ISSN 1022-7954, Russian Journal of Genetics, 2009, Vol. 45, No. 11, pp. 1302–1316. © Pleiades Publishing, Inc., 2009.
Original Russian Text © N.A. Omelyanchuk, V.V. Mironova, N.A. Kolchanov, 2009, published in Genetika, 2009, Vol. 45, No. 11, pp. 1476–1492.

REVIEWS AND THEORETICAL ARTICLES

Plant Developmental Genetics: Integrating Data from Different Experiments in Databases

N. A. Omelyanchuk[a], **V. V. Mironova**[a], **and N. A. Kolchanov**[a,b]
[a] *Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630090 Russia;*
e-mail: nadya@bionet.nsc.ru
[b] *Department of Information Biology, Novosibirsk State University, Novosibirsk, 630090 Russia*

Received January 12, 2009

Abstract—Plant developmental genetics as a scientific discipline integrates data from such different fields of biology as embryology, plant anatomy, molecular biology and genetics, and studies their interactions in the course of plant development. To date, traditional publication of scientific studies in articles is supplemented by presenting in databases the data generated by high-throughoutput methods in genomics, transcriptomics, proteomics, and phenomics. The information burst, caused both by genome-scale research projects and growth in the number of articles, requires the development of general standards of annotating data from different sources for their integration and comparison. In this review, we present classification and analysis of existing databases, in which the user can find various data on plant developmental genetics, and discuss problems of these data integration both within informational resources and among them..

DOI: 10.1134/S1022795409110052

INTRODUCTION

From the time of its appearance as a branch of developmental biology, developmental genetics encompassed evidence on genetic aspects of development gathered in different biological disciplines [1]. Developmental processes are based on differential gene activity and on the regulation of this activity at the level of gene networks [2]. During development, cells exchange information via signal transmission, resulting in coordinated gene expression. To understand causal relationships among molecular events, growth, differentiation, cell division, and later formation of tissues and organs, all aspects of development as an intricate complex system should be considered. To this end, one should integrate information on the participants involved in development (external and internal signals, genes, mRNAs, proteins, and protein complexes) and on their interactions in various cell types, tissues, and organs.

The results of experiments on diverse developmental processes are published as scientific articles in various journals, forming a multitude of scattered and heterogeneous data, the volume of which increases each year. Of course, each experimental paper, and more so a review, deals with a definite area of knowledge, so that information from different publications is compared, yet the size of this area and the specification of the data are always limited. Moreover, the rapid advance and wide employment of new methods for investigating gene activity resulted in the accumulation of huge arrays of experimental data from genome-scale experiments, which require understanding, interpretation, and comparison both with one another and with other observations [3, 4]. In modern bioinformatics, databases are developed to integrate information from different observations and experiments. Related to this, a wide range of methods have been developed, aimed at formation of databases, maintaining their functioning, and effective use [2–4], including:

(i) ontologies and knowledge bases on developmental stages and anatomical elements;

(ii) annotation technologies and standard formats for presentation of data on gene expression and phenotypes;

(iii) methods of meta-analysis of the data, based on their comparison, systematization, and classification and aimed at deriving from them non-contradictory and mutually supplementing facts that can be used for generating hypotheses on mechanisms of development (for their simulation and further experimental verification).

The aim of this review is discussing methods and approaches used in various bioinformatics resources for classification, systematization, and integration of the data. We describe here different types of databases, in which the user can find information on (1) genomics, transcriptomics, and proteomics of various plant species; (2) anatomy, morphology, and developmental stages of wild type plants; (3) phenotypic abnormalities caused by perturbations in gene expression; (4) changes in the gene expression in different genotypes and under the action of various factors.

The review was not aimed at including all databases on the above issues, because structure and data types in many databases are similar. Moreover, the number and

Table 1. Databases on genomics, transcriptomics, and proteomics, accumulating information on various plant species along with other organisms

Name	Description	www.site
INSD	International Nucleotide Sequence Database. International consortium on genome sequence data. Contains three major repositories on genomics (DDBJ, EMBL, and GenBank)	[10]
GEO	Gene Expression Omnibus. Database on Gene expression. Database repository of high throughput gene expression data and hybridization arrays, chips, microarrays	[11]
UniProt	The Universal Protein Resource. Major repository on proteomics	[13]
NCBI	National Center for Biotechnology Information. In addition to maintaining archive of nucleotide sequences GenBank, NCBI provides programs and informational resources for integration of data with other biological data	[17]
InterPro	InterPro. Database of protein families, domains, repeats, and protein sequence sites	[18]
wwPDB	The Worldwide Protein Data Bank	[20]

composition of information resources on the Internet are constantly changing—new databases appear, while others are terminated or stop updating.

DATABASES ON GENOMICS, TRANSCRIPTOMICS, AND PROTEOMICS OF VARIOUS SPECIES

First and foremost, we shall briefly describe several well-known and widely used informational portals that accumulate information on genes and proteins of living organisms, including plants (Table 1).

Databases are constructed for providing accumulation, storage, systematization, and analysis of large information flows [4, 5]. A main source of new information for databases is annotation, i.e., formal description and characterization of biological objects, structures, processes, etc. Annotation can be automated (employing special computer software, e.g., procedures for gene search in genomic nucleotide sequences), semi-automated (the results of automated annotation are verified by experts), and manual (the information is put in the database by an expert). Expert annotation typically implies that the experimental information is pulled from published research papers and conformed to a standard format to be placed into a database.

In most current databases, all three annotation types are combined. Improvement of the data search system usually starts from a simple interface providing separate access to each stored data types, and then develops to an interface for systemic search queries, allowing integration of different types of data both within and among the information resources [4–6].

Most current biological databases are genomic ones. They can be classified into archives, or data banks, where the results of experiments, annotated according to a general standard, are submitted by various research teams (Table 1). Such information resources accumulating experimental data on all living organisms primarily include (i) international data banks of nucleotide sequences DDBJ [7], EMBL [8], and GenBank [9], which are united by the INSD consortium (International Sequence Databases, DDBJ/EMBL/GenBank [10]; (ii) GEO database on transcriptomics (Gene Expression Omnibus), containing the results of gene expression experiments using microarrays and other high-throughput genome-scale technologies [11, 12]; (iii) UniProt databank on proteomics, annotating data on protein sequences and on the structural and functional organization of proteins [13, 14].

The next step in studying the information stored in centralized archives is re-annotation [6, 15, 16], which for genome sequences means eliminating primary annotation errors and supplementing gene descriptions with homologous sequences and with inferred or experimentally confirmed gene functions. Investigation of sequence homology in the genomes and detection of gene functions are the fields of two “post-genomic” scientific disciplines, comparative and functional genomics, respectively. Comparative genomics examines homology of and evolutionary relationships among sequences of macromolecules. The objective of functional genomics is deduction and description of gene function based on integrating data from genomic projects (including data of comparative genomics) and the data on gene activity from transcriptomics, proteomics, and metabolomics [6, 15].

The data obtained at this stage may be submitted to an archive [16, 17] or accumulated at particular information resources (Table 1). For instance, the UniProt database, containing information on sequences of proteins and their functions [13, 14], includes expert-annotated Swiss-Prot, automatically annotated TrEMBL, base of clusters of homologous proteins UniRef, and the original archive of protein sequences UniPark. InterPro [18, 19] accumulated information on protein families, their functional domains, and sites from numerous databases, including Gene3D, Pfam, PROSITE, SMART, and others. International Protein Data Bank

Table 2. Databases integrating genomic date in various plant species

Name	Description	www.site
TIGR Plant TA	TIGR Plant Transcription Assemblies	[22]
PlantGDB	Plant Genome Database. A resource on comparative plant genomics. Annotates genome sequences and transcript assemblies	[24]
Gramene	Resource for Comparative Grass Genomics. In addition to genome annotation, contains information on genes, proteins, chromosome maps, molecular markers, ontology, etc.	[26]
PlantTribes	The Plant Tribe Database. A classification system for plant proteins based on cluster analysis of the inferred proteomes of the sequenced genomes *Arabidopsis thaliana*, *Carica papaya*, *Medicago truncatula*, *Oryza sativa*, and *Populus trichocarpa.* Contains also information on gene expression and protein domains	[28]
GrainGenes	Database for Triticeae and Avena. Database on genes of cultivated and wild species of the tribe Triticeae and the genus *Avena.* Contains information on loci, genes, alleles, chromosome maps, and molecular markers	[31]
RAPESEED	Shanghai RAPESEED Database. Information resource on functional genomics of seed development and fatty acid metabolism in Brassica plants	[33]
LIS	LIS. The Legume Information System. Information resource for integration of data on comparative biology (information on genomes, transcripts, genetics, etc.) in legumes	[35]

wwPDB reannotates the Protein Data Bank (PDB) archive on the structure of proteins, nucleic acids, and complexes produced by these macromolecules [20, 21].

