Published online 6 December 2013 *Nucleic Acids Research, 2014, Vol. 42, Database issue* ***D1–D6***
doi:10.1093/nar/gkt1282

The 2014 *Nucleic Acids Research* Database Issue and an updated NAR online Molecular Biology Database Collection

Xosé M. Fernández-Suárez[1], Daniel J. Rigden[2] and Michael Y. Galperin[3,*]

[1]Life Technologies, Inchinnan Business Park, Paisley PA4 9RF, UK, [2]Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK and [3]National Center for Biotechnology Information (NCBI), National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

Received November 17, 2013; Accepted November 18, 2013

ABSTRACT

The 2014 Nucleic Acids Research Database Issue includes descriptions of 58 new molecular biology databases and recent updates to 123 databases previously featured in *NAR* or other journals. For convenience, the issue is now divided into eight sections that reflect major subject categories. Among the highlights of this issue are six databases of the transcription factor binding sites in various organisms and updates on such popular databases as CAZy, Database of Genomic Variants (DGV), dbGaP, DrugBank, KEGG, miRBase, Pfam, Reactome, SEED, TCDB and UniProt. There is a strong block of structural databases, which includes, among others, the new RNA Bricks database, updates on PDBe, PDBsum, ArchDB, Gene3D, ModBase, Nucleic Acid Database and the recently revived iPfam database. An update on the NCBI's MMDB describes VAST+, an improved tool for protein structure comparison. Two articles highlight the development of the Structural Classification of Proteins (SCOP) database: one describes SCOPe, which automates assignment of new structures to the existing SCOP hierarchy; the other one describes the first version of SCOP2, with its more flexible approach to classifying protein structures. This issue also includes a collection of articles on bacterial taxonomy and metagenomics, which includes updates on the List of Prokaryotic Names with Standing in Nomenclature (LPSN), Ribosomal Database Project (RDP), the SILVA/LTP project and several new metagenomics resources. The NAR online Molecular Biology Database Collection, http://www.oxfordjournals.org/nar/database/c/, has been expanded to 1552 databases. The entire Database Issue is freely available online on the Nucleic Acids Research website (http://nar.oxfordjournals.org/).

NEW AND UPDATED DATABASES

The 21st annual Nucleic Acids Research Database Issue is the largest ever. It includes 185 articles that provide (i) descriptions of the database resources at the NCBI, European Bioinformatics Institute (EBI) and the US Department of Energy Joint Genome Institute (JGI); (ii) 58 new molecular biology databases (Table 1); (iii) updates on 100 databases previously featured in NAR; and (iv) updated descriptions of 23 databases that had been previously described in other journals (Table 2). For the past several years, the order of articles in the Database Issue reflected the categorization of the databases in the NAR online Molecular Biology Database Collection (http://www.oxfordjournals.org/nar/database/c/). Acting on the advice of many readers, we have now made the categories visible and divided the entire Database Issue into the following eight sections: (i) nucleic acid sequence and structure, transcriptional regulation; (ii) protein sequence and structure, motifs and domains, protein–protein interactions; (iii) metabolic and signalling pathways, enzymes, protein modification; (iv) viruses, bacteria, protozoa and fungi; (v) human genome, model organisms, comparative genomics; (vi) genomic variation, diseases and drugs; (vii) plant databases; and (viii) other molecular biology databases. Although each of these sections unifies several of the categories and/or subcategories of the NAR online Database Collection, we believe that they provide an easy-to-use guide to navigate this huge volume and help placing related databases next to each other.

