal for Solanaceae research that provides broad information on genomic resources for Solanaceae species and their close relatives (Bombarely et al. 2011).

**Poaceae (Gramineae) species**

The Poaceae family includes staple food crops such as rice, maize, wheat and barley, as well as grasses used for lignocellulose biomass production, such as switchgrass and Miscanthus (Somerville et al. 2010, Lobell et al. 2011). Since the completion of the japonica rice (*Oryza sativa*) genome project (International Rice Genome Sequencing Project 2005), whole-genome sequences have been completed in sorghum (*Sorghum bicolor*), maize (*Zea mays*) and *Brachypodium* (*Brachypodium distachyon*) (Paterson et al. 2009, Schnable et al. 2009, International Brachypodium Initiative 2010). Rice is a model species of monocots plants as well as one of the three major staple cereals in the world. So far, the genome sequence of japonica rice with high quality gene annotations has played a significant role in promoting the development of a number of genomic resources for the discovery and isolation of important genes for application in molecular breeding. The sorghum genome was sequenced as a representative species of the Saccharinae that includes biomass source plants for starch, sugar and cellulose. Maize is another of the major staple cereals for food and animal feed and is a model organism for fundamental research in complex inheritance and genomic properties such as domestication, epigenetics, evolution, chromosome structure and transposable elements (Walbot 2009). Accompanying the release of the maize B73 genome sequence, the ‘2009 Maize Genome Collection’ was edited (Walbot 2009).

*Brachypodium* is an emerging plant species of the Pooideae subfamily, a model plant for Triticeae crops such as wheat and barley, as well as for understanding systems for cellulose biomass production in grass species. Since the release of the *Brachypodium* Bd21 genome sequence, *Brachypodium* has garnered attention, and a number of genomic resource projects have been initiated at various institutions (Brkljacic et al. 2011). We thus have access to published reference genome sequences of four species from each of the three important Poaceae subfamilies. Wheat and barley have been the subjects of ongoing genome sequencing attempts to regenerate their highly complex genomes. By incorporation of chromosome sorting, NGS, array hybridization and conserved synteny with the *Brachypodium* genome, a tentative linear order of 32,000 barley genes was recently regenerated (Mayer et al. 2011). A special issue on barley was recently published to introduce genetic and genomic resources and their applications (Saisho and Takeda 2011).

The large-scale collection of ESTs and cDNA clones has been performed in Poaceae species such as rice, maize, wheat and barley (Mochida and Shinozaki 2010). Those sequence and clone resources have facilitated gene discovery and structural annotation, large-scale expression analysis, genome-scale interspecific and intraspecific comparative analysis of expressed genes, and the design of expressed gene-oriented molecular markers and probes for microarrays (Mochida et al. 2003, Ogihara et al. 2003, Zhang et al. 2004, Kawaura et al. 2006, Mochida et al. 2006, Mochida et al. 2008, Sato et al. 2009a). Large-scale analyses of full-length cDNA sequences and clone resources have been performed in rice, maize, wheat and barley (Kikuchi et al. 2003, Liu et al. 2007, Kawaura et al. 2009, Sato et al. 2009b, Soderlund et al. 2009, Matsumoto et al. 2011). Full-length cDNA resources have been crucial for the annotation of the sequenced genome (Tanaka et al. 2008), large-scale gene discovery and functional analyses by creating transgenic plants (Kondou et al. 2009), and comparative analysis based on the entire sequence of transcripts (Mochida et al. 2009b) in grass species.

Data sets of transcriptome profiles are available for several Poaceae species. For example, the current version of Genevestigator provides processed transcriptome data from GeneChip hybridization data sets (1,626 for barley, 1275 for rice, 1000 for wheat and 458 for maize (https://www.genevestigator.com/gv/). The transcriptome throughout the reproductive process from primordia development through pollination/fertilization to zygote formation was analyzed in rice using an oligomicroarray as a rice expression atlas (Fujita et al. 2010). Accumulated transcriptome data sets have been made for co-expression analysis of the transcriptome in Poaceae species. The RiceArrayNet and OryzaExpress databases provide web-accessible co-expression data for rice (Lee et al. 2009, Hamada et al. 2011). The ATTED-II database also provides co-expression data sets for rice in addition to those for Arabidopsis (Obayashi et al. 2011). A co-expressed barley gene network was recently generated and then applied to comparative analysis to discover potential Triticeae-specific gene expression networks (Mochida et al. 2011a).