Databases on Plant Genomics

In the postgenomic era, special databases have been constructed on individual species or groups of plant species (Table 2). Depending on the degree of knowledge of the species, such databases can contain, along with nucleotide (genomic, cDNA, EST, etc.) and protein sequences, genetic maps saturated with various markers (restriction fragment length polymorphism RFLP, single nucleotide polymorphism SNP, etc.), data on protein families, regulatory motifs, binding sites, the results of transcriptomic and proteomic experiments, and other information. The database content differs among the databases depending on currently available information and on the aims of the database (Table 2). A significant number of databases contain EST sequences of various plant species. An example is TIGR plant Transcript Assemblies [22, 23], which is a database of transcript assemblies (consensus transcript sequences constructed via clustering EST and cDNA sequences).

Many informational resources present data on comparative genomics (Table 2). Databases Plant GDB [24, 25], Gramene [26, 27], PlantTribes [28, 29] are major resources accumulating such information. Although plants species significantly differ in genome size and in chromosome number and morphology, long chromosome region with conserved gene order (macro-collinearity) were detected [30]. Sequences of homologous genes exhibit different extent of conservatism (micro-collinearity). To date, these characters are widely used for comparing the genomes of various plant species, their structures, function and evolution. There are databases on selective plant families, which can be used in comparative analysis of the genomes of closely related species. These databases include, e.g., GrainGenes [31, 32], RAPESEED [33, 34], and LIS [35, 36] (Table 2).

Model experimental objects of plant genomics, *Arabidopsis* and rice, are fundamental for comparative genomics studies. The specific feature of such models is a completely sequenced genome with high annotation and integration of significant information in many databases. The key informational resource on *Arabidopsis* is the TAIR database [37, 38] (Table 3). In addition, two databases on functional genomics of *Arabidopsis*—MatDB [39, 40] and TIGR *Arabidopsis thaliana* Database—-are widely known [41, 42]. These three informational resources contain cross references based on identifiers of *Arabidopsis* loci (AGI codes). Each locus in the *Arabidopsis* genome has its AGI code, which indicates the number of the chromosome carrying this locus and the ordinal number of the locus on the chromosome. For instance, the AGI code of gene *WUS* is At2g17950. It means that this gene is located in chromosome 2, and its ordinal number on this chromosome is 17950. The *WUS* locus is located in the reverse strand. Upstream of it, but on the forward strand, is At2g17940, the locus of a gene with unknown function. Downstream of the *WUS* locus is the locus of the At2g17960 gene, whose function is also unknown. If a new gene is detected between the At2g17940 and At2g17950 loci, it will have the At2g17945 AGI code. The AGI codes are currently used as identifiers of *Arabidopsis* genes in all databases. Along with the TAIR–MatDB–TIGR triad, reannotation of the *Arabidopsis* genes is presented in some other databases; for

Table 3. Databases integrating information on model plants

Name	Description	www.site
TAIR	The Arabidopsis Information Resource. The basic resource on genomics, transcriptomics, and proteomics of *Arabidopsis*. Integrated information on organization of the genome, genes, proteins, gene expression, molecular markers, chromosome maps, biochemical pathways, mutant and transgenic lines, phenotypes, etc.	[37]
MATDB	MIPS *Arabidopsis thaliana* database. Database of Munich information center of gene sequences. Integrates information on functional and comparative genomics of *Arabidopsis*.	[39]
TIGR *A. thaliana* DB	The TIGR *Arabidopsis thaliana* Database. Database of the Institute for Genomic Research. Integrates information on functional and comparative genomics of *Arabidopsis*.	[41]
GeneFarm	Genoplante project. Structural and functional expert reannotation of genes and gene families of *Arabidopsis*. The database contains data on genes, transcripts, gene expression, and phenotypes of mutants of various plant species	[43]
RAP-DB	The Rice Annotation project Database. Integrates information on structural and functional rice genomics. Contains data on MRSS analysis of gene expression, flanking sequences of insertion lines, etc.	[45]
MOsDB	The MIPS *Oryza sativa* Database. Database on *Oryza sativa* of Munich information center of gene sequences. Integrates information on structural and functional rice genomics	[47]
Rice Genome Annotation	Rice Genome Annotation Project. Integrates information on structural and functional rice genomics	[49]
MaizeGDB	Maize Genetics and Genomics Database. The basic informational portal on maize genetics and genomics. Along with data on functional genomics, contains chromosome maps, information on loci and their alleles, molecular markers, description of phenotypes	[51]
Medicago.org	www site integrating data on structural, functional, and comparative genomics and transcriptomics of alfalfa *Medicago truncatula*, a model for studying legumes	[53]

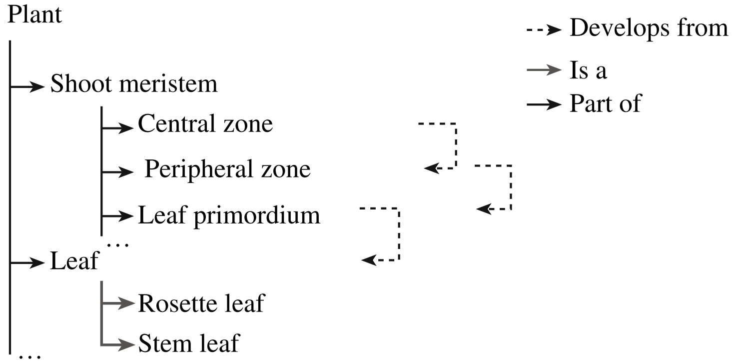
instance, GeneFarm accumulates information on automated and expert reannotation of this genome [43, 44].

Among a relatively high number of databases on structural and functional rice genomics (Table 3), we can cite as an example RAP-DB [45, 46], MOsDB [47, 48], and Rice Genome Annotation [49, 50]. The maize genome is annotated in MaizeGDB [51, 52] and is also represented in databases TIGR [42] and MIPSPlantsDB [48]. The data on various fields of genomics and transcriptomics of the model legume species *Medicago truncatula* are integrated in the medicago.org resource [53, 54].

As a rule, databases on functional genomics of plant species that are not model objects contain cDNA and EST sequences, their assemblies, and chromosome maps. Chromosome maps can be genetic, physical, and integrated [55, 56]. Genetic maps or linkage maps are based on frequencies of recombination between different chromosome regions [57]. Chromosome regions could be marked with mutant gene alleles, chromosome structure (e.g., centromere), and molecular markers (nucleotide sequences of different length). Molecular markers could be restriction fragments (RFLP), microsatellite sequences (SSR, simple sequence repeats), EST sequences, etc. Mapping involves establishment not only of the relative marker positions, but also their association with loci for quantitative or qualitative characters.

The physical maps are constructed on the basis of sequencing of chromosome regions and the distances between individual genes or markers on this map are measured by the number of base pairs [57]. Juxtaposition of genetic and physical maps is an important stage of genomic projects [55, 56]. Users can find integrative maps, combining genetic and physical maps of one or several species, in many databases (Table 2). EST and cDNA sequences are used for constructing functional molecular markers, transcriptional maps, and microarrays. Saturating genetic and physical maps by molecular markers and increasing the EST sequence sets permits detecting genome regions that are rich or poor in genes, finding molecular markers linked with important agronomic traits, and accelerating gene cloning [58, 59].

The development of projects on functional genomics of selective plant species is accompanied by the appearance of gene expression detected by microarrays, MPSS, and other genome-scale technologies, in the corresponding databases. Many information resources (Table 3) present description of collections of introgressive, mutant, and transgenic lines. Thus, information resources on different plant genomes are transformed from collections of nucleotide and protein sequences into tools of functional and comparative genomics for determination of gene functions [60].



