

Genomics and Bioinformatics Resources for Crop Improvement

Keiichi Mochida and Kazuo Shinozaki*

RIKEN Plant Science Center, Yokohama, 230-0045 Japan

*Corresponding author: E-mail, sinozaki@rtc.riken.jp; Fax, +81-45-503-9580

(Received February 8, 2010; Accepted March 1, 2010)

Recent remarkable innovations in platforms for omics-based research and application development provide crucial resources to promote research in model and applied plant species. A combinatorial approach using multiple omics platforms and integration of their outcomes is now an effective strategy for clarifying molecular systems integral to improving plant productivity. Furthermore, promotion of comparative genomics among model and applied plants allows us to grasp the biological properties of each species and to accelerate gene discovery and functional analyses of genes. Bioinformatics platforms and their associated databases are also essential for the effective design of approaches making the best use of genomic resources, including resource integration. We review recent advances in research platforms and resources in plant omics together with related databases and advances in technology.

Keywords: Bioinformatics • Database • Omics resource.

Abbreviations: AT, activation tagging; CE, capillary electrophoresis; CRES-T, chimeric repressor silencing technology; DIGE, difference gel electrophoresis; ESI, electrospray ionization; EST, expressed sequence tag; FT, Fourier transfer; FT-ICR MS, Fourier transform ion cyclotron resonance mass spectrometry; GC, gas chromatography; GST, gene-specific sequence tag; LC, liquid chromatography; MALDI, matrix-assisted laser desorption ionization; miRNA, microRNA; MPSS, massively parallel signature sequencing; MS, mass spectrometry; NMR, nuclear magnetic resonance; QTL, quantitative trait locus; SAGE, serial analysis of gene expression; RNAi, RNA interference; sRNA, small RNA; siRNA, short interfering RNA; SNP, single nucleotide polymorphism; SSR, simple sequence repeat; STS, sequence-tagged site; ta-siRNA, trans-acting siRNA; TF, transcription factor; TILLING, targeting induced local lesions in genomes; TOF, time of flight; TSS, transcription start sites; VIGS, virus-induced gene silencing.

Introduction

Sustainable agricultural production is an urgent issue in response to global climate change and population increase (Brown and

Funk 2008, Turner et al. 2009). Furthermore, recent increased demand for biofuel crops has created a new market for agricultural commodities. One potential solution is to increase plant yield by designing plants based on a molecular understanding of gene function and on the regulatory networks involved in stress tolerance, development and growth (Takeda and Matsuoka 2008). Recent progress in plant genomics has allowed us to discover and isolate important genes and to analyze functions that regulate yields and tolerance to environmental stress.

The whole genome sequencing of *Arabidopsis thaliana* was completed in 2000 (The Arabidopsis Genome Initiative 2000). Subsequently, the National Science Foundation (NSF) Arabidopsis 2010 project in the USA was launched with the stated goal of determining the functions of 25,000 genes of Arabidopsis by 2010 (Somerville and Dangl 2000). Technological advances in each omics research area have become essential resources for the investigation of gene function in association with phenotypic changes. Some of these advances include the development of high-throughput methods for profiling expressions of thousands of genes, for identifying modification events and interactions in the plant proteome, and for measuring the abundance of many metabolites simultaneously. In addition, large-scale collections of bioresources, such as mass-produced mutant lines and clones of full-length cDNAs and their integrative relevant databases, are now available (Brady and Provart 2009, Kuromori et al. 2009, Seki and Shinozaki 2009). The genome sequencing project of *japonica* rice was completed in 2005, and the Rice Annotation Project (RAP), which was orchestrated via 'jamboree-style' annotation meetings, aimed to provide an accurate annotation of the rice genome (International Rice Genome Sequencing Project 2005, Itoh et al. 2007). In conjunction with the rice genome sequence and its related genomic resources, advanced development of mapping populations and molecular marker resources has allowed researchers to accelerate the isolation of agronomically important quantitative trait loci (QTLs) (Ashikari et al. 2005, Konishi et al. 2006, Ma et al. 2006, Kurakawa et al. 2007, Ma et al. 2007).

The aforementioned recent high-throughput technological advances have provided opportunities to develop collections of sequence-based resources and related resource platforms for

Plant Cell Physiol. 51(4): 497–523 (2010) doi:10.1093/pcp/pcq027, available online at www.pcp.oxfordjournals.org

© The Author 2010. Published by Oxford University Press on behalf of Japanese Society of Plant Physiologists.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5>), which permits unrestricted non-commercial use distribution, and reproduction in any medium, provided the original work is properly cited.

specific organisms. A schematic representation of each relevant omics resource is shown in Fig. 1, together with the current status of their availabilities for Arabidopsis, rice and soybean as examples. Each biological element that has been measured comprehensively by a high-throughput method is represented in a corresponding plane in a conceptual model with layers ranging from genome to phenome, a model termed 'omic space' (Toyoda and Wada 2004). Such comprehensive models often provide an excellent starting point for designing experiments, generating hypotheses or conceptualizing based on the integrated knowledge found in the omic space of a particular organism. Furthermore, development of such omic resources and data sets for various species allows the comparison of omic properties among species, which promises to be an efficient way to find collateral evidence for conserved gene functions that might be evolutionarily supported. Bioinformatics platforms have become essential tools for accessing omics data sets for the efficient mining and integration of biologically significant knowledge.

We provide an overview of several representative resources available for use in omics plant research, with particular emphasis on recent progress related to crop species. We describe sequence-related resources such as whole genome, protein coding and non-coding transcripts, and provide sequencing technology updates. We then review resources for genetic map-based approaches such as QTL analyses and population studies. We also describe the current status of resources and technologies for transcriptomics, proteomics and metabolomics, and review the comprehensive profiling of elements in each omics field as well as instances of their combinatorial uses in the investigation of particular biological systems. Mutant resources for use in phenome research will also be discussed. Finally, the integration of omics data sets across plant species and the progress of comparative genomics will be introduced. Throughout this review, we provide examples of applications of such resources and available databases.

Sequence resources in plants

Comprehensively collected sequence data provide essential genomic resources for accelerating molecular understanding of biological properties and for promoting the application of such knowledge. The recent accumulation of nucleotide sequences of model plants, as well as of applied species such as crops and domestic animals, has provided fundamental information for the design of sequence-based research applications in functional genomics. In this section, we describe recently developed plant sequence resources. Species-specific nucleotide sequence collections also provide opportunities to identify the genomic aspects of phenotypic characters based on genome-wide comparative analyses and knowledge of model organisms (Cogburn et al. 2007, Flicek et al. 2008, Paterson 2008, Tanaka et al. 2008).

Genome sequencing projects

The first genome sequence of a plant was completed for *A. thaliana*, which is now used as a model species in plant

molecular biology due to its small size, short generation time and high efficiency of transformation. The Arabidopsis genome sequence project was performed as a cooperative project among scientists in Japan, Europe and the USA (Bevan 1997). The genome sequencing was completed and published in 2000 by the Arabidopsis Genome Initiative (AGI) (The Arabidopsis Genome Initiative 2000). The draft genome sequence of rice, both *japonica* and *indica*, an important staple food as well as a model monocotyledon, was published in 2002 (Goff et al. 2002, Yu et al. 2002). Subsequently, the genome sequence of *japonica* rice was completed and published by the International Rice Genome Sequencing Project in 2005 (International Rice Genome Sequencing Project 2005). To date, several genome sequencing projects involving various plant species have been completed (Table 1).

There are a number of providers for plant genome sequences and annotations. Phytozome is a Web-accessible information resource providing genome sequences and annotations of various plant species. This resource is a joint project of the Department of Energy's Joint Genome Institute (DOE-JGI) and the Center for Integrative Genomics, and is intended to facilitate comparative genomic studies among green plants (http://www.phytozome.net/Phytozome_info.php). The current version of Phytozome (ver. 5.0, January 2010) consists of 18 plant species that were sequenced by JGI and other sequencing projects. Gramene (<http://www.gramene.org/>) is an information resource established as a portal site for grass species, and it provides various kinds of information related to grass genomics, including genome sequences (Ware 2007, Liang et al. 2008). The current version of Gramene (#30, October 2009) provides genome sequence information for 15 plant species, including five wild rice genome assemblies.

According to data provided on the Entrez Genome Project Web site (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj>), as of November 2009, >150 instances of genome projects in species of Viridiplantae have been tracked, including agronomically important crops such as staple foods, fruit trees, medical plants and a number of green alga species. With the ongoing innovations in next-generation sequencing technologies, the release of sequenced genomes is expected to accelerate (Ossowski et al. 2008, Ansorge 2009). Whole-genome sequence information allows us to derive sets of important genomic features, including the identification of protein-coding or non-coding genes and constructs such as gene families, regulatory elements, repetitive sequences, simple sequence repeats (SSRs) and guanine-cytosine (GC) content. These data sets have become primary sequence material for the design of genome sequence-based platforms such as microarrays, tiling arrays or molecular markers, as well as for reference data sets for the integration of omics elements into a genome sequence. Chromosome-scale comparisons identifying conserved similarities of gene coordinates facilitate documentation of segmental and tandem duplications in related species (Haas et al. 2004, De Bodt et al. 2005). Whole-genome comparisons identifying chromosomal duplication and conserved synteny among

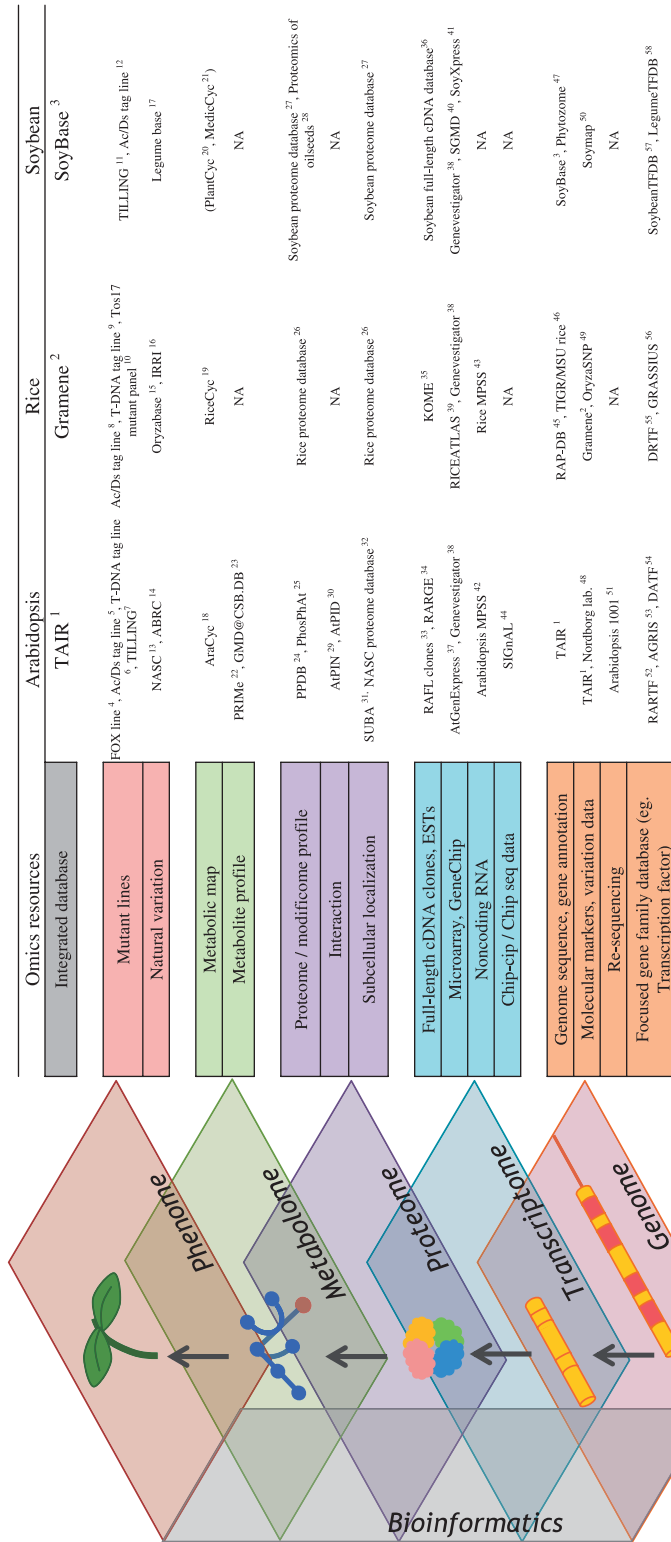


Fig. 1 Omic space and related resources in plants. Examples of resources related to each omics instance are represented in Arabidopsis, rice and soybean, as the model plant, as a model monocot and a sequenced crop, and as an important crop recently sequenced, respectively. These resources are accessible from the following URLs or citations. 1. <http://www.arabidopsis.org/>, 2. <http://www.gramene.org/>, 3. <http://soybase.org/>, 4. <http://nazunafox.psc.database.riken.jp/>, 5. <http://range.gsc.riken.jp/dsmutant/index.pl>, 6. <http://signalsalk.edu/about.html>, 7. <http://tilling.fhcrc.org/>, 8. Kolesnik et al. (2004), 9. <http://www.postech.ac.kr/life/pfg/risd/>, 10. <http://tos.nias.affrc.go.jp/>, 11. <http://www.soybeantilling.org/psearch.jsp>, 12. <http://mulch.cropsoil.uga.edu/~parrottlab/Mutagenesis/acds/index.php>, 13. <http://arabidopsis.org.uk/home.html>, 14. <http://abrc.osu.edu/>, 15. <http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp>, 16. <http://www.irri.org/grc/GRChome/home.htm>, 17. <http://www.legumebase.agr.miyazaki-u.ac.jp/index.jsp>, 18. <http://www.plantcyc.org:1555/ARA/server.html>, 19. <http://pathway.gramene.org/gramene/ricecyc.shtml>, 20. <http://www.plantcyc.org/>, 21. <http://mediccc.noble.org/>, 22. <http://prime.psc.riken.jp/>, 23. <http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html>, 24. <http://ppdb.cc.cornell.edu/>, 25. <http://phosphat.mpimp-golm.mpg.de/>, 26. http://cdna01.dna.affrc.go.jp/RPD/main_en.html, 27. <http://proteome.dc.affrc.go.jp/Soybean/>, 28. <http://oilseedproteomics.missouri.edu/>, 29. <http://bioinfo.esalq.usp.br/cgi-bin/atpin.pl>, 30. <http://atpid.biosino.org/>, 31. <http://suba.plantenergy.uwa.edu.au/>, 32. <http://proteomics.arabidopsis.info/>, 33. <http://www.brc.riken.go.jp/lab/epd/catalog/cdnacode.html>, 34. <http://range.gsc.riken.jp/>, 35. <http://cdna01.dna.affrc.go.jp/cDNA/>, 36. <http://rsoy.psc.riken.jp/>, 37. <http://www.arabidopsis.org/portals/expression/microarray/ATGenExpress.jsp>, 38. <https://www.geneinvestigator.com/gv/index.jsp>, 39. <http://bioinformatics.med.yale.edu/riceatlas/>, 40. <http://bioinformatics.towson.edu/SGMD/Default.htm>, 41. <http://soyexpress.agrenv.mcgill.ca/cgi-bin/soy/soybean.cgi>, 42. <http://impss.ucl.edu/at/>, 43. <http://impss.ucl.edu/rice/>, 44. <http://signal.salk.edu/>, 45. <http://rapdb.dna.affrc.go.jp/>, 46. <http://rice.plantbiology.msu.edu/>, 47. <http://www.phytozome.net/>, 48. <http://walnut.usc.edu/>, 49. <http://www.oryzasnp.org/>, 50. <http://www.soymap.org/>, 51. <http://1001genomes.org/>, 52. <http://range.gsc.riken.jp/rartf/>, 53. <http://arabidopsis.med.ohio-state.edu/>, 54. <http://datf.cbi.pku.edu.cn/>, 55. <http://drtf.cbi.pku.edu.cn/>, 56. <http://grassius.org/>, 57. <http://soybeanTFDB.psc.riken.jp/>, 58. <http://legumetfdb.psc.riken.jp/>.

Table 1 Whole genome sequencing projects in plants (http://www.arabidopsis.org/portals/genAnnotation/other_genomes/index.jsp)

Common name	Latin name	Sequencing group	References	URL
Dicots				
Mouse ear cress	<i>Arabidopsis thaliana</i>	Consortium (AGI)	The Arabidopsis Genome Initiative (2000)	http://www.arabidopsis.org/
Poplar	<i>Populus trichocarpa</i>	JGI	Tuskan et al. (2006)	http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html
Lyre-leaf rockcress	<i>Arabidopsis lyrata</i>	JGI		http://genome.jgi-psf.org/Araly1/Araly1.home.html
Pink shepherd's purse	<i>Capsella rubella</i>	JGI		(http://www.jgi.doe.gov/sequencing/why/3066.html)
Rapeseed	<i>Brassica rapa</i>	Consortium (MGBP)		http://www.brassica-rapa.org/BRGP/index.jsp
Tomato	<i>Solanum lycopersicum</i>	Consortium (ITGSP)		http://solgenomics.net/
Potato	<i>Solanum tuberosum</i>	Consortium (PGSC)		http://www.potatogenome.net/index.php/Main_Page
Barrel medic	<i>Medicago truncatula</i>	Consortium (IMGAG)		http://www.medicago.org/genome/
Monkey flower	<i>Lotus japonicus</i>	Consortium	Sato et al. (2008)	http://www.kazusa.or.jp/lotus/
Soybean	<i>Mimulus guttatus</i>	JGI	Schmutz et al. (2010)	(http://www.jgi.doe.gov/sequencing/why/3062.html)
Cassava	<i>Glycine max</i>	JGI		http://www.phytozome.net/soybean.php
Grape	<i>Manihot esculenta</i>	JGI		http://www.phytozome.org/cassava.php
Columbine	<i>Vitis vinifera</i>	Consortium	Jaillon et al. (2007)	http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/
Eucalyptus	<i>Aquilegia formosa</i>	JGI		(http://www.jgi.doe.gov/sequencing/why/51280.html)
Papaya	<i>Eucalyptus grandis</i>	JGI		http://bioinformatics.psb.ugent.be/genomes/view/Eucalyptus-grandis
Castor bean	<i>Carica papaya</i>	Consortium	Ming et al. (2008)	http://asgpb.mhpcch.hawaii.edu/papaya/
Monocots				
Rice	<i>Ricinus communis</i>	TIGR		http://castorbean.jcvi.org/
Rice	<i>Oryza sativa japonica</i>	Consortium (IRGSP)	International Rice Genome Sequencing Project (2005)	http://rgp.dna.affrc.go.jp/E/IRGSP/index.html
Rice	<i>Oryza sativa indica</i>	Beijing Genomics Institute	Yu et al. (2002)	http://rice.genomics.org.cn/rice/index2.jsp
Maize	<i>Zea mays</i>	Consortium	Schnable et al. (2009)	http://www.maizegdb.org/
Sorghum	<i>Sorghum bicolor</i>	JGI	Paterson et al. (2009)	http://genome.jgi-psf.org/Sorbi1/Sorbi1.home.html
Wheat	<i>Brachypodium distachyon</i>	JGI, Consortium (IBI)	The International Brachypodium Initiative (2010)	http://www.brachypodium.org/
Barley	<i>Triticum aestivum</i>	Consortium (IWGSC)		http://www.wheatgenome.org/
Other	<i>Hordeum vulgare</i>	Consortium (IBSC)		http://www.public.iastate.edu/~imagefpc/IBSC%20Webpage/IBSC%20TEmplate-home.html
Moss	<i>Physcomitrella patens</i>	JGI	Rensing et al. (2008)	http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html
Gemmiferous spike moss	<i>Selaginella moellendorffii</i>	JGI		http://genome.jgi-psf.org/Selmo1/Selmo1.home.html
	<i>Cyanidioschyzon merolae</i>	Consortium	Matsuzaki et al. (2004)	http://merolae.biol.s.u-tokyo.ac.jp/

related species provide evidence for hypotheses on comparative evolutionary histories with regard to the diversification of species in a related lineage (Paterson et al. 2009, Schnable et al. 2009).