The first section, in addition to the annual descriptions of GenBank, the European Nucleotide Archive and the DNA Data Bank of Japan, includes update papers on

*To whom correspondence should be addressed. Tel: +1 301 435 5910; Fax: +1 301 435 7793; Email: nardatabase@gmail.com or galperin@ncbi.nlm.nih.gov

Table 1. Descriptions of new online databases in the 2014 NAR Database issue

Database name	URL	Brief description
1000 Genomes Selection Browser	http://hsb.upf.edu	Signatures of selection in the human genomes
AgeFactDB	http://agefactdb.jenage.de	Ageing Factors, phenotypes and lifespan data
AVPdb	http://crdd.osdd.net/servers/avpdb	A database of experimentally validated AntiViral Peptides
BacDive	http://bacdive.dsmz.de	Bacterial Diversity metadatabase
BacMet	http://bacmet.biomedicine.gu.se	Antibacterial biocide and Metal resistance Genes
BloodChIP	http://149.171.101.136/python/BloodChIP	Transcription factor binding profiles in human haematopoietic stem/progenitor cells
bNAber	http://bnabs.org	A database of broadly Neutralizing HIV-1 Antibodies
CellFinder	http://www.cellfinder.org	Gene and protein expression, phenotype and images mapped to the cell types
ClinVar	http://www.ncbi.nlm.nih.gov/clinvar	Genomic Variation of potential Clinical importance
CollecTF	http://collectf.umbc.edu	Collection of verified bacterial Transcription Factor binding sites
CR Cistrome	http://compbio.tongji.edu.cn/cr	Chromatin Regulators and histone modifications in human and mouse
dbPSHP	http://jjwanglab.org/dbpshp	A database of recent Positive Selection across Human Populations
DPRP	http://syslab.nchu.edu.tw/DPRP	Phenotype-specific Regulatory Programs derived from TF binding data
DriverDB	http://ngs.ym.edu.tw/driverdb/	Cancer driver genes/mutations deduced from cancer exome-seq results
EBI metagenomics	https://www.ebi.ac.uk/metagenomics/	An automated pipeline for the analysis and archiving of metagenomic data
EKPD	http://ekpd.biocuckoo.us	Eukaryotic protein Kinase and Phosphatase Database
ExoLocator	http://exolocator.eopsf.org	Protein-coding exons from complete vertebrate genomes
GoMapMan	http://www.gomapman.org	Unified plant-specific gene ontology
GWIPS-viz	http://gwips.ucc.ie	Genome-Wide Information on Protein Synthesis *in vivo* using ribosome profiling
Hemolytik	http://crdd.osdd.net/raghava/hemolytik	Haemolytic and non-haemolytic peptides
HoPaCI-DB	http://mips.helmholtz-muenchen.de/HoPaCI/	Host–Pathogen Interactions of *Pseudomonas aeruginosa* and *Coxiella* spp.
HRaP	http://bioinfo.protres.ru/hrap	HomoRepeats and Patterns
InvFEST	http://invfestdb.uab.cat	Polymorphic inversions in the human genome
IUPHAR/BPS guide to pharmacology	http://www.guidetopharmacology.org	Properties of established and potential drug targets: GPCRs, ion channels, nuclear hormone receptors, catalytic receptors, transporters and enzymes
LenVarDB	http://caps.ncbs.res.in/lenvardb	Length Variation in protein domains
LoQAtE	http://www.weizmann.ac.il/molgen/loqate	Localization and Quantitation Atlas of the yeast proteome
Lynx	http://lynx.ci.uchicago.edu	Genomic and clinical data on complex heritable disorders
Manteia	http://manteia.igbmc.fr	Embryonic development of the mouse, chicken, zebrafish and human
MCDRiceProt	http://www.genomeindia.org/biocuration	Manually Curated Database of Rice Proteins
MetaRef	http://metaref.org	Reference clade-specific microbial genes for Metagenomic studies
MitoBreak	http://mitobreak.portugene.com	Mitochondrial DNA Breakpoints in human, mouse and rat
MP:PD	http://proteinformatics.charite.de/mppd	Membrane Proteins: Packing Densities, packing defects and internal water molecules
MultiTaskDB	http://wallace.uab.es/multitask	Moonlighting proteins database
mVOC	http://bioinformatics.charite.de/mvoc	Microbial Volatile Organic Compounds
NECTAR	http://cardiodb.org/nectar	Disease-related non-synonymous mutations
Network Portal	http://networks.systemsbiology.net	A database of gene transcription regulatory networks
NeXO	http://nexontology.org/	Network Extracted gene Ontology database
NHGRI GWAS Catalog	http://www.genome.gov/gwastudies, http://www.ebi.ac.uk/fgpt/gwas	A catalog of published Genome-Wide Association Studies, maintained at the NHGRI and EBI
OnTheFly	http://bhapp.c2b2.columbia.edu/OnTheFly	DNA-binding specificities of transcription factors in *Drosophila*
pE-DB	http://pedb.vib.be	Protein Ensemble DataBase: ensembles of intrinsically disordered and unfolded proteins
P-MITE	http://pmite.hzau.edu.cn/django/mite	Plant Miniature Inverted-repeat Transposable Elements (MITEs)
POGO-DB	http://pogo.ece.drexel.edu	Pairwise comparisons Of Genomes and universal Orthologous genes
PortEco	http://porteco.org	*Escherichia coli* K-12 knowledgebase Portal
RADAR	http://rnaedit.com	A Rigorously Annotated Database of A-to-I RNA editing
RepeatsDB	http://repeatsdb.bio.unipd.it	Repeats in protein structures
RhizoBase	http://genome.microbedb.jp/rhizobase	Manually curated annotations for rhizobial genomes
RiceWiki	http://ricewiki.big.ac.cn	Wiki-based open-content platform for community curation of rice genes
RNA Bricks	http://iimcb.genesilico.pl/rnabricks	RNA structural modules and their interactions
rSNPBase	http://rsnp.psych.ac.cn	Annotated SNPs within regulatory DNA elements
SAbDab	http://opig.stats.ox.ac.uk/webapps/sabdab	Structural Antibody Database
SMMRNA	http://www.smmrna.org	Small Molecule inhibitors of RNA
SporeWeb	http://sporeweb.molgenrug.nl	Regulatory pathways during the sporulation cycle of *Bacillus subtilis*
SuperPain	http://bioinformatics.charite.de/superpain	Compounds that stimulate or relieve pain
TFBSshape	http://rohslab.cmb.usc.edu/TFBSshape	DNA shape features of Transcription Factor Binding Sites
TISdb	http://tisdb.human.cornell.edu	Alternative Translation Initiation Sites
Transformer	http://bioinformatics.charite.de/transformer	Biotransformation of drugs and food ingredients by human enzymes
uORFdb	http://cbdm.mdc-berlin.de/tools/uorfdb	Upstream ORFs and their effect of translation of downstream CDSs
WormQTL^{HD}	http://www.wormqtl-hd.org	Links from human disease to natural variation data in *Caenorhabditis elegans*