Types of links between terms in *Plant Ontology.*

ONTOLOGIES AS TOOLS FOR SYSTEMATIZING KNOWLEDGE ON A SUBJECT AREA

Annotation of Gene Function. GO Ontologies

Integration of information on gene from different databases requires standardization of terms describing these data, i.e., development of controlled vocabularies and ontologies. Controlled vocabularies determine terms explicitly and consistently, giving an exact definition of the term and its synonyms [61]. By adding links between terms in controlled vocabularies, ontology is constructed on their basis. Ontology is a domain of knowledge represented by facts and associations among them, based on formal logical rules [62]. Ontology is a generally accepted means for classification, systematization, integration, and formalization of data. The subject area formalizing using ontologies becomes accessible for computer analysis. Using the same ontologies by different databases enables cross references and data integration among them. The International Gene Ontology (GO) Consortium develops terms required for annotating the information on genes and their products [63, 64]. In GO, terms are organized in three knowledge domains, describing the molecular function of the gene, the biological process in which this gene is involved, and the cell components with which it is associated. The GO information resource [65] also presents descriptions of genes and their products using GO vocabulary terms for various plant and animal species [64].

Description of Plant Structure and Developmental Stages in Databases

Standardization of terms describing developmental stages and structural elements in databases is required for description of gene expression dynamics in space and time and phenotype changes in mutant and transgenic plants. In view of this, the *Plant Ontology* Consortium of international experts has constructed *Plant Ontology* [66–68]. *Plant Ontology* consists of two sections: ontology of plant structure and ontology of growth and development stages. *Plant Ontology* was developed on the basis of ontologies of structure and developmental stages of *Arabidopsis*, cereals (maize, rice, and others), as well as plants from the families Leguminaceae and Solanaceae. The structure of *Plant Ontology* can be presented as a graph (in mathematics, a set of objects with connections among them), whose nodes are vocabulary terms, and connections are associations of the following three types (figure):

(1) Object denoted by term B, is a specific case of the object denoted by term A {Is_a (or Instance _of)}. This connection indicates that the specific term is related to the more general one, e.g., rosette leaf is a specific term of the more general term leaf.

(2) Object denoted by term B, is a part the object denoted by term A {Part_of}, e.g., the central zone is a part of the shoot meristem.

(3) Object denoted by term B, develops from its precursor denoted by term A {Develops_from}, e.g., leaf develops from leaf primordium.

Terms of the ontology and connections between them are recognized by computer programs, which provide automated processing of and search for information [67, 68]. At present, *Plant Ontology*

enables cross references and general queries for various plant information resources;

provides a general basis for annotating data of various experiments and promotes integration of information on plant development from classical plant developmental genetics, genomics, transcriptomics, proteomics, and phenomics;

solves previous contradictions such as use of one term to denote different anatomical structures in different plant species or, conversely, use of different terms for similar anatomical structures;

standartizes terms denoting developmental stages and structural elements of plants of different angiosperm species and promotes the advance of plant comparative developmental genetics.

TAIR [37], Gramene [26], MaizeGDB [51], and other databases use *Plant Ontology* for describing gene

Table 4. Databases on whole-genome plant transcriptomics

Name	Description	Method	www.site
Genevestigator	Genevestigator is a reference expression database and a meta analysis system for various organisms, including *Arabidopsis*, rice, and wheat	Microarrays	[72]
BAR	The BioArray Resource. Resource for functional genomics of *Arabidopsis* and poplar. Contains several applications on gene expression analysis	Microarrays	[74]
AtGenExpress	Atlas of *Arabidopsis* gene expression	Microarrays	[76]
MPSS	Database on results of MPSS experiments in *Arabidopsis*, rice, and grapevine	MPSS	[79]
PLEXdb	Plant Expression Database. Resource for description of results of microarray experiments in 14 plant species and plant–pathogen systems	Microarrays	[80]
VitisExpDB	Vitis gene Expression Database. Database for gene expression in various grapevine cultivars and expression changes upon interaction with *Xylella fastidiosa*	EST, microarrays	[82]
Wheat EST database and Wheat big picture	Wheat EST database and Wheat big picture. Integrated resources on changes in the wheat transcriptome during development. Wheat bp contains description of wheat life cycle from germination to harvest	EST, microarrays	[85, 86]
TFGD	Tomato Functional Genomics Database. Accumulates data on changes during fruit maturation in expression of key genes controlling this process in wild type and mutants	EST, microarrays	[87]

expression, mutant phenotypes, and natural diversity of various species.

DATABASES ON PLANT GENE EXPRESSION

Databases integrating information on gene expression (Tables 4, 5) are aimed to describe gene expression patterns in the wild type, in different genetic backgrounds, and after various treatments. Information resources annotate gene expression patterns produced in four types of experiments:

(1) transcriptome studies using high throughoutput technologies e.g. microarrays, MPSS (Massive Parallel Signature Sequencing), and EST (Expressed sequence tags);

(2) genome-scale studies of collections of transgenic lines carrying reporter constructions;

(3) expression patterns of specific genes, detected by their mRNA levels using PCR, in situ hybridization, reporter genes, etc., and

(4) expression patterns of genes, detected by their protein levels.

In the type 3 experiments, lines from genetic collections carrying reporter constructs are also often used. The main difference lies in the fact that in databases on type 3 experiments, the information is annotated from published papers.

Databases on Whole Genome Studies of Plant Gene Expression

Currently, studies of plant gene expression using microchips are widespread. Depending on the type of microchip used, there are several varieties of this technique: cDNA membrane arrays (macroarrays), cDNA microarrays, oligonucleotide microarrays. The most widely used are oligo and cDNA microarrays, which exhibit highest sensitivity. Many databases on plant genomics contain information on the results of microarray studies of gene expression (Tables 2, 3). In addition, specialized databases have been developed (Table 4), which consist of (1) a depositary of raw data with description of experiments according to the MIAME (minimum information about a microarray experiment) standards [70] and (2) software for analysis of the raw data and visualization of the results. According to the MIAME, the experiment description lists the genotype, the structural element in which the expression was measured, and the stage of plant development [70]. The description uses terms from the existing ontologies, which significantly facilitates integration and analysis of data from different experiments [71].

Table 5. Database accumulating the results of experiments on detecting expression of specific genes

Name	Description	www.site
AREX	The Arabidopsis Gene Expression Database. Database on *Arabidopsis* gene expression in the root from genome-wide (microarrays) and gene-specific (in situs, promoter :: reporter constructs, etc.) sources	[89]
TIGR	Section "Expression Profiling of *Arabidopsis* Transcriptome" presents a profile catalog of expression of *Arabidopsis* genes, generated by Quantitative Real Time PCR and reporter genes	[92]
Oryza Tag Line	Integrated database on rice functional genomics. Contains information on rice gene expression from experiments with reporter genes	[95]
RMD	Rice Mutant Database. Contains information on reporter-gene expression patterns in rice ET lines	[97]
AGNS	The Arabidopsis GeneNet Supplementary Database. Annotates from published articles data on the expression of key developmental *Arabidopsis* genes	[99]
Oryzabase	Integrated Rice Science Database. Database on functional rice genomics. Contains also annotations of data on rice gene expression from published articles	[100]
GeneFarm	Along with other data on functional plant genomics (Table 3), contains information on gene expression from published articles	[43]

Let us consider a number of databases, whose primary content is information of gene expression, produced in whole genome experiments (Table 4). Genevestigator [72, 73] accumulates and analyses the information on gene expression obtained in microarray experiments in *Arabidopsis*, wheat, and rice. The application Digital Northern retrieves for a chosen experiment the expression of gene groups including up to ten genes. The Gene Correlation constructs diagrams of correlation of the gene expression with the expression of other genes. Gene Atlas estimates the gene expression in different tissues and detects tissue-specific genes. The application Gene Chronologer visualizes changes in the gene expression during the development. Mutant Surveyor enables detection of mutants with altered expression of the specific gene. The Response Viewer program builds gene expression profiles in response to various stimuli and detects genes whose expression change in response to the chosen factor. The results of the experiments are presented as histograms.

In the BAR (Bio-Array Resource) database [74, 75] (Table 4), the Expression Angler program helps finding gene groups with the same stress response or tissue and stage-specific genes from any five experiments. The Sampler Angler identifies in the database the expression profile that is the most similar to the given one, which can be used for typing mutant lines and the results of various treatments. The Expression Browser shows the gene expression level in the particular experiment. In the e_FP Browser (electronic Fluorescent Pictograph Browser), a query with the gene name presents a picture of the anatomical elements of *Arabidopsis* and poplar with the relative expression level of the gene.

The AtGenExpress [76, 77] (Table 4) presents the results of various types of experiments (changes of gene expression during development, by treatment with hormones, abiotic stress, illumination, pathogens) in tabular forms or as plots. The information on the gene expression in different anatomical elements, at different developmental stages, in different genotypes, under different growth conditions or stresses is presented in one plot. In plots of the AtGE DEVELOPMENT section, the abscissa gives the regions of expression presentation in the root, stem, leaf, whole plant, shoot apex, flower, floral organs, seeds, and the ordinate, the expression intensity. Each point on the plot is provided with an annotation, presenting the identifier of the original experiment, the expression level, the description of the experiment, the genotype, the anatomical element and its developmental stage, the plant age, the light intensity and soil substrate, the type of treatment or acting stress factor, and the duration of treatment or the exposure to the factor.

