The 2016 database issue of Nucleic Acids Research and an updated molecular biology database collection

Daniel J. Rigden1,* Xosé M. Fernández-Suárez2 and Michael Y. Galperin3,*

1Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK, 2Thermo Fisher Scientific, Inchinnan Business Park, Paisley, Renfrew PA4 9RF, UK and 3National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

Received November 22, 2015; Accepted November 23, 2015

ABSTRACT
The 2016 Database Issue of Nucleic Acids Research starts with overviews of the resources provided by three major bioinformatics centers, the U.S. National Center for Biotechnology Information (NCBI), the European Bioinformatics Institute (EMBL-EBI) and Swiss Institute for Bioinformatics (SIB). Also included are descriptions of 62 new databases and updates on 95 databases that have been previously featured in NAR plus 17 previously described elsewhere. A number of papers in this issue deal with resources on nucleic acids, including various kinds of non-coding RNAs and their interactions, molecular dynamics simulations of nucleic acid structure, and two databases of super-enhancers. The protein database section features important updates on the EBI’s Pfam, PDBe and PRIDE databases, as well as a variety of resources on pathways, metabolomics and metabolic modeling. This issue also includes updates on popular metagenomics resources, such as MG-RAST, EBI Metagenomics, and probeBASE, as well as a newly compiled Human Pan-Microbe Communities database. A significant fraction of the new and updated databases are dedicated to the genetic basis of disease, primarily cancer, and various aspects of drug research, including resources for patented drugs, their side effects, withdrawn drugs, and potential drug targets. A further six papers present updated databases of various antimicrobial and anticancer peptides. The entire Database Issue is freely available online on the Nucleic Acids Research website (http://nar.oxfordjournals.org/). The NAR online Molecular Biology Database Collection, http://www.oxfordjournals.org/nar/database/c/, has been updated with the addition of 88 new resources and removal of 23 obsolete websites, which brought the current listing to 1685 databases.

NEW AND UPDATED DATABASES
The 2016 Nucleic Acids Research Database Issue is the 23rd annual collection of descriptions of various molecular biology databases. It includes 178 papers, of which 62 describe newly created databases (Table 1), 95 papers provide updates on databases that have been described in the previous NAR Database Issues and 17 contain updates on databases whose descriptions have previously been published in other journals (Table 2).

This year’s issue is again divided into eight sections that deal with (i) nucleic acid sequence and structure; (ii) protein sequence and structure; (iii) metabolic and signaling pathways; (iv) viruses, bacteria, protozoa and fungi; (v) genomes of human and model organisms; (vi) human diseases and drugs; (vii) plants and (viii) other topics, including mitochondrial databases and databases of chemical compounds. It should be noted, however, that these general categories may only partly reflect the database scope, so we encourage the reader to browse the entire table of contents: a useful database might be found in a totally unexpected bin. As an example, a researcher interested in G-protein coupled receptors would obviously be drawn to the dedicated resource GPCRdb (1), but would also find value in the broader IUPHAR/BPS Guide to Pharmacology (2), the two databases assigned to different sections based on their slightly different foci. The Nucleic Acids Research online Molecular Biology Database Collection, which is available at http://www.oxfordjournals.org/nar/database/a/, retains the same 15 categories and 41 subcategories as it did before.

The current issue opens with brief overviews of the resources provided by three major bioinformatics centers, the U.S. National Center for Biotechnology Information (NCBI), the European Bioinformatics Institute (EMBL-EBI), and Swiss Institute for Bioinformatics (SIB). These papers cover the recent developments and ongoing efforts at

*To whom correspondence should be addressed. Tel: +1 301 435 5910; Fax: +1 301 435 7793; Email: nardatabase@gmail.com
Correspondence may also be addressed to Daniel J. Rigden. Tel: +44 151 795 4467; Fax: +44 151 795 4406; Email: drigden@liv.ac.uk

