

# PlasmID: a centralized repository for plasmid clone information and distribution

Dongmei Zuo<sup>1,2</sup>, Stephanie E. Mohr<sup>2</sup>, Yanhui Hu<sup>1</sup>, Elena Taycher<sup>1</sup>, Andreas Rolfs<sup>1</sup>, Jason Kramer<sup>2</sup>, Janice Williamson<sup>1</sup> and Joshua LaBaer<sup>1,2,\*</sup>

<sup>1</sup>Harvard Institute of Proteomics, Harvard Medical School, 320 Charles Street, Cambridge, MA 02141, USA and <sup>2</sup>DF/HCC DNA Resource Core, Harvard Medical School, 320 Charles Street, Cambridge, MA 02141, USA

Received August 15, 2006; Accepted October 11, 2006

## ABSTRACT

The Plasmid Information Database (PlasmID; <http://plasmid.hms.harvard.edu>) was developed as a community-based resource portal to facilitate search and request of plasmid clones shared with the Dana-Farber/Harvard Cancer Center (DF/HCC) DNA Resource Core. PlasmID serves as a central data repository and enables researchers to search the collection online using common gene names and identifiers, keywords, vector features, author names and PubMed IDs. As of October 2006, the repository contains >46 000 plasmids in 98 different vectors, including cloned cDNA and genomic fragments from 26 different species. Moreover, the clones include plasmid vectors useful for routine and cutting-edge techniques; functionally related sets of human cDNA clones; and genome-scale gene collections for *Saccharomyces cerevisiae*, *Pseudomonas aeruginosa*, *Yersinia pestis*, *Francisella tularensis*, *Bacillus anthracis* and *Vibrio cholerae*. Information about the plasmids has been fully annotated in adherence with a high-quality standard, and clone samples are stored as glycerol stocks in a state-of-the-art automated –80°C freezer storage system. Clone replication and distribution is highly automated to minimize human error. Information about vectors and plasmid clones, including downloadable maps and sequence data, is freely available online. Researchers interested in requesting clone samples or sharing their own plasmids with the repository can visit the PlasmID website for more information.

## INTRODUCTION

Plasmids are useful for a wide range of molecular genetic, genomic and proteomic approaches. In recent years, plasmid clone production has increased dramatically in response to

the availability of genome information and new technologies. For example, various laboratories and research centers have recently produced genome-scale gene and shRNA collections (1–10). The Harvard Institute of Proteomics is among those groups engaged in the production of such clone collections. This group launched the FLEXGene (Full Length EXpression-ready) project several years ago to build comprehensive, fully sequence-verified gene collections in recombinational vectors (1,11,12). Plasmid resources, including large-scale gene collections, are necessary for full exploitation of genomic information, as they facilitate genome-scale experimentation at the level of functional and phenotypic assays, as well as protein expression, purification and analysis (13–15).

Circulation of plasmids in the scientific community is essential for scientific progress, as it facilitates both original research and independent verification of reported results (15,16). Distribution of plasmid clones has traditionally fallen to individual researchers, who receive requests for clones in response to publications or presentations, and then retrieve, prepare and send out the clones. This *ad hoc* system lacks standardized quality control and reliable delivery, causing inefficiency and delays between clone requests and fulfillment. Moreover, equipment failure, poor record keeping, personnel turnover and organizational changes can lead to loss of clone samples and/or supporting information over time. Given the dramatic increase in the number and scale of plasmid clone collections, the traditional method of sharing plasmids has become insufficient to handle storage, maintenance and distribution of plasmid clones and supporting information.

Centralization of plasmid resources in a repository can provide standardized quality control and faster delivery of clones, in addition to relieving individual investigators of the need to spend time and resources on clone storage, maintenance and distribution. Centralized repositories are better equipped to manage large clone collections in an accurate and cost effective manner (15,16). In addition, a repository can speed delivery of clones by standardizing and simplifying the Materials Transfer Agreement (MTA) process, which can otherwise delay the sharing of resources (16,17). Finally, a

\*To whom correspondence should be addressed. Tel: +1 6173240816; Fax: +1 6173240824; Email: [joshua\\_labaer@hms.harvard.edu](mailto:joshua_labaer@hms.harvard.edu)

centralized repository serves as a reliable archive to prevent loss of samples and information over time.