Large-scale collections of expressed sequence tags (ESTs) and cDNA clones

ESTs are created by partial 'one-pass' sequencing of randomly picked gene transcripts that have been converted into cDNA (Adams et al. 1993). Since cDNA and EST collections can be acquired regardless of genomic complexity, this approach has been applied not only to model species but also to a number of applied species with large genome sizes due to polyploidy and/or to their number of repetitive sequences. As of November 2009, there are >63 million ESTs in the National Center for Biotechnology Information (NCBI)'s dbEST, a public domain EST database (<http://www.ncbi.nlm.nih.gov/dbEST/>) that includes a number of plant species (Table 2) (Boguski et al. 1993).

Because EST data collected from the cDNA libraries of a particular organism consist of redundant sequence data derived from the same gene locus or transcription unit, it is often necessary to perform EST grouping by transcription units and to assemble these groups in order to create a consolidated alignment and representative sequence of each transcript before further analysis. Such steps are performed computationally: a typical work flow consists of 'base calling', i.e. converting the output trace of a sequencer to identified nucleotide data, followed by a cleaning step involving identification and removal of contaminated sequences, the masking out of cloning vector sequences, clustering of identical sequences and alignment of clustered sequences (Ewing et al. 1998, Huang and Madan 1999, Masoudi-Nejad et al. 2006). Then, the obtained data sets of representative transcripts can be used as unified transcript data. There are several data resources that provide such unified data sets of plants, such as NCBI-UniGene, PlantGDB, TIGR Plant Gene Index and HarvEST (Feolo et al. 2000, Lee et al. 2005, Close et al. 2007, Duvick et al. 2008).

The comprehensive and rapid accumulation of cDNA clones together with mass volume data sets of their sequence tags have become significant resources for functional genomics. ESTs derived from various kinds of tissues, including tissues from organisms in a range of developmental stages or under stress, could significantly facilitate gene discovery as well as gene structural annotation, large-scale expression analysis, genome-scale intraspecific and interspecific comparative analysis of expressed genes and the design of expressed gene-oriented molecular markers and probes for microarrays (Ogihara et al. 2003, Zhang et al. 2004, Kawaura et al. 2006, Mochida et al. 2006).

Full-length cDNA projects

Although partial cDNAs are useful for rapidly creating catalogs of expressed genes, they are not suitable for further study of gene function. This is because the most popular method for preparing a cDNA library does not provide a full-length cDNA

Table 2 Numbers of ESTs and unified transcripts in plants (November 2009)

Species	No. of ESTs (dbEST)	No. of entries (UniGene)
<i>Physcomitrella patens</i>	382,584	18,870
<i>Picea glauca</i> (white spruce)	299,455	22,472
<i>Picea sitchensis</i> (Sitka spruce)	175,662	18,838
<i>Pinus taeda</i> (loblolly pine)	328,628	18,921
<i>Aquilegia formosa</i> × <i>Aquilegia pubescens</i>	85,039	8,046
<i>Arabidopsis thaliana</i> (thale cress)	1,527,298	30,579
<i>Artemisia annua</i> (sweet wormwood)	85,402	9,462
<i>Brassica napus</i> (rape)	643,601	26,733
<i>Brassica oleracea</i>	59,946	5,617
<i>Brassica rapa</i> (field mustard)	44,570	14,497
<i>Capsicum annuum</i>	116,541	8,868
<i>Citrus clementina</i>	118,365	9,123
<i>Citrus sinensis</i> (Valencia orange)	208,909	15,808
<i>Glycine max</i> (soybean)	1,422,604	33,001
<i>Gossypium hirsutum</i> (upland cotton)	268,786	21,738
<i>Gossypium raimondii</i>	63,577	3,297
<i>Helianthus annuus</i> (sunflower)	133,682	12,216
<i>Lactuca sativa</i> (garden lettuce)	80,781	7,940
<i>Lotus japonicus</i>	195,385	14,493
<i>Malus × domestica</i> (apple)	324,308	23,731
<i>Medicago truncatula</i> (barrel medic)	269,237	18,098
<i>Nicotiana tabacum</i> (tobacco)	317,190	24,069
<i>Populus tremula</i> × <i>Populus tremuloides</i> (hybrid aspen)	76,160	9,652
<i>Populus trichocarpa</i> (western balsam poplar)	89,943	14,965
<i>Prunus persica</i> (peach)	79,203	7,620
<i>Raphanus raphanistrum</i> (wild radish)	164,119	18,788
<i>Raphanus sativus</i> (radish)	83,034	17,649
<i>Solanum lycopersicum</i> (tomato)	296,848	18,228
<i>Solanum tuberosum</i> (potato)	236,568	18,784
<i>Theobroma cacao</i>	159,320	24,958
<i>Vigna unguiculata</i> (cowpea)	187,443	15,740
<i>Vitis vinifera</i> (wine grape)	357,856	22,083
<i>Selaginella moellendorffii</i>	93,806	8,810
<i>Hordeum vulgare</i> (barley)	501,614	23,595
<i>Oryza sativa</i> (rice)	1,249,110	40,978
<i>Panicum virgatum</i> (switchgrass)	436,535	20,973
<i>Saccharum officinarum</i> (sugarcane)	246,892	15,594
<i>Sorghum bicolor</i> (sorghum)	209,814	13,899
<i>Triticum aestivum</i> (wheat)	1,067,290	40,349
<i>Zea mays</i> (maize)	2,018,798	97,123
<i>Chlamydomonas reinhardtii</i>	204,076	11,310
<i>Volvox carteri</i>	132,038	5,638

that includes the capped site sequence. The biotinylated cap trapper method, which uses trehalose-thermostabilized reverse transcriptase and is an efficient method for constructing full-length cDNA-enriched libraries, was developed by Hayashizaki's group at RIKEN about 10 years ago. Full-length cDNA libraries and large-scale sequence data sets of clones have become invaluable resources for life science projects studying various species (Hayashizaki 2003, Imanishi et al. 2004, Maeda et al. 2006, Tanaka et al. 2008, Yamasaki et al. 2008a).

The sequence resources derived from full-length cDNAs can also help substantially in identifying transcribed regions in completed or draft genome sequences. In *Arabidopsis* and rice, full-length cDNA sequences have been used to identify genomic structural features such as transcription units, transcription start sites (TSSs) and transcriptional variants (Seki et al. 2002b, Iida et al. 2004, Itoh et al. 2007, Yamamoto et al. 2009). In species for which we have draft genomes, such as *Physcomitrella*, soybean and poplar, full-length cDNA clones have been sequenced to help consolidate genomic infrastructure; this should also contribute to gene discovery (Nanjo et al. 2007, Ralph et al. 2008a, Umezawa et al. 2008) (Table 3). Full-length cDNAs are also useful for determining the three-dimensional (3D) structures of proteins by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy and for functional biochemical analyses of expressed proteins in the molecular interactions of protein–ligands, protein–proteins and protein–DNAs. Furthermore, recent advances in proteomics infrastructure require comprehensive data sets of the full-length amino acid sequences used to assign peptides to a protein. These advances also necessitate functional annotations to support systematic knowledge mined for proteins corresponding to identified peptides and for residues modified by, for example, phosphorylation, for use in combination with comparative

analyses of modificomic events among species. The full-length cDNA library has also contributed importantly to functional analysis by creating overexpressors used in reverse genetics. The advent of function-based gene discovery by systems such as full-length cDNA overexpressor (FOX) gene hunting, which use full-length cDNA transgenic plants as overexpressors, has provided an effective approach to high-throughput discovery of functional genes associated with phenotypic changes (Ichikawa et al. 2006, Fujita et al. 2007, Kondou et al. 2009).

Recently, full-length enriched cDNA libraries have been constructed for non-sequenced crops or forestry species, such as wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), cassava (*Manihot esculenta*), Japanese cedar (*Cryptomeria japonica*) and Sitka spruce (*Picea sitchensis*), as well as for plant species showing specific biological characters such as salt tolerance in salt cress (*Thellungiella halophila*) (Table 3). These full-length cDNA libraries have been used to identify biological features through comparisons of target sequences with those of model organisms such as *Arabidopsis*, rice and poplar. These libraries also serve as primary sequence resources for designing microarray probes and as clone resources for genetic engineering to improve crop efficiency (Sakurai et al. 2007, Futamura et al. 2008, Ralph et al. 2008b, Taji et al. 2008).

Because of the various key functionalities of full-length cDNA resources in omic space, it is also essential to establish relevant information resources that provide gateways to these resources as well as to integrate related data sets derived from other omics fields and species (Sakurai et al. 2005, Mochida et al. 2009b).

Ultrahigh-throughput DNA sequencing

During the past decade, the Sanger sequencing method has been used to complete sequencing of microbial and higher

Table 3 Large-scale collections of full-length cDNA clones in plants

Species	Database	References
<i>Arabidopsis thaliana</i>	http://rarge.gsc.riken.jp/	Seki et al. (2002b)
Citrus species		Marques et al. (2009)
<i>Cryptomeria japonica</i>		Futamura et al. (2008)
<i>Glycine max</i>	http://rsoy.psc.riken.jp/	Umezawa et al. (2008)
<i>Hordeum vulgare</i>	http://www.shigen.nig.ac.jp/barley/	Sato et al. (2009)
<i>Manihot esculenta</i>	http://amber.gsc.riken.jp/cassava/	Sakurai et al. (2007)
<i>Oryza rufipogon</i>		Lu et al. (2008)
<i>Oryza sativa</i> , (<i>japonica</i>)	http://cdna01.dna.affrc.go.jp/cDNA/	Kikuchi et al. (2003)
<i>Oryza sativa</i> , (<i>indica</i>)	http://www.ncgr.ac.cn/ricd	Liu et al. (2007)
<i>Physcomitrella patens</i>	http://www.brc.riken.go.jp/lab/epd/catalog/p_patens.html	Nishiyama et al. (2003)
<i>Populus nigra</i>	http://rpop.psc.riken.jp/index.pl	Nanjo et al. (2004); Nanjo et al. (2007)
<i>Populus trichocarpa</i>		Ralph et al. (2008a)
<i>Thellungiella halophila</i>		Taji et al. (2008)
<i>Triticum aestivum</i>	http://trifldb.psc.riken.jp/	Kawaura et al. (2009); Ogihara et al. (2004)
<i>Zea mays</i>	http://www.maizecdna.org/	Soderlund et al. (2009)

eukaryote genomes. In recent years, a number of alternative technologies, which are adaptations of methods such as pyrosequencing procedures, massively parallel DNA sequencing or single molecule sequencing, have become available (Margulies et al. 2005, Ansorge 2009). Such new sequencing technologies have provided us with new opportunities to be addressed at the entire genome level in the fields of comparative genomics, meta-genomics and evolutionary genomics (Varshney et al. 2009).

Whole-genome resequencing

Next-generation sequencing technology coupled with reference genome sequence data allows us to discover variations among individuals, strains and/or populations. Nucleotide polymorphisms are effectively identified by mapping sequence fragments onto a particular reference genome data set, a capability that is of immense importance in all genetic research. A whole-genome resequencing project to discover whole-genome sequence variations in 1,001 strains (accessions) of *Arabidopsis* will result in a data set that will become a fundamental resource for promoting future genetics studies to identify alleles in association with phenotypic diversity across the entire genome and across the entire species range (<http://1001genomes.org/>) (Weigel and Mott 2009). In rice, a high-throughput method for genotyping recombinant populations that used whole-genome resequencing data generated by the Illumina Genome Analyzer was performed (Huang et al. 2009). One of the most anticipated innovations for next-generation sequencers is the application to whole-genome de novo sequencing. Although, to date, this approach has been realized only in bacterial genomes (Farrer et al. 2009, Moran et al. 2009), a number of attempts are being made to realize this advance in higher species.

Comprehensive discovery of small RNAs (sRNAs)

In plants, sRNAs, including microRNAs (miRNAs), short interfering RNAs (siRNAs) and trans-acting siRNAs (ta-siRNAs), are also playing roles as crucial components of epigenetic processes and gene networks involved in development and homeostasis (Ruiz-Ferrer and Voinnet 2009). These RNA molecules are important targets that should be comprehensively identified and their expression should be analyzed using next-generational genomic technologies (Nobuta et al. 2007, Chellappan and Jin 2009). In maize, sRNAs in the wild type and in the isogenic *mop1-1* loss-of-function mutant were analyzed by deep sequencing using Illumina's sequencing-by-synthesis (SBS) technology to characterize the complement of maize sRNA (Nobuta et al. 2008). In poplar, expressed sRNAs from leaves and vegetative buds were also discovered using high-throughput Roche 454 pyrosequencing, subsequently the genes of miRNA families, including the novel ones, were identified (Barakat et al. 2007). Deep sequencing of *Brachypodium* sRNAs at the global genome level has also been performed, resulting in identification of miRNAs involved in the cold stress response (J. Zhang et al. 2009). The plant miRNA database (PMRD) is a useful information resource on plant miRNA and is

available on the Web (<http://bioinformatics.cau.edu.cn/PMRD/>) (Z. Zhang et al. 2009).

Resources for variation analysis

Recent innovations related to DNA sequencing technology and the rapid growth of genome and cDNA sequence resources allow us to design various types of molecular markers covering entire genomes (Feltus et al. 2004). For high-throughput genotyping, a number of platforms have been developed that have been applied to genetic map construction, marker-assisted selection and QTL cloning using multiple segregation populations (Hori et al. 2007). Such genotyping systems have also been used in post-genome sequencing projects such as genotyping of genetic resources, accessions to evaluate population structure and association studies to identify genetic loci involved in phenotypic changes of species. This recent expansion of analysis platforms addressing genome-wide polymorphisms provides an essential resource in the 'variome' study of plants.

Molecular markers

Accumulation and saturation of available genetic markers contribute to advances in marker-assisted genetic studies and are important resources with a wide range of applications. Genetic markers designed to cover a genome extensively allow not only identification of individual genes associated with complex traits by QTL analysis but also the exploration of genetic diversity with regard to natural variations (Feltus et al. 2004, Varshney et al. 2005, Caicedo et al. 2007). With the progress of genome sequencing and large-scale EST analysis in various species, these sequence data sets have become quite efficient sequence resources for designing molecular markers. A number of attempts to design polymorphic markers from accumulated sequence data sets have been made for various species. Several genome-wide rice (*Oryza sativa*) DNA polymorphism data sets have been constructed based on alignment between *japonica* and *indica* rice genomes (Han and Xue 2003, Shen et al. 2004). Large-scale EST data sets are also important resources for discovery of sequence polymorphisms, especially for allocating expressed genes onto a genetic map. Therefore, computational discovery of ESTbase single-nucleotide polymorphisms (SNPs) and/or EST-SNP markers for the purpose of identifying sequence-tagged site (STS) markers has progressed for numerous species, including barley, wheat, maize, melon, *Brassica*, common bean and sunflower (Kota et al. 2001, Kantety et al. 2002, Kota et al. 2003, Torada et al. 2006, Heesacker et al. 2008, Kota et al. 2008, Blair et al. 2009, Deleu et al. 2009, Kaur et al. 2009, Li et al. 2009).

Several databases provide information on molecular markers in plant species. PlantMarkers is a genetic marker database that contains predicted molecular markers, such as SNP, SSR and conserved ortholog set (COS) markers, from various plant species (Heesacker et al. 2008). GrainGenes is a popular site for Triticeae genomics; it also provides genetic markers and linkage map data on wheat, barley, rye and oat (Carollo et al. 2005).

Gramene is a database for plant comparative genomics that provides genetic maps of various plant species (Liang et al. 2008). The Triticeae Mapped EST database (TriMEDB) provides information regarding mapped cDNA markers that are related to barley and their wheat homologs (Mochida et al. 2008).

Platforms for variation analyses

High-throughput polymorphism analysis is an essential tool for facilitating any genetic map-based approach. So far, genome-wide genotyping using a hybridization-based SNP typing method has been used to analyze representative Arabidopsis ecotypes and rice strains, and the data sets containing the calculated genome-wide variation patterns for each species have been released. As typified by the Arabidopsis 1,001 project, genome-wide variation study is a key analysis that should be performed after genome sequencing has been completed for a particular reference strain. Therefore, the demand for high-throughput and cost-effective platforms for comprehensive variation analysis (also called variome analysis) has rapidly increased.

As we have already mentioned, whole-genome resequencing approaches are already being realized as a direct solution for variome analysis in species whose reference genome sequence data are available. Diversity Array Technology (DArT) is a high-throughput genotyping system that was developed based on a microarray platform (<http://www.diversityarrays.com/index.html>) (Jaccoud et al. 2001, Wenzl et al. 2007). In various crop species such as wheat, barley and sorghum, DArT markers have been used together with conventional molecular markers to construct denser genetic maps and/or to perform association studies (Crossa et al. 2007, Peleg et al. 2008, Mace et al. 2009). In barley and wheat, Affymetrix GeneChip Arrays have been used to discover nucleotide polymorphisms as single-feature polymorphisms based on the differential hybridization of GeneChip probes (Rostoks et al. 2005, Bernardo et al. 2009). The Illumina GoldenGate Assay allows the simultaneous analysis of up to 1,536 SNPs in 96 samples and has been used to analyze genotypes of segregation populations in order to construct genetic maps allocating SNP markers in crops such as barley, wheat and soybean (Hyten et al. 2008, Akhunov et al. 2009, Close et al. 2009).

Transcriptome resources in plants

Comprehensive and high-throughput analysis of gene expression, called transcriptome analysis, is also a significant approach to screen candidate genes, predict gene function and discover *cis*-regulatory motifs. The hybridization-based method, such as that used in microarrays and GeneChips, has been well established for acquiring large-scale gene expression profiles for various species. The recent rapid accumulation of data sets containing large-scale gene expression profiles and the ability of related databases to support the availability of such large repositories of data has provided us with access to large amounts of information in the public domain. This public domain data are

an efficient and valuable resource for many secondary uses, such as co-expression and comparative analyses. Furthermore, as a next-generation DNA sequencing application, deep sequencing of short fragments of expressed RNAs, including sRNAs, is quickly becoming an efficient tool for use with genome-sequenced species (Harbers and Carninci 2005, de Hoon and Hayashizaki 2008).

Sequence tag-based platforms in transcriptomics

Large-scale sequencing of ESTs from cDNA libraries was an early approach for acquiring transcriptome profiles. In this approach, ESTs that are randomly sequenced in an unbiased cDNA library are classified into clusters of transcript sequences using sequence-clustering and/or assembling methods. Then, the abundance of transcripts expressed in each tissue is estimated by counting the number of ESTs with identifiers for each cDNA library and/or each sequence cluster. The same methodological principle has been applied in human and mouse in the form of a 'body map' to derive the transcriptome in various organs (Hishiki et al. 2000, Kawamoto et al. 2000, Ogasawara et al. 2006). Moreover, this principle has also been used in the digital differential display (DDD) tool, which is a component of NCBI's UniGene database and has been applied in large-scale cDNA projects for various species, including plants (Mochida et al. 2003, Fei et al. 2004, Sterky et al. 2004, Zhang et al. 2004). Although this approach, coupled with cDNA clone resources, has facilitated gene discovery and expression profiling, its disadvantages include cost and limited resolution due to large-scale sequencing.