Published by Oxford University Press on behalf of Nucleic Acids Research 2016.
This work is written by (a) US Government employee(s) and is in the public domain in the US.
| Database name         | URL                             | Brief description                                                                 |
|-----------------------|---------------------------------|-----------------------------------------------------------------------------------|
| AgingChart            | http://www.agingchart.org/      | Pathways of age-related processes                                                   |
| Assembly              | http://www.ncbi.nlm.nih.gov/assembly | Status of whole-genome shotgun assemblies                                           |
| BacWGSTdb             | http://bacdb.org/BacWGSTdb/     | Bacterial whole genome sequence typing database                                     |
| BIGNASim              | mmmb.ir.brcarlos.org/BIGNASim/  | Molecular dynamics simulations of nucleic acids                                     |
| BrecAN-DB             | brecandb.igib.res.in/           | Breakpoint profiles of cancer genomes                                               |
| Cancer RNA-Seq        | syadb4.nchu.edu.tw/cRN          | Transcriptome profiling in cancer cells                                             |
| Nexus                 |                                 |                                                                                     |
| Caulobrowser          | http://www.caulobrowser.org     | Biology of Caulobacter crescentus                                                   |
| ccmGDB                | bioinfo.mc.vanderbilt.edu/ccmGDB/ | Cancer cell metabolism gene database                                                |
| CEGA                  | cega.ezlab.org/                 | Conserved elements from genomic alignments                                          |
| CincNet               | http://cincnet.mbc.nctu.edu.tw  | Tissue-specific expression profiles of circular RNA                                  |
| Colorectal Cancer     | http://www.colonatlas.org       | Genes and proteins of colorectal cancer cells                                       |
| Atlas                 |                                 |                                                                                     |
| CRISPRz               | research.nhgrni.nih.gov/crisprz | CRISPR single guide RNAs to zebrafish genes                                         |
| CSDB                  | csdb.glycoscience.ru/database   | Carbohydrate structure database                                                     |
| DASHR                 | issanwanglab.org/DASHR          | Database of human small non-coding RNA                                              |
| dbMAE                 | maec.hms.harvard.edu            | Database of monoallelic gene expression                                              |
| dbSUPER               | bioinfo.au.tsinghua.edu/cdbuper/ | A database of super-enhancers                                                      |
| DESM                  | www.ebcre.kaut.edu.sa/desm     | Microbial knowledge exploration systems                                              |
| DIDA                  | dida.ibsqueare.be               | DiGeorge disease database                                                           |
| Digital               | cell-lineage.org                | C. elegans development and cell differentiation                                       |
| Database              |                                 |                                                                                     |
| DMDD                  | dmdd.org.uk                     | Deciphering the mechanisms of developmental disorder                                |
| EK3D                  | http://www.ith.ac.in/EK3D/      | Capsular polysaccharide (K antigen) structures of various E. coli serotypes          |
| ENCODE DCC            | encodeproject.org               | ENCODE (Encyclopedia of DNA Elements) consortium data portal                        |
| FLOR-ID               | flor-id.org/                    | Flowering interactive database                                                      |
| GEnestATION           | genestation.org                 | Genes in gestation: genomics of pregnancy-related tissues                           |
| GlyToucan             | glytoucan.org                   | International glycan Structure Repository                                           |
| GreeNC                | greenc.sciencedesigners.com/   | Green non-coding: plant lncRNAs                                                    |
| HGTree                | hgtree.snu.ac.kr                | Horizontally transferred genes identified by tree-based methods                     |
| HPMCD                 | hpmcd.org/                      | Human pan-microbial communities database                                            |
| hPSCreg               | hpscreg.eu                      | Human pluripotent stem cell registry                                                |
| IC4R                  | ic4r.org                        | Information commons for rice                                                       |
| InsectBase            | insect-genome.com/              | Insect genomes and transcriptomes                                                    |
| InterRNA              | mfnl.org/interma/               | Base interactions in RNA structures                                                  |
| JuncDB                | juncdb.carmelab.huji.ac.il/     | Exon-exon junction database                                                         |
| Lnc2Cancer            | bio-bigdata.com/lnc2cancer/      | Human lncRNA and cancer associations                                                |
| MERAV                 | merav.wi.mit.edu               | Metabolic gene rapid visualizer                                                      |
| Metabolomics          | metabolomicsworkbench.org/      | Metabolomics data, standards and protocols                                          |
| Workbench             |                                 |                                                                                     |
| MitoAge               | mitoage.org                     | Mitochondrial DNA properties and aging                                              |
| MutationAligner       | mutationaligner.org/            | Mutation hotspots in protein domains in cancer                                       |
| NBDB                  | nbdb.bi.a-star.edu.sg           | Nucleotide binding protein motifs                                                    |
| OpenTein              | opentein.hgc.je/                | Open teratology investigation: images                                               |
| PCOSKB                | pcoskb.bicnirh.res.in/          | Polysaccharide osyndrome knowledgebase                                              |
| PDBflex               | pdbflex.org                     | Flexibility in protein structures                                                   |
| PhytoPath             | phytopathdb.org                 | Genomics of fungal, oomycete and bacterial phytopathogens                           |
| pRNAClu sterDB        | http://www.smallrnagroup-mainz.de/ | Clusters of priRNAs                                                                 |
| PlanMine              | planmine.mpi-cbg.de/            | Planarian genomics                                                                  |
| PlantDHS              | plantdhs.org                    | Plant D Nase I- hypersensitive Sites                                                |
| RBP-Var               | rbp-vari.biols.ac.cn/           | Variation that can affect RNA-protein interactions                                   |
| RMBase                | mirlab.sysu.edu.cn/rmbase/      | RNA modification database                                                            |
| RPMdb                 | sysbio.sysu.edu.cn/rpdb/        | Ribosome profiling database                                                          |
| SATPdb                | crdd.osdd.net/raghava/satpdb/   | Structurally annotated therapeutic peptides                                          |
| SBR-Blood             | sbrblood.nhgr.nih.gov/          | Systems biology repository for hematopoietic cells                                  |
| SEA                   | sea.edbc.org                    | Super enhancer archive                                                              |
| SigMol                | bioinfo.imtech.res.in/mancoj/sigmol | Quorum sensing signalling molecules                                               |
| SIGNOR                | signor.uniriona2.it/            | Signaling network open resource                                                    |
| sORFs                 | www.sorfs.org                   | Small ORFs identified by ribosome profiling                                         |
| Start2Fold            | start2fold.eu                   | Hydrogen/deuterium exchange data on protein folding and stability                   |
| SureChEMBL            | surechembl.org                  | Chemical compounds in patent documents                                              |
| SynLethDB             | histone.sce.ntu.edu.sg/SynLethDB/ | Synthetic lethality gene pairs as potential anticancer drug targets                |
Table 1. Continued