The MPSS (Massive Parallel Signature Sequencing) method is an alternative to microarray technologies in investigation of plant gene expression. The MPSS enables to concurrently examine more than one million transcripts, which can be presented in the pool of interest by less than 10 or more than 50000 copies [78]. This method quantitatively describes gene expression profiles in cells or tissues by "counting" all transcripts in them; its sensitivity is higher than that of the microarray method, enabling to record low expression levels. Using special software applications, the number of transcripts from each gene is counted, so that the expression level of the gene is presented as an integer.

To compare transcript frequency in different libraries, their number per million transcripts (TPM, transcripts per million) is estimated. This standardization is performed because the total transcript number significantly varies among the libraries (such as libraries from different plant organs). The data for a number of plant species generated using this method are presented in the MPSS database [79].

There are numerous other databases on transcriptome analysis. As additional examples, Table 4 lists a number of databases that accumulate data on several [80, 81] or individual [82–88] plant species. In a transcriptome database, the user can find a list of genes expressed in a particular organ, tissue, or cell type, at a particular developmental stage, in the wild type or an altered genotype, or in response to stressors (biotic or abiotic). The systems for data analysis implemented in these databases enable identifying tissue- or stage-specific genes, genes of specific response to particular stressors, or genes changing their expression in a particular genotype. These data should be further verified in gene-specific experiments (in situ and blot hybridization, reporter constructs, PCR, etc.).

The information on gene expression in the root of *Arabidopsis* is integrated in the APEX database [89–91] (Table 5). The gene expression is described according to *Plant Ontology*. This database provides search for: (1) expression pattern of the given gene (the search is performed using the AGI gene code); (2) genes expressed in particular anatomical elements; (3) the expression of specific gene in particular anatomical elements; (4) the gene expression pattern in selected experiment. The TIGR database presents the results of a research project on investigation of gene expression using quantitative PCR in various *Arabidopsis* tissues in the wild type and after treatments with hormones (or other biologically active substances), as well as under different growth conditions [92, 93]. This database contains information on more than 4000 genes, whose expression cannot be revealed using cDNA microarrays (because of the absence of the probes or low expression of these genes).

Databases on Gene Expression in Collections of Transgenic Lines with Reporter Constructs

Transformation by means of the so-called enhancer trap (ET) [94], which is T-DNA with an inserted reporter gene, e.g., GUS (β-glucoronidase) or GFP (green fluorescent protein), may lead to incorporation of this genetic construct in the physical and functional vicinity of the regulatory region of a gene and by this way induces the expression of the reporter gene. The next stage is identification of the gene in whose regulatory region the transgene is inserted, and description of the reporter expression pattern as the expression pattern of this gene. The TIGR database contains a section on the expression of *Arabidopsis* genes in the ET line collection for over 1000 genes with unknown function [92] (Table 5). In the Oryza Tag Line database [95, 96] for ET lines of rice, along with the images of the gene expression patterns, there are their description as the following: line identificator, reporter (GUS or GFP), stage of development, organ, tissue, expression level, and additional characteristics (e.g., tissue specificity). RMD (Rice Mutant Database) is another database, which describes rice gene expression using lines carrying reporter constructs [97, 98]. RMD employs a special text format: the presence or absence of the reporter expression is indicated at each item across the list of plant organs for each ET line.

Databases Integrating Information on Gene Expression from Published Papers

Published articles are another very important source of knowledge of gene function in development. Since gene expression data from published papers are annotated by experts, it is more labor-consuming. The TAIR database [37] gives a brief description of the gene expression and regulation of this expression in the Description section (textual, non-formalized description of the gene function) and the complete description of the gene expression in structured format based on *Plant Ontology*, listing the experiment type and references to the publication in the Annotations section. Using the structured format and ontologies is necessary for establishing connections between the data and underlies the work of automated archive systems, data search and analysis [67, 68]. The main format fields for describing the gene expression patters in many databases are fields: organ/tissue and stage of development. In addition to these main fields, other fields may describe experimental conditions (exposure to various factors), type of experiment, source of information (reference) [43, 99]. Using the structured format, rice gene expression is annotated in Oryzabase [100, 101]. In the Institute of Cytology and Genetics, Novosibirsk, Russia, database AGNS (*Arabidopsis* GeneNet Supplementary Database) [99, 102] has been developed, in which data on *Arabidopsis* gene expression from published papers are presented in the structured format (Table 5). In the GeneFarm database [43], the gene expression is annotated for rice, *Arabidopsis*, *Brassica napus, Glycine max, Hordeum vulgare, Lycopersicon esculentum, Medicago sativa, Phaseolus vulgaris, Triticum aestivum, Zea mays, Nicotiana alata, Vigna unguiculata, Daucus carota, Nicotiana plumbaginifolia*, and *Beta vulgaris.* In all databases listed in this section the expression of genes in wild-type is annotated. In databases GeneFarm and AGNS, in addition to search for the gene name [99], one can also conduct the search in other format fields, i.e., search for gene expression in a particular anatomical element or at a particular developmental stage. GeneFarm [43] provides additional search by the type of the experiment and its conditions. AGNS [99] also enables finding genes coexpressed with the given gene. AGNS also annotates gene expression in mutant and transgenic plants and has queries, using which one can learn (1) which mutants and transgenic plants exhibit altered this gene expression and (2) the expression of which genes is regulated by other genes in this anatomical element or at this stage of development. For annotation of gene expression, AGNS develops its own ontology, in which, along with terms from *Plant Ontology*, many additional terms are introduced for anatomical elements and developmental stages, allowing for more precise description of gene expression domains and their changes during development. The AGNS Ontology provides a detailed description of each term by taking all details from published papers with corresponding references. As a result, the AGNS Ontology can be regarded as a knowledge base on developmental changes in *Arabidopsis* anatomy and morphology.

Databases on the Expression of Plant Proteins

Information on plant proteins can be found in many of the databases listed above and in specialized databases on plant proteomics. Effective techniques for studying protein concentrations by mass spectroscopy (MS) and software programs for MS data analysis on plants with sequenced genome enabled quantitative examination of gene expression at proteomic level. Such data are presented in a number of databases. ProMEX [103, 104], is a database on peptides mapping to *Arabidopsis* and alfalfa proteins for quantitative analysis of protein concentrations and protein identification. AtProteome [105, 106] is a proteomic map of *Arabidopsis*, based on MS data on protein extracts from various organs and different developmental stages and in undifferentiated cell cultures. Along with these, there are databases on particular aspects of protein functioning, accumulating information from various experiments on one or several plant species. These databases include, for example, SUBA [107, 108] on intracellular protein localization; PhosPhAt [109, 110] on phosphorylation sites of *Arabidopsis* proteins; PlantsP [111] and PlantsT [112] on phosphorylation and dephosphorylation of plant proteins and on transport proteins [113], respectively; and ARAMEMNON, a database on membrane proteins [114, 115]. Thus, protein databasing is extending from the accumulation of protein structure data to the accumulation of data on measuring their concentrations and characterization of their activity in different cells, i.e., description of the functional dynamics of these molecules.

DATABASES ON PHENOTYPIC ABNORMALITIES IN PLANTS

On the Internet there is a vast diversity of databases describing plant phenotypes. Most of them contain descriptions of phenotypes of lines from various genetic collections. In this review, we consider only the databases that present phenotypes of plants from hybrid populations and collections of mutants and transgenic plants. Some lines from these collections were thoroughly studied experimentally; the results of these experiments are reported in publications. The largest portals on plant biology, TAIR [37] and MaizeGDB [51], in these cases extend the descriptions of phenotypes by including the data from published articles. In all databases, information about a phenotype can be obtained by a unique identifier of a gene or an allele. Information on phenotypes may be in text format or structured by fields corresponding to anatomic elements, developmental stages, and phenotypic abnormalities [26, 31, 37, 51].