Serial analysis of gene expression (SAGE) is a method based on deep sequencing of short read cDNA tags. SAGE allows the identification of a large number of transcripts present in tissues and enables quantitative comparison of transcriptomes (Velculescu et al. 1995). SAGE is designed to generate a short specific tag (13–15 bp) from the 3' end of each mRNA present in a sample, after which >10 tags are concatenated and cloned to generate a SAGE library. The sequencing of selected clones from the SAGE library allows efficient collection of transcript tag sequences. A data set of genome sequences or large-scale ESTs is required to identify genes corresponding to each SAGE tag. Some derivatives of the original protocol (MAGE, SADE, microSAGE, miniSAGE, longSAGE, superSAGE, deepSAGE, 5' SAGE, etc.) have been developed to improve and expand the utility of SAGE (Hashimoto et al. 2004, Anisimov 2008). For example, superSAGE is an improved version of SAGE that produces 26 bp fragment tags from cDNAs. This method has been applied to simultaneous and quantitative gene expression profiling of both host cells and their eukaryotic pathogens in rice (Matsumura et al. 2003). The 26 bp superSAGE tags have also been used to design probes directly for oligo microarrays (Matsumura et al. 2008).

Another sequencing-based technology is massively parallel signature sequencing (MPSS). MPSS uses a unique method to quantify gene expression levels; it generates millions of short sequence tags per library by sequencing 16–20 bp from the

3' side of cDNA using a microbead array (Brenner et al. 2000). Databases containing MPSS data on plant species, including *Arabidopsis*, rice, grape and *Magnaporthe grisea* (the rice blast fungus), are available online (<http://mpss.udel.edu>) (Nakano et al. 2006). In addition, the MPSS method has also been used to perform genome-scale discovery and expression profiling of sRNAs in *Arabidopsis* and rice (Lu et al. 2006, Nobuta et al. 2007). The CT-MPSS was a method recently developed for quantitative analysis of the 5' end of transcripts coupled with cap-trapper method for full-length cDNA cloning. This method has been applied to perform high-density mapping of TSS in *Arabidopsis* to figure out genome-scale instances of plant promoters (Yamamoto et al. 2009). The data set of *Arabidopsis* CT-MPSS tags is accessible from ppdb (<http://www.ppdb.gene.nagoya-u.ac.jp>), a plant promoter database that provides promoter annotation of *Arabidopsis* and rice (Yamamoto and Obokata 2008).

Hybridization-based platforms in transcriptomics

The history of DNA microarray began with a paper from the P. O. Brown laboratory at Stanford University in 1995 (Schena et al. 1995). Since then, microarray- and DNA chip-related technologies have advanced rapidly and their application has expanded to a wide variety of life sciences disciplines. The methodological principle of DNA microarray or GeneChip analysis is to acquire a comprehensive data set of the molecular abundance of each molecule in a given sample based on its simultaneous hybridization with large numbers of DNA molecular species immobilized on a glass slide or on a silicon chip used as a probe set.

DNA microarrays can be classified into two major types: (i) the 'spotting' type, which was developed at Stanford University; and (ii) the 'on-chip synthesis' type based on manufactured probes. The spotting type was widely used during the early years of transcriptome research. This method entailed preparing DNA microarrays by spotting a cDNA solution onto a glass slide. The on-chip (in situ) oligo synthesis-based method is a light-directed chemical synthesis process that combines solid-phase chemical synthesis with photolithographic fabrication techniques. Initially, this method was employed only in conjunction with the Affymetrix-manufactured GeneChip Array system. In the Affymetrix GeneChip system, a known gene or potentially expressed sequence is represented on the chip by 11–20 unique oligomeric probes that are each 25 bases in length. Roche NimbleGen and Agilent Technology offer platforms to manufacture high-density DNA arrays based, respectively, on Roche's proprietary Maskless Array Synthesizer (MAS) technology and on a non-contact industrial inkjet printing process, both of which are also used for in situ oligo synthesis.

With the recent and rapid increase in the number of sequenced species in whole-genome and/or large-scale cDNA clones, a number of DNA microarrays have also been developed for transcriptome analysis in various plant species. For example, Seki and co-workers designed a custom DNA microarray that uses 7,000 full-length cDNA clones of *Arabidopsis* as probes and then successfully screens genes in response to

abiotic stresses using a two-color method (Seki et al. 2002a). With the recent increase in commercially available DNA microarrays, laboratories are able to use a particular DNA microarray design to obtain transcriptome data from many experiments in order to accumulate a more comprehensive resource for organism-specific transcriptome data. AtGenExpress was a multinational effort designed to uncover the transcriptome of *A. thaliana*. The data sets collected in AtGenExpress have been one of the most comprehensive resources for the *Arabidopsis* transcriptome to date (Kilian et al. 2007, Goda et al. 2008).

NCBI's Gene Expression Omnibus (GEO) and the European Bioinformatics Institute (EBI)'s ArrayExpress have been serving as the primary archives of transcriptome data in the public domain (Parkinson et al. 2007, Barrett et al. 2009). There are also several more focused databases that provide calculated transcriptome data with user-friendly interfaces and annotations on probes. ATTED II (<http://atted.jp/>) is a database that provides co-expression analysis data calculated from publicly available *Arabidopsis* ATH1 GeneChip data (Obayashi et al. 2007, Obayashi et al. 2009). Co-expression analysis data sets generated from comprehensively collected transcriptome data sets have become an efficient resource capable of facilitating the discovery of genes closely correlated in their expression patterns. Genevestigator (<https://www.genevestigator.com/gv/index.jsp>), which is a reference expression database and meta-analysis system, also provides summary information from hundreds of microarray experiments on various organisms, including *Arabidopsis*, barley and soybean, with easily interpretable results (Zimmermann et al. 2004). The electronic fluorescent pictograph (eFP) browser provides gene expression patterns collected from *Arabidopsis*, poplar, *Medicago*, rice and barley via a user-friendly interface on the Web (<http://www.bar.utoronto.ca/>) (Winter et al. 2007). The *Arabidopsis* Gene Expression Database AREX is a database that provides data sets of high-resolution gene expression patterns of root tissues in *Arabidopsis* (<http://www.arexdb.org/index.jsp>) (Birnbaum et al. 2003, Brady et al. 2007). The RICEATLAS is a database housing rice transcriptome data covering various types of tissues (<http://bioinformatics.med.yale.edu/riceatlas/>) (Jiao et al. 2009).

Tiling arrays, which are high-density oligonucleotide probes spanning the entire genome in a particular organism, are a platform for analyzing expressed regions throughout a whole genome; it is an effective method by which to discover novel genes and elucidate their structure. Seki and co-workers performed transcriptome analysis in *Arabidopsis* under abiotic stress conditions using a whole-genome tiling array and discovered a number of antisense transcripts induced by abiotic stresses (Matsui et al. 2008). The *A. thaliana* Tiling Array Express (At-TAX) is a whole-genome tiling array resource for developmental expression analysis and transcript identification in *Arabidopsis* (Laubinger et al. 2008, Zeller et al. 2009). The usefulness of tiling arrays has recently been extended by coupling this platform with the immunoprecipitation method. For example, the binding sites of AGAMOU-Like15, AGL15,

a MADS domain transcriptional regulator promoting somatic embryogenesis, were identified using a chromatin immunoprecipitation (ChIP) approach coupled with the Affymetrix tiling array for Arabidopsis. This method found approximately 2,000 sites (Zheng et al. 2009). Using the methylcytosine immunoprecipitation (mCIP) method in combination with the Arabidopsis tiling array, a comprehensive DNA methylation map of the genome was constructed as an Arabidopsis methylome data set (Zhang et al. 2006). Sequencing of co-precipitated DNAs together with a protein using the next generation sequencer, 'ChIP-seq', has also become an alternative approach (Park 2009).

Platforms and resources in proteomics

As genome sequencing projects for several organisms have been completed, proteome analysis, which is the detailed investigation of the functions, functional networks and 3D structures of proteins, has gained increasing attention. Large-scale proteome data sets are also an important resource for the better understanding of protein functions in cellular systems, which are controlled by the dynamic properties of proteins. These properties reflect cell and organ states in terms of growth, development and response to environmental changes. The primary objective of functional proteomics was the high-throughput identification of all of the proteins appeared in cells and/or tissues. Recent, rapid technical advances in proteomics (e.g. protein separation and purification methods, advances in mass spectrometry equipment and methodological developments in protein quantification) have allowed us to progress to the second generation of functional proteomics, including quantitative proteomics, subcellular proteomics and various modifications and protein-protein interactions (Rossignol et al. 2006, Jorin-Novo et al. 2009, Yates et al. 2009). The different Web-accessible plant proteome-related databases are summarized on the proteomics subcommittee of the Multinational Arabidopsis Steering Committee (MASCPC) Web site (<http://www.mascpc.org/>) under the heading of "Proteomic Databases and Resources".

Proteome profiling

The typical experimental workflow of protein profiling can be summarized as protein sample preparation, separation and detection, then identification. Various technical advances for each step of the process have greatly increased the overall performance of plant proteomics (Jorin-Novo et al. 2009).

Sample preparation is the most critical step in any proteomics experiment. The method that uses trichloroacetic acid (TCA) and acetone is the most commonly used procedure for protein precipitation. A method using phenol and NH_4OAc /MeOH is also popular for plant tissues. Sample fractionation effectively improves protein detection and increases proteome coverage by reducing sample complexity. Sequential solubilization is an efficient method for fractionating protein samples based on solubility, molecular mass and isoelectric point.

By using a series of different reagents to separate proteins by their different solubilities, sequential solubilization is also an effective way to reduce the complexity of proteins in each fraction and to enrich rare proteins (Agrawal et al. 2005).

One-dimensional SDS-PAGE has been widely used to fractionate complex proteins based on their molecular masses. For high-resolution separation of proteins, two-dimensional gel electrophoresis (2-DE), which uses isoelectric focusing (IEF) as the first dimension and SDS-PAGE as the second dimension, is an effective method. Furthermore, the later development of the immobilized pH gradient (IPG)-IEF as the first dimension has improved reproducibility and resolution. The 2-DE methods have been widely used in proteomics in various species (Islam et al. 2004, Mechin et al. 2004, Chen and Harmon 2006), and databases housing 2-DE information have been developed and released [e.g. the Swiss Institute of Bioinformatics' ExPASy SWISS-2DPAGE database (<http://au.expasy.org/ch2d/>) and the Kazusa DNA Research Institute's Cyano2DBase (http://bacteria.kazusa.or.jp/cyano_legacy/Synechocystis/cyano2D/index.html)]. Chromatography-based separation methods such as gel filtration chromatography, ion exchange chromatography and affinity chromatography are also effective in separating proteins based on their physicochemical properties.

To identify each protein found in a prepared sample, peptide mass fingerprinting has been widely employed. Currently the most efficient method available consists of two steps: (i) enzymatic digestion of separated proteins into peptides; and (ii) accurate mass measurements of peptides using mass spectrometry (MS). In-gel digestion methods have been widely used to separate protein samples by using 2-DE. With its rapid technical advances, MS continues to play an important role in proteomics. MS equipment consists of a source to ionize samples and a mass spectrometer(s) to detect the ionized samples. In proteomics, usually the matrix-assisted laser desorption ionization (MALDI) method or the electrospray ionization (ESI) method is applied to ionize sample peptides. The MALDI method is typically used in combination with time of flight (TOF) MS as MALDI-TOF-MS while the ESI method is usually used in combination with quadrupole (Q) or ion trap (IT) MS. Recently, MS, such as Q-TOF MS, IT-TOF MS or MALDI Q-TOF MS, has become popular. Furthermore, ion fragmentation by collision-induced dissociation (CID) using tandem MS such as Q-TOF MS or by post-source decay (PSD) using MALDI-TOF MS have been applied to determine peptide amino acid sequences. To identify target proteins, obtained peptide mass fingerprint data are searched against a database of theoretically predicted masses of known amino acid sequences (Hirano et al. 2004, Newton et al. 2004).

In addition to conventional gel electrophoresis-based separation, the gel-free separation method is often used, particularly in the 'shotgun proteomics' approach. In the gel-free method, the protein mixture is directly digested into peptides and separated by the multidimensional separation method. The multidimensional separation method is a combination of different online separation methods including multidimensional protein

identification technology (MudPIT). The shotgun approach is suitable for the analysis of proteins that are difficult to separate by 2-DE as well as for high-throughput analysis by automated analytical instruments (Yates et al. 2009).

Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) possesses high resolution, high sensitivity, high dynamic range and high mass measurement accuracy. The high resolution and precision of FT-ICR MS allows us to carry out 'top-down proteomics' in which an intact protein mixture is analyzed directly, without separation (Bogdanov and Smith 2005).

Quantitative proteomics

Comprehensive quantification of each protein's abundance is quite important for a better understanding of the protein dynamics regulated in response to cellular state and environmental changes. A quantitative proteome approach also plays a crucial role in the discovery of key proteomic changes, including expression, interaction and modification, that are associated with genetic variations and/or visible phenotypic changes (Gstaiger and Aebersold 2009).

Difference gel electrophoresis (DIGE) is a popular method for differential display of proteins for quantitative protein comparison. In DIGE, protein samples are labeled with different fluorescent dyes before 2-D electrophoresis, enabling accurate analysis of differences in protein abundance between samples (Rossignol et al. 2006). This method is an effective way to remove gel to gel variation while significantly increasing accuracy and reproducibility. Isotope-coded affinity tags (ICATs), isobaric tags for relative and absolute quantitation (iTRAQ) and stable isotope labeling with amino acids in cell culture (SILAC) are widely used methods for protein differential display using stable isotope labeling (Jorin-Novo et al. 2009). Using a single MS/MS analysis, corresponding peptides from each sample are differentially detectable based on mass shift caused by the differential isotopes; this allows comparison of the relative abundance of the two samples. Recently, label-free quantitative techniques have been developed to facilitate high-throughput comparisons of proteomic expression. For label-free quantification, the proteomes from each of two samples are separately analyzed using liquid chromatography (LC)-MS/MS. Then, each MS1 spectrum is aligned to calculate relative protein abundance changes based on ion intensity changes such as peptide peak areas or peak heights in chromatography. Finally, MS/MS analysis is used to identify peptides (Gstaiger and Aebersold 2009).

Subcellular proteomics

Large-scale proteome analysis of cell organelles is essential for understanding the enzymatic inventory of a cell organelle; the compartmentalization of metabolic pathways; cellular logistics such as protein targeting, trafficking and regulation; and proteomic dynamics at the organelle level caused by changes in cellular systems (Andersen and Mann 2006, Chen and Harmon 2006, Baginsky 2009). A number of approaches

have been applied to analyze the proteome of organelles or subcellular compartments of plant cells such as chloroplasts, etioplasts, amyloplasts, chromoplasts, mitochondria, vacuoles, plasma membranes, nucleus, peroxisomes, cytosolic ribosome and cell wall (Baginsky 2009). Proteomic analyses of chloroplasts, mitochondria and further fractionations have been carried out to determine detailed localizations of protein in different suborganelle compartments. Techniques for quantitative proteomics, such as the ICAT and iTRAQ methods described above, are also effective for acquiring quantitative data on proteomes in each organelle. In Arabidopsis, rice and alga, differential proteome profiles of plant plasma membranes were monitored to identify those proteins differentially expressed in response to environmental factors such as cold acclimation, salt stress and bacterial elicitor (Benschop et al. 2007, Katz et al. 2007, Cheng et al. 2009, Minami et al. 2009).

Several databases provide subcellular proteome information. The rice proteome database (<http://gene64.dna.affrc.go.jp/RPD/>) is a 2-DE image database for rice that contains data from various tissues as well as subcellular compartments (Komatsu 2005). The Nottingham Arabidopsis Stock Centre (NASCC) Proteomics database (<http://proteomics.arabidopsis.info/>) and the SUB-cellular location database for Arabidopsis proteins (SUBA) (<http://suba.plantenergy.uwa.edu.au/>) provides subcellular proteome analysis data for Arabidopsis (Dunkley et al. 2006). The soybean proteome database (http://proteome.dc.affrc.go.jp/cgi-bin/2d/2d_view_map.cgi) also provides 2-DE data for various tissues as well as for subcellular compartments (Sakata et al. 2009).

Post-translational protein modifications

Comprehensive approaches to investigate various kinds of post-translational protein modifications also play a key role in the current study of proteomics. It (is also called modifome research) aims to identify modified proteins and to elucidate and coordinate the role of each protein functional modification with its associated biological event (Kwon et al. 2006).

Protein phosphorylation is a critical regulatory step in signaling networks and is a widespread protein modification affecting most basic cellular processes in eukaryotic organisms (Schmelzle and White 2006). Advances in MS-based technologies, accompanied by phosphopeptide enrichment techniques, have allowed us to perform high-throughput, large-scale *in vivo* phosphorylation site mapping. So far, several plant phosphoproteome studies have been reported (Nuhse et al. 2003, Nuhse et al. 2004, de la Fuente van Bentem et al. 2006, Benschop et al. 2007, Nuhse et al. 2007, Sugiyama et al. 2008). For example, the proteome-wide mapping of *in vivo* phosphorylation sites in Arabidopsis has recently been achieved by using complementary phosphopeptide enrichment techniques coupled with high-accuracy LC-MS/MS with a Finnigan LTQ-Orbitrap (Sugiyama et al. 2008). The Arabidopsis Protein Phosphorylation Site Database (PhosPhAt) provides information on Arabidopsis phosphorylation sites which were identified by MS by different research groups (<http://phosphat.mpimp-golm.mpg.de/>).

The Plant Protein Phosphorylation Database (P3DB) (<http://www.p3db.org/>), an information resource for plant phosphoproteomes, provides a resource for protein phosphorylation data from multiple plants (Gao et al. 2009).

Ubiquitination of protein is also one of the major post-translational modifications occurring in eukaryotic cells. Protein ubiquitination is a key regulatory mechanism that controls protein abundance, localization and activity. Several large-scale analyses of protein ubiquitination in plants have been reported (Maor et al. 2007, Manzano et al. 2008, Igawa et al. 2009). For example, in *Arabidopsis*, affinity purification using an anti-ubiquitin antibody and the subsequent use of MS/MS analysis has been performed to identify ubiquitinated proteins (Igawa et al. 2009).

Platform advances in structural proteomics

Large-scale data sets of protein 3D structures are also crucial information resources for elucidating relationships between protein functions and structures or for analyzing molecules in protein complexes. The International Structural Genomics Organization (ISGO, <http://www.isgo.org>) was formed to facilitate global structural genomics research efforts (Stevens et al. 2001). The key centers for structural genomics have been the RIKEN Structural Genomics/Proteomics Initiative (RSGI) in Japan, the Protein Structure Initiative (PSI) in the USA and the structural genomics centers of Europe (Yokoyama et al. 2000). International efforts to determine protein structures have contributed to increases in the number of solved protein structures. Thus, the number of solved protein structures appearing in the protein data bank, PDB (<http://www.pdb.org/pdb/home/home.do>), which is the most popular resource for biomolecule structure data sets, has dramatically increased during the past decade (Kouranov et al. 2006).