The Dana-Farber/Harvard Cancer Center (DF/HCC) DNA Resource Core recently launched an effort to store, maintain and distribute plasmid clones shared by researchers at academic and non-profit institutions. The initial repository collection includes clone sets from DF/HCC members, a group of ~800 principal investigators at Harvard-affiliated institutions whose research focuses on the biology and treatment of cancer. Clones received by the repository are stored in an automated freezer storage system with an integrated database that ensures accurate sample tracking and long-term viability, and are made immediately available to researchers across the globe upon request.

The Plasmid Information Database (PlasmID) was developed to facilitate the effort to collect and distribute plasmid clones. As described here, PlasmID tracks and manages all relevant information related to clone intake, storage, maintenance and distribution.

## DATABASE DESCRIPTION

### Database design

PlasmID was developed with a J2EE compliant three-tier architecture consisting of a relational database as the data storage layer, a suite of Java modules hosted by an application server as the business logic layer and a web-based presentation layer. The database was designed to manage every step from clone collection, replication and storage to clone distribution. It includes four major tracking components: core data, sample and process data, clone order data, and user account management.

Core data tracking includes all relevant aspects of plasmid information, including vector backbone, inserts, growth conditions, host strains, selectable markers, tags, mutations, associated authors or contributors, and relevant publications. Individual clones are defined as vector plus insert(s). This design obviates the need to store duplicate information for individual clones that share the same vector or insert; keeps track of the lineage from parent to child vectors; and also makes it possible to store maps, sequences and other information associated with clones, vectors and inserts separately. In addition, this design is flexible enough to accommodate vectors with any number of inserts. Most clones in the repository collection contain either no insert ('empty vector' clones) or a single gene insert. For the latter, the clone is linked in the database to data pertinent to the single gene insert. When the insert contains more than one gene (e.g. genomic fragments), the clone is linked to all of the genes present in the insert, such that search with any one of the genes will retrieve the clone record from the database.

PlasmID was also designed to track clone sample processing steps via integration with robotic liquid-handling instruments and the freezer storage system. This enforces use of barcode labels and automation and thus, helps ensure accurate tracking and clone sample integrity.

### Quality control

Research success demands that clone repositories deliver exactly what was requested. To establish and maintain a

high-quality standard, several quality control procedures are enforced by PlasmID, including a stringent clone annotation process; automated clone storage and retrieval; and automated sample handling.

When plasmid clones are received from researchers, information about the clones is carefully annotated and imported into the database. Annotation of clone data from different sources may require specialized handling to ensure the best clone information quality. For clones shared by most laboratories, information is curated from user-provided files by a PhD-level scientist. For large collections, information can be extracted and parsed automatically from other databases. The annotation process relies on controlled vocabularies and standardized terms used by public databases such as EntrezGene, FlyBase, WormBase or the *Saccharomyces* Genome Database (18–22). The annotated information is then uploaded into the database via formatted files. Once the clone records are in the database, the contributing researchers are encouraged to review the records and make changes where needed. In addition, feedback from users who have requested clones is solicited and reviewed to help validate clone information.

Immediately upon intake, plasmids are replicated in a T1, T5 phage-resistant bacterial host strain, and working and archival glycerol samples are created. The archival copies are stored in 96- or 384-well format in a standard  $-80^{\circ}\text{C}$  freezer. Working copies are stored in individually 2D barcode-labeled tubes (Micronic BV) in an automated BioBank freezer (Thermo Electron Co.). The BioBank comprises a pair of  $-80^{\circ}\text{C}$  storage units with liquid  $\text{CO}_2$  back-up and a centralized automated picking station that reads the 2D barcodes on all tubes and re-arrays any requested samples at  $-15^{\circ}\text{C}$ . The system has a capacity of 160 000 samples, which is expandable. Once in storage, the integrated software alerts the operator after a clone has experienced six freeze-thaw cycles or four years in storage, to allow for re-growth, and creation of a new working sample to maintain clone viability. Retrieval of samples from the BioBank is fully automated and integrated with PlasmID. Upon clone request in PlasmID, work lists are automatically generated and stored in a common file server. The operator uploads the work lists directly to the BioBank software to retrieve the clones. Clones can be retrieved and arranged in a default format that maximizes efficiency, or in a user-defined format that facilitates downstream experimentation.

