Databases for Microbiologists

Igor B. Zhulin

Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, and Department of Microbiology, University of Tennessee, Knoxville, Tennessee, USA

Databases play an increasingly important role in biology. They archive, store, maintain, and share information on genes, genomes, expression data, protein sequences and structures, metabolites and reactions, interactions, and pathways. All these data are critically important to microbiologists. Furthermore, microbiology has its own databases that deal with model microorganisms, microbial diversity, physiology, and pathogenesis. Thousands of biological databases are currently available, and it becomes increasingly difficult to keep up with their development. The purpose of this minireview is to provide a brief survey of current databases that are of interest to microbiologists.

A database is an organized collection of data. Very few biologists cared about databases 25 years ago, simply because there was no need to organize biological data—it was relatively scarce, except for biological literature. It all has changed with the arrival of genome sequencing technologies. The amount of biological data began to grow quickly, as sequenced genomes became commonplace and then skyrocketed when next-generation sequencing (NGS) (1) spread the "omics" revolution. Since January 2008, the speed of DNA sequencing is beating the infamous "Moore's law," and all of a sudden, not only astrophysicists but also biologists are facing big data, and big data needs to be organized big time (2). As of March 2015, the genomes of more than 45,000 strains of bacteria and archaea have been sequenced or are in the process of being sequenced (3). Consequently, the role of databases in microbiology has increased dramatically in recent years and will become even more important in the foreseeable future. The purpose of this minireview is to focus on databases that are specific to microbiology rather than to provide a comprehensive view on thousands of biological databases available today. On the other hand, interdisciplinary boundaries are blurry, and some of the comprehensive databases must be mentioned in the context of microbiology although they serve much broader biological communities. Inevitably, many important databases will be missed from this survey, and I would like to direct a more meticulous reader to a more comprehensive source—the annual *Nucleic Acids Research* database issues and molecular biology database collection. As stated by Michael Galperin and his colleagues (4), the 2015 *Nucleic Acids Research* database issue contains 172 papers describing new and updated databases, and the journal online database collection (http://www.oxfordjournals.org/our_journals/nar/database/a/) provides links to more than 1,550 biological databases.

There is no single, unified way of classifying databases. For the purpose of this minireview, I classify them into three categories: global resources, comprehensive databases, and special-purpose or community databases. "Shopping for information" is quite similar to shopping for goods and services that we do in our everyday lives. Accordingly, one can compare global resources to Wal-Mart or Amazon. Comprehensive databases are reminiscent of large retailers delivering a wide variety of goods in a certain category, such as IKEA or iTunes. Special-purpose databases could be viewed as fashion boutiques that target a very specific clientele. Doing away with analogies and metaphors, I hope that this minireview might serve as a database "consumer digest" or "shopping list" for *Journal of Bacteriology* readers. As shopping lists are rarely logical, this minireview lacks a single logical scheme. For example, some databases that are parts of large comprehensive resources can be described in sections devoted to special-purpose databases, when it seems more appropriate to do so. Links to key databases for microbiologists are provided in Table 1, and many more are briefly described in the text. One important note about biological databases to keep in mind is that although many of them contain very similar information (e.g., the same genomes), they were developed for various purposes, contain heterogeneous types of data accessible by different tools, and curated by different people. Therefore, no two databases are alike.

GLOBAL RESOURCES

Global resources consist of many interconnected databases and tools in order to provide "one-stop shopping" for a vast majority of users. The National Center for Biotechnology Information (NCBI) at the National Institutes of Health in the United States and the European Molecular Biology Laboratory/European Bioinformatics Institute (EMBL-EBI) are undisputed leaders that offer the most comprehensive suites of genomic and molecular biology data collections in the world. The key features of these resources are described by their developers and curators in corresponding publications (5, 6). Here, I will focus specifically on a question asked by a microbiologist: "What's in there for me?" First and foremost, all genomes of bacteria, archaea, eukaryotic microorganisms, and viruses that have been deposited to GenBank, EMBL Bank or DNA Data Bank of Japan (DDBJ) become an integral part of the NCBI and EMBL-EBI collection of databases and therefore

Accepted manuscript posted online 26 May 2015

Citation Zhulin IB. 2015. Databases for microbiologists. J Bacteriol 197:2458–2467. doi:10.1128/JB.00330-15.