The PSI has promoted large-scale attempts to determine the 3D structure of protein folding. The PSI-1 was started as a pilot phase in 2000 with the aim of promoting the strategic development of tools and infrastructures necessary for large-scale determination of protein structures. The PSI-1 centers developed a systematic workflow pipeline encompassing target selection, cloning, expression and purification, followed by crystallization, X-ray crystallography or NMR methods to solve structures. The final step of this pipeline consisted of structure deposition into a database. In 2005, the PSI shifted to its second phase, which was known as PSI-2. The goal of PSI-2's 'production phase' is to solve more challenging structures such as protein complexes and integral membrane proteins (Fox et al. 2008). The RIKEN SGPI has solved >2,700 protein structures, including 33 from *Arabidopsis* that appear in the PDB (http://www.rsgi.riken.go.jp/rsgi_e/index.html).

Although methodological bottlenecks still exist in structural proteomics, some methodological advances have played an important role in this field. One of the major bottlenecks is the production of soluble and folded proteins. Most centers for structural genomics use *Escherichia coli* cells for protein production in their automated pipelines as an application of

the cell-based method. Cell-free expression systems have also become important as a way to address several limitations of cell-based methods, such as protein quality and quantity as well as throughput issues. The *E. coli* cell-free system has been applied to amino acid-selective stable isotope labeling of proteins for NMR spectroscopy (Yabuki et al. 1998, Kigawa et al. 1999). The wheat germ embryo cell-free system has also been developed as a eukaryotic cell-free system and has the advantage of producing multidomain proteins (Madin et al. 2000, Endo and Sawasaki 2003, Endo and Sawasaki 2006). The wheat germ cell-free system has since been incorporated into a robotic automation platform (Cell-Free Sciences Co. Ltd.). A comparative study of protein production from 96 *Arabidopsis* open reading frames (ORFs) using cell-based and cell-free systems was reported by the Center for Eukaryotic Structural Genomics (CESG) group in 2005 (Tyler et al. 2005).

The technology and platform of NMR spectroscopy has also played an important role in structural proteomics. The cell-free systems of *E. coli* and wheat germ embryo in combination with selected amino acid labeling, as described above, have produced synergistic advances in the promotion of automated protein structure determination using solution NMR methods (Yokoyama 2003). Furthermore, high-resolution multidimensional solid state NMR methods used in combination with cross polarization (CP), magic angle spinning (MAS) and dipolar decoupling (DD) are also becoming the methods of choice for structural analysis of membrane proteins by NMR platforms (Castellani et al. 2002, McDermott 2009). Furthermore, recent hardware improvements in NMR probes, including the cryoprobe for improved sensitivity, the micro-coil probe for sample mitigation and the flow-probe designed to shorten preparation time, provide us with the opportunity to use NMR methods to screen ligands that bind to a particular protein.

X-ray crystallography has been used to determine the protein structures of almost 90% of the protein entries in the PDB (http://www.rcsb.org/pdb/static.do?p=general_information/pdb_statistics/index.html). In particular, the beamlines of third-generation X-ray synchrotrons have become an essential infrastructure for macromolecular crystallography (MX), which is used to determine the 3D structures of macromolecules (such as large proteins and protein complexes) (Samatey et al. 2001). For example, the synchrotron SPring-8 of RIKEN in Japan is used to determine the structures of important membrane proteins and protein complex supermolecules such as Ca²⁺-ATPase, rhodopsin and flagellin (Palczewski et al. 2000, Samatey et al. 2001, Toyoshima et al. 2003).

By using existing analytical platforms for structural proteomics, the structures of many representative DNA-binding domains (DBDs), namely AP2/ERF, NAC, WRKY, B3 and SBP, of plant-specific transcription factor (TF) families have thus far been determined (Yamasaki et al. 2004, Yamasaki et al. 2008b).

Information resources in structural proteomics

Bioinformatics and related databases are also essential tools for advancing the study of structural proteomics. The methods of

computational prediction of protein 3D structure are mainly classified into two methods: template-based modeling (TBM) and free modeling (FM) (Zhang 2008, Zhang 2009b). Free modeling, which is also called 'ab initio' or 'de novo' modeling, is used to predict the 3D structure of proteins, without any information on previously solved structures. A number of Web server and computational tools for free and/or template-based modeling have recently been made available; for example, the I-TASSER internet service, which is used in Critical Assessment of Techniques for Protein Structure Prediction (CASP), was released (Zhang 2009a). Template-based modeling method is a comparative method for matching proteins using evolutionarily related proteins of known structure as a template. There are many Web services and tools (e.g. Swiss Institute of Bioinformatics' SWISS-MODEL server) to support template-based modeling (Schwede et al. 2003). Databases housing previously predicted structures from amino acid sequences by template-based modeling for a wide range of species also exist: the Genomes TO Protein structures and functions (GTOP) database (<http://spock.genes.nig.ac.jp/~genome/gtop.html>) provides information on protein structures and functions obtained through the application of various computational tools for structure prediction and annotation from the amino acid sequences deduced from annotated genes in sequenced genomes (Fukuchi et al. 2009).

The database for structure-based protein classification, as typified by CATH (<http://www.cathdb.info/>) and the Structural Classification of Proteins (SCOP) database (<http://scop.mrc-lmb.cam.ac.uk/scop/>), has provided important clues to the relationships between protein structures, protein functions and protein evolution (Greene et al. 2007, Andreeva et al. 2008). Databases of protein families based on conserved protein domains, such as Pfam, Superfamily and Protein ANalysis THrough Evolutionary Relationships (PANTHER), are important resources for classifying proteins into families. These types of databases are often used for the functional prediction and classification of proteins; for example, such resources are used in genome-wide identification of genes putatively encoding specific TFs (Mi et al. 2005, Wilson et al. 2007, Finn et al. 2008). Many of these protein family databases can be simultaneously searched using the EBI's Interpro (Mulder and Apweiler 2008, Hunter et al. 2009).

Platforms and resources in metabolomics

Metabolomics aims to understand metabolic systems based on comprehensive and integrated approaches by taking advantage of various measurement instruments to characterize metabolites. Metabolomic approaches allow us to conduct parallel assessments of multiple metabolites and to undertake quantitative analysis of particular metabolites in ways that provide major advantages over chemical-level phenotyping and diagnostic analysis. It is notable that the plant metabolome represents an enormous chemical diversity due to the complex set of metabolites produced in each plant species

(Bino et al. 2004, von Roepenack-Lahaye et al. 2004). Therefore, plant metabolomics is not only a great analytical challenge, but is also quite important. In its ability to elucidate plant cellular systems, metabolomics permits us to engineer molecular breeding to improve the productivity and functionality of plants in areas such as stress tolerance, pharmaceutical production, functional foods, biomaterials and energy (Trethewey 2004, Oksman-Caldentey and Saito 2005, Fernie and Schauer 2009). In this section, we introduce metabolomic analytical platforms for plants, metabolic profiling, and their applications in combination with other omics. We also describe plant metabolomics-related computational tools and databases.

Instruments for metabolomics

Many remarkable technological advances have recently been made in instrumentation related to metabolomics. Metabolomics experiments start with the acquisition of metabolic fingerprints using various analytical instruments such as GC-MS, LC-MS, FT-MS, FT-IR and NMR (Fiehn et al. 2000, Roessner et al. 2001, Fernie et al. 2004).

Methods for sample separation, such as gas chromatography (GC), high-performance or ultraperformance LC and capillary electrophoresis (CE), are typically used in conjunction with various types of MS, as detailed below. CE-MS is an especially effective, high-sensitivity method for separating and analyzing polarized molecules in samples (Ramautar et al. 2009). QMS and TOF MS are also well regarded for use in metabolomics. Triple Q (QqQ) MS (a tandem-type MS) and Q-TOF (a hybrid-type MS) are also used. Methods that do not involve pre-separation of samples, e.g. FT-ICR MS, are also being used, allowing for MSn analysis (Werner et al. 2008).

NMR-based methods are also used in metabolomic analysis (Dixon et al. 2006, Schripsema 2009). These methods can be broadly classified into solution NMR and insoluble or solid-state NMR, according to sample solubility. Using high-resolution (hr)-MAS techniques, it is possible to acquire metabolic fingerprints from insoluble samples and solid-state samples (Bertocchi and Paci 2008). In one-dimensional NMR, protons (^1H) are usually observed (^1H -NMR) due to the sensitivity and common occurrence of this magnetic nucleus. More detailed analyses, such as metabolite identification or flux analysis, can be obtained with other nuclei, particularly ^{13}C and ^{15}N that are coupled with ^1H nuclei in two-dimensional or multidimensional NMR analysis (Kikuchi et al. 2004, Sekiyama and Kikuchi 2007).

The resulting fingerprints, MS or NMR spectra, are pre-processed, including background noise suppression, peak alignment and peak picking. Pre-processed data sets are subsequently used to identify metabolites corresponding to each spectrum signal by searching against compound databases. In non-target analyses, spectrum data sets that include spectra of unknown compounds are subjected to statistical analyses, such as multivariate analysis, to mine the data for biological significance (Tikunov et al. 2005). In target analyses, spectrum data sets that are associated with particular compounds are

used as metabolic profiles for each compound in further analyses.

Data analysis is important in the determination of biological significance in metabolomics. Statistical analyses using multivariate analysis, such as principal component analysis (PCA), hierarchical clustering analysis (HCA) and self-organization mapping (SOM), are typically used to classify samples and/or metabolites (Kose et al. 2001, Hirai et al. 2004, Jonsson et al. 2004, Matsuda et al. 2009). The visualization of metabolic profiles on metabolic pathway maps is also often used and is combined with other omics methods, including gene expression profiles of genes encoding enzymes involved in particular pathways (Thimm et al. 2004, Tokimatsu et al. 2005).

Metabolite profiling in plants

The systematic collection of metabolite profiles is the initial step in metabolomics. This step can be performed with various instruments capable of high throughput and simultaneous measurement, as we mentioned above. Comprehensive metabolic profile data sets can contribute to the understanding of the cellular system in response to changes in intracellular and extracellular environments. Furthermore, the changes in metabolic profiles associated with genetic variations can be evaluated as chemical phenotypes to identify genes involved in particular metabolic pathways.

A number of studies of metabolic profiling in plant species have been performed that have resulted in the publication of related databases. For example, in the case of Arabidopsis, an NSF-funded multi-institutional project aimed at development of the metabolomics database, Plantmetabolomics, has recently been undertaken (http://lab.bcb.iastate.edu/sandbox/pbais05/alpha/plantmetabolomics_trimmed/index.php). Several databases for Solanaceae species are already available. The Metabolome Tomato Database (MoTo DB) was developed as an LC-MS-based metabolome database (<http://appliedbioinformatics.wur.nl/moto/>) (Moco et al. 2006). The KOMICS (Kazusa-omics) database collects annotations of metabolite peaks detected by LC-FT-ICR-MS and contains a representative metabolome data set for the tomato cultivar, Micro-tom (Iijima et al. 2008). The ArmeC Repository Project provides metabolome data on the potato and serves as a data repository for metabolite peaks detected by ESI-MS (<http://www.armec.org/MetaboliteLibrary/index.jsp>). The Golm Metabolome Database (GMD) provides public access to custom mass spectra libraries and metabolite profiling experiments as well as to additional information and related tools (<http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html>) (Kopka et al. 2005). The MS/MS spectral tag (MS2T) libraries at the Platform for Riken Metabolomics (PRIME) website provides access to libraries of phytochemical LC-MS2 spectra obtained from various plant species by using an automatic MS2 acquisition function of LC-ESI-Q-TOF/MS (<http://prime.psc.riken.jp/lcms/ms2tview/ms2tview.html>) (Matsuda et al. 2009). These databases play crucial roles as information resources and repositories of large-scale data sets and also serve as tools for further integration of metabolic

profiles containing comprehensive data acquired from other omics research (Akiyama et al. 2008).

Combinatorial approaches in metabolomics and other omics resources

Metabolome approaches also support the understanding of global relationships among cellular metabolic systems in combination with other omics instances such as profiles of the transcriptome and proteome, and also genetic variations. So far, these combinatorial approaches have been successfully demonstrated in the well-studied Arabidopsis by taking advantage of the many other omics resources that currently exist, including the whole-genome sequence with mature annotations, large-scale transcriptome data sets and related co-expression data, and bioresources such as collections of mutants and full-length cDNA clones. A conceptual scheme for systematic elucidation of gene-to-metabolites molecular networks through a combinatorial approach using transcriptome and metabolome resources has been demonstrated by Saito's group in the RIKEN Plant Science Center (Saito et al. 2008). Data sets containing transcriptome and metabolome changes of Arabidopsis under stress conditions induced by deficiency of sulfur and nitrogen were analyzed using a batch-learning, self-organizing map (BL-SOM) analysis, enabling identification of genes involved in glucosinolate biosynthesis (Hirai et al. 2004). An integrated approach that comprised metabolome and transcriptome analysis was conducted for investigation of an activation-tagged mutant and overexpressors of an MYB TF, PAP1 gene in order to identify genes involved in anthocyanin biosynthesis in Arabidopsis (Tohge et al. 2005). Co-expression data of the Arabidopsis transcriptome provided by the ATTED-II database have been applied to the investigation of key genes involved in specific metabolic pathways and then to the configuration of a metabolome analysis coupled with mutant lines of the targeted genes (Obayashi et al. 2009). The ATTED-II database was used to identify novel genes involved in lipid metabolism, leading to identification of a novel gene, *UDP-glucose pyrophosphorylase3* (*UGP3*) that is required for the first step of sulfolipid biosynthesis (Okazaki et al. 2009). Co-expression analysis was also used to identify all of the genes related to flavonoid biosynthesis, which led to further detailed analysis of two flavonoid pathway genes *UGT78D3* and *RHM1* (Yonekura-Sakakibara et al. 2008).

Approaches that integrated metabolome and transcriptome data have also elucidated regulatory networks that act in response to environmental stresses in plants. The metabolic pathways that act in response to cold and dehydration conditions in Arabidopsis were investigated by metabolome analysis using various types of MS coupled with microarray analysis of overexpressors of genes encoding two TFs, DREB1A/CBF3 and DREB2A (Maruyama et al. 2009). Metabolomic profiling was also used to investigate chemical phenotypic changes between wild-type Arabidopsis and a knockout mutant of the *NCED3* gene under dehydration stress conditions. The metabolic data were then integrated with transcriptome data to reveal ABA-dependent regulatory networks (Urano et al. 2009).

Metabolome profiling has also been used to evaluate chemical phenotypes of natural variations and/or segregation populations simultaneously. A comprehensive exploration of the association between metabolic and genomic diversity will enable the discovery of key genes involved in metabolic changes and would also aid in the identification of genetic associations between metabolic and/or visible phenotypes (Schauer et al. 2008, Fu et al. 2009). Metabolite QTL (mQTL) analysis using segregated populations has been applied to various plant species such as *Arabidopsis*, poplar and tomato in a popular forward genetics approach (Morreel et al. 2006, Schauer et al. 2006, Liseč et al. 2008, Rowe et al. 2008; Schauer et al. 2008). Furthermore, along with the recent availability of data sets of genome-wide variation acquired by high-throughput genotyping methods including resequencing, interest in the discovery of the genetic association between nucleotide variation and phenotypic changes has also increased, especially with regard to the identification of key genes that play significant roles in evolutionary histories. The attempts to mine correlative patterns between metabolic and genomic diversities have recently been applied to sesame and rice using seed stocks of natural variations (Laurentin et al. 2008, Mochida et al. 2009a).

Information resources for metabolomics

Various information resources related to metabolomics have played crucial roles not only in metabolome research but also in synergistic integration with other omics data. The Web site of metabolome resources at TAIR (<http://www.arabidopsis.org/portals/metabolome/index.jsp>) provides a summarized list of Web hyperlinks to resources that facilitate metabolome research. In addition, a data set of biological pathway maps is available via the Kyoto Encyclopedia of Genes and Genomes (KEGG) by using a popular database for information on life sciences called the KEGG PATHWAY Database (<http://www.genome.jp/kegg/pathway.html>) (Kanehisa and Goto 2000, Kanehisa et al. 2008). The Plant Metabolic Network (PMN) is a collaborative project that aims to build plant metabolic pathway databases (<http://www.plantcyc.org/>). One of its main components, PlantCyc, is a comprehensive plant biochemical pathway database that contains curated information from the literature and from computational analyses of genes, enzymes, compounds, reactions and pathways involved in primary and secondary plant metabolism (<http://www.plantcyc.org:1555/PLANT/server.html>) that can be visualized using a pathway tool (<http://bioinformatics.ai.sri.com/ptools/>). AraCyc and PoplarCyc are also available at the PMN Web site and provide manually curated or reviewed information about metabolic pathways in *Arabidopsis* and poplar, respectively (Mueller et al. 2003). There are also metabolic pathway databases for several other plant species generated by PMN collaborators. MapMan is a tool to project omics datasets including gene expression data onto diagrams of metabolic pathways or other processes (<http://mapman.gabipd.org/web/guest>) (Thimm et al. 2004). KaPPA-View is another Web-based analysis tool that can be used to superimpose transcriptome and metabolome

data onto plant metabolic pathway maps (<http://kpv.kazusa.or.jp/kappa-view/>) (Tokimatsu et al. 2005). PRIME is a Web-based service that provides data sets of metabolites measured by multidimensional NMR spectroscopy, GC-MS, LC-MS and CE-MS together with analytical tools to promote integrated approaches using comprehensive data sets within the metabolome and transcriptome (<http://prime.psc.riken.jp/>) (Akiyama et al. 2008).

Mutant resources for phenome analysis

Analysis of mutants is an effective approach for investigation of gene function (Springer 2000, Stanford et al. 2001). Comprehensive collections of mutant lines are also essential biorepositories for radically accelerating forward and reverse genetics. The available mutant resources for phenome analysis in plant species have been well described in a recent review by Kuromori et al. (2009). As described above, various analytical platforms have rapidly evolved, allowing us to discover genes involved in particular phenotypic changes. Along with these analysis platforms, the demands for comprehensive collections of mutants and related information resources have dramatically increased, encouraging high-throughput and genome-wide phenome analysis in plant species (Alonso and Ecker 2006).