Table 2. Updated descriptions of online databases not previously featured in the NAR Database issue

Database name	URL	Brief description
COLOMBOS	http://colombos.net	Collections of Microarrays for Bacterial Organisms
Consensus CDS	http://www.ncbi.nlm.nih.gov/projects/CCDS	A collaborative effort to identify a core set of human protein coding regions
CottonGen	http://www.cottongen.org	Cotton Genomics, genetics and breeding
Database of Genomic Variants	http://dgv.tcag.ca/dgv/app/home	Curated catalog of human genomic structural variation
dbGaP	http://www.ncbi.nlm.nih.gov/gap	Database of Genotypes and Phenotypes
DECIPHER	http://decipher.sanger.ac.uk	Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources
GEISHA	http://geisha.arizona.edu	*Gallus* Expression In Situ Hybridization Analysis
GeneProf	http://www.geneprof.org	Human and mouse gene expression data from RNA-seq and ChIP-seq
GGBN	http://www.ggbn.org/dataportal	The Global Genome Biodiversity Network portal
HMDD	http://cmbi.bjmu.edu.cn/hmdd	Human MicroRNA and Disease Associations Database
Human Phenotype Ontology	http://www.human-phenotype-ontology.org	Standardized vocabulary of phenotypic abnormalities in human disease
IMPC	http://www.mousephenotype.org	International Mouse Phenotyping Consortium portal
iPfam	http://www.ipfam.org	Protein family interactions mapped to Pfam domains
KBDOCK	http://kbdock.loria.fr	Knowledge-Based Docking: protein domain interfaces
Locus Reference Genomic	http://www.lrg-sequence.org	Each LRG is a stable genomic DNA sequence for a region of the human genome
LPSN	http://www.bacterio.net	List of Prokaryotic names with Standing in Nomenclature
NDB	http://ndbserver.rutgers.edu	Nucleic Acid DataBase, nucleic acids-containing structures
Plasma Proteome Database	http://www.plasmaproteomedatabase.org	Quantitative information on proteins in human plasma and serum
Progenetix	http://www.progenetix.org	Copy number abnormalities in human cancer
SEED	http://pubseed.theseed.org or http://www.theseed.org	Genome annotations based on curated functional systems
SFLD	http://sfld.rbvi.ucsf.edu	Structure-Function Linkage Database: sequence conservation in enzyme superfamilies
SoyKB	http://soykb.org	Soybean Knowledge Base
Virus variation	http://www.ncbi.nlm.nih.gov/genomes/VirusVariation	Variation data on influenza, dengue and West Nile viruses
YeastNet	http://www.inetbio.org/yeastnet	Functional gene networks for *Saccharomyces cerevisiae*