Editor: W. W. Margolin

Address correspondence to ijouline@utk.edu.

doi:10.1128/JB.00330-15

TABLE 1 Key databases for microbiologists

Main subject	Database name	URL	Brief description
Microbial genomic resources	IMG	http://img.jgi.doe.gov/	Comprehensive platform for annotation and analysis of microbial genomes and metagenomes
	MicrobesOnline	http://www.microbesonline.org/	Portal for comparative and functional microbial genomics
	SEED	http://www.theseed.org/	Portal for curated genomic data and automated annotation of microbial genomes
	GOLD	https://gold.jgi-psf.org/	Resource for comprehensive information about genome and metagenome sequencing projects
Protein families	CDD	http://www.ncbi.nlm.nih.gov/cdd/	Conserved domain database
	Pfam	http://pfam.xfam.org/	Database of protein families
Protein-protein interactions	STRING	http://string-db.org/	Database of protein association networks
Microbial diversity	RDP	http://rdp.cme.msu.edu/	16S rRNA gene database
	SILVA	http://www.arb-silva.de/	rRNA gene database
	GREENGENES	http://greengenes.lbl.gov/	16S rRNA gene database
	BIGSdb	http://pubmlst.org/software/database/bigsdb/	Bacterial isolate genome sequence database
	EBI metagenomics	www.ebi.ac.uk/metagenomics/	Portal for submission and analysis of metagenomics data
Model organisms	EcoCyc	http://EcoCyc.org	*E. coli* genome and metabolism knowledge base
	RegulonDB	http://regulondb.ccg.unam.mx/	*E. coli* transcriptional regulation resource
	Pseudomonas	http://pseudomonas.com	*Pseudomonas* genome database
Pathogenesis	PATRIC	https://www.patricbrc.org/	Portal for many prokaryotic pathogens
	EuPathDB	http://eupathdb.org/	Portal for many eukaryotic pathogens
	TBDB	http://www.tbdb.org/	Integrated platform for tuberculosis research
Transport and metabolism	TCDB	http://www.tcdb.org/	Transporter classification database
	TransportDB	http://www.membranetransport.org/	Transporter protein analysis database
	MetaCyc	http://metacyc.org/	Metabolic pathway database
	KEGG	http://www.genome.jp/kegg/	Genome database with emphasis on metabolism
Signal transduction and gene regulation	MiST	http://mistdb.com/	Microbial signal transduction database
	SwissRegulon	http://swissregulon.unibas.ch/	Genome-wide annotations of regulatory sites in model organisms
	RegPrecise	http://regprecise.lbl.gov/RegPrecise/	Database of regulons in prokaryotic genomes

are accessible to everyone through a variety of text-based and sequence-based search engines.

NCBI. Like all other biologists, microbiologists depend on NCBI literature resources—PubMed and PubMed Central, which provides the full text of peer-reviewed journal articles, as well as NCBI Bookshelf, which provides free access to the full text of books and reports. The central features of the NCBI collection are nonredundant (NR) databases of nucleotide and protein sequences and their curated subset, known as Reference Sequences or RefSeq (7). The NCBI Genome database collects genome sequencing projects, including all sequenced microbial genomes, and provides links to corresponding records in NR databases and BioProject, which is a central access point to the primary data from sequencing projects. NCBI also maintains the Sequence Read Archive (SRA), which is a public repository for next-generation sequence data (8) and GEO, the archive for functional genomics data sets, which provides an R-based web application to help users analyze its data (9). BLAST (10) is the most popular sequence database search tool, and it now offers an option to search for sequence similarity against any taxonomic group from its NCBI web page. For example, a user may choose to search for similarity only in *Proteobacteria* or *Firmicutes*, or even in a single organism, such as *Escherichia coli* or *Bacillus subtilis*. Alternatively, any taxonomic group or an organism can be excluded from the search. NCBI BLAST also allows its users to search genomic data from environmental samples, thus providing a way to explore vast metagenomics data. NCBI Primer-BLAST (11) helps bench microbiologists design and analyze PCR primers. NCBI Taxonomy database (12) is another useful resource for microbiologists, because it contains information for each taxonomic node, from superkingdoms to subspecies, for virtually all of the formally described species of prokaryotes, and about 10% of eukaryotes. The NCBI Virus Variation resource (13) links viral genome sequence data from influenza, dengue, and West Nile viruses to the corresponding literature, sequences, structures, and population studies.