Insertion mutant

With the completion of genome sequencing in plants, insertion mutant resources with index data that document the inserted genomic position have become extremely beneficial resources by which to promote functional analysis of annotated genes that are disrupted by a reverse genetics approach. Transferred DNA-tagged (T-DNA-tagged) lines and transposon-tagged lines have become popular resources for the investigation of insertion mutants in plants. T-DNA-tagged lines have emerged as a popular mutant resource due to the rapid generation of large-scale populations in *Arabidopsis* (Krysan et al. 1999). The two-component maize transposon, *Activator (Ac) / Dissociation (Ds)*, is a popular system for inducing transposon-based insertion that enables the generation of mutants with a high proportion of single-copy insertions (Long et al. 1993). In rice, the endogenous retrotransposon *Tos17*, which is activated in particular conditions, is also available for the study of the insertion mutant lines of a *japonica* rice cultivar, Nipponbare (Miyao et al. 2003, Miyao et al. 2007), and the Web resource that provides information on the rice *Tos17* mutant panel with flanking sequences of insertion is available at <http://tos.nias.affrc.go.jp/index.html.en>. Additionally, the maize *Enhancer/Suppressor Mutator (En/Spm)* element has also been used as an effective tool for the study of functional genomics in plants (Kumar et al. 2005). The enhancer trap (ET) and the gene trap (GT) constructs have been coupled with T-DNA and *Ac/Ds* transposons, which additionally facilitates entrapment of genes in monitoring of adjacent promoter or enhancer activity (Sundaresan et al. 1995, An et al. 2005). OryGenesDB (<http://orygenesdb.cirad.fr/>) is a database that integrates information of available insertion

mutant resources in rice (Droc et al. 2009). There are a number of resources for insertion mutant populations with insertion site index-tagged data available for various plant species (Kuromori et al. 2009)

Activation tagging

Activation tagging (AT) is a popular method for generating gain-of-function mutant populations. The method uses T-DNA or a transposable element containing cauliflower mosaic virus 35S enhancer multimers (Weigel et al. 2000). With transcriptional activation of genes near the insertion, novel phenotypes are expected to appear that will identify genes that are redundant or essential for survival. Mutant resources have then been used to isolate genes from Arabidopsis, rice, petunia and tomato (Kakimoto 1996, Zubko et al. 2002, Mathews et al. 2003, Mori et al. 2007). Recently, AT systems using a transposon of maize *En/Spm* or *Ac/Ds* have been developed in Arabidopsis and rice, respectively (Schneider et al. 2005, Qu et al. 2008). A number of AT projects have been performed in various plant species such as Arabidopsis, rice and soybean (Weigel et al. 2000, An et al. 2005, Kuromori et al. 2009).

The FOX hunting system

The FOX gene hunting system is a recently developed novel gain-of-function system that combines a transformation algorithm with large-scale resources for full-length cDNA clones (Ichikawa et al. 2006). A normalized full-length cDNA library of Arabidopsis was first used to generate the Arabidopsis FOX lines, which are available at <http://nazunafox.psc.database.riken.jp>. The system was also used to screen salt stress-resistant lines in the T₁ generation produced by the transformation of 43 focused stress-inducible TFs of Arabidopsis (Fujita et al. 2007). Then, the system was applied to a set of full-length rice cDNA clones aiming for in planta high-throughput screening of rice functional genes, with Arabidopsis as the host species (<http://ricefox.psc.riken.jp>) (Kondou et al. 2009). The FOX lines with rice full-length cDNA transformed into rice plants has been also generated as a gain-of-function mutant resource of rice (Nakamura et al. 2007). A similar technique (in terms of overexpressors using cDNA libraries) has also been carried out in tobacco (Lein et al. 2008).

Chemical and physical mutagenesis

Chemical mutagenic agents, such as ethyl methanesulfonate (EMS), sodium azide and methylnitrosourea (MNU), and physical mutagens, such as fast-neutrons, gamma rays and ion-beam irradiation, have been used to generate mutant populations for many years for forward genetics in various plant species. Targeting induced local lesions in genomes (TILLING) was developed as a general reverse-genetic strategy that provides an allelic series of induced point mutations in genes of interest (Till et al. 2004, Till et al. 2006). Because high-throughput TILLING permits the rapid and low-cost discovery of induced point mutations in populations of chemically mutagenized individuals, the method has been applied to various animal and

plant species. The TILLING technology can also be used to explore allelic variations that are appeared in natural variation; this technology is called EcoTILLING (Comai et al. 2004, Wang et al. 2006). Several laboratory sites have established TILLING and/or EcoTILLING centers for communities of users as a public service (Barkley and Wang 2008). TILLING projects in rice, tomato and Arabidopsis have been performed at the University of California Davis Genome Center (http://tilling.ucdavis.edu/index.php/Main_Page). A soybean mutation database provides soybean mutagenized lines together with data on TILLED genes and their phenotypes (<http://www.soybeantilling.org/index.jsp>). RevGenUK at the John Innes Center provides TILLING service for TILLING populations of *Medicago truncatula*, *Lotus japonicus* and *Brassica rapa* (<http://revgenuk.jic.ac.uk/about.htm>). UTILLdb of INRA is another database for TILLING populations of pea and tomato that provides an interface to search for TILLED lines based on phenotypes (<http://urgv.evry.inra.fr/UTILLdb>).

Gene silencing technologies

Although insertion mutagenesis is an effective method for generating loss-of-function mutants, it also has limitations in the case of redundant genes and lethal mutants. To overcome these limitations, methods to interrupt gene expression have been developed and applied to the functional analysis of plant genes.

RNA interference (RNAi) is a popular method for RNA-mediated gene silencing by sequence-specific degradation of homologous mRNA triggered by double-stranded RNA (dsRNA), which is also known as post-transcriptional gene silencing (PTGS) (Waterhouse et al. 1998, Chuang and Meyerowitz 2000). Constitutive expression of an intron-containing self-complementary hairpin RNA (ihpRNA) has been an effective method for silencing target genes in plants. With demands for conditional silencing of target genes (the silencing of which results in prevention of plant regeneration or embryonic lethality), conditional RNAi systems using a chemical-inducible Cre/loxP recombination system or a promoter of heat shock-inducible genes have been recently developed (Guo et al. 2003, Masclaux et al. 2004). In Arabidopsis, the Complete Arabidopsis Transcriptome MicroArray (CATMA) project has been started to design and produce high-quality gene-specific sequence tags (GSTs) covering most Arabidopsis genes (<http://www.catma.org/>). Using the GST data set of the CATMA project, the Arabidopsis Genomic RNAi Knock-out Line Analysis (AGRIKOLA) project has also been started with the goal of systematically analyzing Arabidopsis genes by RNAi-based technology (<http://www.agrikola.org/index.php?o=/agrikola/html/index>) (Hilson et al. 2003). The *M. truncatula* RNAi Database is also available on the Web as an information resource for RNAi-based gene silencing (<https://mtrnai.msi.umn.edu/>). Virus-induced gene silencing (VIGS) is a derivative method that takes advantage of the plant RNAi-mediated antiviral defense mechanism. The VIGS system was used to assess the function of almost 5000 random *Nicotiana benthamiana* cDNAs in disease resistance (Lu et al. 2003a, Lu et al. 2003b).

The chimeric repressor silencing technology (CRES-T) system was developed as a novel method for gene silencing; it takes advantage of the fact that TFs fused to the EAR motif, a plant-specific repression domain, act as dominant repressors in transgenic plants and therefore suppress the expression of target genes (Hiratsu et al. 2003). The CRES-T system has been applied to the TFs annotated for Arabidopsis in order to analyze their biological function and to obtain transgenic plants with agronomically preferable traits. An associated database, FioreDB, is available at http://www.cres-t.org/fiore/public_db/index.shtml (Mitsuda and Ohme-Takagi 2009).

Plant comparative genomics and databases

The recent accumulation of nucleotide sequences for agricultural species, including crops and domestic animals, now allows us to perform genome-wide comparative analyses of model organisms with the goal of discovering key genes involved in phenotypic characteristics (Sato and Tabata 2006, Itoh et al. 2007, Neale and Ingvarsson 2008). The integration of genomic resources derived from various related species, such as large-scale collections of cDNAs and data from whole-genome sequencing projects, should facilitate sharing of information about gene function between models and applied organisms. This will also accelerate molecular elucidation of cellular systems related to agronomically important traits. A number of information resources for plant genomics accessible on the

Web have appeared, along with appropriate analytical tools. Here we highlight integrative databases promoting plant comparative genomics that we have not described previously. The URLs of each integrative database in plant genomics are shown in **Table 4**.

Portal information resources in plants

TAIR is one of the most popular and integrated information resource in plant science, and it plays an essential role as a portal site in Arabidopsis research (<http://www.arabidopsis.org/>) (Swarbreck et al. 2008). The Salk Institute Genomic Analysis Laboratory (SIGnAL) is also an information resource that integrates various data sets of significant omics results mainly related to Arabidopsis (<http://signal.salk.edu/>). The RIKEN Arabidopsis Genome Encyclopedia (RARGE) provides information on various genomic resources built at RIKEN for Arabidopsis research (http://rarge.gsc.riken.jp/db_home.pl) (Sakurai et al. 2005). Such portal sites have provided gateways for access to comprehensive omics data and/or bioresources. These sites also house cross-referenced data sets built between each annotated gene and its associated instances, such as gene–full-length cDNA clones, gene–mutants, gene–expression patterns and gene–homologous genes. Therefore, to visualize an annotated gene along with genome sequences and associated information, genome browsers such as Gbrowse have been implemented on Web sites (Donlin 2007).

Table 4 Integrative databases in plants

Database name	Species	URL
TAIR	Arabidopsis	http://www.arabidopsis.org/
SIGnAL	Arabidopsis	http://signal.salk.edu/
RARGE	Arabidopsis	http://rarge.psc.riken.jp/
Rice Genome Annotation Project	Rice	http://rice.plantbiology.msu.edu/
RAP-DB	Rice	http://rapdb.dna.affrc.go.jp/
SOL genomics network	Solanaceae	http://solgenomics.net/
Gramene	Gramineae	http://www.gramene.org/
GrainGenes	Triticeae and Avena	http://wheat.pw.usda.gov/GG2/index.shtml
SoyBase	Soybean	http://www.soybase.org/
MazieGDB	Maize	http://www.maizegdb.org/
CyanoBase	Cyanobacteria	http://genome.kazusa.or.jp/cyanobase/
GDR (Genome Database for Rosaceae)	Rosaceae	http://www.bioinfo.wsu.edu/gdr/
Brassica Genome Gateway	Brassica	http://brassica.bbsrc.ac.uk/
Cucurbit Genomics Database	Cucurbitaceae	http://www.icugi.org/
Phytozome	Plant species (whole genome data available)	http://www.phytozome.net/
PlantGDB	Plant species (whole genome and/or large-scale EST data available)	http://www.plantgdb.org/
EnsemblPlants	Plant species (whole genome data available)	http://plants.ensembl.org/index.html
ChloroplastDB	Plant species (Chloroplast genome data available)	http://chloroplast.cbio.psu.edu/
KEGG PLANT	Plant species (whole genome and/or large-scale EST data available)	http://www.genome.jp/kegg/plant/

Gramene is a popular portal site that is not only an integrated rice information resource but also a portal for promoting plant comparative genomics (<http://www.gramene.org/>). Gramene offers integrated genome-oriented data including gene annotation and molecular markers, and also a QTL database mainly for Gramineae species (Liang et al. 2008).

Along with the launch of genome sequencing projects, portal sites to share the progression of outcomes and to integrate related resources have appeared for various species. The Sol genomics network is a portal site for Solanaceae genome resources that includes information on the tomato genome sequencing project (<http://solgenomics.net/>) (Mueller et al. 2005). SoyBase is a resource portal site for genomic soybean research, and it includes released whole-genome sequence data (<http://soybase.org/>). The MaizeGDB is the community database for biological information about *Zea mays*, and includes genetic and genomic data sets and related information (<http://www.maizegdb.org/>) (Lawrence et al. 2004).

Genome-wide comparisons among plants species

With the completion of genome sequencing in a number of plant species, genome-scale comparative analyses can be used to produce and publish data sets that facilitate identification of conserved and/or characteristic properties among plant species.

Using modeled proteome data sets deduced from sequenced genomes in plants, several efforts have been completed to construct comprehensive gene families with the aim of establishing platforms to verify gene content and elucidating the process of gene duplication and functional diversification among species (Sterck et al. 2007). Comprehensive gene family data sets are usually produced by computational procedures including a step that conducts an all-against-all sequence similarity search and then a step for building clusters of protein families by methods such as Markov Clustering (MCL) or consideration of protein domain structures (Hulsen et al. 2006). The results of

such studies can themselves yield databases that are useful for further phylogenetic studies (Horan et al. 2005, Conte et al. 2008, Wall et al. 2008).

Correlated gene arrangements among taxa along with chromosomal allocation, also known as synteny and collinearity, have become valuable frameworks for inference of shared ancestry of genes and for transfer of knowledge from a species to another related species (Tang et al. 2008a). The plant genome duplication database (PGDD) provides a data set of intragenome or cross-genome syntenic relationships identified throughout genome-sequenced plant species (<http://chibba.agtec.uga.edu/duplication/>) (Tang et al. 2008b).

Focused database for plant genomics

Databases housing focused data sets together with rich annotations and well interrelated cross-references are also quite useful for the better understanding of focused issues in particular gene families and/or particular cellular processes.

Sequence-specific DNA-binding TFs are key molecular switches that control or influence many biological processes, such as development or responses to environmental changes. In plants, the genome-wide identification of repertoires of genes encoding TFs of the Arabidopsis genome was reported first, and comparisons with other organisms revealed the properties of plant-specific TFs (Riechmann et al. 2000). In the past decade, with the availability of complete genome sequences, we have been able to compile catalogs describing the function and organization of TF regulatory systems in a number of organisms. There are many databases that provide data sets of genes putatively encoding TFs in many plant species; these are usually predictions based on computational methods such as sequence similarity search and/or hidden Markov model search of conserved DNA-binding domains (Table 5). Recently, further integration of data sets of TF-encoding genes has been performed, thus establishing an integrative, knowledge-based resource of

Table 5 Transcription factor database in plants

Database	URL	Species	References
RARTF	http://rarge.gsc.riken.jp/rartf/	Arabidopsis	Iida et al. (2005)
AGRIS, AtTFDB	http://arabidopsis.med.ohio-state.edu/AtTFDB/	Arabidopsis	Palaniswamy et al. (2006)
DATF	http://datf.cbi.pku.edu.cn/	Arabidopsis	Guo et al. (2005)
DRTF	http://drtf.cbi.pku.edu.cn/	Rice	Gao et al. (2006)
DPTF	http://dptf.cbi.pku.edu.cn/	Poplar	Zhu et al. (2007)
TOBFAC	http://compsysbio.achs.virginia.edu/tobfac/	Tobacco	Rushton et al. (2008)
SoybeanTFDB	http://soybeantfdb.psc.riken.jp/	Soybean	Mochida et al. (2009c)
PlantTFDB	http://planttfdb.cbi.pku.edu.cn/	22 plant species	Guo et al. (2008)
PlnTFDB	http://plntfdb.bio.uni-potsdam.de/v3.0/	20 plant species	Riano-Pachon et al. (2007)
GRASSIUS, GrassTFDB	http://grassius.org/grasstfdb.html	Maize, rice, sorghum, sugarcane	Yilmaz et al. (2009)
LegumeTFDB	http://legumetfdb.psc.riken.jp/	Soybean, <i>Lotus japonicus</i> , <i>Medicago truncatula</i>	Mochida et al. (2010)
DBD	http://dbd.mrc-lmb.cam.ac.uk/DBD/index.cgi?Home	>700 species	Wilson et al. (2008)

TFs across related plant species in terms of comparative genomics of transcriptional regulatory networks. GRASSIUS provides the first step toward building a comprehensive platform for integration of information, tools and resources for comparative regulatory genomics across the grass species (Yilmaz et al. 2009). The Grass Transcription Factor Database (GrassTFDB) of GRASSIUS houses integrated information on MaizeTFDB, RiceTFDB, SorghumTFDB and CaneTFDB (<http://grassius.org/grasstfdb.html>). The LegumeTFDB provides predicted TF-encoding genes annotated in the genome sequences of three major legume species: soybean, *L. japonicus* and *M. truncatula* (<http://legumetfdb.psc.riken.jp/>). This database is an extended version of the SoybeanTFDB (<http://soybeantfdb.psc.riken.jp/>) and is aimed at integrating knowledge on legume TFs and providing a public resource for comparative genomics of the TFs of legumes, non-legume plants and other organisms (Mochida et al. 2009c, Mochida et al. 2010).

Acknowledgments

The authors thank Takuhiro Yoshida, Lam-Son Phan Tran and Tetsuya Sakurai of RIKEN Plant Science Center for their valuable suggestions and comments.

References

- Adams, M.D., Soares, M.B., Kerlavage, A.R., Fields, C. and Venter, J.C. (1993) Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nat. Genet.* 4: 373–380.
- Agrawal, G.K., Yonekura, M., Iwahashi, Y., Iwahashi, H. and Rakwal, R. (2005) System, trends and perspectives of proteomics in dicot plants. Part I: technologies in proteome establishment. *J. Chromatogr. B* 815: 109–123.
- Akhunov, E., Nicolet, C. and Dvorak, J. (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theor. Appl. Genet.* 119: 507–517.
- Akiyama, K., Chikayama, E., Yuasa, H., Shimada, Y., Tohge, T., Shinozaki, K., et al. (2008) PRiME: a Web site that assembles tools for metabolomics and transcriptomics. *In Silico Biol.* 8: 339–345.
- Alonso, J.M. and Ecker, J.R. (2006) Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. *Nat. Rev. Genet.* 7: 524–536.
- An, G., Jeong, D.H., Jung, K.H. and Lee, S. (2005) Reverse genetic approaches for functional genomics of rice. *Plant Mol. Biol.* 59: 111–123.
- Andersen, J.S. and Mann, M. (2006) Organellar proteomics: turning inventories into insights. *EMBO Rep.* 7: 874–879.
- Andreeva, A., Howorth, D., Chandonia, J.M., Brenner, S.E., Hubbard, T.J., Chothia, C., et al. (2008) Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res.* 36: D419–D425.
- Anisimov, S.V. (2008) Serial analysis of gene expression (SAGE): 13 years of application in research. *Curr. Pharm. Biotechnol.* 9: 338–350.
- Anson, W.J. (2009) Next-generation DNA sequencing techniques. *Nat. Biotechnol.* 25: 195–203.
- Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., et al. (2005) Cytokinin oxidase regulates rice grain production. *Science* 309: 741–745.
- Baginsky, S. (2009) Plant proteomics: concepts, applications, and novel strategies for data interpretation. *Mass Spectrom. Rev.* 28: 93–120.
- Barakat, A., Wall, P.K., Diloreto, S., Depamphilis, C.W. and Carlson, J.E. (2007) Conservation and divergence of microRNAs in Populus. *BMC Genomics* 8: 481.
- Barkley, N.A. and Wang, M.L. (2008) Application of TILLING and EcoTILLING as reverse genetic approaches to elucidate the function of genes in plants and animals. *Curr. Genomics* 9: 212–226.
- Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., et al. (2009) NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.* 37: D885–D890.
- Benschop, J.J., Mohammed, S., O'Flaherty, M., Heck, A.J., Slijper, M. and Menke, F.L. (2007) Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. *Mol. Cell. Proteomics* 6: 1198–1214.
- Bernardo, A.N., Bradbury, P.J., Ma, H., Hu, S., Bowden, R.L., Buckler, E.S., et al. (2009) Discovery and mapping of single feature polymorphisms in wheat using Affymetrix arrays. *BMC Genomics* 10: 251.
- Bertocchi, F. and Paci, M. (2008) Applications of high-resolution solid-state NMR spectroscopy in food science. *J. Agric. Food. Chem.* 56: 9317–9327.
- Bevan, M. (1997) Objective: the complete sequence of a plant genome. *Plant Cell* 9: 476–478.
- Bino, R.J., Hall, R.D., Fiehn, O., Kopka, J., Saito, K., Draper, J., et al. (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci.* 9: 418–425.
- Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., et al. (2003) A gene expression map of the Arabidopsis root. *Science* 302: 1956–1960.
- Blair, M.W., Torres, M.M., Giraldo, M.C. and Pedraza, F. (2009) Development and diversity of Andean-derived, gene-based microsatellites for common bean (*Phaseolus vulgaris* L.). *BMC Plant Biol.* 9: 100.
- Bogdanov, B. and Smith, R.D. (2005) Proteomics by FTICR mass spectrometry: top down and bottom up. *Mass Spectrom. Rev.* 24: 168–200.
- Boguski, M.S., Lowe, T.M. and Tolstoshev, C.M. (1993) dbEST—database for 'expressed sequence tags'. *Nat. Genet.* 4: 332–333.
- Brady, S.M., Orlando, D.A., Lee, J.Y., Wang, J.Y., Koch, J., Dinneny, J.R., et al. (2007) A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* 318: 801–806.
- Brady, S.M. and Provart, N.J. (2009) Web-queryable large-scale data sets for hypothesis generation in plant biology. *Plant Cell* 21: 1034–1051.
- Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., et al. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* 18: 630–634.
- Brown, M.E. and Funk, C.C. (2008) Climate. Food security under climate change. *Science* 319: 580–581.
- Caicedo, A.L., Williamson, S.H., Hernandez, R.D., Boyko, A., Fledel-Alon, A., York, T.L., et al. (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genet.* 3: 1745–1756.
- Carollo, V., Matthews, D.E., Lazo, G.R., Blake, T.K., Hummel, D.D., Lui, N., et al. (2005) GrainGenes 2.0. an improved resource for the small-grains community. *Plant Physiol.* 139: 643–651.
- Castellani, F., van Rossum, B., Diehl, A., Schubert, M., Rehbein, K. and Oschkinat, H. (2002) Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* 420: 98–102.
- Chellappan, P. and Jin, H. (2009) Discovery of plant microRNAs and short-interfering RNAs by deep parallel sequencing. *Methods Mol. Biol.* 495: 121–132.