Microarray analyses coupled with laser microdissection have been applied to analyze transcriptomes of the male gametophyte and tapetum in rice (Hobo et al. 2008, Suwabe et al. 2008). Transcriptome data were recently collected from rice grown in field environments (Y. Sato et al. 2011). As a proteomics resource in Poaceae, for example, the plant proteome database (http://ppdb.tc.cornell.edu/) provides information on the maize and Arabidopsis proteomes. The RIKEN Plant Phosphoproteome Database (RIPP-DB, http://phosphoproteome.psc.database.riken.jp) was updated with a data set of large-scale identification of rice phosphorylated proteins.
The large Fabaceae family includes economically important legume crops such as the soybean, common bean and alfalfa. The symbiotic nitrogen fixation that is produced by communication between plants and nitrogen-fixing bacteria is a biological phenomenon, particularly in legume species. Understanding this symbiosis is important for plant and microbial biology as well as for agriculture. Recent progress in symbiosis research between plants and microbes, including nitrogen-fixing symbiosis in legume plants, was presented in a special issue on plant–microbe symbiosis that appeared in a special issue on plant–microbe symbiosis in legume plants, was presented in a special issue on plant–microbe symbiosis (Ikeda et al. 2010) and Kouchi et al. (2010). It was also favorably applied for the homology-based gene prediction (Umezawa et al. 2008). Several genomic resources are available for soybean genomics as well as for molecular breeding for improved productivity and stress tolerance (Manavalan et al. 2009). A recent review was published by Tran and Mochida (2010). By using the soybean genome sequence and annotated gene models, genome-scale exploration of gene families and those functional analyses have been performed to identify genes for molecular breeding (Mochida et al. 2009c, Mochida et al. 2010a, Mochida et al. 2010b, Le et al. 2011a, Le et al. 2011b). A transcriptome atlas of the soybean (http://digbio.missouri.edu/soybean_atlas/) has been developed using an NGS platform to perform RNA-seq of samples from 14 distinct conditions (Libault et al. 2010). NGS-based approaches were also applied to a genome-scale survey of sRNAs (Joshi et al. 2011, Song et al. 2011). As an information portal for soybean research, Soybase (http://soybase.org/) has played a significant role in integrating various resources and analytical platforms for soybean research (Grant et al. 2010).
Data integration to progress plant omics research

Recent remarkable innovations in omics platform research have produced a wealth of genome-scale data in the life sciences. The key problem in omics research now is developing ways to deal with such huge and heterogeneous data sets. A conceptual relationship of omics data sets inter-related with a gene is represented as an example for the integration of omics data sets (Fig. 2). Information resources such as databases and computational tools are becoming more and more important for effectively handling genome-scale data sets. Data storage for omics data sets must ensure persistence and retrieval functionalities for shared use. To integrate heterogeneous data sets effectively, the formation of procedures to generate inter-relationships between data sets and consolidate the data structure for each feature is another goal. User interface and visualization techniques to allow users to perform heuristic data mining and to be inspired by integrated genome-scale data sets are also important. Many information resources are available to facilitate plant science (Mochida and Shinozaki 2010). Here, we describe representative methods and visualizations of genome-scale data sets together with examples of currently available resources.

Creating relationships using various kinds of data sets

With the completion of a number of plant genome sequences, genome-scale evolutionary and comparative analyses have allowed us to identify conserved and/or characteristic properties among plant species. Comparative genomics resources integrate genome sequences from multiple organisms and associated information with evolutionary insights. For example, PLAZA, an online platform for plant comparative genomics (http://bioinformatics.psb.ugent.be/plaza/), provides structural and functional annotations of published plant genomes and analytical tools (Proost et al. 2009). The schematic representation of the relationships between the data types and tools of the PLAZA platform is illustrative of the current state of data integration in comparative genomics. As another example, the SALAD database (http://salad.dna.affrc.go.jp/salad/) provides comparative information on 10 sequenced plant species based on the organization of shared sequence motifs (Mihara et al. 2010). The SALAD database was used to discover candidate cis-regulatory elements by synergistic use of the phylogenetic relationships of genes with gene expression profiles and upstream flanking sequence features (Mihara et al. 2008).