Databases Describing Phenotypes of Lines from Genetic Collections

The possibility of creating complete genetic collections (representing a large part of genes in the genome) and describing phenotypes of lines from these collections in databases makes a new field in functional genomics, namely, phenomics [116]. In the series of functional genomics field (structural genomics, transcriptomics, proteomics), phenomics is the last element, investigating phenotypic changes at the level of genome-scale studies. The highest number of plant lines used in phenomics was generated by insertional mutagenesis. Insertional mutagenesis by T-DNA or transposons provides for not only suppression or activation of gene function, but also for gene cloning and identification. The existence of databases containing gene sequences allowed researchers to develop more targeted method of suppressing gene functions using RNA interference [117]. RNA interference and most insertion techniques produce lines with complete or partial loss of the gene function. Enhanced gene function (activation tagging) appears when the transforming construct contains a promoter providing consistently high (constitutive) gene expression (as a rule, such promoters are various modifications of the 35S promoter of cauliflower mosaic virus) [116]. Databases often list phenotypes from several genetic collections of different origin [26, 31, 118]. Along with collections of transgenic lines, phenotypes of mutants generated otherwise, ecotypes, and hybrid populations are described. Description of hybrid population phenotypes is generally related to mapping and tagging of quantitative trait loci.

A large number of *Arabidopsis* genetic collections (ecotypes, mutants, recombinant inbred lines, hybrid populations, and insertional lines) are maintained in the Nottingham *Arabidopsis* Stock Center (NASC) [118]. The NASC browser, atEnsembl [118, 119], in addition to presenting data on genes, transcripts, and proteins of *Arabidopsis* from various resources, provides access to description of genetic lines. The line description is presented in both (1) text format and (2) using *Plant Ontology* and ontology of phenotypic abnormalities developed in NASC (PATO, Phenotype, Attribute and Trait

Ontology), for phenotypes of mutants in various plant species. Phenotype is described using three term types: object, trait, and value. The object is annotated in terms of *Plant Ontology*; trait and value, in PATO terms. For instance, the phenotype of increased leaf width is formally annotated as follows: object, leaf; trait, width; value: broad.

Databases Gramene [26], GrainGenes [31], and Panzea [120, 121] describe phenotypic (quantitative and qualitative) traits in cereal lines and hybrids, employing *Plant Ontology* and Trait Ontology, developed in Gramene [122]. Trait Ontology has nine sections, which include traits connected with (1) anatomy and morphology, (2) growth and development, (3) sterility or fertility, (4) viability, (5) stress resistance, (6) biochemistry, (7) productivity, (8) qualitative characteristics, and (9) other parameters. Anatomy and morphology traits are grouped into characteristic of shoot, seed, fruit, vascular tissues, plant cells, and cell organelles. Growth and developmental traits include periods of vegetative and reproductive growth, development of embryo, seed, shoot, sporophyte, fruit, root, tissue, cells, cell organelles, and others.

Database MaizeGDB [51] uses, on one hand, *Plant Ontology*, and on the other, its own ontology of anatomical elements and vocabulary of phenotypic abnormalities and describes, along with hybrid lines, mutant phenotypes. The vocabulary of phenotypic abnormalities is applied for searching for information on mutant phenotypes. Inquire for any phenotypic trait provides a list of abnormalities having this trait. For instance, the query “purple” provides all possible anomalies associated with altering the organ color for purple. Selecting any of these anomalies, user will get a list of alleles resulting in it. Click on any of these alleles gives access to description of mutant phenotype for this allele with references to the source of information (laboratory, author, or articles); in some cases, images of the mutants are provided. In MaizeGDB, also more complex queries can be implemented, e.g., finding mutation leading to abnormalities in a particular organ or changing a particular trait. In Oryzabase [100], its own classification of phenotypic abnormalities is used along with textual description and images of mutant, inbred or recombinant rice lines.

In ET lines [94], insertion of T-DNA with a reporter gene generally results in loss of function of the gene located in the insertion site. In the Oryza Tag Lines database [95], in addition to images of ET lines, description of their phenotypic abnormalities is presented in the following format: line identifier, name of the mutant, class of phenotypic abnormality (e.g., physiology), trait (according to Trait Ontology [26, 122], organ and developmental stage according to *Plant Ontology*. Using this format enables to search for data both by individual fields and by their combination. Classifications of phenotypic abnormalities have been also developed for rice ET lines in database RMD [97], and for Ds-insertional *Arabidopsis* lines, in database RAPID (RIKEN *Arabidopsis* Phenome Information Database) [123, 124]. Database AGRICOLA (*Arabidopsis* Genomic RNAi Knockout Line Analysis) [125] contains information on phenotypes of plants, in which gene activity is suppressed through RNA interference. The phenotype description in AGRICOLA is standardized by format using two controlled vocabularies: of anatomical elements and of phenotypic abnormalities. In addition to query by gene name, AGRICOLA provides for searching lines with defects in particular organs or having the same phenotypic abnormality.

The SeedGenes database [126, 127] systematizes information on knockout mutant lines with abnormal seed development. The phenotype description is performed with the use of four controlled vocabularies: terminal phenotype, special characteristics, seed color, and embryo color. The user is able to make integrative queries with each of these controlled vocabularies. The terminal phenotype controlled vocabulary is a standardized tool for description of preliminary terminated embryogenesis including the morphology of the mutant embryo and the stage in embryogenesis at which the development was arrested. The special characteristics include an increase in cell and nucleus size in various seed parts. Controlled vocabularies “seed color” and “embryo color” enable to describe in detail the color shades of the corresponding organs.

Databases Annotating Mutant and Transgenic Phenotypes Using Data from Published Articles

To date, some databases on plant genetic collections started to input published data on phenotypic abnormalities in their lines. For instance, large-scale annotation of phenotypic abnormalities from publications is conducted in the TAIR database [37]. In this database, as well as in database GeneFarm [43], phenotypic abnormalities are described by non-formalized text. With this approach, the automated query is limited by query for a gene or an allele. The AGNS database [99] describes phenotypes using ontology and this allows searching for abnormalities of certain anatomical elements or those characteristic of certain development stages. The AGNS format includes also the additional Treatment field, which enables annotating information about the effect of particular factors (various stresses, hormones, and other substances) on mutant and transgenic plants.

DATABASES FOR GENE NETWORKS AND REGULATORY INTERACTIONS

Changes in gene activity during development occur in coordination, i.e., they are implemented through the network of genes and their products [128]. Gene networks determine cell differentiation at different developmental stages and transition from one stage to the other. Reconstruction of gene networks is based on

(1) experimental data on direct interactions between the gene network elements (binding of transcription factors with promoters, formation of protein complexes, phosphorylation and dephosphorylation of proteins, etc.) and (2) hypotheses on regulatory interactions in molecular processes, based on indirect evidence, e.g., a change in gene expression in a plant carrying a mutation at another gene, differences in phenotypic abnormalities between single and double mutants, etc. In the former case, links in the network are called direct, in the latter, indirect and subject to further experimental verification.

Databases for Molecular Interactions in Gene Networks

Most databases describing molecular interactions in gene networks are devoted to binding of transcription factors with regulatory elements in promoters. These databases include the following information resources and annotated published data: TRRD [129, 130], experimentally confirmed data on transcriptional regulatory regions of eukaryotic genes; EPD [131, 132] and PlantProm [133, 134], Pol II promoters with experimentally shown transcription start of eukaryotic and plant genes; PlantCARE [135, 136] and PLACE [137, 138], cis-regulatory plant elements; AtcisDB [139, 140], cis-regulatory elements in *Arabidopsis*. PlnTFDB [141, 142], RARTF [143, 144], and TOBFAC [145, 146] integrate information from different databases on transcriptional factors of various plant species (PlnTFDB), *Arabidopsis* (RARTF), and tobacco (TOBFAC). Many databases on functional protein interactions annotate and integrate data on plant proteins. Databases STRING [147, 148] and DIP [149, 150] can serve as examples. In information resource Brenda [151, 152], data on plant enzymes are annotated along with data on other organisms.

Databases for Reconstructed Gene Networks

Plant metabolic pathways (e.g., hormone synthesis pathways), as well as metabolic pathways of other organisms, are annotated from published articles in databases KEGG [153, 154] and MetaCyc [155, 156]. Metabolic pathways of *Arabidopsis* are presented separately in the AraCyc database [157]. AtRegNet visualizes regulatory circuits of *Arabidopsis* gene networks on the basis of information from AtcisDB and data on transcription factors [139]. Complex gene networks underlying some processes, e.g., functioning of plant meristem, seed germination, photomorphogenesis, are reconstructed by experts in computer system GeneNet [128, 158]. As genes belonging to one gene network are expressed in a coordinated fashion, analysis of many microarray experiments enables reconstruction of a network of coexpressed genes [159]. In *Arabidopsis*, for instance, such networks are stored in several databases [159–162]. In many of them, special programs align these networks relative to pathways annotated in KEGG or to networks of coexpressed genes of other organisms.