- Chen, S. and Harmon, A.C. (2006) Advances in plant proteomics. *Proteomics* 6: 5504–5516.
- Cheng, Y., Qi, Y., Zhu, Q., Chen, X., Wang, N., Zhao, X., et al. (2009) New changes in the plasma-membrane-associated proteome of rice roots under salt stress. *Proteomics* 9: 3100–3114.
- Chuang, C.F. and Meyerowitz, E.M. (2000) Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* 97: 4985–4990.
- Close, T.J., Bhat, P.R., Lonardi, S., Wu, Y., Rostoks, N., Ramsay, L., et al. (2009) Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics* 10: 582.
- Close, T.J., Wanamaker, S., Roose, M.L. and Lyon, M. (2007) HarVEST: an EST database and viewing software. *Methods Mol. Biol.* 406: 161–178.
- Cogburn, L.A., Porter, T.E., Duclos, M.J., Simon, J., Burgess, S.C., Zhu, J.J., et al. (2007) Functional genomics of the chicken—a model organism. *Poult. Sci.* 86: 2059–2094.
- Comai, L., Young, K., Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., et al. (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *Plant J.* 37: 778–786.
- Conte, M.G., Gaillard, S., Lanau, N., Rouard, M. and Perin, C. (2008) GreenPhylDB: a database for plant comparative genomics. *Nucleic Acids Res.* 36: D991–D998.
- Crossa, J., Burgueno, J., Dreisigacker, S., Vargas, M., Herrera-Foessel, S.A., Lillemo, M., et al. (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. *Genetics* 177: 1889–1913.
- De Bodt, S., Maere, S. and Van de Peer, Y. (2005) Genome duplication and the origin of angiosperms. *Trends Ecol. Evol.* 20: 591–597.
- de Hoon, M. and Hayashizaki, Y. (2008) Deep cap analysis gene expression (CAGE): genome-wide identification of promoters, quantification of their expression, and network inference. *Biotechniques* 44: 627–628, 630, 632.
- de la Fuente van Bentem, S., Anrather, D., Roitinger, E., Djamei, A., Hufnagl, T., Barta, A., et al. (2006) Phosphoproteomics reveals extensive in vivo phosphorylation of *Arabidopsis* proteins involved in RNA metabolism. *Nucleic Acids Res.* 34: 3267–3278.
- Deleu, W., Esteras, C., Roig, C., Gonzalez-To, M., Fernandez-Silva, I., Gonzalez-Ibeas, D.D., et al. (2009) A set of EST-SNPs for map saturation and cultivar identification in melon. *BMC Plant Biol.* 9: 90.
- Dixon, R.A., Gang, D.R., Charlton, A.J., Fiehn, O., Kuiper, H.A., Reynolds, T.L., et al. (2006) Applications of metabolomics in agriculture. *J. Agric. Food. Chem.* 54: 8984–8994.
- Donlin, M.J. (2007) Using the Generic Genome Browser (GBrowse). *Curr. Protoc. Bioinformatics* Chapter 9: Unit 9.9.
- Droc, G., Perin, C., Fromentin, S. and Larmande, P. (2009) OryGenesDB 2008 update: database interoperability for functional genomics of rice. *Nucleic Acids Res.* 37: D992–D995.
- Dunkley, T.P., Hester, S., Shadforth, I.P., Runions, J., Weimar, T., Hanton, S.L., et al. (2006) Mapping the *Arabidopsis* organelle proteome. *Proc. Natl Acad. Sci. USA* 103: 6518–6523.
- Duvick, J., Fu, A., Muppirala, U., Sabharwal, M., Wilkerson, M.D., Lawrence, C.J., et al. (2008) PlantGDB: a resource for comparative plant genomics. *Nucleic Acids Res.* 36: D959–D965.
- Endo, Y. and Sawasaki, T. (2003) High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system. *Biotechnol. Adv.* 21: 695–713.
- Endo, Y. and Sawasaki, T. (2006) Cell-free expression systems for eukaryotic protein production. *Curr. Opin. Biotechnol.* 17: 373–380.
- Ewing, B., Hillier, L., Wendl, M.C. and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8: 175–185.
- Farrer, R.A., Kemen, E., Jones, J.D. and Studholme, D.J. (2009) De novo assembly of the *Pseudomonas syringae* pv. *syringae* B728a genome using Illumina/Solexa short sequence reads. *FEMS Microbiol. Lett.* 291: 103–111.
- Fei, Z., Tang, X., Alba, R.M., White, J.A., Ronning, C.M., Martin, G.B., et al. (2004) Comprehensive EST analysis of tomato and comparative genomics of fruit ripening. *Plant J.* 40: 47–59.
- Feltus, F.A., Wan, J., Schulze, S.R., Estill, J.C., Jiang, N. and Paterson, A.H. (2004) An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. *Genome Res.* 14: 1812–1819.
- Feolo, M., Helmsberg, W., Sherry, S. and Maglott, D.R. (2000) NCBI genetic resources supporting immunogenetic research. *Rev. Immunogenet.* 2: 461–467.
- Fernie, A.R. and Schauer, N. (2009) Metabolomics-assisted breeding: a viable option for crop improvement? *Trends Genet.* 25: 39–48.
- Fernie, A.R., Trethewey, R.N., Krotzky, A.J. and Willmitzer, L. (2004) Metabolite profiling: from diagnostics to systems biology. *Nat. Rev. Mol. Cell Biol.* 5: 763–769.
- Fiehn, O., Kopka, J., Dormann, P., Altmann, T., Trethewey, R.N. and Willmitzer, L. (2000) Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* 18: 1157–1161.
- Finn, R.D., Tate, J., Mistry, J., Coghill, P.C., Sammut, S.J., Hotz, H.R., et al. (2008) The Pfam protein families database. *Nucleic Acids Res.* 36: D281–D288.
- Flicek, P., Aken, B.L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., et al. (2008) Ensembl 2008. *Nucleic Acids Res.* 36: D707–D714.
- Fox, B.G., Goulding, C., Malkowski, M.G., Stewart, L. and Deacon, A. (2008) Structural genomics: from genes to structures with valuable materials and many questions in between. *Nat. Methods* 5: 129–132.
- Fu, J., Keurentjes, J.J., Bouwmeester, H., America, T., Verstappen, F.W., Ward, J.L., et al. (2009) System-wide molecular evidence for phenotypic buffering in *Arabidopsis*. *Nat. Genet.* 41: 166–167.
- Fujita, M., Mizukado, S., Fujita, Y., Ichikawa, T., Nakazawa, M., Seki, M., et al. (2007) Identification of stress-tolerance-related transcription-factor genes via mini-scale Full-length cDNA Over-eXpressor (FOX) gene hunting system. *Biochem. Biophys. Res. Commun.* 364: 250–257.
- Fukuchi, S., Homma, K., Sakamoto, S., Sugawara, H., Tateno, Y., Gojobori, T., et al. (2009) The GTOP database in 2009: updated content and novel features to expand and deepen insights into protein structures and functions. *Nucleic Acids Res.* 37: D333–D337.
- Futamura, N., Totoki, Y., Toyoda, A., Igasaki, T., Nanjo, T., Seki, M., et al. (2008) Characterization of expressed sequence tags from a full-length enriched cDNA library of *Cryptomeria japonica* male strobili. *BMC Genomics* 9: 383.
- Gao, G., Zhong, Y., Guo, A., Zhu, Q., Tang, W., Zheng, W., et al. (2006) DRTF: a database of rice transcription factors. *Bioinformatics* 22: 1286–1287.
- Gao, J., Agrawal, G.K., Thelen, J.J. and Xu, D. (2009) P3DB: a plant protein phosphorylation database. *Nucleic Acids Res.* 37: D960–D962.
- Goda, H., Sasaki, E., Akiyama, K., Maruyama-Nakashita, A., Nakabayashi, K., Li, W., et al. (2008) The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J.* 55: 526–542.

- Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296: 92–100.
- Greene, L.H., Lewis, T.E., Addou, S., Cuff, A., Dallman, T., Dibley, M., et al. (2007) The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution. *Nucleic Acids Res.* 35: D291–D297.
- Gstaiger, M. and Aebersold, R. (2009) Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat. Rev. Genet.* 10: 617–627.
- Guo, A., He, K., Liu, D., Bai, S., Gu, X., Wei, L., et al. (2005) DATF: a database of Arabidopsis transcription factors. *Bioinformatics* 21: 2568–2569.
- Guo, A.Y., Chen, X., Gao, G., Zhang, H., Zhu, Q.H., Liu, X.C., et al. (2008) PlantTFDB: a comprehensive plant transcription factor database. *Nucleic Acids Res.* 36: D966–D969.
- Guo, H.S., Fei, J.F., Xie, Q. and Chua, N.H. (2003) A chemical-regulated inducible RNAi system in plants. *Plant J.* 34: 383–392.
- Haas, B.J., Delcher, A.L., Wortman, J.R. and Salzberg, S.L. (2004) DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics* 20: 3643–3646.
- Han, B. and Xue, Y. (2003) Genome-wide intraspecific DNA-sequence variations in rice. *Curr. Opin. Plant Biol.* 6: 134–138.
- Harbers, M. and Carninci, P. (2005) Tag-based approaches for transcriptome research and genome annotation. *Nat. Methods* 2: 495–502.
- Hashimoto, S., Suzuki, Y., Kasai, Y., Morohoshi, K., Yamada, T., Sese, J., et al. (2004) 5'-End SAGE for the analysis of transcriptional start sites. *Nat. Biotechnol.* 22: 1146–1149.
- Hayashizaki, Y. (2003) RIKEN mouse genome encyclopedia. *Mech. Ageing Dev.* 124: 93–102.
- Heesacker, A., Kishore, V.K., Gao, W., Tang, S., Kolkman, J.M., Gingle, A., et al. (2008) SSRs and INDELS mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. *Theor. Appl. Genet.* 117: 1021–1029.
- Hilson, P., Small, I. and Kuiper, M.T. (2003) European consortia building integrated resources for Arabidopsis functional genomics. *Curr. Opin. Plant Biol.* 6: 426–429.
- Hirai, M.Y., Yano, M., Goodenow, D.B., Kanaya, S., Kimura, T., Awazuhara, M., et al. (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. *Proc. Natl Acad. Sci. USA* 101: 10205–10210.
- Hirano, H., Islam, N. and Kawasaki, H. (2004) Technical aspects of functional proteomics in plants. *Phytochemistry* 65: 1487–1498.
- Hiratsu, K., Matsui, K., Koyama, T. and Ohme-Takagi, M. (2003) Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in Arabidopsis. *Plant J.* 34: 733–739.
- Hishiki, T., Kawamoto, S., Morishita, S. and Okubo, K. (2000) BodyMap: a human and mouse gene expression database. *Nucleic Acids Res.* 28: 136–138.
- Horan, K., Lauricha, J., Bailey-Serres, J., Raikhel, N. and Girke, T. (2005) Genome cluster database. A sequence family analysis platform for Arabidopsis and rice. *Plant Physiol.* 138: 47–54.
- Hori, K., Sato, K. and Takeda, K. (2007) Detection of seed dormancy QTL in multiple mapping populations derived from crosses involving novel barley germplasm. *Theor. Appl. Genet.* 115: 869–876.
- Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., et al. (2009) High-throughput genotyping by whole-genome resequencing. *Genome Res.* 19: 1068–1076.
- Huang, X. and Madan, A. (1999) CAP3: A DNA sequence assembly program. *Genome Res.* 9: 868–877.
- Hulsen, T., de Vlieg, J. and Groenen, P.M. (2006) PhyloPat: phylogenetic pattern analysis of eukaryotic genes. *BMC Bioinformatics* 7: 398.
- Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., et al. (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res.* 37: D211–D215.
- Hyten, D.L., Song, Q., Choi, I.Y., Yoon, M.S., Specht, J.E., Matukumalli, L.K., et al. (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. *Theor. Appl. Genet.* 116: 945–952.
- Ichikawa, T., Nakazawa, M., Kawashima, M., Iizumi, H., Kuroda, H., Kondou, Y., et al. (2006) The FOX hunting system: an alternative gain-of-function gene hunting technique. *Plant J.* 48: 974–985.
- Igawa, T., Fujiwara, M., Takahashi, H., Sawasaki, T., Endo, Y., Seki, M., et al. (2009) Isolation and identification of ubiquitin-related proteins from Arabidopsis seedlings. *J. Exp. Bot.* 60: 3067–3073.
- Iida, K., Seki, M., Sakurai, T., Satou, M., Akiyama, K., Toyoda, T., et al. (2004) Genome-wide analysis of alternative pre-mRNA splicing in Arabidopsis thaliana based on full-length cDNA sequences. *Nucleic Acids Res.* 32: 5096–5103.
- Iida, K., Seki, M., Sakurai, T., Satou, M., Akiyama, K., Toyoda, T., et al. (2005) RARTF: database and tools for complete sets of Arabidopsis transcription factors. *DNA Res.* 12: 247–256.
- Iijima, Y., Nakamura, Y., Ogata, Y., Tanaka, K., Sakurai, N., Suda, K., et al. (2008) Metabolite annotations based on the integration of mass spectral information. *Plant J.* 54: 949–962.
- Imanishi, T., Itoh, T., Suzuki, Y., O'Donovan, C., Fukuchi, S., Koyanagi, K.O., et al. (2004) Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS Biol.* 2: e162.
- International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436: 793–800.
- Islam, N., Lonsdale, M., Upadhyaya, N.M., Higgins, T.J., Hirano, H. and Akhurst, R. (2004) Protein extraction from mature rice leaves for two-dimensional gel electrophoresis and its application in proteome analysis. *Proteomics* 4: 1903–1908.
- Itoh, T., Tanaka, T., Barrero, R.A., Yamasaki, C., Fujii, Y., Hilton, P.B., et al. (2007) Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with Arabidopsis thaliana. *Genome Res.* 17: 175–183.
- Jaccoud, D., Peng, K., Feinstein, D. and Kilian, A. (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res.* 29: E25.
- Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463–467.
- Jiao, Y., Tausta, S.L., Gandotra, N., Sun, N., Liu, T., Caly, N.K., et al. (2009) A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nat. Genet.* 41: 258–263.
- Jonsson, P., Gullberg, J., Nordstrom, A., Kusano, M., Kowalczyk, M., Sjostrom, M., et al. (2004) A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Anal. Chem.* 76: 1738–1745.
- Jorrin-Novo, J.V., Maldonado, A.M., Echevarria-Zomeno, S., Villedor, L., Castillejo, M.A., Curto, M., et al. (2009) Plant proteomics update (2007–2008): second-generation proteomic techniques, an appropriate experimental design, and data analysis to fulfill MIAPE standards, increase plant proteome coverage and expand biological knowledge. *J. Proteomics* 72: 285–314.