| Database name | URL | Brief description |
|---------------|-----|-------------------|
| TCGA SpliceSeq | http://projects.insilico.us.com/TCGASpliceSeq | Alternative splicing patterns in cancer cells |
| UET | http://mammoth.bcm.tmc.edu/uet/ | Universal evolutionary trace: protein motifs important for function |
| WeGET | http://coexpression.cmbi.umcn.nl/ | Weighted gene co-expression tool |
| WITHDRAWN | http://cheminfo.charite.de/withdrawn/ | Withdrawn and discontinued drugs |

Table 2. Updated description of databases most recently published elsewhere

| Database name | URL | Brief description |
|---------------|-----|-------------------|
| ANISEED | http://www.aniseed.cnrs.fr | Ascidian network for in situ expression and embryological data |
| BiGG Models | http://bigg.ucsd.edu | Biologically, genetically and genomically structured metabolic network models |
| CPPsite | http://crdd.osdd.net/taghava/cppsite/ | Validated cell penetrating peptides |
| DBAASP | http://dbaasp.org | Database of antimicrobial activity and structure of peptides |
| DGIdb | http://dgIdb.genome.wustl.edu | Drug-gene interaction database |
| iGNM | http://gnmdb.csb.pitt.edu/ | Protein functional motions based on Gaussian network model |
| IID* | http://ophid.utoronto.ca/iid | Integrated interactions database: tissue-specific protein-protein interactions |
| iPPIDB | http://www.iepibdb.cdithem.fr/ | Inhibitors of protein-protein interactions |
| KLIFS | http://klifs.vu-compmedchem.nl | Kinase-ligand interaction fingerprints and structures |
| MG-RAST | http://metagenomics.anl.gov/ | Data portal for processing, analyzing, sharing and disseminating metagenomic data sets |
| MitoCarta | http://www.broadinstitute.org/pubs/MitoCarta | Mouse and human mitochondrial proteins |
| MitoCarta | http://www.mitanetx.org | Genome-scale metabolic networks |
| MNXref/MetaNetX | | |
| MouseNet | http://www.inetbio.org/mousenet/ | Functional network of mouse genes |
| PlantPAN | http://PlantPAN2.tips.ncku.edu.tw | Plant promoter analysis navigator |
| SIDER | http://sidefetcs.embl.de/ | Side effect resource: adverse drug reactions |
| sRNATarBase* | http://ccb1.bmi.ac.cn/srnatarbase/ | sRNA-target interactions in bacteria |
| SugarBindDB | http://sugarbind.expasy.org | Host-pathogen interactions mediated by glycans |

*IID and sRNATarBase have been previously listed in the NAR Database Collection as entries nos. 897 and 1832, respectively.

these centers and provide a general introduction into their activities that should be useful for both experienced and novice users. One more introductory paper describes the web resources that are supported by ELIXIR, the European life-sciences infrastructure for biological information, and presents a listing of their providers. This ELIXIR Tools and Data Services Registry aims to be a comprehensive and consistent registry of information about (mostly) European bioinformatics databases and tools.