five microRNA databases: miRBase, miRNEST, mirTarBase, PolymiRTS and starBASE, and on the NONCODE database of various types of non-coding RNA. There are also several databases of transcription factor (TF) binding sites (TFBSs), including updates on JASPAR and YEASTRACT and new databases of TFBS in *Escherichia coli*, *Drosophila* and human haematopoietic stem cells (1–5). An interesting work, chosen by the reviewers and NAR editors as a 'Breakthrough paper', describes TFBSshape (6), a database of DNA structural features (minor groove width, roll, propeller twist and helix twist) of TFBSs for various TFs, which have been collected from the JASPAR (1) and UniProbe (7) databases. The TFBSshape website, http://rohslab.cmb.usc.edu/TFBSshape/, allows the users to submit their own aligned TFBS sequences, which could be used, for example, to compare the DNA binding specificities of closely related TFs (6).

The protein database section includes annual updates on UniProt and KEGG, as well as updates of such popular databases as Pfam, eggNOG, ELM, FireDB, SEED, SIMAP and Transporter protein Classification DataBase (TCDB). Two new databases, HRaP and RepeatsDB, collect information on protein repeats, the former at the sequence level (runs of the same amino acid residue) and the latter at the structural level (8,9).

As in previous years, this Database Issue features an impressive selection of structural databases. Two of them deal with nucleic acid structure: an update on the well-known Nucleic Acids Database (NDB) and RNA Bricks, a new database of RNA 3D motifs and their contacts (10,11). The block of protein structure databases includes, among others, updates on Protein Data Bank in Europe (PDBe), PDBsum, ArchDB, Gene3D, ModBase, SCOP and the recently revived iPfam database. Diverse improvements at PDBe include comprehensive visualization and analysis of the rapidly growing collection of electron microscopy-derived structures, whereas PDBsum now offers facilities to browse domain architectures and new connections to ligand and SNP data (12). For the past 18 years, the NCBI's Molecular Modeling Database (MMDB) displayed the lists of structural neighbours of a given protein, calculated using the Vector Alignment Search Tool (VAST) (13). The MMDB update paper describes VAST+, a recent extension of that tool, which allows calculation of structural similarity at the level of 'biological assemblies' (hetero- or homo-oligomeric protein complexes). Accordingly, for macromolecular complexes, MMDB now displays precalculated lists of similar protein complexes ranked by the extent of similarity (14). Several databases, including iPfam, 3did and UniHI, reflect current interest in the structural basis of protein interaction networks and take on the challenge of presenting complex protein–protein and protein–ligand interaction data in clear and useful ways (15–17). The aptly named Negatome database (18) provides a useful

benchmark, documenting protein pairs that definitely do not interact, and could be used as negative control for the constantly growing protein 'interactome'. A pair of papers published back-to-back highlight two different directions in the development of the Structural Classification of Proteins (SCOP) database: one of them describes SCOPe, an extension of SCOP that focuses on regularly and automatically assigning new structures to the existing SCOP hierarchy, whereas the other one describes the birth of SCOP2, with its more flexible graph-based approach to classifying protein structures and documenting the subtleties of their relationships (19,20).