Clone replication is managed with workstation automation. Each step is tracked by PlasmID via direct integration with the liquid-handling instruments to avoid human error by enforcing automated cross-check of barcode labels and by collecting appropriate log files as verification of each process step.

### Database content

PlasmID currently contains information pertaining to  $\geq 46\,000$  plasmids in 98 different vectors (Table 1). The set of 'empty vector' clones include plasmids useful for routine and cutting-edge cloning, and plasmids useful for cell-free and cell-based expression. The set of insert-containing clones includes genes from 26 distinct species of viruses, bacteria and eukaryotes. The insert types represented include

**Table 1.** Clones and genes currently represented in PlasmID

	Number of clones (vectors)	Number of genes
<i>Homo sapiens</i>	13 775 (33)	4674
<i>Saccharomyces cerevisiae</i>	6675 (5)	4953
<i>Yersinia pestis</i>	3968 (1)	3968
<i>Vibrio cholerae</i>	7463 (3)	3776
<i>Pseudomonas aeruginosa</i>	2383 (1)	2383
<i>Francisella tularensis</i>	4503 (3)	1585
<i>Bacillus anthracis</i>	7608 (2)	3804
Other species (19)	386 (43)	311
Empty vectors (no insert)	39	NA
Total	46 800 (98)	25 454

cDNAs, shRNAs, genomic fragments and whole viral genomes. Moreover, a large group of plasmids in the repository comprises fully sequence-verified genes in recombinational cloning vectors that were shared with the repository by the Harvard Institute of Proteomics. These plasmids include complete or near-complete gene collections for *Saccharomyces cerevisiae*, *Pseudomonas aeruginosa*, *Yersinia pestis*, *Francisella tularensis*, *Bacillus anthracis* and *Vibrio cholerae*, in addition to a large collection of human genes. Related clones have been organized into specific ‘clone collections’ that include sets of  $\geq 500$  human kinases and sets of  $\geq 1000$  breast cancer-associated genes (4,23). Detailed information about clones and their component parts, vectors and inserts, are freely available to all researchers via the PlasmID website. Moreover, users can download maps, gene lists and sequence data relevant to specific vectors, inserts and clones.

### Database queries

PlasmID can be searched online using four basic search types: gene identifiers, clone identifiers, vector properties, and a combined text string search by gene, vector, author and/or publication.

Search by gene or clone identifiers are useful for users who know specifically what genes or clones they are interested in. Searches can be performed using common gene identifiers such as Gene IDs, GenBank accession or GI numbers, Gene Symbols (including official symbols and aliases), specific clone identifiers (e.g. PlasmID clone IDs and FLEXGene IDs), and in some cases, specialized database identifiers relevant to well-annotated model organisms (e.g. SGD, FlyBase or WormBase IDs). Searches can be done in batch mode using a list of identifiers.

For researchers who wish to find vectors that carry specific functionalities (e.g. mammalian expression and epitope tags), a more complex query tool for vectors that harbor a broad variety of features was developed. Extensive annotation has been done for the vector backbones using controlled vocabularies for consistency. Each vector is assigned one or more properties that are further organized into broader categories such as assay type, cloning system and expression (Table 2). In response to a search, the system identifies all the vectors and clones in those vectors that meet the criteria selected by the researcher. The user then has the option to view the list of clones or to further limit the search results by selecting specific vectors.