In addition to the annual papers from the International Nucleotide Sequence Database collaboration (INSDC), which comprises the DNA Data Bank of Japan, the European Nucleotide Archive, and GenBank, this issue introduces the NCBI’s new Assembly database (http://www.ncbi.nlm.nih.gov/assembly/), which helps track the progress of the genome assembly data in GenBank as the genome sequence progresses from a set of unordered contigs to a draft genome assembly and finally to a complete genome that includes either a single chromosome or multiple chromosomes (3).

Among newly created nucleic acid sequence resources, it is worth noting the Conserved Elements from Genomic Alignments (CEGA) database, a collection of non-coding sequences that are poorly characterized but highly conserved within various groups of vertebrates and include potential promoters, enhancers, and other regulatory elements (4), and a database of exon-exon junction sequences, aptly named JuncDB (5). Two more new databases, dbSUPER and SEA (6,7), collect the sequences of super-enhancers, the recently discovered regulatory elements that consist of clusters of transcriptional enhancers and regulate gene expression in a cell- and tissue-specific fashion (8). Other noteworthy contributions include updates on Pfam, a database of human DNA repeat families; ARESite, a resource on AU-rich elements in vertebrate UTRs; NPIDB, a nuclear-protein interaction database which proposes a new classification of DNA-protein complexes, and such popular databases of transcriptional regulation as JASPAR, HOCOMOCO, ORegAnno and RegulonDB. A potentially important new contribution is the BIGNAsim database of DNA dynamics based on molecular dynamics simulations using the ParmBSC1 force field (9). A separate block of papers features various RNA databases, including resources on 5S rRNA, tRNA, piRNA, circular RNA, long non-coding RNA and their interactions.

The protein sequence section features, among others, updates on such popular protein families databases as Pfam, PANTHER, eggNOG, GPCRdb, Transporter Classification database (TCDB), and two databases of proteases and protease inhibitors, MEROPS, which is now in its 20th year, and Degradome. The Pfam update paper deserves a particularly careful reading because it provides a detailed descrip-
tion of the recent and upcoming changes in this popular database as it attempts to cope with the rapidly increasing amount of sequence data. The authors see the solution in transitioning Pfam from attempting to incorporate the entire UniProt sequence database to focusing instead on the UniProt reference proteomes (at least for seed alignments), a much smaller set of higher-quality protein sequences (10).

With respect to protein sequence motifs, there is an update on the Eukaryotic Linear Motif (ELM) database and two new resources: the Nucleotide Binding Database (NBDB) of nucleotide-binding motifs and the Universal Evolutionary Trace (UET) database of predicted protein functional sites (11–13). The proteomics databases are represented by sORF, a collection of small ORFs identified by ribosome profiling (14), and updates on the widely used databases on proteomic peptide identification (PRIDE) and post-translational modifications (dbPTM) (15,16).

The protein structure-related papers include an update from PDB (17) reporting significant improvements to the value added to and accessibility of structure reports. A trio of papers cover different aspects of protein folding, flexibility and dynamics: Start2Fold collates experimental hydrogen/deuterium exchange data, PDBFlex provides statistics of and animations between pairs of homologous structures in the PDB, and iGNN offers improved computationally predicted flexibility information for most PDB entries (18–20). Two databases use CATH structural domain classifications to shed light on protein function, Gen3D by assigning domain annotations and associated function predictions to proteomes and FunTree by attempting to better understand the evolution of protein function in superfamilies. Finally, the biological and medicinal interest in kinases fully justifies the effort spent in updating KLIFS, a database dedicated to a detailed understanding of kinase-ligand interactions.

The next section includes updated resources on metabolic pathways, such as KEGG, MetaCyc, Reactome, WikiPathways and the Escherichia coli metabolism database (ECMDB), and databases of metabolic network modeling, such as BiGG Models and MNXref/MetaNetX. A new arrival here is the Metabolomics Workbench (21), which strives to be a one-stop repository for all kinds of metabolomics data, including metabolite standards, protocols, tutorials and analysis tools.