CONCLUSIONS

To understand how coordinated gene activity results in the formation of complex plant organism, it is essential to integrate classical knowledge on developmental processes, their mechanisms and disturbances in different genotypes with modern data from molecular genetics and functional genomics. Such integration by juxtaposition of experimental facts using computed methods and algorithms for data analysis reveals new knowledge. On the base of new relationships among the data found by this way, hypotheses may be formulized and experiments designed to prove them. Consequently, databases on genetics and genomics, providing access both to experimental data and to their analysis, become essential tools of research. Along with the basic portals, such as TAIR [37] and MaizeGDB [51], integrating data on model plants, *Arabidopsis* and maize (from gene sequence to phenotype), there are a multitude of databases on particular fields with their specific tasks. The scientific community actively uses databases both for finding information and planning experiments and for systematization and interpretation of research results in light of the data generated by other authors in different fields. Now, system biology approach to developmental processes not only guarantees the completeness of knowledge, but also becomes a method for producing it. In the near future, this approach will be implemented not only in databases, but also in quantitative and qualitative simulations that reproduce dynamics of the mechanisms underlying cell, tissue and organ growth and differentiation and enabling to conduct precise calculations for predictions on functioning of these mechanisms at various developmental stages and under different conditions.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, grant no. 08-04-01214a; SS-2447.2008.4; Programs of the Russian Academy of Sciences 23.29 and 22.8; and the Interdisciplinary Integrative Project of Siberian Branch of Russian Academy of Sciences no. 119.

REFERENCES

1. Gilbert, S.F., *Developmental Biology*, Sunderland: Sinauer, 2006.
2. Kitano, H., Computational Systems Biology, *Nature*, 2002, vol. 420, no. 6912, pp. 206–210.
3. Hennig, L., Patterns of Beauty-omics Meets Plant Development, *Trends Plant Sci.*, 2007, vol. 12, no. 7, pp. 287–293.

4. Ng, A., Bursteinas, B., Gao, Q., et al., Resources for Integrative Systems Biology: From Data through Databases to Networks and Dynamic System Models, *Brief Bioinform.*, 2006, vol. 4, pp. 318–330.
5. Stein, L., Beavis, W., Gessler, D., et al., The Plant Genome Database Working Group: Save Our Data!, *Scientist*, 2006, vol. 20, no. 4, pp. 24–25.
6. Mizrachi, I.K., Managing Sequence Data Methods, *Mol. Biol.*, 2008, vol. 452, pp. 3–27.
7. http://www.ddbj.nig.ac.jp
8. http://www.ebi.ac.uk/embl
9. http://www.ncbi.nlm.gov/Genbank
10. http://www.insdc.org
11. http://www.ncbi.nlm.gov/geo/
12. Barrett, T., Troup, D.B., and Wilhite, S.E., NCBI GEO: Archive for High-Throughput Functional Genomic Data, *Nucleic Acids Res.,* 2008, vol. 36, pp. D885–D890.
13. http://www.uniprot.org
14. UniProt Consortium: The Universal Protein Resource (UniProt), *Nucleic Acids Res.*, 2008, vol. 36, pp. D190–D195.
15. Terryn, N., Rouzé, P.Van., and Montagu, M., Plant Genomics, *FEBS Lett.*, 1999, vol. 452, pp. 3–6.
16. Wheeler, D.L., Barrett, T., Benson, D.A., et al., Database Resources of the National Center for Biotechnology Information, *Nucleic Acids Res.*, 2008, vol. 36, pp. D13–D21.
17. http://www.ncbi.nlm.gov
18. http://www.ebi.ac.uk/interpro
19. Mulder, N.J., Apweiler, R., and Attwood, T.K., New Developments in the InterPro Database, *Nucleic Acids Res.*, 2007, vol. 35, pp. D224–D228.
20. http://www.wwpdb.org
21. Henrick, K., Feng, Z., Bluhm, W.F., et al., Remediation of the Protein Data Bank Archive, *FEBS Lett.*, 1999, vol. 452, nos. 1–2, pp. 3–6.
22. http://plantta.jcvi.org/
23. Childs, K.L., Hamilton, J.P., Zhu, W., et al., The TIGR Plant Transcript Assemblies Database, *Nucleic Acids Res.*, 2007, vol. 35, pp. D846–D851.
24. http://www.plantgdb.org/
25. Duvick, J., Fu, A., Muppirala, U., et al., PlantGDB: A Resource for Comparative Plant Genomics, *Nucleic Acids Res.*, 2008, vol. 36, pp. D959–D965.
26. http://www.gramene.org
27. Liang, C., Jaiswal, P., Hebbard, C., et al., Gramene: A Growing Plant Comparative Genomics Resource, *Nucleic Acids Res.*, 2008, vol. 36, pp. D947–D953.
28. http://fgp.huck.psu.edu/tribe.htlm
29. Wall, P.K., Leebens-Mack, J., Müller, K.F., et al., PlantTribes: A Gene and Gene Family Resource for Comparative Genomics in Plants, *Nucleic Acids Res.*, 2008, vol. 36, pp. D970–D976.
30. Xu, Y., McCouch, S.R., and Zhang, Q., How Can We Use Genomics to Improve Cereals with Rice as a Reference Genome?, *Plant Mol. Biol.*, 2005, vol. 59, no. 1, pp. 7–26.
31. http://wheat.pw.usda.gov/GG2
32. O'Sullivan, H., GrainGenes: A Genomic Database for Triticeae and *Avena*, *Methods Mol. Biol.*, 2007, vol. 406, pp. 301–314.
33. http://rapeseed.plantsignal.cn
34. Wu, G.Z., Shi, Q.M., Niu, Y., et al., Shanghai RAPESEED Database: A Resource for Functional Genomics Studies of Seed Development and Fatty Acid Metabolism of *Brassica*, *Nucleic Acids Res.*, 2008, vol. 36, pp. D1044–D1047.
35. http://lis.comparative-legumes.org
36. Gonzales, M.D., Archuleta, E., Farmer, A., et al., The Legume Information System (LIS): An Integrated Information Resource for Comparative Legume Biology, *Nucleic Acids Res.*, 2005, vol. 33, pp. D660–D665.
37. http://arabidopsis.org
38. Swarbreck, D., Wilks, C., Lamesch, P., et al., The *Arabidopsis* Information Resource (TAIR): Gene Structure and Function Annotation, *Nucleic Acids Res.*, 2008, vol. 36, pp. D1009–D1014.
39. http://mips.helmholtz-muenchen.de/plant/athal/
40. Schoof, H., Ernst, R., Nazarov, V., et al., MIPS *Arabidopsis thaliana* Database (MAtDB): An Integrated Biological Knowledge Resource for Plant Genomics, *Nucleic Acids Res.*, 2004, vol. 32, pp. D373–D376.
41. http://www.tigr.org/tdb/e2k1/ath1/
42. Chan, A.P., Rabinowicz, P.D., Quackenbush, J., et al., Plant Database Resources at the Institute for Genomic Research, *Methods Mol. Biol.*, 2007, vol. 406, pp. 113–136.
43. http://urgi.versailles.inra.fr/Genefarm/index.htpl
44. Aubourg, S., Brunaud, V., and Bruyere, C., GeneFarm, Structural and Functional Annotation of *Arabidopsis* Gene and Protein Families by a Network of Experts, *Nucleic Acids Res.*, 2005, vol. 33, pp. D641–D646.
45. http://rapdb.dna.affrc.go.jp
46. Rice Annotation Project: The Rice Annotation Project Database (RAP-DB): 2008 Update, *Nucleic Acids Res.*, 2008, vol. 36, pp. D1028–D1033.
47. http://mips.helmholtz-muenchen.de/plant/rice/index.jsp
48. Spannagl, M., Noubibou, O., Haase, D., et al., MIPSPlantsDB-Plant Database Resource for Integrative and Comparative Plant Genome Research, *Nucleic Acids Res.*, 2007, vol. 35, pp. D834–D840.
49. http://rice.plantbiology.msu.edu
50. Ouyang, S., Zhu, W., Hamilton, J., et al., The TIGR Rice Genome Annotation Resource: Improvements and New Features, *Nucleic Acids Res.*, 2007, vol. 35, pp. D883–D887.
51. http://www.maizegdb.org/
52. Lawrence, C.J., Schaeffer, M.L., Seigfried, T.E., et al., MaizeGDB's New Data Types, Resources and Activities, *Nucleic Acids Res.*, 2007, vol. 35, pp. D895–D900.
53. http://www.medicago.org/
54. Retzel, E.F., Johnson, J.E., Crow, J.A., et al., Legume Resources: MtDB and Medicago.Org, *Methods Mol. Biol.*, 2007, vol. 406, pp. 261–274.
55. Sasaki, T., Matsumoto, T., Antonio, B.A., and Nagamura, Y., From Mapping to Sequencing, Post-Sequenc-

ing and Beyond, *Plant Cell Physiol.*, 2005, vol. 46, no. 1, pp. 3–13.