- Kakimoto, T. (1996) CK11, a histidine kinase homolog implicated in cytokinin signal transduction. *Science* 274: 982–985.
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36: D480–D484.
- Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28: 27–30.
- Kantety, R.V., La Rota, M., Matthews, D.E. and Sorrells, M.E. (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol. Biol.* 48: 501–510.
- Katz, A., Waridel, P., Shevchenko, A. and Pick, U. (2007) Salt-induced changes in the plasma membrane proteome of the halotolerant alga *Dunaliella salina* as revealed by blue native gel electrophoresis and nano-LC-MS/MS analysis. *Mol. Cell. Proteomics* 6: 1459–1472.
- Kaur, S., Cogan, N.O., Ye, G., Baillie, R.C., Hand, M.L., Ling, A.E., et al. (2009) Genetic map construction and QTL mapping of resistance to blackleg (*Leptosphaeria maculans*) disease in Australian canola (*Brassica napus* L.) cultivars. *Theor. Appl. Genet.* 120: 71–83.
- Kawamoto, S., Yoshii, J., Mizuno, K., Ito, K., Miyamoto, Y., Ohnishi, T., et al. (2000) BodyMap: a collection of 3' ESTs for analysis of human gene expression information. *Genome Res.* 10: 1817–1827.
- Kawaura, K., Mochida, K., Enju, A., Totoki, Y., Toyoda, A., Sakaki, Y., et al. (2009) Assessment of adaptive evolution between wheat and rice as deduced from full-length common wheat cDNA sequence data and expression patterns. *BMC Genomics* 10: 271.
- Kawaura, K., Mochida, K., Yamazaki, Y. and Ogihara, Y. (2006) Transcriptome analysis of salinity stress responses in common wheat using a 22k oligo-DNA microarray. *Funct. Integr. Genomics* 6: 132–142.
- Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T., et al. (1999) Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett.* 442: 15–19.
- Kikuchi, J., Shinozaki, K. and Hirayama, T. (2004) Stable isotope labeling of *Arabidopsis thaliana* for an NMR-based metabolomics approach. *Plant Cell Physiol.* 45: 1099–1104.
- Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., et al. (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *Science* 301: 376–379.
- Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., et al. (2007) The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J.* 50: 347–363.
- Kolesnik, T., Szevenyi, I., Bachmann, D., Kumar, C.S., and Jiang, S. (2004) Establishing an efficient Ac/Ds tagging system in rice: large-scale analysis of Ds flanking sequences. *Plant J.* 37:301–314.
- Komatsu, S. (2005) Rice proteome database: a step toward functional analysis of the rice genome. *Plant Mol. Biol.* 59: 179–190.
- Kondou, Y., Higuchi, M., Takahashi, S., Sakurai, T., Ichikawa, T., Kuroda, H., et al. (2009) Systematic approaches to using the FOX hunting system to identify useful rice genes. *Plant J.* 57: 883–894.
- Konishi, S., Izawa, T., Lin, S.Y., Ebana, K., Fukuta, Y., Sasaki, T. and Yano, M. (2006) An SNP caused loss of seed shattering during rice domestication. *Science* 312: 1392–1396.
- Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Burgmüller, E., et al. (2005) GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* 21: 1635–1638.
- Kose, F., Weckwerth, W., Linke, T. and Fiehn, O. (2001) Visualizing plant metabolomic correlation networks using clique-metabolite matrices. *Bioinformatics* 17: 1198–1208.
- Kota, R., Rudd, S., Facius, A., Kolesov, G., Thiel, T., Zhang, H., et al. (2003) Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L.). *Mol. Genet. Genomics* 270: 24–33.
- Kota, R., Varshney, R.K., Prasad, M., Zhang, H., Stein, N. and Graner, A. (2008) EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome. *Funct. Integr. Genomics* 8: 223–233.
- Kota, R., Varshney, R.K., Thiel, T., Dehmer, K.J. and Graner, A. (2001) Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). *Hereditas* 135: 145–151.
- Kouranov, A., Xie, L., de la Cruz, J., Chen, L., Westbrook, J., Bourne, P.E., et al. (2006) The RCSB PDB information portal for structural genomics. *Nucleic Acids Res.* 34: D302–305.
- Krysan, P.J., Young, J.C. and Sussman, M.R. (1999) T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell* 11: 2283–2290.
- Kumar, C.S., Wing, R.A. and Sundaresan, V. (2005) Efficient insertional mutagenesis in rice using the maize *En/Spm* elements. *Plant J.* 44: 879–892.
- Kurakawa, T., Ueda, N., Maekawa, M., Kobayashi, K., Kojima, M., Nagato, Y., et al. (2007) Direct control of shoot meristem activity by a cytokinin-activating enzyme. *Nature* 445: 652–655.
- Kuromori, T., Takahashi, S., Kondou, Y., Shinozaki, K. and Matsui, M. (2009) Phenome analysis in plant species using loss-of-function and gain-of-function mutants. *Plant Cell Physiol.* 50: 1215–1231.
- Kwon, S.J., Choi, E.Y., Choi, Y.J., Ahn, J.H. and Park, O.K. (2006) Proteomics studies of post-translational modifications in plants. *J. Exp. Bot.* 57: 1547–1551.
- Laubinger, S., Zeller, G., Henz, S.R., Sachsenberg, T., Widmer, C.K., Naouar, N., et al. (2008) At-TAX: a whole genome tiling array resource for developmental expression analysis and transcript identification in *Arabidopsis thaliana*. *Genome Biol.* 9: R112.
- Laurentin, H., Ratzinger, A. and Karlovsky, P. (2008) Relationship between metabolic and genomic diversity in sesame (*Sesamum indicum* L.). *BMC Genomics* 9: 250.
- Lawrence, C.J., Dong, Q., Polacco, M.L., Seigfried, T.E. and Brendel, V. (2004) MaizeGDB, the community database for maize genetics and genomics. *Nucleic Acids Res.* 32: D393–D397.
- Lee, Y., Tsai, J., Sunkara, S., Karamycheva, S., Perlea, G., Sultana, R., et al. (2005) The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. *Nucleic Acids Res.* 33: D71–D74.
- Lein, W., Usadel, B., Stitt, M., Reindl, A., Ehrhardt, T., Sonnwald, U., et al. (2008) Large-scale phenotyping of transgenic tobacco plants (*Nicotiana tabacum*) to identify essential leaf functions. *Plant Biotechnol. J.* 6: 246–263.
- Li, F., Kitashiba, H., Inaba, K. and Nishio, T. (2009) A Brassica rapa linkage map of EST-based SNP markers for identification of candidate genes controlling flowering time and leaf morphological traits. *DNA Res.* 16: 311–323.
- Liang, C., Jaiswal, P., Hebbard, C., Avraham, S., Buckler, E.S., Casstevens, T., et al. (2008) Gramene: a growing plant comparative genomics resource. *Nucleic Acids Res.* 36: D947–D953.
- Lisec, J., Meyer, R.C., Steinfath, M., Redestig, H., Becher, M., Witucka-Wall, H., et al. (2008) Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. *Plant J.* 53: 960–972.
- Liu, X., Lu, T., Yu, S., Li, Y., Huang, Y., Huang, T., et al. (2007) A collection of 10,096 indica rice full-length cDNAs reveals highly expressed sequence divergence between *Oryza sativa indica* and japonica subspecies. *Plant Mol. Biol.* 65: 403–415.

- Long, D., Martin, M., Sundberg, E., Swinburne, J., Puangsomlee, P. and Coupland, G. (1993) The maize transposable element system Ac/Ds as a mutagen in Arabidopsis: identification of an albino mutation induced by Ds insertion. *Proc. Natl Acad. Sci. USA* 90: 10370–10374.
- Lu, C., Kulkarni, K., Souret, F.F., MuthuValliappan, R., Tej, S.S., Poethig, R.S., et al. (2006) MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. *Genome Res.* 16: 1276–1288.
- Lu, R., Malcuit, I., Moffett, P., Ruiz, M.T., Peart, J., Wu, A.J., et al. (2003a) High throughput virus-induced gene silencing implicates heat shock protein 90 in plant disease resistance. *EMBO J.* 22: 5690–5699.
- Lu, R., Martin-Hernandez, A.M., Peart, J.R., Malcuit, I. and Baulcombe, D.C. (2003b) Virus-induced gene silencing in plants. *Methods* 30: 296–303.
- Lu, T., Yu, S., Fan, D., Mu, J., Shangquan, Y., Wang, Z., et al. (2008) Collection and comparative analysis of 1888 full-length cDNAs from wild rice *Oryza rufipogon* Griff. W1943. *DNA Res.* 15: 285–295.
- Ma, J.F., Tamai, K., Yamaji, N., Mitani, N., Konishi, S., Katsuhara, M., et al. (2006) A silicon transporter in rice. *Nature* 440: 688–691.
- Ma, J.F., Yamaji, N., Mitani, N., Tamai, K., Konishi, S., Fujiwara, T., et al. (2007) An efflux transporter of silicon in rice. *Nature* 448: 209–212.
- Mace, E.S., Rami, J.F., Bouchet, S., Klein, P.E., Klein, R.R., Kilian, A., et al. (2009) A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. *BMC Plant Biol.* 9: 13.
- Madin, K., Sawasaki, T., Ogasawara, T. and Endo, Y. (2000) A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. *Proc. Natl Acad. Sci. USA* 97: 559–564.
- Maeda, N., Kasukawa, T., Oyama, R., Gough, J., Frith, M., Engström, P.G., et al. (2006) Transcript annotation in FANTOM3: mouse gene catalog based on physical cDNAs. *PLoS Genet.* 2: e62.
- Manzano, C., Abraham, Z., Lopez-Torres, G. and Del Pozo, J.C. (2008) Identification of ubiquitinated proteins in Arabidopsis. *Plant Mol. Biol.* 68: 145–158.
- Maor, R., Jones, A., Nuhse, T.S., Studholme, D.J., Peck, S.C. and Shirasu, K. (2007) Multidimensional protein identification technology (MudPIT) analysis of ubiquitinated proteins in plants. *Mol. Cell. Proteomics* 6: 601–610.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bembien, L.A., et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376–380.
- Marques, M.C., Alonso-Cantabrana, H., Forment, J., Arribas, R., Alamar, S., Conejero, V., et al. (2009) A new set of ESTs and cDNA clones from full-length and normalized libraries for gene discovery and functional characterization in citrus. *BMC Genomics* 10: 428.
- Maruyama, K., Takeda, M., Kidokoro, S., Yamada, K., Sakuma, Y., Urano, K., et al. (2009) Metabolic pathways involved in cold acclimation identified by integrated analysis of metabolites and transcripts regulated by DREB1A and DREB2A. *Plant Physiol.* 150: 1972–1980.
- Masclaux, F., Charpentreau, M., Takahashi, T., Pont-Lezica, R. and Galaud, J.P. (2004) Gene silencing using a heat-inducible RNAi system in Arabidopsis. *Biochem. Biophys. Res. Commun.* 321: 364–369.
- Masoudi-Nejad, A., Tonomura, K., Kawashima, S., Moriya, Y., Suzuki, M., Itoh, M., et al. (2006) EGASSEMBLER: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. *Nucleic Acids Res.* 34: W459–W462.
- Mathews, H., Clendennen, S.K., Caldwell, C.G., Liu, X.L., Connors, K., Matheis, N., et al. (2003) Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. *Plant Cell* 15: 1689–1703.
- Matsuda, F., Yonekura-Sakakibara, K., Niida, R., Kuromori, T., Shinozaki, K. and Saito, K. (2009) MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. *Plant J.* 57: 555–577.
- Matsui, A., Ishida, J., Morosawa, T., Mochizuki, Y., Kaminuma, E., Endo, T.A., et al. (2008) Arabidopsis transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. *Plant Cell Physiol.* 49: 1135–1149.
- Matsumura, H., Kruger, D.H., Kahl, G. and Terauchi, R. (2008) SuperSAGE: a modern platform for genome-wide quantitative transcript profiling. *Curr. Pharm. Biotechnol.* 9: 368–374.
- Matsumura, H., Reich, S., Ito, A., Saitoh, H., Kamoun, S., Winter, P., et al. (2003) Gene expression analysis of plant host–pathogen interactions by SuperSAGE. *Proc. Natl Acad. Sci. USA* 100: 15718–15723.
- Matsuzaki, M., Misumi, O., Shin, I.T., Maruyama, S., Takahara, M., Miyagishima, S.Y., et al. (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428: 653–657.
- McDermott, A. (2009) Structure and dynamics of membrane proteins by magic angle spinning solid-state NMR. *Annu. Rev. Biophys.* 38: 385–403.
- Mechin, V., Balliau, T., Chateau-Joubert, S., Davanture, M., Langella, O., Negroni, L., et al. (2004) A two-dimensional proteome map of maize endosperm. *Phytochemistry* 65: 1609–1618.
- Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejarawal, A., Vandergriff, J., Rabkin, S., et al. (2005) The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* 33: D284–D288.
- Minami, A., Fujiwara, M., Furuto, A., Fukao, Y., Yamashita, T., Kamo, M., et al. (2009) Alterations in detergent-resistant plasma membrane microdomains in Arabidopsis thaliana during cold acclimation. *Plant Cell Physiol.* 50: 341–359.
- Ming, R., Hou, S., Feng, Y., Yu, Q., Dionne-Laporte, A., Saw, J.H., et al. (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452: 991–996.
- Mitsuda, N. and Ohme-Takagi, M. (2009) Functional analysis of transcription factors in Arabidopsis. *Plant Cell Physiol.* 50: 1232–1248.
- Miyao, A., Iwasaki, Y., Kitano, H., Itoh, J., Maekawa, M., Murata, K., et al. (2007) A large-scale collection of phenotypic data describing an insertional mutant population to facilitate functional analysis of rice genes. *Plant Mol. Biol.* 63: 625–635.
- Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., et al. (2003) Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell* 15: 1771–1780.
- Mochida, K., Furuta, T., Ebana, K., Shinozaki, K. and Kikuchi, J. (2009a) Correlation exploration of metabolic and genomic diversities in rice. *BMC Genomics* 10: 568.
- Mochida, K., Kawaura, K., Shimosaka, E., Kawakami, N., Shin-I, T., Kohara, Y., et al. (2006) Tissue expression map of a large number of expressed sequence tags and its application to in silico screening of stress response genes in common wheat. *Mol. Gen. Genomics* 276: 304–312.
- Mochida, K., Saisho, D., Yoshida, T., Sakurai, T. and Shinozaki, K. (2008) TriMEDB: a database to integrate transcribed markers and facilitate genetic studies of the tribe Triticeae. *BMC Plant Biol.* 8: 72.
- Mochida, K., Yamazaki, Y. and Ogihara, Y. (2003) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis

- of contigs grouped from a large number of expressed sequence tags. *Mol. Genet. Genomics* 270: 371–377.
- Mochida, K., Yoshida, T., Sakurai, T., Ogihara, Y. and Shinozaki, K. (2009b) TriFLDB: a database of clustered full-length coding sequences from Triticeae with applications to comparative grass genomics. *Plant Physiol.* 150: 1135–1146.
- Mochida, K., Yoshida, T., Sakurai, T., Yamaguchi-Shinozaki, K., Shinozaki, K. and Tran, L.S. (2009c) In silico analysis of transcription factor repertoire and prediction of stress responsive transcription factors in soybean. *DNA Res.* 16: 353–369.
- Mochida, K., Yoshida, T., Sakurai, T., Yamaguchi-Shinozaki, K., Shinozaki, K. and Tran, L.S. (2010) LegumeTFDB: an integrative database of Glycine max, Lotus japonicus and Medicago truncatula transcription factors. *Bioinformatics* 26: 290–291.
- Moco, S., Bino, R.J., Vorst, O., Verhoeven, H.A., de Groot, J., van Beek, T.A., et al. (2006) A liquid chromatography–mass spectrometry-based metabolome database for tomato. *Plant Physiol.* 141: 1205–1218.
- Moran, N.A., McLaughlin, H.J. and Sorek, R. (2009) The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science* 323: 379–382.
- Mori, M., Tomita, C., Sugimoto, K., Hasegawa, M., Hayashi, N., Dubouzet, J.G., et al. (2007) Isolation and molecular characterization of a Spotted leaf 18 mutant by modified activation-tagging in rice. *Plant Mol. Biol.* 63: 847–860.
- Morreel, K., Goeminne, G., Storme, V., Sterck, L., Ralph, J., Coppieters, W., et al. (2006) Genetical metabolomics of flavonoid biosynthesis in Populus: a case study. *Plant J.* 47: 224–237.
- Mueller, L.A., Solow, T.H., Taylor, N., Skwarecki, B., Buels, R., Binns, J., et al. (2005) The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. *Plant Physiol.* 138: 1310–1317.
- Mueller, L.A., Zhang, P. and Rhee, S.Y. (2003) AraCyc: a biochemical pathway database for Arabidopsis. *Plant Physiol.* 132: 453–460.
- Mulder, N.J. and Apweiler, R. (2008) The InterPro database and tools for protein domain analysis. *Curr. Protoc. Bioinformatics* Chapter 2: Unit 2.7.
- Nakamura, H., Hakata, M., Amano, K., Miyao, A., Toki, N., Kajikawa, M., et al. (2007) A genome-wide gain-of function analysis of rice genes using the FOX-hunting system. *Plant Mol. Biol.* 65: 357–371.
- Nakano, M., Nobuta, K., Vemmaraju, K., Tej, S.S., Skogen, J.W. and Meyers, B.C. (2006) Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. *Nucleic Acids Res.* 34: D731–D735.
- Nanjo, T., Futamura, N., Nishiguchi, M., Igasaki, T., Shinozaki, K. and Shinohara, K. (2004) Characterization of full-length enriched expressed sequence tags of stress-treated poplar leaves. *Plant Cell Physiol.* 45: 1738–1748.
- Nanjo, T., Sakurai, T., Totoki, Y., Toyoda, A., Nishiguchi, M., Kado, D., et al. (2007) Functional annotation of 19,841 Populus nigra full-length enriched cDNA clones. *BMC Genomics* 8: 448.
- Neale, D.B. and Ingvarsson, P.K. (2008) Population, quantitative and comparative genomics of adaptation in forest trees. *Curr. Opin. Plant Biol.* 11: 149–155.
- Newton, R.P., Brenton, A.G., Smith, C.J. and Dudley, E. (2004) Plant proteome analysis by mass spectrometry: principles, problems, pitfalls and recent developments. *Phytochemistry* 65: 1449–1485.
- Nishiyama, T., Fujita, T., Shin, I.T., Seki, M., Nishide, H., Uchiyama, I., et al. (2003) Comparative genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: implication for land plant evolution. *Proc. Natl Acad. Sci. USA* 100: 8007–8012.
- Nobuta, K., Lu, C., Shrivastava, R., Pillay, M., De Paoli, E., Accerbi, M., et al. (2008) Distinct size distribution of endogenous siRNAs in maize: evidence from deep sequencing in the mop1-1 mutant. *Proc. Natl Acad. Sci. USA* 105: 14958–14963.
- Nobuta, K., Venu, R.C., Lu, C., Belo, A., Vemmaraju, K., Kulkarni, K., et al. (2007) An expression atlas of rice mRNAs and small RNAs. *Nat. Biotechnol.* 25: 473–477.
- Nuhse, T.S., Bottrill, A.R., Jones, A.M. and Peck, S.C. (2007) Quantitative phosphoproteomic analysis of plasma membrane proteins reveals regulatory mechanisms of plant innate immune responses. *Plant J.* 51: 931–940.
- Nuhse, T.S., Stensballe, A., Jensen, O.N. and Peck, S.C. (2003) Large-scale analysis of in vivo phosphorylated membrane proteins by immobilized metal ion affinity chromatography and mass spectrometry. *Mol. Cell. Proteomics* 2: 1234–1243.
- Nuhse, T.S., Stensballe, A., Jensen, O.N. and Peck, S.C. (2004) Phosphoproteomics of the Arabidopsis plasma membrane and a new phosphorylation site database. *Plant Cell* 16: 2394–2405.
- Obayashi, T., Hayashi, S., Saeki, M., Ohta, H. and Kinoshita, K. (2009) ATTED-II provides coexpressed gene networks for Arabidopsis. *Nucleic Acids Res.* 37: D987–D991.
- Obayashi, T., Kinoshita, K., Nakai, K., Shibaoka, M., Hayashi, S., Saeki, M., et al. (2007) ATTED-II: a database of co-expressed genes and cis elements for identifying co-regulated gene groups in Arabidopsis. *Nucleic Acids Res.* 35: D863–D869.
- Ogasawara, O., Otsuji, M., Watanabe, K., Iizuka, T., Tamura, T., Hishiki, T., et al. (2006) BodyMap-Xs: anatomical breakdown of 17 million animal ESTs for cross-species comparison of gene expression. *Nucleic Acids Res.* 34: D628–D631.
- Ogihara, Y., Mochida, K., Kawaura, K., Murai, K., Seki, M., Kamiya, A., et al. (2004) Construction of a full-length cDNA library from young spikelets of hexaploid wheat and its characterization by large-scale sequencing of expressed sequence tags. *Genes Genet. Syst.* 79: 227–232.
- Ogihara, Y., Mochida, K., Nemoto, Y., Murai, K., Yamazaki, Y., Shin, I.T., et al. (2003) Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags. *Plant J.* 33: 1001–1011.
- Okazaki, Y., Shimojima, M., Sawada, Y., Toyooka, K., Narisawa, T., Mochida, K., et al. (2009) A chloroplastic UDP-glucose pyrophosphorylase from Arabidopsis is the committed enzyme for the first step of sulfolipid biosynthesis. *Plant Cell* 21: 892–909.
- Oksman-Caldentey, K.M. and Saito, K. (2005) Integrating genomics and metabolomics for engineering plant metabolic pathways. *Curr. Opin. Biotechnol.* 16: 174–179.
- Ossowski, S., Schneeberger, K., Clark, R.M., Lanz, C., Warthmann, N. and Weigel, D. (2008) Sequencing of natural strains of Arabidopsis thaliana with short reads. *Genome Res.* 18: 2024–2033.
- Palaniswamy, S.K., James, S., Sun, H., Lamb, R.S., Davuluri, R.V. and Grotewold, E. (2006) AGRIS and AtRegNet: a platform to link cis-regulatory elements and transcription factors into regulatory networks. *Plant Physiol.* 140: 818–829.
- Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., et al. (2000) Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* 289: 739–745.
- Park, P.J. (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* 10: 669–680.
- Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, N., Coulson, R., Farne, A., et al. (2007) ArrayExpress—a public database