The combined text string search (‘Advanced Text Search’) is the most versatile search tool at PlasmID. This tool

**Table 2.** Vector feature categories

Category	Vector features
Assay	Red/RecET mutagenesis, BAC mutagenesis, I-SceI site, blue/white selection, with tag/fusion/marker, baculovirus/insect cell transposition, fluorescent marker, RecA mutagenesis, two-hybrid
Cloning system	loxP, Red recombination, CpoI enzyme-based directional cloning, recombinational cloning, multiple cloning site, acceptor (destination), Creator™, Gateway®, MAGIC mating system, donor (entry)
Expression	Bacterial expression, baculovirus/insect cell expression, inducible expression, retroviral vector, ssDNA production, <i>in vitro</i> transcription, yeast expression, yeast shuttle, mammalian expression, mammalian transduction, lentiviral infection

facilitates full-text search for all possible matches to gene or vector name(s), vector feature(s), author or contributor name(s), or PubMed ID(s), either alone or in combination. To further increase the flexibility of the tool, partial terms can be used. For example, users can search with ‘CDK’ to identify all the CDK-related genes, including CDK2, CDK3, etc. Moreover, search results can be limited to display only those results that are relevant to one species.

In all cases, search results are displayed in tabular format with active links to Clone Detail and Vector Detail pages, and to external databases (Figure 1). Results can be sorted alphabetically by specific columns (e.g. vector name and gene name) and if desired, results can be downloaded as an MS-Excel spreadsheet for local storage and analysis.

### PLASMID SHARING AND DISTRIBUTION

Plasmids are currently accepted as purified DNA in solution, which is then transformed into an appropriate phage-resistant bacterial host strain. At a minimum, researchers provide growth conditions, vector and insert-related information, and the PubMed IDs of any relevant publications. Vector and/or clone maps and sequences are also included when they are available.

Most plasmids received by the repository are distributed world-wide to researchers at any academic or non-profit laboratory. However, for some clones, intellectual property or biological safety concerns restrict distribution to specific sub-groups and/or geographical regions. In most cases, plasmid clone samples are distributed as glycerol stocks. In some cases, legal restrictions or shipping concerns require that plasmids are shipped as purified DNA spotted to filter paper. In keeping with accepted practices and in compliance with granting agencies, the DNA Resource Core charges a minimal handling fee for plasmid clone samples and asks that requestors cover shipping charges (16).

Most clones are distributed under the terms of a standardized and simple MTA. The DF/HCC DNA Resource Core has had considerable success in gaining pre-approval for the MTA from specific institutions in order to streamline clone distribution. Members of institutions that have granted pre-approval do not need to obtain signatures on an MTA to receive most clones. We encourage additional institutions to participate in this effort.

**PlasmID Database - Microsoft Internet Explorer**

Address: <http://plasmid.hms.harvard.edu/AdvancedSearch.do>

Welcome To PlasmID  
DF/HCC DNA Resource Core

Sign Out | My Account | FAQ

Home Plasmid Submission Plasmid Request Contact Us

**Plasmid Request** Plasmid Request > Search Plasmids > Search Results

**Search Results**  
Here are the plasmids found in the database.

List of search terms found

Page: 1 Display Download [Explanation of Terms](#)

Clone ID	Clone Type	Species Specific ID	Gene Symbol	Gene Name	Reference Sequence	Mutation	Discrepancy	Insert Format	Vector	Selectable Markers	Special MTA
<a href="#">HCDD0000613</a>	cDNA	<a href="#">T187</a>	TP53	tumor protein p53 (Li-Fraumeni syndrome)	<a href="#">NM_000546</a>	No	Yes	FUSION	<a href="#">pCMV-Dual</a>	bacterial: ampicillin	
2	cDNA	<a href="#">T187</a>	TP53	tumor protein p53 (Li-Fraumeni syndrome)	<a href="#">NM_000546</a>	No	Yes	CLOSED	<a href="#">pCMV-Dual</a>	bacterial: ampicillin	

**Clone Detail - Microsoft Internet Explorer**

Address: <http://plasmid.hms.harvard.edu/CloneDetail.do?cloneid=4112&access=>

Clone: HCDD0000613

Clone ID: HCDD0000613  
Type: cDNA  
Is Verified: Y  
Status: AVAILABLE  
Source: HP  
Description: HP  
Comments: Map