Comparative analyses of plant species have also been performed on the transcriptome. PlaNet (http://aranet.mpimp-golm.mpg.de/), a database of co-expression networks for Arabidopsis and six plant crop species, uses a comparative network algorithm, NetworkComparer, to estimate similarities between network structures (Mutwil et al. 2011). The PlaNet platform integrates gene expression patterns, associated functional annotations and MapMan term-based ontology, and facilitates knowledge transfer from Arabidopsis to crop species for the discovery of conserved co-expressed gene networks.

The metabolic pathway map could be a knowledge foundation to superimpose profiles of each compound, proteome and transcriptome of the genes involved in each pathway. The Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.jp/kegg/) is one of the most long established integrated information resources. The KEGG PLANT Resource provides information on biosynthetic pathways, sequenced plant genomes and phytochemicals, and it aims to integrate genomic information resources with the biosynthetic pathways of natural plant products (Masoudi-Nejad et al. 2007). Another information resource for biosynthetic pathways, PlantCyc, contains curated information from the literature and from computational analyses of the genes, enzymes, compounds, reactions and pathways involved in primary and secondary plant metabolism. PlantCyc recently added PoplarCyc for Populus trichocarpa in addition to AraCyc for Arabidopsis (Zhang et al. 2010), which is based on the BioCyc platform (Caspi et al. 2010). This platform has been used for a number of plant species. The pathways section in the Gramene databases provides RiceCyc, MaizeCyc, BrachyCyc and SorghumCyc, for rice, maize, Brachypodium and sorghum, respectively (http://www.gramene.org/pathway/). The web page also provides mirrors of species-specific pathway databases from Arabidopsis, tomato, potato, coffee, Medicago, Escherichia coli, and the MetaCyc and PlantCyc reference databases, enabling us to perform interspecific comparison between pathways (Youens-Clark et al. 2010).

Sequence identity- and/or similarity-based integration has been widely used to build cross-reference data sets between individual genes and their associated references, such as gene–genomic structure, gene–transcript/cDNA clones–protein, gene–transcript expression pattern and gene–polymorphism/mutants–phenotype. For example, an integrative database for the genes encoding transcription factors of the Gramineae species, GramineaeTFDB (http://gramineaeftdb.psc.riken.jp/) provides genomic sequence features, promoter regions, domain alignments, GO assignments, full-length cDNAs and gene expression profiles, and cross-references to various public databases and genetic resources (Mochida et al. 2011b). Genome sequence-oriented data integration is essential to the representation of genomic characteristics such as whole-genome transcription, genome-wide methylene patterns and genome-wide polymorphisms (Fig. 3). The ARabidopsis Tiling-Array-based Detection of Exons database 2 (http://artade.org) provides precisely predicted gene structures by tiling array-based transcriptome data sets (Iida et al. 2011). The database also infers gene function based on over-representation in gene model co-expression analyses. The SiGnAL Arabidopsis Methylome Mapping Tool (http://signal.salk.edu/cgi-bin/methylome) integrates existing Arabidopsis gene annotation
and methylome information at single base pair resolution. The PhosPhAt database of Arabidopsis phosphorylation sites (http://phosphat.mpimp-golm.mpg.de/index.html) is built on mass spectrometry data (Durek et al. 2009). The database contains a collection of phosphopeptides and phosphorylation sites assigned to annotated Arabidopsis genes, and hyperlinks to external information resources such as The Arabidopsis Information Resource (TAIR). The RNA-editing site on protein 3D structure is a unique database of RNA-editing sites found in plant organelle genes with the results mapped onto amino acid sequences and 3D structures (Yura et al. 2009).

Literature mining for gene associations is also an essential approach to knowledge integration in the life sciences (Krallinger et al. 2008, Winnenburg et al. 2008). PLAN2L (plant annotation to literature, http://zope.bioinfo.cnio.es/plan2l) is a web-based online application that integrates literature-derived bioentities and associated information in Arabidopsis (Krallinger et al. 2009). PosMed-plus (positional Medline for plant upgrading science, http://omicspace.riken.jp/PosMed-plus/) is a web-accessible tool that assists in candidate selection for positional cloning in plants (Makita et al. 2009). The accuracy of PosMed-plus is correlated with its ability to make correct associations between genes and documents that are based on direct searches, inference searches by co-citation and manual curation.

**Visualization of plant omics data sets**

Observation is the most basic approach in biology. Therefore, visualization of genome-scale data sets is an important task in bioinformatics. Gehlenborg et al. (2010) reviewed a number of practical tools for omics data visualization and integration. The goal of omics data visualization should be to create clear, meaningful and integrated resources without being overwhelmed by the intrinsic complexity of the data (Gehlenborg et al. 2010). We must remove scale dependency and implement
seamless navigation between associated data to uncover their true value as a heuristic information resource.