Coverage of organismal genome diversity is provided by the updated Ensembl Genomes and Bacterial Diversity (BacDive) databases (22,23), as well as specialized resources dedicated to Caulobacter, Pseudomonas and Bacillus subtilis (24–26). The current issue also includes updates on popular metagenomics resources, such as MG-RAST, EBI Metagenomics and probeBASE (27–29), as well as the newly compiled Human Pan-Microbe Communities database (30). Another new arrival, the bacterial whole-genome sequence typing database BacWGST, aims to simplify the important task of identifying the bacterial strains in samples isolated from infection (31).

As in previous years, this Database Issue includes a selection of genome resources for human and model organisms (Ensembl, RefSeq, UCSC Genome Browser, ENCODE portal), including yeast (SGD), Caenorhabditis elegans (WormBase), Drosophila (FlyBase), ants, bees and wasps (Hymenoptera Genome Database), cow and mouse. The new arrivals include a collection of insect genome resources and genome databases of planaria and ascidians (32–34). A very interesting entry Deciphering the Mechanisms of Developmental Disorders (DMD) database collects phenotypic data of mouse mutant embryos (35). This section also includes the database of autosomal monoallelic gene expression (dbMAE, (36)), which has been chosen by the NAR editors as one of the two Breakthrough papers in this issue. dbMAE provides manually curated data on allele-specific expression of autosomal genes, whereby the transcriptional activity of two allelles is epigenetically controlled and maintained in a clonal cell lineage, resulting in diversification of cells within the same tissue (37). dbMAE promises to become a useful resource that will help researchers achieve a better understanding of this recently emerged epigenetic phenomenon.

A significant fraction of the databases profiled in this issue (including ClinVar, GWAdb, HaplOreg and others) are dedicated to human genetic variation as it relates to disease, primarily cancer, and various aspects of drug research. These include resources on patented drugs, their side effects, withdrawn drugs, and potential drug targets (38–40). Six papers in this section present updated databases of various antimicrobial and anticancer peptides. An interesting work, also chosen by the NAR editors as a Breakthrough paper, describes the newly compiled Database of Digenic Diseases (DIDA), which collects data on such diseases as Bardet-Biedl and Kallmann syndromes that are caused by single nucleotide variants or small indels in specific pairs of genes (41).

This issue also presents updates on the widely used databases of small molecules, NCBI’s PubChem and EBI’s ChEBI, and introduces SureChEMBL, the recently created Structure Database (CSDB) and the International Glycan Structure Repository (GlyTouCan), collect knowledge and facilitate further research on these important but often-overlooked compounds (45,46). Ten papers describe various plant databases, including an update on the popular Plant Promoter Analysis Navigator (47) and Information Commons for Rice (IC4R), a compendium of Chinese databases on all aspects of rice research (48). Finally, there are three databases on mitochondrial research: MitoCarta and MitoMiner, two excellent databases of mitochondrial proteins, and MitoAge, a database of mitochondrial DNA properties from various organisms (49–51).

### UPDATED NAR ONLINE MOLECULAR BIOLOGY DATABASE COLLECTION

This year’s update of the NAR online Molecular Biology Database Collection (which is freely available at [http://www.oxfordjournals.org/nar/database/e/](http://www.oxfordjournals.org/nar/database/e/)) involved inclusion of 62 new databases (Table 1) and 15 databases that have been previously described elsewhere and were not part of this Collection (Table 2). In addition, the Collection has been expanded by including such databases as Integrative Cancer Genomics (IntOGen) and Disease Variant Store (DIVAS) (52,53). Our curation checks revealed 121 non-responsive databases, of which 23 obsolete entries have been removed.
from the Collection and the rest marked for potential removal next year. In addition, 26 entries in the Collection have been updated with respect to their URLs, descriptions, and/or author contact information.

We welcome suggestions for inclusion in the Collection of additional databases that have been published in other journals. Such suggestions should be addressed to XMFS at xose.m.fernandez@gmail.com and should include database summaries in plain text, organized in accordance with the http://www.oxfordjournals.org/nar/database/summary/1 template.

ACKNOWLEDGEMENTS

We thank NAR Editorial Administrator Dr Martine Bernardes-Silva and the Oxford University Press team led by Jennifer Boyd and Caoimhe Ni Dhonáill for their great efforts in compiling this issue.

FUNDING

The NIH Intramural Research Program at the National Library of Medicine [to M.Y.G.]. The open access publication charge for this paper has been waived by Oxford University Press - NAR.

Conflict of interest statement. The authors’ opinions do not necessarily reflect the views of their respective institutions. XMFS is an employee of Thermo Fisher Scientific Inc.