56. Wei, F., Coe, E., Nelson, W., et al., Physical and Genetic Structure of the Maize Genome Reflects Its Complex Evolutionary History, *PLoS Genet*, 2007, vol. 3, no. 7, p. e123.
57. Semagn, K., Bjornstad, A., and Ndjiondjop, M.N., Principles, Requirements and Prospects of Genetic Mapping in Plants, *African J. Biotechnol.*, 2006, vol. 5, no. 25, pp. 2569–2587.
58. Gupta, P.K., Mir, R.R., Mohan, A., and Kumar, J., Wheat Genomics: Present Status and Future Prospects, *Int. J. Plant Genomics*, 2008, no. 896451.
59. Zhang, H.B., Li, Y., Wang, B., and Chee, P.W., Recent Advances in Cotton Genomics, *Int. J. Plant Genomics*, 2008, no. 742304.
60. Stein, L., Genome Annotation: From Sequence to Biology, *Nat. Rev. Genet.*, 2001, vol. 2, no. 7, pp. 493–503.
61. Köhler, J., Philippi, S., and Lange, M., SEMEDA: Ontology Based Semantic Integration of Biological Databases, *Bioinformatics*, 2003, vol. 19, no. 18, pp. 2420–2427.
62. Strizh, I.G., Ontologies for Data and Knowledge Sharing in Biology: Plant ROS Signaling as a Case Study, *BioEssays*, 2006, vol. 28, no. 2, pp. 199–210.
63. Ashburner, M., Ball, C.A., Blake, J.A., et al., Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium. *Nat. Genet.*, 2000, vol. 25, no. 1, pp. 25–29.
64. Gene Ontology Consortium: The Gene Ontology Project in 2008, *Nucleic Acids Res.*, 2008, vol. 36, pp. D440–D444.
65. http://www.geneontology.org/
66. http://www.plantontology.org/
67. Avraham, S., Tung, C.W., Ilic, K., et al., The Plant Ontology Database: A Community Resource for Plant Structure and Developmental Stages Controlled Vocabulary and Annotations, *Nucleic Acids Res.*, 2008, vol. 36, pp. D449–D454.
68. Ilic, K., Kellogg, E.A., Jaiswal, P., et al., The Plant Structure Ontology, a Unified Vocabulary of Anatomy and Morphology of a Flowering Plant, *Plant Physiol.*, 2007, vol. 143, no. 2, pp. 587–599.
69. Pujar, A., Jaiswal, P., Kellogg, E.A., et al., Whole-Plant Growth Stage Ontology for Angiosperms and Its Application in Plant Biology, *Plant Physiol.*, 2006, vol. 142, pp. 414–428.
70. Brazma, A., Hingamp, P., Quackenbush, J., et al., Minimum Information about a Microarray Experiment (MIAME)-toward Standards for Microarray Data, *Nat. Genet.*, 2001, vol. 29, no. 4, pp. 365–371.
71. Zimmermann, P., Schildknecht, B., Craigon, D., et al., MIAME/Plant—Adding Value to Plant Microarrray Experiments, *Plant Methods*, 2006, vol. 2, pp. 1–3.
72. https://www.genevestigator.ethz.ch/
73. Hruz, T., Laule, O., Szabo, G., et al., Genevestigator V3: A Reference Expression Database for the Meta-Analysis of Transcriptomes, *Adv. Bioinform.*, 2008, no. 420747.
74. http://www.bar.utoronto.ca/
75. Winter, D., Vinegar, B., Nahal, H., et al., An 'Electronic Fluorescent Pictograph' Browser for Exploring and Analyzing Large-Scale Biological Data Sets, *PLoS One*, 2007, vol. 2, no. 8, p. e718.
76. http://www.weigelworld.org/expviz/expviz.jsp
77. Schmid, M., Davison, T.S., and Henz, S.R., A Gene Expression Map of *Arabidopsis thaliana* Development, *Nat. Genet.*, 2005, vol. 37, no. 5, pp. 501–506.
78. Nakano, M., Nobuta, K., Vemaraju, K., et al., Plant MPSS Databases: Signature-Based Transcriptional Resources for Analyses of mRNA and Small RNA, *Nucleic Acids Res.*, 2006, vol. 34, pp. D731–D735.
79. http://mpss.udel.edu
80. http://www.plexdb.org/
81. Wise, R.P., Caldo, R.A., and Hong, L., BarleyBase/PLEXdb: A Unified Expression Profiling Database for Plants and Plant Pathogens, *Methods Mol. Biol.*, 2007, vol. 406, pp. 347–364.
82. http://cropdisease.ars.usda.gov/vitis_at/main-page.htm
83. Doddapaneni, H., Lin, H., Walker, M.A., et al., VitisExpDB: A Database Resource for Grape Functional Genomics, *BMC Plant Biol.*, 2008, vol. 8, p. 23.
84. Wilson, I.D., Barker, G.L., and Beswick, R.W., A Transcriptomics Resource for Wheat Functional Genomics, *Plant Biotechnol. J.*, 2004, vol. 2, no. 6, pp. 495–506.
85. http://www.cerealsdb.uk.net
86. http://www.wheatbp.net
87. http://ted.bti.cornell.edu/
88. Fei, Z., Tang, X., Alba, R., and Giovannoni, J., Tomato Expression Database (TED): A Suite of Data Presentation and Analysis Tools, *Nucleic Acids Res.*, 2006, vol. 34, pp. D766–D770.
89. http://www.arexdb.org/database.jsp
90. Birnbaum, K., Shasha, D.E., Wang, J.Y., et al., A Gene Expression Map of the *Arabidopsis* Root, *Science*, 2003, vol. 302, no. 5652, pp. 1956–1960.
91. Brady, S.M., Orlando, D.A., Lee, J.-Y., et al., A High-Resolution Root Spatiotemporal Map Reveals Dominant Expression Patterns, *Science*, 2007, vol. 318, no. 5851, pp. 801–806.
92. http://www.tigr.org/tdb/e2k1/qpcr/index.shtml
93. Underwood, B., Xiao, Y., Redman, J., et al., Completing the Expression Catalog of the *Arabidopsis* Transcriptome by Quantitative Real Time PCR, in *Proceedings of Botany and Plant Biology Joint Congress*, Chicago, 2007, no. P42010.
94. Acosta-García, G., Autran, D., and Vielle-Calzada, J.P., Enhancer Detection and Gene Trapping as Tools for Functional Genomics in Plants, *Methods Mol. Biol.*, 2004, vol. 267, pp. 397–414.
95. http://urgi.versailles.inra.fr/OryzaTagLine/
96. Larmande, P., Gay, C., and Lorieux, M., *Oryza* Tag Line, a Phenotypic Mutant Database for the Genoplante Rice Insertion Line Library, *Nucleic Acids Res.*, 2008, vol. 36, pp. D1022–D1027.
97. http://rmd.ncpgr.cn
98. Zhang, J., Li, C., and Wu, C., RMD: A Rice Mutant Database for Functional Analysis of the Rice Genome, *Nucleic Acids Res.*, 2006, vol. 34, pp. D745–D748.
99. http://wwwmgs2.bionet.nsc.ru/agns