The section on enzymes and metabolism includes updates on three metabolic pathway databases, MetaCyc, Reactome and the Small Molecule Pathway Database (21–23). This issue also features updates of two excellent databases of the active sites in various enzyme superfamilies, the Catalytic Site Atlas and the Structure-Function Linkage Database (SFLD) (24,25). There are also updates of the Carbohydrate-Active enzymes database (CAZy) and the protease database MEROPS, as well as new databases: EKPD, a database of eukaryotic protein kinases and protein phosphatases, and MultiTaskDB, a database of 'moonlighting' enzymes (26–29).

The increased interest in microbial genomics, fuelled in part by the Human Microbiome Project, led to several important developments in database construction. Many databases now emphasize improved genome annotation for a variety of microbes, including human pathogens (IMG, PATRIC, SEED), and for selected free-living microorganisms (CyanoBase, PortEco, Rhizobase, SubtiWiki). A number of databases, such as JGI'S IMG/M (30) and the newly created EBI metagenomics resource (31), strive to capture microbial diversity in natural environments. The rapid growth of completely or partially sequenced microbial genomes makes particularly important their proper classification, which increasingly relies on such resources as the Ribosomal Database Project (RDP), the SILVA/LTP project, BacDive at the DSMZ-German Collection of Microorganisms and Cell Cultures and the List of Prokaryotic Names with Standing in Nomenclature (LPSN) (32–35). The new MetaRef database collects from reference microbial genomes clade-specific genes that could be useful for taxonomic assignments of metagenomic reads (36).

One of the major highlights of this issue is the block of articles on the improved annotation of human genome and detailed analysis of genome variation and its potential clinical significance. These articles include, among others, updates on the Consensus CDS project, a collaborative effort to identify a core set of human protein-coding regions, and on the dbGaP, a database of genotyping results and related clinically relevant phenotypes (37,38). dbGaP contains openly available study data but requires pre-authorization for access to personal health information, such as individual-level data including phenotypic data tables and genotypes (see http://www.ncbi.nlm.nih.gov/gap for details). This issue also includes descriptions of several related databases: Locus Reference Genomic, a set of reference sequences for reporting of clinically relevant sequence variants; the NCBI's ClinVar, a database documenting these clinically relevant sequence variants; the NHGRI GWAS Catalog, a curated resource of SNP-trait associations; Sanger Institute's DECIPHER, a database of pathogenic single nucleotide variants, indels and copy-number variants; and the Database of Genomic Variants (DGV) at the Toronto's Centre for Applied Genomics (39–43). There are also several more specialized databases (canSAR, DriverDB, FINDBase, HbVar, Lynx, NECTAR, Progenetix) that cover genetic defects leading to various human diseases, including cancer. In addition, three separate databases, Selectome, dbPSHP and 1000 Genomes Selection Browser, report the sites of likely positive selection in human genomes.

Model organism databases featured in this issue include regular updates from *Saccharomyces* Genome Database (SGD), WormBase, FlyBase, Mouse Genome Database (MGD), Mouse Gene Expression Database, Mouse Phenome Database and Vertebrate Genome Annotation (VEGA) database, as well as a description of the International Mouse Phenotyping Consortium (IMPC) web portal.

NAR ONLINE MOLECULAR BIOLOGY DATABASE COLLECTION

The NAR online Molecular Biology Database Collection (freely available at http://www.oxfordjournals.org/nar/database/a/) has been updated by including the databases introduced in the 2014 NAR Database Issue. This list has been expanded by including such databases as CREDO, DoSA, DBATE, RedoxDB and TMBB-DB, whose descriptions had been published in our sister journals Bioinformatics and Database (Oxford) and are freely available online, as well as selected databases published elsewhere. The entire collection has been carefully curated by checking all non-responsive database websites; coordinators of such databases have been asked to confirm their commitment to maintaining their resources. Based on the received responses (or a lack thereof), URLs of 193 databases have been corrected and 24 obsolete databases have been removed from the list. As a result of these changes, the online collection now includes 1552 databases that are sorted into 14 categories and 41 subcategories.