REFERENCES

1. Isberg,V., Mordalski,S., Munk,C., Rataj,K., Harpsoe,K., Hauser,A.S., Vrøling,B., Bojarai,A.J., Friend,G. and Gloriam,D.E. (2016) GPCRDb: an information system for G protein-coupled receptors. Nucleic Acids Res., 44, doi:10.1093/nar/gkw1178.
2. Southan,C., Sharman,J.L., Benson,H.E., Faccenda,E., Pawson,A.J., Alexander,S.P., Buneman,O.P., Davenport,A.P., McGrath,J.C., Peters,J.A. et al. (2016) The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: towards curated quantitative interactions between 1300 protein targets and 6000 ligands. Nucleic Acids Res., 44, doi:10.1093/nar/gkt1037.
3. Kitts,P.A., Church,D.M., Thibaud-Nissen,F., Choi,J., Hem,V., Sapojnikov,V., Smith,R.G., Tatusova,T., Xiang,C., Zherikov,A. et al. (2016) Assembly: a resource for assembled genomes at NCBI. Nucleic Acids Res., 44, doi:10.1093/nar/gkx1226.
4. Dousse,A., Junier,T. and Zdobnov,E.M. (2016) CEGA-a catalog of conserved elements from genomic alignments. Nucleic Acids Res., 44, doi:10.1093/nar/gkx1163.
5. Chorev,M., Guy,L. and Carmel,L. (2016) JuncDB: an exon-exon junction database. Nucleic Acids Res., 44, doi:10.1093/nar/gkv1142.
6. Khan,A. and Zhang,X. (2016) dbSUPER: a database of super-enhancers in mouse and human genome. Nucleic Acids Res., 44, doi:10.1093/nar/gkv1002.
7. Wei,Y., Zhang,S., Zhang,B., Li,S., Wang,X., Wang,F., Wu,Q., Liu,H. et al. (2016) SEA: a super-enhancer archive. Nucleic Acids Res., 44, doi:10.1093/nar/gkv1243.
8. Downen,J.M., Fan,Z., Hnisz,D., Ren,G., Abraham,B.J., Zhang,L.N., Weintraub,A.S., Schuierer,J., Lee,T.I., Zhao,K. et al. (2014) Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell, 159, 374–387.
9. Hospital,A., Andriolo,P., Cugnasco,C., Codo,L., Becerra,Y., Dans,P.D., Battistini,F., Gorji,R., Orozco,M. et al. (2016) BIGNASim: A NoSQL database structure and analysis portal for nucleic acids simulation data. Nucleic Acids Res., 44, doi:10.1093/nar/gkv1301.
10. Finn,R.D., Coghill,L., Eberhardt,R.Y., Mistry,J., Mitchell,A., Potter,S.C., Qureshi,M., Sangrador-Vegas,A., Salazar,G.A. et al. (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res., 44, doi:10.1093/nar/gkv1344.
11. Dinkel,H., Van Roey,K., Michael,S., Kumar,M., Uyar,B., Altenberg,B., Milchevskaya,V., Schneider,M., Kühn,H., Behrendt,A. et al. (2016) ELM 2016 - data update and new functionality of the eukaryotic linear motif resource. Nucleic Acids Res., 44, doi:10.1093/nar/gkw1291.
12. Zheng,Z., Goncarenco,A. and Berezovsky,I.N. (2016) Nucleotide binding database NBDB - a collection of sequence motifs with specific protein-ligand interactions. Nucleic Acids Res., 44, doi:10.1093/nar/gkw1124.
13. Luo,R.C., Wilson,S.J., Konecki,D.M., Wilkins,A.D., Venner,E., Morgan,D.H. and Lichtarge,O. (2016) UET: A database of evolutionarily-predicted functional determinants of protein sequences that cluster as functional sites in protein structures. Nucleic Acids Res., 44, doi:10.1093/nar/gkv1279.