100. http://www.shigen.nig.ac.jp/rice/oryzabase
101. Kurata, N. and Yamazaki, Y., Oryzabase: An Integrated Biological and Genome Information Database for Rice, *Plant Physiol.*, 2006, vol. 140, no. 1, pp. 12–17.
102. Omel'yanchuk, N.A., Mironova, V.V., Zalevskii, E.M., et al., Systemic Approach to the Study of *Arabidopsis thaliana* Morphogenesis: I. AGNS Data Base, *Biofizika*, 2006, vol. 51, no. 7, pp. 75–82.
103. http://promex.mpimp-golm.mpg.de
104. Hummel, J., Niemann, M., Wienkoop, S., et al., ProMEX: A Mass Spectral Reference Database for Proteins and Protein Phosphorylation Sites, *BMC Bioinform.*, 2007, vol. 8, p. 216.
105. http://fgcz-atproteome.unizh.ch/index.php
106. Baerenfaller, K., Grossmann, J., and Grobei, M.A., Genome-Scale Proteomics Reveals *Arabidopsis thaliana* Gene Models and Proteome Dynamics, *Science*, 2008, vol. 320, no. 5878, pp. 938–941.
107. http://www.plantenergy.uwa.edu.au/suba2
108. Heazlewood, J.L., Verboom, R.E., Tonti-Filippini, J., et al., SUBA: The *Arabidopsis* Subcellular Database, *Nucleic Acids Res.*, 2007, vol. 35, pp. D213–D218.
109. http://phosphat.mpimp-golm.mpg.de
110. Heazlewood, J.L., Durek, P., Hummel, J., et al., PhosPhAt: A Database of Phosphorylation Sites in *Arabidopsis thaliana* and a Plant-Specific Phosphorylation Site Predictor, *Nucleic Acids Res.*, 2008, vol. 36, pp. D1015–D1021.
111. http://plantsp.genomics.purdue.edu/html/
112. http://plantst.sdsc.edu
113. Tchieu, J.H., Fana, F., Fink, J.L., et al., The PlantsP and PlantsT Functional Genomics Databases, *Nucleic Acids Res.*, 2003, vol. 31, no. 1, pp. 342–344.
114. http://aramemnon.botanik.uni-koeln.de/
115. Schwacke, R., Schneider, A., and van der Graaff, E., ARAMEMNON, a Novel Database for *Arabidopsis* Integral Membrane Proteins, *Plant Physiol.*, 2003, vol. 131, no. 1, pp. 16–26.
116. Holtorf, H., Guitton, M.C., and Reski, R., Plant Functional Genomics, *Naturwissenschaften*, 2002, vol. 89, no. 6, pp. 235–249.
117. Hilson, P., Allemeersch, J., and Altmann, T., Versatile Gene-Specific Sequence Tags for *Arabidopsis* Functional Genomics: Transcript Profiling and Reverse Genetics Applications, *Genome Res.*, 2004, vol. 10B, pp. 2176–2189.
118. http://www. arabidopsis.info/
119. James, N., Graham, N., Clements, D., et al., AtEnsEMBL: A Post-Genomic Resource Browser for *Arabidopsis*, *Methods Mol. Biol.*, 2007, vol. 406, pp. 213–228.
120. http://www.panzea.org.
121. Canaran, P., Buckler, E.S., Glaubitz, J.C., et al., Panzea: An Update on New Content and Features, *Nucleic Acids Res.*, 2008, vol. 36, pp. D1041–D1043.
122. Yamazaki, Y. and Jaiswal, P., Biological Ontologies in Rice Databases: An Introduction to the Activities in Gramene and Oryzabase, *Plant Cell Physiol.*, 2005, vol. 46, no. 1, pp. 63–68.
123. http://rarge.gsc.riken.jp/phenome/
124. Kuromori, T., Wada, T., Kamiya, A., et al., A Trial of Phenome Analysis Using 4000 Ds-Insertional Mutants in Gene-Coding Regions of *Arabidopsis*, *Plant J.*, 2006, vol. 47, no. 4, pp. 640–651.
125. http://www.agrikola.org
126. http://www.seedgenes.org
127. Tzafrir, I., Dickerman, A., Brazhnik, O., et al., The *Arabidopsis* SeedGenes Project, *Nucleic Acids Res.*, 2003, vol. 31, no. 1, pp. D90–D93.
128. Ananko, E.A., Podkolodny, N.L., Stepanenko, I.L., et al., GeneNet in 2005, *Nucleic Acids Res.*, 2005, vol. 33, pp. D425–D427.
129. http://wwwmgs.bionet.nsc.ru/mgs/gnw/trrd/
130. Kolchanov, N., Ignatieva, E., Podkolodnaya, O., et al., Transcription Regulatory Region Database (TRRD): A Source of Experimentally Confirmed Data on Transcription Regulatory Regions of Eukaryotic Genes, *Bioinformatics of Genome Regulation and Structure II*, Kolchanov, N., Hofestaedt, R., Milanesi, L., Eds., Springer Verlag, 2006, pp. 43–53.
131. http://www.epd.isb-sib.ch/
132. Schmid, C.D., Perier, R., Praz, V., and Bucher, P., EPD in Its Twentieth Year: Towards Complete Promoter Coverage of Selected Model Organisms, *Nucleic Acids Res.*, 2006, vol. 34, pp. D82–D85.
133. http://linux1.softberry.com/berry.phtml?topic=plantprom&subgroup=plantprom
134. Shahmuradov, I.A., Gammerman, A.J., Hancock, J.M., et al., PlantProm: A Database of Plant Promoter Sequences, *Nucleic Acids Res.*, 2003, vol. 31, no. 1, pp. 114–117.
135. http://www.bioinformatics.psb.ugent.be/webtools/plantcare/html/
136. Lescot, M., Déhais, P., Thijs, G., et al., PlantCARE, a Database of Plant Cis-Acting Regulatory Elements and a Portal to Tools for in Silico Analysis of Promoter Sequences, *Nucleic Acids Res.*, 2002, vol. 30, no. 1, pp. 325–327.
137. http://www.dna.affrc.go.jp/PLACE/
138. Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T., Plant Cis-Acting Regulatory DNA Elements (PLACE) Database: 1999, *Nucleic Acids Res.*, 1999, vol. 27, no. 1, pp. 297–300.
139. http://www.arabidopsis.med.ohio-state.edu/AtcisDB/
140. Palaniswamy, S.K., James, S., Sun, H., et al., AGRIS and AtRegNet: A Platform to Link Cis-Regulatory Elements and Transcription Factors into Regulatory Networks, *Plant Physiol.*, 2006, vol. 140, no. 3, pp. 818–829.
141. http://www.plantfdb.bio.uni-potsdam.de/v2.0/
142. Riano-Pachon, D.M., Ruzicic, S., Dreyer, I., and Mueller-Roeber, B., PlnTFDB: An Integrative Plant Transcription Factor Database, *BMC Bioinform.*, 2007, vol. 8, no. 42.
143. http://rarge.gsc.riken.jp/rartf/
144. Iida, K., Seki, M., and Sakurai, T., RARTF: Database and Tools for Complete Sets of *Arabidopsis* Transcription Factors, *DNA Res.*, 2005, vol. 12, no. 4, pp. 247–256.
145. http://compsysbio.achs.virginia.edu/tobfac/
146. Rushton, P.J., Bokowiec, M.T., and Laudeman, T.W., TOBFAC: The Database of Tobacco Transcription Factors, *BMC Bioinform.*, 2008, vol. 9, p. 53.

147. http://string.embl.de/
148. von Mering, C., Jensen, L.J., and Kuhn, M., STRING7 Recent Developments in the Integration and Prediction of Protein Interactions, *Nucleic Acids Res.*, 2007, vol. 35, pp. D358–D362.
149. http://dip.doe-mbi.ucla.edu
150. Salwinski, L., Miller, C.S., Smith, A.J., et al., The Database of Interacting Proteins: 2004 Update, *Nucleic Acids Res.*, 2004, vol. 32, pp. D449–D451.
151. http://www.brenda-enzymes.info/
152. Barthelmes, J., Ebeling, C., Chang, A., et al., BRENDA, AMENDA and FRENDA: The Enzyme Information System in 2007, *Nucleic Acids Res.*, 2007, vol. 35, pp. D511–D514.
153. http://www.genome.jp/kegg/
154. Kanehisa, M., Araki, M., Goto, S., et al., KEGG for Linking Genomes to Life and the Environment, *Nucleic Acids Res.*, 2008, vol. 36, pp. D480–D484.
155. http://metacyc.org/
156. Zhang, P., Foerster, H., Tissier, C., et al., MetaCyc and AraCyc: Metabolic Pathway Databases for Plant Research, *Plant Physiol.*, 2005, vol. 138, pp. 27–37.
157. http://www. arabidopsis.org/biocyc/index.jsp
158. http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/
159. Wei, H., Persson, S., Mehta, T., et al., Transcriptional Coordination of the Metabolic Network in *Arabidopsis*, *Plant Physiol.*, 2006, vol. 142, no. 2, pp. 762–774.
160. http://www.transvar.org/results/at_coexpress/
161. http://www.atted.jp/
162. Obayashi, T., Kinoshita, K., Nakai, K., et al., ATTED-II: A Database of Co-Expressed Genes and Cis Elements for Identifying Co-Regulated Gene Groups in *Arabidopsis*, *Nucleic Acids Res.*, 2007, vol. 35, pp. D863–D869.