- of microarray experiments and gene expression profiles. *Nucleic Acids Res.* 35: D747–D750.
- Paterson, A.H. (2008) Genomics of sorghum. *Int. J. Plant Genomics* 2008: 362451.
- Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., et al. (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* 457: 551–556.
- Peleg, Z., Saranga, Y., Suprunova, T., Ronin, Y., Roder, M.S., Kilian, A., et al. (2008) High-density genetic map of durum wheat × wild emmer wheat based on SSR and DaRT markers. *Theor. Appl. Genet.* 117: 103–115.
- Qu, S., Desai, A., Wing, R. and Sundaresan, V. (2008) A versatile transposon-based activation tag vector system for functional genomics in cereals and other monocot plants. *Plant Physiol.* 146: 189–199.
- Ralph, S.G., Chun, H.J., Cooper, D., Kirkpatrick, R., Kolosova, N., Gunter, L., et al. (2008a) Analysis of 4,664 high-quality sequence-finished poplar full-length cDNA clones and their utility for the discovery of genes responding to insect feeding. *BMC Genomics* 9: 57.
- Ralph, S.G., Chun, H.J., Kolosova, N., Cooper, D., Oddy, C., Ritland, C.E., et al. (2008b) A conifer genomics resource of 200,000 spruce (*Picea* spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (*Picea sitchensis*). *BMC Genomics* 9: 484.
- Ramautar, R., Somsen, G.W. and de Jong, G.J. (2009) CE-MS in metabolomics. *Electrophoresis* 30: 276–291.
- Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., et al. (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* 319: 64–69.
- Riano-Pachon, D.M., Ruzicic, S., Dreyer, I. and Mueller-Roeber, B. (2007) PlnTFDB: an integrative plant transcription factor database. *BMC Bioinformatics* 8: 42.
- Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., et al. (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290: 2105–2110.
- Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., et al. (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* 13: 11–29.
- Rosignol, M., Peltier, J.B., Mock, H.P., Matros, A., Maldonado, A.M. and Jorin, J.V. (2006) Plant proteome analysis: a 2004–2006 update. *Proteomics* 6: 5529–5548.
- Rostoks, N., Borevitz, J.O., Hedley, P.E., Russell, J., Mudie, S., Morris, J., et al. (2005) Single-feature polymorphism discovery in the barley transcriptome. *Genome Biol.* 6: R54.
- Rowe, H.C., Hansen, B.G., Halkier, B.A. and Kliebenstein, D.J. (2008) Biochemical networks and epistasis shape the Arabidopsis thaliana metabolome. *Plant Cell* 20: 1199–1216.
- Ruiz-Ferrer, V. and Voinnet, O. (2009) Roles of plant small RNAs in biotic stress responses. *Annu. Rev. Plant Biol.* 60: 485–510.
- Rushton, P.J., Bokowiec, M.T., Laudeman, T.W., Brannock, J.F., Chen, X. and Timko, M.P. (2008) TOBFAC: the database of tobacco transcription factors. *BMC Bioinformatics* 9: 53.
- Saito, K., Hirai, M.Y. and Yonekura-Sakakibara, K. (2008) Decoding genes with coexpression networks and metabolomics—‘majority report by precogs’. *Trends Plant Sci.* 13: 36–43.
- Sakata, K., Ohyanagi, H., Nobori, H., Nakamura, T., Hashiguchi, A., Nanjo, Y., et al. (2009) Soybean proteome database: a data resource for plant differential omics. *J. Proteome Res.* 8: 3539–3548.
- Sakurai, T., Plata, G., Rodriguez-Zapata, F., Seki, M., Salcedo, A., Toyoda, A., et al. (2007) Sequencing analysis of 20,000 full-length cDNA clones from cassava reveals lineage specific expansions in gene families related to stress response. *BMC Plant Biol.* 7: 66.
- Sakurai, T., Satou, M., Akiyama, K., Iida, K., Seki, M., Kuromori, T., et al. (2005) RARGE: a large-scale database of RIKEN Arabidopsis resources ranging from transcriptome to phenome. *Nucleic Acids Res.* 33: D647–D650.
- Samatey, F.A., Imada, K., Nagashima, S., Vonderviszt, F., Kumasaka, T., Yamamoto, M., et al. (2001) Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling. *Nature* 410: 331–337.
- Sato, K., Shin, I.T., Seki, M., Shinozaki, K., Yoshida, H., Takeda, K., et al. (2009) Development of 5006 full-length cDNAs in barley: a tool for accessing cereal genomics resources. *DNA Res.* 16: 81–89.
- Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., et al. (2008) Genome structure of the legume, Lotus japonicus. *DNA Res.* 15: 227–239.
- Sato, S. and Tabata, S. (2006) Lotus japonicus as a platform for legume research. *Curr. Opin. Plant Biol.* 9: 128–132.
- Schauer, N., Semel, Y., Balbo, I., Steinfath, M., Reipsilber, D., Selbig, J., et al. (2008) Mode of inheritance of primary metabolic traits in tomato. *Plant Cell* 20: 509–523.
- Schauer, N., Semel, Y., Roessner, U., Gur, A., Balbo, I., Carrari, F., et al. (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol.* 24: 447–454.
- Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467–470.
- Schmelzle, K. and White, F.M. (2006) Phosphoproteomic approaches to elucidate cellular signaling networks. *Curr. Opin. Biotechnol.* 17: 406–414.
- Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178–183.
- Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326: 1112–1115.
- Schneider, A., Kirch, T., Gigolashvili, T., Mock, H.P., Sonnewald, U., Simon, R., et al. (2005) A transposon-based activation-tagging population in Arabidopsis thaliana (TAMARA) and its application in the identification of dominant developmental and metabolic mutations. *FEBS Lett.* 579: 4622–4628.
- Schripsema, J. (2009) Application of NMR in plant metabolomics: techniques, problems and prospects. *Phytochem. Anal.* 21: 14–21.
- Schwede, T., Kopp, J., Guex, N. and Peitsch, M.C. (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* 31: 3381–3385.
- Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., et al. (2002a) Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J.* 31: 279–292.
- Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T., et al. (2002b) Functional annotation of a full-length Arabidopsis cDNA collection. *Science* 296: 141–145.
- Seki, M. and Shinozaki, K. (2009) Functional genomics using RIKEN Arabidopsis thaliana full-length cDNAs. *J. Plant Res.* 122: 355–366.
- Sekiyama, Y. and Kikuchi, J. (2007) Towards dynamic metabolic network measurements by multi-dimensional NMR-based fluxomics. *Phytochemistry* 68: 2320–2329.
- Shen, Y.J., Jiang, H., Jin, J.P., Zhang, Z.B., Xi, B., He, Y.Y., et al. (2004) Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant Physiol.* 135: 1198–1205.

- Soderlund, C., Descour, A., Kudrna, D., Bomhoff, M., Boyd, L., Currie, J., et al. (2009) Sequencing, mapping, and analysis of 27,455 maize full-length cDNAs. *PLoS Genet.* 5: e1000740.
- Somerville, C. and Dangl, J. (2000) Genomics. *Plant biology in 2010. Science* 290: 2077–2078.
- Springer, P.S. (2000) Gene traps: tools for plant development and genomics. *Plant Cell* 12: 1007–1020.
- Stanford, W.L., Cohn, J.B. and Cordes, S.P. (2001) Gene-trap mutagenesis: past, present and beyond. *Nat. Rev. Genet.* 2: 756–768.
- Sterck, L., Rombauts, S., Vandepoele, K., Rouze, P. and Van de Peer, Y. (2007) How many genes are there in plants (... and why are they there)? *Curr. Opin. Plant Biol.* 10: 199–203.
- Sterky, F., Bhalerao, R.R., Unneberg, P., Segerman, B., Nilsson, P., Brunner, A.M., et al. (2004) A *Populus* EST resource for plant functional genomics. *Proc. Natl Acad. Sci. USA* 101: 13951–13956.
- Stevens, R.C., Yokoyama, S. and Wilson, I.A. (2001) Global efforts in structural genomics. *Science* 294: 89–92.
- Sugiyama, N., Nakagami, H., Mochida, K., Daudi, A., Tomita, M., Shirasu, K., et al. (2008) Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in *Arabidopsis*. *Mol. Syst. Biol.* 4: 193.
- Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J.D., Dean, C., et al. (1995) Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. *Genes Dev.* 9: 1797–1810.
- Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T.Z., Garcia-Hernandez, M., Foerster, H., et al. (2008) The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.* 36: D1009–D1014.
- Taji, T., Sakurai, T., Mochida, K., Ishiwata, A., Kurotani, A., Totoki, Y., et al. (2008) Large-scale collection and annotation of full-length enriched cDNAs from a model halophyte, *Thellungiella halophila*. *BMC Plant Biol.* 8: 115.
- Takeda, S. and Matsuoka, M. (2008) Genetic approaches to crop improvement: responding to environmental and population changes. *Nat. Rev. Genet.* 9: 444–457.
- Tanaka, T., Antonio, B.A., Kikuchi, S., Matsumoto, T., Nagamura, Y., Numa, Y., et al. (2008) The Rice Annotation Project Database (RAP-DB): 2008 update. *Nucleic Acids Res.* 36: D1028–D1033.
- Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M. and Paterson, A.H. (2008a) Synteny and collinearity in plant genomes. *Science* 320: 486–488.
- Tang, H., Wang, X., Bowers, J.E., Ming, R., Alam, M. and Paterson, A.H. (2008b) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* 18: 1944–1954.
- The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
- The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463: 763–768.
- Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., et al. (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* 37: 914–939.
- Tikunov, Y., Lommen, A., de Vos, C.H., Verhoeven, H.A., Bino, R.J., Hall, R.D., et al. (2005) A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. *Plant Physiol.* 139: 1125–1137.
- Till, B.J., Colbert, T., Codomo, C., Enns, L., Johnson, J., Reynolds, S.H., et al. (2006) High-throughput TILLING for *Arabidopsis*. *Methods Mol. Biol.* 323: 127–135.
- Till, B.J., Reynolds, S.H., Weil, C., Springer, N., Burtner, C., Young, K., et al. (2004) Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biol.* 4: 12.
- Tohge, T., Nishiyama, Y., Hirai, M.Y., Yano, M., Nakajima, J., Awazuhara, M., et al. (2005) Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. *Plant J.* 42: 218–235.
- Tokimatsu, T., Sakurai, N., Suzuki, H., Ohta, H., Nishitani, K., Koyama, T., et al. (2005) KaPPA-view: a web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. *Plant Physiol.* 138: 1289–1300.
- Torada, A., Koike, M., Mochida, K. and Ogihara, Y. (2006) SSR-based linkage map with new markers using an intraspecific population of common wheat. *Theor. Appl. Genet.* 112: 1042–1051.
- Toyoda, T. and Wada, A. (2004) Omic space: coordinate-based integration and analysis of genomic phenomic interactions. *Bioinformatics* 20: 1759–1765.
- Toyoshima, C., Nomura, H. and Sugita, Y. (2003) Crystal structures of Ca²⁺-ATPase in various physiological states. *Ann. NY Acad. Sci.* 986: 1–8.
- Trethewey, R.N. (2004) Metabolite profiling as an aid to metabolic engineering in plants. *Curr. Opin. Plant Biol.* 7: 196–201.
- Turner, W.R., Oppenheimer, M. and Wilcove, D.S. (2009) A force to fight global warming. *Nature* 462: 278–279.
- Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Helsten, U., et al. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596–1604.
- Tyler, R.C., Aceti, D.J., Bingman, C.A., Cornilescu, C.C., Fox, B.G., Frederick, R.O., et al. (2005) Comparison of cell-based and cell-free protocols for producing target proteins from the *Arabidopsis thaliana* genome for structural studies. *Proteins* 59: 633–643.
- Umezawa, T., Sakurai, T., Totoki, Y., Toyoda, A., Seki, M., Ishiwata, A., et al. (2008) Sequencing and analysis of approximately 40 000 soybean cDNA clones from a full-length-enriched cDNA library. *DNA Res.* 15: 333–346.
- Urano, K., Maruyama, K., Ogata, Y., Morishita, Y., Takeda, M., Sakurai, N., et al. (2009) Characterization of the ABA-regulated global responses to dehydration in *Arabidopsis* by metabolomics. *Plant J.* 57: 1065–1078.
- Varshney, R.K., Graner, A. and Sorrells, M.E. (2005) Genomics-assisted breeding for crop improvement. *Trends Plant Sci.* 10: 621–630.
- Varshney, R.K., Nayak, S.N., May, G.D. and Jackson, S.A. (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* 27: 522–530.
- Velculescu, V.E., Zhang, L., Vogelstein, B. and Kinzler, K.W. (1995) Serial analysis of gene expression. *Science* 270: 484–487.
- von Roepenack-Lahaye, E., Degenkolb, T., Zerjeski, M., Franz, M., Roth, U., Wessjohann, L., et al. (2004) Profiling of *Arabidopsis* secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry. *Plant Physiol.* 134: 548–559.
- Wall, P.K., Leebens-Mack, J., Muller, K.F., Field, D., Altman, N.S. and dePamphilis, C.W. (2008) PlantTribes: a gene and gene family resource for comparative genomics in plants. *Nucleic Acids Res.* 36: D970–D976.
- Wang, D.K., Sun, Z.X. and Tao, Y.Z. (2006) Application of TILLING in plant improvement. *Yi Chuan Xue Bao* 33: 957–964.

- Ware, D. (2007) Gramene: a resource for comparative grass genomics. *Methods Mol. Biol.* 406: 315–330.
- Waterhouse, P.M., Graham, M.W. and Wang, M.B. (1998) Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc. Natl Acad. Sci. USA* 95: 13959–13964.
- Weigel, D., Ahn, J.H., Blazquez, M.A., Borevitz, J.O., Christensen, S.K., Fankhauser, C., et al. (2000) Activation tagging in Arabidopsis. *Plant Physiol.* 122: 1003–1013.
- Weigel, D. and Mott, R. (2009) The 1001 genomes project for Arabidopsis thaliana. *Genome Biol.* 10: 107.
- Wenzl, P., Raman, H., Wang, J., Zhou, M., Huttner, E. and Kilian, A. (2007) A DArT platform for quantitative bulked segregant analysis. *BMC Genomics* 8: 196.
- Werner, E., Heilier, J.F., Ducruix, C., Ezan, E., Junot, C. and Tabet, J.C. (2008) Mass spectrometry for the identification of the discriminating signals from metabolomics: current status and future trends. *J. Chromatogr. B* 871: 143–163.
- Wilson, D., Charoensawan, V., Kummerfeld, S.K. and Teichmann, S.A. (2008) DBD—taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res.* 36: D88–92.
- Wilson, D., Madera, M., Vogel, C., Chothia, C. and Gough, J. (2007) The SUPERFAMILY database in 2007: families and functions. *Nucleic Acids Res.* 35: D308–D313.
- Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V. and Provart, N.J. (2007) An 'electronic fluorescent pictograph' browser for exploring and analyzing large-scale biological data sets. *PLoS One* 2: e718.
- Yabuki, T., Kigawa, T., Dohmae, N., Takio, K., Terada, T., Ito, Y., et al. (1998) Dual amino acid-selective and site-directed stable-isotope labeling of the human c-Ha-Ras protein by cell-free synthesis. *J. Biomol. NMR* 11: 295–306.
- Yamamoto, Y.Y. and Obokata, J. (2008) ppdb: a plant promoter database. *Nucleic Acids Res.* 36: D977–D981.
- Yamamoto, Y.Y., Yoshitsugu, T., Sakurai, T., Seki, M., Shinozaki, K. and Obokata, J. (2009) Heterogeneity of Arabidopsis core promoters revealed by high-density TSS analysis. *Plant J.* 60: 350–362.
- Yamasaki, C., Murakami, K., Fujii, Y., Sato, Y., Harada, E., Takeda, J., et al. (2008a) The H-Invitational Database (H-InvDB), a comprehensive annotation resource for human genes and transcripts. *Nucleic Acids Res.* 36: D793–D799.
- Yamasaki, K., Kigawa, T., Inoue, M., Tateno, M., Yamasaki, T., Yabuki, T., et al. (2004) Solution structure of the B3 DNA binding domain of the Arabidopsis cold-responsive transcription factor RAV1. *Plant Cell* 16: 3448–3459.
- Yamasaki, K., Kigawa, T., Inoue, M., Watanabe, S., Tateno, M., Seki, M., et al. (2008b) Structures and evolutionary origins of plant-specific transcription factor DNA-binding domains. *Plant Physiol. Biochem.* 46: 394–401.
- Yates, J.R., Ruse, C.I. and Nakorchevsky, A. (2009) Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* 11: 49–79.
- Yilmaz, A., Nishiyama, M.Y., Jr., Fuentes, B.G., Souza, G.M., Janies, D., Gray, J., et al. (2009) GRASSIUS: a platform for comparative regulatory genomics across the grasses. *Plant Physiol.* 149: 171–180.
- Yokoyama, S. (2003) Protein expression systems for structural genomics and proteomics. *Curr. Opin. Chem. Biol.* 7: 39–43.
- Yokoyama, S., Hirota, H., Kigawa, T., Yabuki, T., Shirouzu, M., Terada, T., et al. (2000) Structural genomics projects in Japan. *Nat. Struct. Biol.* 7 Suppl: 943–945.
- Yonekura-Sakakibara, K., Tohge, T., Matsuda, F., Nakabayashi, R., Takayama, H., Niida, R., et al. (2008) Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene–metabolite correlations in Arabidopsis. *Plant Cell* 20: 2160–2176.
- Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 296: 79–92.
- Zeller, G., Henz, S.R., Widmer, C.K., Sachsenberg, T., Ratsch, G., Weigel, D., et al. (2009) Stress-induced changes in the Arabidopsis thaliana transcriptome analyzed using whole-genome tiling arrays. *Plant J.* 58: 1068–1082.
- Zhang, H., Sreenivasulu, N., Weschke, W., Stein, N., Rudd, S., Radchuk, V., et al. (2004) Large-scale analysis of the barley transcriptome based on expressed sequence tags. *Plant J.* 40: 276–290.
- Zhang, J., Xu, Y., Huan, Q. and Chong, K. (2009a) Deep sequencing of Brachypodium small RNAs at the global genome level identifies microRNAs involved in cold stress response. *BMC Genomics* 10: 449.
- Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., et al. (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. *Cell* 126: 1189–1201.
- Zhang, Y. (2008) Progress and challenges in protein structure prediction. *Curr. Opin. Struct. Biol.* 18: 342–348.
- Zhang, Y. (2009a) I-TASSER: fully automated protein structure prediction in CASP8. *Proteins* 77 Suppl 9: 100–113.
- Zhang, Y. (2009b) Protein structure prediction: when is it useful? *Curr. Opin. Struct. Biol.* 19: 145–155.
- Zhang, Z., Yu, J., Li, D., Liu, F., Zhou, X., Wang, T., Ling, Y. and Su, Z. (2009b) PMRD: plant microRNA database. *Nucleic Acids Res.*
- Zheng, Y., Ren, N., Wang, H., Stromberg, A.J. and Perry, S.E. (2009) Global identification of targets of the Arabidopsis MADS domain protein AGAMOUS-Like15. *Plant Cell* 21: 2563–2577.
- Zhu, Q.H., Guo, A.Y., Gao, G., Zhong, Y.F., Xu, M., Huang, M., et al. (2007) DPTF: a database of poplar transcription factors. *Bioinformatics* 23: 1307–1308.
- Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L. and Gruissem, W. (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. *Plant Physiol.* 136: 2621–2632.
- Zubko, E., Adams, C.J., Machaekova, I., Malbeck, J., Scollan, C. and Meyer, P. (2002) Activation tagging identifies a gene from *Petunia hybrida* responsible for the production of active cytokinins in plants. *Plant J.* 29: 797–808.