The networks formed by associated biomolecules such as protein–protein interactions and co-expressed genes can be visualized to gain biological insights from genome-scale data sets of experimentally confirmed and/or computationally predicted molecular associations. For example, the node graph-based network visualization of co-expressed genes has been widely used in various plant species and has been implemented in web-accessible information resources such as ATTED-II, OryzaExpress and PlaNet (Hamada et al. 2011, Mutwil et al. 2011, Obayashi et al. 2011). SeedNet and SCoPNet, at The Virtual Seed Web Resource (http://vseed.nobum.ac.uk/), are genome-wide network models of transcriptional interactions in dormant and germinating Arabidopsis seeds (Bassel et al. 2011a, Bassel et al. 2011b). AtCAST (Arabidopsis thaliana: DNA Microarray Correlation Analysis Tool, http://atcast.sd.gkoma.su/cgi/network/home.cgi) is a unique information resource that combines multiple Arabidopsis microarray experiments into a single network (Sasaki et al. 2010).

Visualization of the spatio-temporal accumulation of biomolecules allows us to infer the biological significance of accumulated molecules and their function in different developmental stages, environmental conditions and tissues. The electronic fluorescent pictograph (eFP) browser (http://www.bar.utoronto.ca/) illustrates the visualization of spatio-temporal accumulation patterns of biomolecules (Winter et al. 2007). The current version of the Arabidopsis eFP browser consists of gene expression patterns of different developmental stages, tissues, stresses and natural variations. In addition to Arabidopsis, eFP browsers are available for poplar, M. truncatula, rice, barley, maize and soybean.

The genome browser is a key tool that integrates sequence-based information with genomic position. Several tools such as the genome browser are available in various web-accessible resources. For example, Gbrowse is a popular genome browser that is widely used on a number of web sites, such as TAIR, to visualize genome annotations (Podicheti and Dong 2011). With the increase in NGS-based sequencing data, the types of genomic features that can be visualized throughout genome sequences have been expanded and include, for example, transcript abundances based on RNA-seq, polymorphisms based on whole-genome re-sequencing, and quantitative Chip-seq data (Fig. 3). Circos (http://circos.ca/) is another genome browser that visualizes genome(s) in a circular layout. The circular layout of Circos facilitates visual annotations of the circular genomes of microbes as well as the structural relationships between chromosomal regions such as whole-genome duplications and syntenic relationships (Krzywinski et al. 2009).

Imaging provides immediate visualization of biological information. Along with technological advances in bioimaging, the informatics of bioimaging data sets is an emerging field (Moore et al. 2008). In plant cell biology, advanced imaging technologies allow us to visualize organelles and biomolecules and characterize cellular systems based on large-scale data sets of images and movies (Mano et al. 2009). Image databases in plant science were reviewed by Mano et al. in 2009. For example, the aim of the plant organellome database (http://podb.nibb.ac/organellome) is to promote the understanding of organelle dynamics such as organelle function, biogenesis, differentiation, movement and interactions with other organelles (Mano et al. 2008, Mano et al. 2011). Furthermore, comprehensive acquisitions of cellular images and movies are essential in the development of information resources to address computationally quantitative characterizations of plant cellular properties. Recently, various computational procedures have been developed to analyze cellular image data sets. For example, a computational algorithm based on machine learning and statistical modeling was applied to recognize subcellular localization of proteins in microscope images automatically (Hu et al. 2010). Multi-angle image acquisition, three-dimensional reconstruction and cell segmentation-automated lineage tracking was developed and applied to perform quantitative analysis of Arabidopsis flower development at cell resolution (Fernandez et al. 2010).