Related Identifiers:  
GI: [4112](#)  
GenBank Accession: [4112](#)  
HP Clone ID: [4112](#)  
HP Master Clone ID: 10976  
Original Clone ID: FLH005481.01L

Property:  
Collection:  
Insert Information:  
Insert Size (bp): 1102  
Specified Mutations: None

**Vector Detail - Microsoft Internet Explorer**

Address: <http://plasmid.hms.harvard.edu/VectorDetail.do?vectorid=2>

Vector: pCMV-Dual

Name: pCMV-Dual  
Description: Recombinational donor/master vector with 50-90 tag ORF, T7 and M13 primer sites; ampicillin resistance; restriction enzyme cloning (into) and recombinational cloning (from)  
Synonyms:  
Type: bacterial plasmid  
Form: ccdDNA  
Size (bp): 4938  
Properties: Creator: donor (entry), topP, multiple cloning site, recombinational cloning, with tag/fusion marker  
Vector Map:  
Vector Sequence: [pCMV-Dual\\_F00751](#)  
Comments: The position of features were determined for the empty form of the vector, which is described in detail on the BD-Clontech website. Subcloning into the MCS was performed using the infusion reaction strategy and some restriction sites in the MCS may be lost after an insert is added.  
Features:

Name	Description	Start Position	End Position
MCS	restriction enzyme-type multiple cloning site	45	129
ColE1	ColE1 (pUC-type) origin of replication	4111	4574
SacB	SacB negative selector marker gene (sucrose sensitivity)	1057	2972
M13 primer	M13 primer sequencing site	4939	4955
T7 primer	T7 RNA polymerase promoter	4856	4858
topP	topP reconstruction site	9	42
topP site	topP reconstruction site	1076	1040
ori	ori	896	327
ampR	ampicillin resistance gene	3163	3963
SD site	splice donor site	132	136
kanR tag	kanR tag	141	170

**Nucleotide**

Search: [tp53](#)

Results: 1102

1. [NM\\_000546](#) tumor protein p53 (Li-Fraumeni syndrome) [tp53]

2. [U00096](#) Escherichia coli O157:H7 genome [tp53]

3. [U00096](#) Escherichia coli O157:H7 genome [tp53]

4. [U00096](#) Escherichia coli O157:H7 genome [tp53]

5. [U00096](#) Escherichia coli O157:H7 genome [tp53]

6. [U00096](#) Escherichia coli O157:H7 genome [tp53]

7. [U00096](#) Escherichia coli O157:H7 genome [tp53]

8. [U00096](#) Escherichia coli O157:H7 genome [tp53]

9. [U00096](#) Escherichia coli O157:H7 genome [tp53]

10. [U00096](#) Escherichia coli O157:H7 genome [tp53]

**Figure 1.** Representative search results at PlasmID. Large image, results of a search with 'tp53' (e.g. p53) in the gene field of the Advanced Text Search query tool. Column headers include PlasmID Clone ID, clone type (e.g. no insert or cDNA), species-specific ID (external link to Entrez Gene or databases such as FlyBase or WormBase), gene symbol, gene name, reference sequence, mutation (yes or no), discrepancy (yes or no; discrepancy indicates differences between the cloned sequence and the reference sequence), insert format (where closed means the natural stop codon is present and fusion means that the stop codon is not present, facilitating production of C-terminally tagged proteins), vector, selectable markers (for selection in bacterial or other hosts), and special MTA, which allows us to indicate that a specific clone requires something other than the standard MTA in use for the majority of clones. Insets, examples of pages provided as active links in the results table. The PlasmID Clone ID in the results table is an active link to a Clone Detail Page (left inset), which includes detailed information about the clone (growth conditions, downloadable map, etc.) and the insert (nucleotide sequence, annotation of mutations or discrepancies, etc.). The Clone Detail Page also provides further links to external databases of gene information such as NCBI Entrez Gene (left, lower inset). The vector name in the results table is an active link to a Vector Detail Page (right inset), which includes detailed information about the vector (specific vector features, downloadable map, etc.). For all PlasmID pages, red text is indicative of an active link.