Suggestions for inclusion in the collection of additional databases are welcome. They should include extended databases summaries in plain text, generally formatted according to the http://www.oxfordjournals.org/nar/database/summary/1 template, including references to the published database descriptions freely available online, and should be addressed to XMFS at xose.m.fernandez@gmail.com.

ACKNOWLEDGEMENTS

The authors thank Sir Richard Roberts and Drs Alex Bateman and David Landsman for helpful comments and Dr Martine Bernardes-Silva and the Oxford University Press team led by Jennifer Boyd and Oliver Barham for their help in compiling this issue.

FUNDING

NIH Intramural Research Program at the National Library of Medicine (to M.Y.G.). Funding for open access charge: Waived by Oxford University Press.

Conflict of interest statement. The authors' opinions do not necessarily reflect the views of their respective institutions. X.M.F.S. is an employee of Life Technologies.

REFERENCES

1. Mathelier,A., Zhao,X., Zhang,A.W., Parcy,F., Worsley-Hunt,R., Arenillas,D.J., Buchman,S., Chen,C.Y., Chou,A., Ienasescu,H. *et al*. (2014) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res*., **42**, D142–D147.
2. Teixeira,M.C., Monteiro,P.T., Guerreiro,J.F., Goncalves,J.P., Mira,N.P., Dos Santos,S.C., Cabrito,T.R., Palma,M., Costa,C., Francisco,A.P. *et al*. (2014) The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic Acids Res*., **42**, D161–D166.
3. Kilic,S., White,E.R., Sagitova,D.M., Cornish,J.P. and Erill,I. (2014) CollecTF: a database of experimentally validated transcription factor-binding sites in Bacteria. *Nucleic Acids Res*., **42**, D156–D160.
4. Shazman,S., Lee,H., Socol,Y., Mann,R. and Honig,B. (2014) OnTheFly: a database of *Drosophila melanogaster* transcription factors and their binding sites. *Nucleic Acids Res*., **42**, D167–D171.
5. Chacon,D., Beck,D., Perera,D., Wong,J.W. and Pimanda,J.E. (2014) BloodChIP: a database of comparative genome-wide transcription factor binding profiles in human blood cells. *Nucleic Acids Res*., **42**, D172–D177.
6. Yang,L., Zhou,T., Dror,I., Mathelier,A., Wasserman,W.W., Gordan,R. and Rohs,R. (2014) TFBSshape: a motif database for DNA shape features of transcription factor binding sites. *Nucleic Acids Res*., **42**, D148–D155.
7. Robasky,K. and Bulyk,M.L. (2011) UniPROBE, update 2011: expanded content and search tools in the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res*., **39**, D124–D128.
8. Lobanov,M.Y., Sokolovskiy,I.V. and Galzitskaya,O.V. (2014) HRaP: database of occurrence of HomoRepeats and Patterns in proteomes. *Nucleic Acids Res*., **42**, D273–D278.
9. Di Domenico,T., Potenza,E., Walsh,I., Parra,R.G., Giollo,M., Minervini,G., Piovesan,D., Ihsan,A., Ferrari,C., Kajava,A.V. *et al*. (2014) RepeatsDB: a database of tandem repeat protein structures. *Nucleic Acids Res*., **42**, D352–D357.
10. Coimbatore Narayanan,B., Westbrook,J., Ghosh,S., Petrov,A.I., Sweeney,B., Zirbel,C.L., Leontis,N.B. and Berman,H.M. (2014) The Nucleic Acid Database: new features and capabilities. *Nucleic Acids Res*., **42**, D114–D122.
11. Chojnowski,G., Walen,T. and Bujnicki,J.M. (2014) RNA Bricks—a Database of RNA 3D motifs and their interactions. *Nucleic Acids Res*., **42**, D123–D131.
12. de Beer,T.A., Berka,K., Thornton,J.M. and Laskowski,R.A. (2014) PDBsum additions. *Nucleic Acids Res*., **42**, D292–D296.
13. Gibrat,J.F., Madej,T. and Bryant,S.H. (1996) Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol*., **6**, 377–385.
14. Madej,T., Lanczycki,C.J., Zhang,D., Thiessen,P.A., Geer,R.C., Marchler-Bauer,A. and Bryant,S.H. (2014) MMDB and VAST+: tracking structural similarities between macromolecular complexes. *Nucleic Acids Res*., **42**, D297–D303.
15. Finn,R.D., Miller,B.L., Clements,J. and Bateman,A. (2014) iPfam: a database of protein family and domain interactions found in the Protein Data Bank. *Nucleic Acids Res*., **42**, D364–D373.