**Systems analysis of plant functions**

Systems analysis based on a combination of multiple omics analyses has been an efficient approach to determining the global picture of cellular systems. From the early period of plant metabolomics research, we have achieved a number of significant advances in our understanding of gene function in metabolic systems by integrating metabolome analysis with genome and transcriptome resources (Hirai et al. 2004, Toghe et al. 2005, Hirai et al. 2007, Hirai and Saito 2008, Saito et al. 2008, Watanabe et al. 2008, Yonekura-Sakakibara et al. 2008, Okazaki et al. 2009). Following these successes, multi-omics-based systems analyses have improved our understanding of plant cellular systems. For instance, integrated metabolome and transcriptome analyses were recently applied to analyze rice developing carpoxys under high temperature conditions (Yamakawa and Hakata 2010), molecular events underlying pollination-induced and pollination-independent fruit sets (Wang et al. 2009a) and the effects of **DE-ETIOLATED1** down-regulation in tomato fruits (Enfissi et al. 2010). Integrated metabolome and transcriptome analysis has also been applied to investigate changing metabolic systems in plants growing in field conditions, such as the rice Os**-**GIGANTEA (Os**-**GI) mutant and transgenic barley (Kogel et al. 2010, Izawa et al. 2011). Furthermore, a systems approach combined hormone, metabolome and transcriptome analyses in Arabidopsis transgenic lines, displaying increased leaf growth to gain insight into the molecular mechanisms that control leaf size (Gonzalez et al. 2010).
An integrated proteome and metabolome analysis was applied to compare the differences in response to anoxia between rice and wheat coleoptiles (Shingaki-Wells et al. 2011). An integrated transcriptome, proteome and metabolome analysis was conducted to characterize the cascading changes in UV-B-mediated responses in maize (Casati et al. 2011). These illustrative examples demonstrate the power of multi-omics-based systems analysis for understanding the key components of cellular systems underlying various plant functions.

Fig. 3 An example of integrated biological instances in a genome sequence. A gene annotation and related data, phosphorylation site, cDNA/EST, homology-based mapping, mutants and polymorphisms were retrieved from TAIR. Methylome data sets were retrieved from the Arabidopsis Epigenetics & Epigenomics group’s web site. Each data set has been allocated onto the genomic region defined as a TAIR8 annotation.
Thanks to recent technical advances of high-throughput and genome-scale genotyping platforms such as whole-genome re-sequencing, comprehensive exploration of the association between genomic diversity and quantitative instances in various omics aspects has facilitated the discovery of key genes involved in adaptive changes in various omics levels as another combinatorial approach (Kroymann 2011). Using a large set of Arabidopsis accessions and genome-scale variation data sets, GWAS identified genetic loci associated with enzyme activities, metabolome profiles and biomass (Chan et al. 2010, Sulpic et al. 2010). The hormonal responses of natural variations have been addressed to find relationships between physiologic-al variations of hormonal response and other variations, such as in the genome and transcriptome (Delker et al. 2010). A combinatorial approach to population genomics using hormone profiling would allow us to identify the association between genome polymorphisms and plant hormone abundance as quantitative traits that might be closely related to environmental adaptation. Recently, relational instances of epigenomic modification, coding gene transcription and non-coding RNAs have been coupled with genome-scale nucleotide polymorphism data sets in human population genomics (Sigurdsson et al. 2009, Shoemaker et al. 2010). In similar fashion, plant epigenome analysis can also be integrated with genome-scale variations to provide important clues to the epigenetic and genetic regulation associated with phenotypic diversity.

During the past few years, we have witnessed significant advances in plant omics, as shown in this review. Technological advances in analytical platforms are providing genome-scale outcomes, revitalizing our knowledge and opportunities to address long-standing basic questions in plant science. At the same time, we are also facing various global issues such as water, food and energy security; global warming; and climate changes. Understanding the specific plant functions that appear in particular plant species is also important to discover useful genes to improve plant functions. The integration of a wide spectrum of omics data sets from various plant species is then essential to promote translational research to engineer plant systems in response to the emerging demands of mankind.

Funding

This work was supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan [Grant-in-Aid for Scientific Research on Innovative Areas (23119524) to K.M.].

Acknowledgments

The authors thank Tetsuya Sakurai, Takuhiro Yoshida, Lam-Son Phan Tran and Yuji Sawada of the RIKEN Plant Science Center, and Kei Iida of RIKEN BASE for their valuable suggestions and comments. The authors also thank Daisuke Saisho of Okayama University for critical reading of the manuscript.

References


analysis and transparent dissemination of GC/MS metabolomics datasets. BMC Bioinformatics 11: 376.


Mochida, K., Yoshida, T., Sakurai, T., Yamaguchi-Shinozaki, K., Shinozaki, K. and Tran, L.S. (2011b) In silico analysis of transcription factor repertoires and prediction of stress-responsive transcription factors from six major Gramineae plants. *DNA Res.*