EMBL-EBI. Similar to NCBI, EMBL-EBI implemented a user-centered design for its cross-linked databases and tools (6). Various databases are organized in several areas: DNA and RNA (genes, genomes, and expression data), proteins (sequences, families, and structures), metabolites (chemogenomics and metabolomics), and systems (reactions, interactions, and pathways). Universal Protein Resource (UniProt) and its curated knowledge base UniProtKB (14) are the most highly regarded EMBL-EBI databases. As at NCBI, information for microbiologists scattered throughout this vast collection. One of the most interesting resources for microbiologists at EMBL-EBI is its Metagenomics portal (15). It allows researchers to submit, archive, and analyze genomic information from various environments, and its analysis pipeline enables feature (genes and rRNAs), function (families, structures, and ontologies), and taxonomic (operational taxonomic units) predictions. EMBL-EBI also contributes to the development of several special-purpose databases for microbiologists that will be considered below.

All databases, but global resources in particular, have to deal with the recent flood of genomic and metagenomic data, which is by no means a simple task. Both NCBI and EMBL-EBI separated environmental sequencing samples from the nonredundant database in order to reduce the number of hypothetical and partial hits in BLAST searches. More recently, both resources consolidated redundant entries from different strains (sometimes even species) into single records. For example, such records in the NCBI databases start with a label “MULTISPECIES,” and a single record may contain sequences from hundreds of strains (e.g., in the case of *E. coli*).

COMPREHENSIVE SPECIAL-PURPOSE DATABASES

Various microorganisms have several hundred to more than ten thousand proteins encoded in their genomes. Classifying them into distinct protein families characterized by conserved domains and evolutionary relationships helps define their biological function. Several resources aim at classifying proteins from their sequence or structure. Pfam (16), SMART (17), and TIGRFAM (18) are carefully curated collections of models that identify conserved proteins and protein domains. The Clusters of Orthologous Groups of Proteins (COGs) database (19) provides phylogenetic classification of proteins. The latest version of COGs is substantially improved by expanded microbial genome coverage (20). A Conserved Domain Database (CDD) at NCBI (21) established a unified collection of such models by combining data from all four above-mentioned databases in addition to its own models. The InterPro protein sequence analysis and classification database at EMBL-EBI (22) serves a similar purpose and provides comprehensive information on protein domains, regions, motifs, and functional sites by integrating data from various sources. Understanding the structure of biological macromolecules leads to deeper understanding of their function. In addition to the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB), which is a key central repository of atomic coordinates and relevant information on the three-dimensional (3D) structures of proteins, nucleic acids, and their complexes (23), the SCOP (Structural Classification of Proteins) (24) and CATH (Class, Architecture, Topology, and Homology) (25) databases are comprehensive resources for classification of protein structures. Finding orthologs (genes and their products in different species that evolved from a common ancestral gene by speciation) and paralogs (genes and their products originated by gene duplication) is yet another important aspect in predicting protein function. In addition to COGs, the OrthoDB database provides comprehensive information on gene/protein orthology (26).

Reconstruction of metabolism from the genome sequence is a key process in modern biology, which combines the wealth of knowledge obtained by experimental biochemistry with the power of comparative genome analysis. Kyoto Encyclopedia of Genes and Genomes (KEGG) (27) is one of the leaders in this field. KEGG metabolic maps are part of the annotation platform for nearly every bacterial and archaeal genome sequencing project. MetaCyc is a metabolic arm of a large BioCyc collection (28). This database contains more than 2,000 experimentally elucidated metabolic pathways from more than 2,000 organisms from all three domains of life. Protein-protein interactions are critically important for metabolism, transport, signal transduction, and other cellular functions. The STRING database is the leading resource for functional protein networks (29). The BioGRID database provides information about protein and genetic interactions in many model organisms, including *E. coli*, *B. subtilis*, eukaryotic microorganisms, and viruses (30).