14. Olexiuk,V., Crappe,J., Verbruggen,S., Verhegen,K., Martens,L. and Menschaert,G. (2016) sORFs.org: a repository of small ORFs identified by ribosome profiling. Nucleic Acids Res., 44, doi:10.1093/nar/gkv1175.
15. Vizzaino,J.A., Csordas,A., Del Toro,N., Di Nola,G., Grisse,J., Cavaliere,M. et al. (2016) A NoSQL database structure and analysis portal for nucleic acids simulation data. Nucleic Acids Res., 44, doi:10.1093/nar/gkv1227.
16. Huang,K.Y., Su,M.G., Kao,H.J., Hsieh,Y.C., Hsiong,J.H., Cheng,K.H., Huang,H.D. and Lee,T.Y. (2016) dbPTM 2016: 10-year anniversary of a resource for post-translational modification of proteins. Nucleic Acids Res., 44, doi:10.1093/nar/gkw1240.
17. Velankar,S., van Ginkel,G., Alrouby,Y., Battle,G.M., Berriford,J.M., Conroy,M.J., Dana,J.M., Gore,S.P., Gutmanas,A., Haslam,P. et al. (2016) PDB: improved accessibility of macromolecular structure data from PDB and EMDB. Nucleic Acids Res., 44, doi:10.1093/nar/gkov1047.
18. Pancsa,R., Varadi,M., Tompa,P. and Vranken,W.F. (2016) Start2Fold: a database of hydrogen/deuterium exchange data on protein folding and stability. Nucleic Acids Res., 44, doi:10.1093/nar/gkv1185.
19. Li,H., Chang,Y., Yang,L.W. and Bahar,I. (2016) iGGM 2.0: the Gaussian network model database for biomolecular structural dynamics. Nucleic Acids Res., 44, doi:10.1093/nar/gkv1236.
20. Hrabc,T., Li,Z., Sedova,M., Rigkiewicz,P., Casareski,L. and Godzik,A. (2016) PDBeFlex: exploring flexibility in protein structures. Nucleic Acids Res., 44, doi:10.1093/nar/gk1316.
21. Sud,M., Fahy,E., Cotter,D., Azam,K., Vadivelu,I., Burant,C., Edison,A., Fiehn,O., Higashi,R., Nair,K.S. et al. (2016) Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. Nucleic Acids Res., 44, doi:10.1093/nar/gkx1042.
22. Kersey,P.J., Allen,J.E., Armeen,I., Boddup,S., Bolt,B.J., Carvalho-Silva,D., Christensen,M., Davis,P., Fain,L.J., Grabmauler,C. et al. (2016) Ensembl Genomes: more genomes, more complexity. Nucleic Acids Res., 44, doi:10.1093/nar/gkv1209.
23. Sönhgen,C., Bunk,B., Podstawka,A., Klei,D. and Overmann,J. (2014) BacDive – The Bacterial Diversity metadata database. Nucleic Acids Res., 42, doi:10.1093/nar/gkt1058.
24. Lasker,K., Schrader,J.M., Men,Y., Marahsi,T., Dill,D.L., McAdams,H.H. and Shapiro,L. (2016) CauliBrowser: a systems biology resource for Caulobacter crescentus. Nucleic Acids Res., 44, doi:10.1093/nar/gkv1050.
25. Winsor,G.L., Griffiths,E.J., Lo,R., Dhillon,B.K., Shay,J.A. and Brinkman,F.S. (2016) Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database. Nucleic Acids Res., 44, doi:10.1093/nar/gkw1227.
26. Michna,R.H., Zhu,B., Mader,U. and Stulke,J. (2016) SubtiWiki, a super-enhancer archive. Nucleic Acids Res., 44, doi:10.1093/nar/gkv1322.
27. Mitchell,A., Bucchin,F., Cochrane,G., Denise,H., ten Hoopen,P., Frasset,P., Potier,S., Scheremetjew,M., Sterk,P. et al. (2016) EBI Metagenomics in 2016 - an expanding and evolving database. Nucleic Acids Res., 44, doi:10.1093/nar/gkv1301.
resource for the analysis and archiving of metagenomic data. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1195.