16. Mosca,R., Ceol,A., Stein,A., Olivella,R. and Aloy,P. (2014) 3did: a catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res*., **42**, D374–D379.
17. Kalathur,R.K., Pinto,J.P., Hernandez-Prieto,M.A., Machado,R.S., Almeida,D., Chaurasia,G. and Futschik,M.E. (2014) UniHI 7: an enhanced database for retrieval and interactive analysis of human molecular interaction networks. *Nucleic Acids Res*., **42**, D408–D414.
18. Blohm,P., Frishman,G., Smialowski,P., Goebels,F., Wachinger,B., Ruepp,A. and Frishman,D. (2014) Negatome 2.0: a database of non-interacting proteins derived by literature mining, manual annotation and protein structure analysis. *Nucleic Acids Res*., **42**, D396–D400.
19. Andreeva,A., Howorth,D., Chothia,C., Kulesha,E. and Murzin,A.G. (2014) SCOP2 prototype: a new approach to protein structure mining. *Nucleic Acids Res*., **42**, D310–D314.
20. Fox,N.K., Brenner,S.E. and Chandonia,J.-M. (2014) SCOPe: Structural Classification of Proteins extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res*., **42**, D304–D309.
21. Caspi,R., Altman,T., Billington,R., Dreher,K., Foerster,H., Fulcher,C.A., Holland,T.A., Keseler,I.M., Kothari,A., Kubo,A. *et al*. (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res*., **42**, D459–D471.
22. Croft,D., Fabregat Mundo,A., Haw,R., Milacic,M., Weiser,J., Wu,G., Caudy,M., Garapati,P., Gillespie,M., Kaudar,M.R. *et al*. (2014) The reactome pathway knowledgebase. *Nucleic Acids Res*., **42**, D472–D477.
23. Jewison,T., Su,Y., Disfany,F.M., Liang,Y., Knox,C., Maciejewski,A., Poelzer,J., Huynh,J., Zhou,Y., Arndt,D. *et al*. (2014) SMPDB 2.0: big improvements to the small molecule pathway database. *Nucleic Acids Res*., **42**, D478–D484.
24. Akiva,E., Brown,S., Almonacid,D.E., Barber,A.E., Custer,A.F., Hicks,M.A., Huang,C.C., Lauck,F., Mashiyama,S.T., Meng,E.C. *et al*. (2014) The Structure-Function Linkage Database. *Nucleic Acids Res*., **42**, D521–D530.
25. Furnham,N., Holliday,G.L., de Beer,T.A.P., Jacobsen,J.O.B., Pearson,W.R. and Thornton,J.M. (2014) The catalytic site atlas 2.0: cataloguing catalytic sites and residues identified in enzymes. *Nucleic Acids Res*., **42**, D485–D489.
26. Lombard,V., Ramulu,H.G., Drula,E., Coutinho,P.M. and Henrissat,B. (2009) The Carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*., **42**, D490–D495.
27. Rawlings,N.D., Waller,M., Barrett,A.J. and Bateman,A. (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res*., **42**, D503–D509.
28. Wang,Y., Liu,Z., Cheng,H., Gao,T., Pan,Z., Yang,Q., Guo,A. and Xue,Y. (2014) EKPD: a hierarchical database of eukaryotic protein kinases and protein phosphatases. *Nucleic Acids Res*., **42**, D496–D502.
29. Hernández,S., Ferragut,G., Amela,I., Perez-Pons,J.A., Piñol,J., Mozo-Villarias,A., Cedano,J. and Querol,E. (2014) MultitaskProtDB: a database of multitasking proteins. *Nucleic Acids Res*., **42**, D517–D520.
30. Markowitz,V.M., Chen,I.M., Chu,K., Szeto,E., Palaniappan,K., Pillay,M., Ratner,A., Huang,J., Pagani,I., Tringe,S. *et al*. (2014) IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res*., **42**, D568–D573.
31. Hunter,S., Corbett,M., Denise,H., Fraser,M., Gonzalez-Beltran,A., Hunter,C., Jones,P., Leinonen,R., McAnulla,C., Maguire,E. *et al*. (2014) EBI metagenomics—a new resource for the analysis and archiving of metagenomic data. *Nucleic Acids Res*., **42**, D600–D606.
32. Cole,J.R., Wang,Q., Fish,J.A., Chai,B., McGarrell,D.M., Sun,Y., Brown,C.T., Porras-Alfaro,A., Kuske,C.R. and Tiedje,J.M. (2014) The Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res*., **42**, D633–D642.
33. Yilmaz,P., Parfrey,L.W., Yarza,P., Gerken,J., Prüsse,E., Quast,C., Schweer,T., Peplies,J., Ludwig,W. and Glöckner,F.O. (2014) The SILVA and 'The All Species Living Tree (LTP)' taxonomic frameworks. *Nucleic Acids Res*., **42**, D643–D648.
34. Söhngen,C., Bunk,B., Podstawka,A., Gleim,D. and Overmann,J. (2014) BacDive–The Bacterial Diversity metadatabase. *Nucleic Acids Res*., **42**, D592–D599.
35. Parte,A. (2014) LPSN–list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res*., **42**, D613–D616.