All clone requests are handled online. Users can monitor clone request status and retrieve clone information related to a given request via PlasmID at any time.

## FUTURE PERSPECTIVES

Plasmid clones are tools for a vast array of cloning, expression and DNA manipulation approaches; thus, centralization and open access to plasmid samples and information help facilitate biological research. We have successfully built a robust, highly-quality controlled and flexible database of plasmid clone information. As in the past, clones produced in high-throughput cloning efforts will continue to serve as a major source for the repository. Moreover, recruitment of diverse types of plasmids from the scientific community will help expand the scope and utility of the repository collection. Researchers who wish to submit their plasmids into the repository are invited to visit the PlasmID website (<http://plasmid.hms.harvard.edu>) for more information about how they can help in this effort. At the same time, we will continue to improve the quality of the existing collection based on user feedback, and to improve the web interface in order to best facilitate ease-of-use and enhance search capability of PlasmID.

## NOTE ADDED IN PROOF

Our group was recently named the site of the Protein Structure Initiative (PSI) Materials Repository for storage and distribution of plasmid clones generated at PSI site laboratories (NIH Grant 1U01GM079610-01). This will increase the number of sequence-verified, unique plasmid clones available at PlasmID by approximately 50 000–100 000 over the next five years. The clone inserts in the PSI collections will include open reading frames from several different eukaryotic and prokaryotic species.

## ACKNOWLEDGEMENTS

The authors would like to thank Michael Collins for system support and Mauricio Fernandez for help with integration of automated instruments. This work was supported by the DF/HCC, which is a designated Comprehensive Cancer Center of the National Cancer Institute (5P30 CA06516-06). Additional support was provided by the Harvard Skin Disease Research Center of Brigham and Women's Hospital and Harvard Medical School (P30 AR42689-11). Funding for the automated freezer storage system was provided by the National Institutes of Health (1S10 RR19310-01). Funding to pay the Open Access publication charges for this article was provided by the National Cancer Institute (5P30 CA06516-06).

*Conflict of interest statement.* None declared.