29. Greuter.D, Loy.A., Horn.M. and Rattei.T. (2016) probeBase - an online resource for rRNA-targeted oligonucleotide probes and primers: new features 2016. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1232.

30. Forster,S.C., Browne,H.P., Kumar,N., Hunt,M., Denise,H., Mitchell,A., Finn,R.D. and Lawley,T.D. (2016) HPMCd: the database of human microbial communities from metagenomic datasets and microbial reference genomes. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1216.

31. Ruan.Z. and Feng.Y. (2016) BacWGStdb, a database for genotyping and source tracking bacterial pathogens. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1004.

32. Yin.C., Shen.G., Guo.D., Wang.S., Ma.X., Xiao.H., Liu.J., Zhang.Z., Liu.Y., Zhang.Y. et al. (2016) InsectBase: a resource for insect genomes and transcriptsomes. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1204.

33. Brozovic,M., Martin,C., Dantec.C., Dauga,D., Mendez,M., Simion,P., Percher.M., Laporte,B., Scornavacca,C., Gregorio,A. et al. (2016) ANISEED 2015: a digital framework for the comparative developmental biology of ascidians. Nucleic Acids Res. 44, doi:10.1093/nar/gkv966.

34. Brandl,H., Moon,H., Vila-Farre,M., Liu,S.Y., Henry.I. and Rink,J.C. (2016) PlanMine - a mineable resource of planarian biology and biodiversity. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1148.

35. Wilson.R., McGuire.C. and Mohun.T. (2016) Deciphering the mechanisms of developmental disorders: phenotype analysis of embryos from mutant mouse lines. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1138.

36. Savova,V., Patsenker,J., Vigneau.S. and Gimelbrant,A.A. (2016) dbMAE: the database of autosomal monallelic expression. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1106.

37. Savova,V., Vigneau,S. and Gimelbrant,A.A. (2013) Autosomal monallelic expression: genetics of epigenetic diversity? Curr. Opin. Genet. Dev., 23, 642–648.

38. Siramshetty,V.B., Nickel,J., Omiecinski,C., Gohlke,B.O., Drwal,M.N. and Preisner.R. (2016) WITHDRAWN-a resource for withdrawn and discontinued drugs. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1192.

39. Kuhn,M., Letunic,I., Jensen,L.J. and Bork.P. (2016) The SIDER database of drugs and side effects. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1075.

40. Gilson,M.K., Liu,T., Baituluk,M., Nicola.G., Hwang.L. and Chong.J. (2016) BindingDB in 2015: a public database for medicinal chemistry, computational biology and systems pharmacology. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1072.

41. Gazzo,A.M., Dancel,D., Cilia,E., Bonduelle,M., Abramowicz,M., Van Dooren,S., Smits,G. and Lenaerts,T. (2016) DIDA: A curated and annotated digenic diseases database. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1068.

42. Kim,S., Thiessen,P.A., Bolton,E.E., Chen,J., Fu,G., Gindulyte,A., Han,L., He,J., He.S., Shoemaker,B.A. et al. (2016) PubChem Substance and Compound databases. Nucleic Acids Res. 44, doi:10.1093/nar/gkv951.

43. Hastings,J., Owen,G., Dekker.A., Ennis,M., Kale,N., Muthukrishnan,V., Turner.S., Swainston,N., Mendes,P. and Steinbeck.C. (2016) ChEMBL in 2016: Improved services and an expanding collection of metabolites. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1031.

44. Fersht,A.R., Chapter 5. The structure and stability of proteins. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1253.

45. Toukach,P.V. and Egorova,K.S. (2016) Carbohydrate Structure Database merged from bacterial, archaeal, plant and fungal parts. Nucleic Acids Res. 44, doi:10.1093/nar/gkv640.

46. Aoki-Kinoshita,K., Agrawal,S., Aoki,N.P., Arpinar,S., Cummings,R.D., Fujita,A., Fujita,N., Hart,G.M., Haslam,S.M., Kawasaki,T. et al. (2016) GlyTouCan 1.0 - The international glycan structure repository. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1041.

47. Chow,C.N., Zheng,H.Q., Wu,N.Y., Chien,C.H., Huang,H.D., Lee,T.Y., Chang-Hsieh,Y.F., Hou,P.F., Yang,T.Y. and Chang,W.C. (2016) PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1015.

48. IC4R Project Consortium. (2016) Information Commons for Rice (IC4R). Nucleic Acids Res. 44, doi:10.1093/nar/gkv1141.

49. Calvo,S.E., Clauser,K.R. and Mootha,V.K. (2016) MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1003.

50. Smith,A.C. and Robinson,A.J. (2016) MitoMiner v3.1, an update on the mitochondrial proteomics database. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1001.

51. Toren,D., Barzilay,T., Tacutu,R., Lehmann,G., Muradian,K.K. and Fraifeld,V.E. (2016) MitoAge: a database for comparative analysis of mitochondrial DNA, with a special focus on animal longevity. Nucleic Acids Res. 44, doi:10.1093/nar/gkv1187.

52. Perez-Llamas,C., Gundem,G. and Lopez-Bigas,N. (2011) Integrative cancer genomics (IntOGen) in Biomart. Database (Oxford), 2011, bar039.

53. Cheng,W.Y., Hakenberg,J., Li,S.D. and Chen,R. (2015) DIVAS: a centralized genetic variant repository representing 150 000 individuals from multiple disease cohorts. Bioinformatics, doi:10.1093/bioinformatics/btv511.