36. Huang,K., Brady,A., Mahurkar,A., White,O., Gevers,D., Huttenhower,C. and Segata,N. (2014) MetaRef: a pan-genomic database for comparative and community microbial genomics. *Nucleic Acids Res.*, **42**, D617–D624.
37. Farrell,C.M., O'Leary,N.A., Harte,R.A., Loveland,J.E., Wilming,L.G., Wallin,C., Diekhans,M., Barrell,D., Searle,S.M., Aken,B. *et al.* (2014) Current status and new features of the Consensus Coding Sequence database. *Nucleic Acids Res.*, **42**, D865–D872.
38. Tryka,K.A., Hao,L., Sturcke,A., Jin,Y., Wang,Z.Y., Ziyabari,L., Lee,M., Popova,N., Sharopova,N., Kimura,M. *et al.* (2014) NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res.*, **42**, D975–D979.
39. MacArthur,J.A.L., Morales,J., Tully,R.E., Astashyn,A., Gil,L., Bruford,E.A., Larsson,P., Flicek,P., Dalgleish,R., Maglott,D.R. *et al.* (2014) Locus Reference Genomic: reference sequences for the reporting of clinically relevant sequence variants. *Nucleic Acids Res.*, **42**, D873–D878.
40. Welter,D., MacArthur,J., Morales,J., Burdett,T., Hall,P., Junkins,H., Klemm,A., Flicek,P., Manolio,T., Hindorff,L. *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001–D1006.
41. Landrum,M., Lee,J.M., Riley,G., Jang,W., Rubinstein,W.S., Church,D.M. and Maglott,D. (2014) ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.*, **42**, D980–D985.
42. Bragin,E., Chatzimichali,E.A., Wright,C.F., Hurles,M.E., Firth,H.V., Bevan,A.P. and Swaminathan,G.J. (2014) DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic Acids Res.*, **42**, D993–D1000.
43. Macdonald,J.R., Ziman,R., Yuen,R.K., Feuk,L. and Scherer,S.W. (2014) The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.*, **42**, D986–D992.