## REFERENCES

1. Labaer, J., Qiu, Q., Anumanthan, A., Mar, W., Zuo, D., Murthy, T.V., Taycher, H., Halleck, A., Hainsworth, E., Lory, S. *et al.* (2004) The *Pseudomonas aeruginosa* PA01 gene collection. *Genome Res.*, **14**, 2190–2200.
2. Martzen, M.R., McCraith, S.M., Spinelli, S.L., Torres, F.M., Fields, S., Grayhack, E.J. and Phizicky, E.M. (1999) A biochemical genomics approach for identifying genes by the activity of their products. *Science*, **286**, 1153–1155.
3. Paddison, P.J., Silva, J.M., Conklin, D.S., Schlabach, M., Li, M., Aruleba, S., Balija, V., O'Shaughnessy, A., Gnoj, L., Scobie, K. *et al.* (2004) A resource for large-scale RNA-interference-based screens in mammals. *Nature*, **428**, 427–431.
4. Park, J., Hu, Y., Murthy, T.V., Vannberg, F., Shen, B., Rolfs, A., Hutt, J.E., Cantley, L.C., Labaer, J., Harlow, E. *et al.* (2005) Building a human kinase gene repository: bioinformatics, molecular cloning and functional validation. *Proc. Natl Acad. Sci. USA*, **102**, 8114–8119.
5. Reboul, J., Vaglio, P., Tzellas, N., Thierry-Mieg, N., Moore, T., Jackson, C., Shin-i, T., Kohara, Y., Thierry-Mieg, D., Thierry-Mieg, J. *et al.* (2001) Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in *C.elegans*. *Nature Genet.*, **27**, 332–336.
6. Silva, J.M., Li, M.Z., Chang, K., Ge, W., Golding, M.C., Rickles, R.J., Siolas, D., Hu, G., Paddison, P.J., Schlabach, M.R. *et al.* (2005) Second-generation shRNA libraries covering the mouse and human genomes. *Nature Genet.*, **37**, 1281–1288.
7. Stapleton, M., Carlson, J., Brokstein, P., Yu, C., Champe, M., George, R., Guarin, H., Kronmiller, B., Pacleb, J., Park, S. *et al.* (2002) A Drosophila full-length cDNA resource. *Genome Biol.*, **3**, RESEARCH0080.
8. Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F. *et al.* (2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc. Natl Acad. Sci. USA*, **99**, 16899–16903.
9. Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P. *et al.* (2000) A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature*, **403**, 623–627.
10. Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T. *et al.* (2001) Global analysis of protein activities using proteome chips. *Science*, **293**, 2101–2105.
11. Brizuela, L., Braun, P. and LaBaer, J. (2001) FLEXGene repository: from sequenced genomes to gene repositories for high-throughput functional biology and proteomics. *Mol. Biochem. Parasitol.*, **118**, 155–165.
12. Brizuela, L., Richardson, A., Marsischky, G. and LaBaer, J. (2002) The FLEXGene repository: exploiting the fruits of the genome projects by creating a needed resource to face the challenges of the post-genomic era. *Arch. Med. Res.*, **33**, 318–324.
13. Braun, P. and LaBaer, J. (2003) High throughput protein production for functional proteomics. *Trends Biotechnol.*, **21**, 383–388.
14. Pearlberg, J. and LaBaer, J. (2004) Protein expression clone repositories for functional proteomics. *Curr. Opin. Chem. Biol.*, **8**, 98–102.
15. Hayashizaki, Y. and Kawai, J. (2004) A new approach to the distribution and storage of genetic resources. *Nature Rev. Genet.*, **5**, 223–228.
16. Weaver, T., Maurer, J. and Hayashizaki, Y. (2004) Sharing genomes: an integrated approach to funding, managing and distributing genomic clone resources. *Nature Rev. Genet.*, **5**, 861–866.
17. Cozzarelli, N.R. (2004) UPSIDE: Uniform principle for sharing integral data and materials expeditiously. *Proc. Natl Acad. Sci. USA*, **101**, 3721–3722.
18. Drysdale, R.A. and Crosby, M.A. (2005) FlyBase: genes and gene models. *Nucleic Acids Res.*, **33**, D390–D395.
19. Schwarz, E.M., Antoshechkin, I., Bastiani, C., Bieri, T., Blasiar, D., Canaran, P., Chan, J., Chen, N., Chen, W.J., Davis, P. *et al.* (2006) WormBase: better software, richer content. *Nucleic Acids Res.*, **34**, D475–D478.
20. Chen, N., Harris, T.W., Antoshechkin, I., Bastiani, C., Bieri, T., Blasiar, D., Bradnam, K., Canaran, P., Chan, J., Chen, C.K. *et al.* (2005) WormBase: a comprehensive data resource for *Caenorhabditis* biology and genomics. *Nucleic Acids Res.*, **33**, D383–D389.
21. Dwight, S.S., Balakrishnan, R., Christie, K.R., Costanzo, M.C., Dolinski, K., Engel, S.R., Feierbach, B., Fisk, D.G., Hirschman, J., Hong, E.L. *et al.* (2004) *Saccharomyces* genome database: underlying principles and organisation. *Brief Bioinformatics*, **5**, 9–22.
22. Cherry, J.M., Adler, C., Ball, C., Chervitz, S.A., Dwight, S.S., Hester, E.T., Jia, Y., Juvik, G., Roe, T., Schroeder, M. *et al.* (1998) SGD: *Saccharomyces* Genome Database. *Nucleic Acids Res.*, **26**, 73–79.
23. Witt, A.E., Hines, L.M., Collins, N.L., Hu, Y., Gunawardane, R.N., Moreira, D., Raphael, J., Jepson, D., Koundinya, M., Rolfs, A. *et al.* (2006) Functional proteomics approach to investigate the biological activities of cDNAs implicated in breast cancer. *J. Proteome Res.*, **5**, 599–610.