COMPREHENSIVE MICROBIAL RESOURCES

In this category are databases that aim to deliver comprehensive coverage not for a specific purpose but for a specific group of organisms—those that are of interest to the *Journal of Bacteriology* readership. Ironically, the database known as the Comprehensive Microbial Resource (CMR) (31) is no longer supported. Unfortunately, many other valuable resources also went offline due to lack of funding. For example, the number of obsolete databases removed from their listings last year was similar to that of new databases (4). However, despite the CMR loss, several other large databases still provide comprehensive information on microorganisms centering on comparative genomics. The Integrated Microbial Genomes (IMG) data warehouse (32) is a leading comprehensive resource devoted specifically to microbes. It integrates genomes and relative metadata from bacteria, archaea, eukaryotic microbes, and viruses. IMG provides a way to analyze genome information in a comparative context. It uses a wide variety of bioinformatics tools and curation by domain experts to deliver high-quality annotation across thousands of microbial genomes. In addition, IMG incorporates newly available proteomics and high-throughput RNA sequencing (RNA-seq) data sets and contains information on biosynthetic clusters—sets of genes encoding pathways for secondary metabolite production in selected bacterial genomes. Other resources aiming to provide the community of microbiologists with the means to analyze comparative genomic data include xBase (33), Microbial Genome Database (MBGD) (34), and the MicrobesOnline resource (35). MicrobesOnline not only contains information on thousands of bacterial, archaeal, and fungal genomes but also provides access to gene expression and fitness data.

Annotation and comparative analysis. Genome annotation is a multistep process of taking the raw genomic DNA and adding the layers of analysis and interpretation necessary to extract its biological significance (36). Different annotation platforms use different tools at every step of this process, thus often creating discrepancies starting from gene calling to assignment of biological function. Unfortunately, these problems, which were recognized in the early days of genome sequencing (36), are persistent. While all comprehensive microbial resources implement annotation and allow comparative analysis, some databases make it a priority and/or provide a means to improve its quality. The SEED database implemented the subsystems approach (37) to genomic data, which is also extended to a popular platform for automated microbial genome annotation, RAST (Rapid Annotation using

Subsystems Technology) (38). The MicroScope genome annotation and comparative analysis database (39) aims at improving annotation by enabling groups of investigators to collectively curate various aspects of a given microbial genome and enables cross-genome comparisons. The idea of bringing together the computational and experimental communities in the interest of improving our understanding of microbial gene function is behind the COMBREX database (40). The Gene Ontology resource (41) was designed to unify the representation of genes and proteins across all species, and it is often used as a part of the genome annotation process. The main goal of the POGO-DB database (42) is to allow pairwise comparisons of microbial genomes and identification of orthologous genes. Identification of orthologs in bacterial and archaeal genomes is also the main scope of the Orthologe database (43). Similarly, the ATGC database (44) is the database of orthologous genes, but its distinctive feature is that orthologs are defined in closely related genomes, which is useful for studying microevolution in prokaryotes. Visualization of bacterial chromosome maps at scale delivered by the BacMap database (45) is another valuable tool for comparative analysis. Specialized databases useful for genome annotation and analysis include PSORTdb (46), which contains information on protein subcellular localizations for bacteria and archaea, DoriC (47), which catalogs experimentally identified and computationally predicted *oriC* regions in bacterial and archaeal genomes, ICEberg (48), which does a similar job for bacterial integrative and conjugative elements, and MICAS (49), which contains information on simple sequence repeats and short tandem repeats (microsatellites). For those looking for a quick automated annotation of a bacterial or archaeal genome, Prokka is a good solution. This recent, but already highly popular platform can deliver automated annotation in 10 min on a desktop computer (50).

SPECIAL-PURPOSE DATABASES

Model organisms. *Escherichia coli* is the most widely studied prokaryotic organism. Consequently, there are many resources for the large community of microbiologists, molecular biologists, and biotechnologists interested in this bacterium. Several *E. coli* resources deliver comprehensive coverage of genome information, literature-based curation, and experimental data. EcoGene (51) and GenoBase (52) primarily focus on genome information but also contain other “omics” data, such as microarrays; (ii) EcoCyc (53) provides a powerful synthesis of genome and metabolism information; (iii) PortEco (54) builds a platform where key *E. coli* data can be accessed through the same web portal and serves as a forum for community interactions. Finally, the *E. coli* genome project at ASAP (A Systematic Annotation Package) database (55) stores and distributes genome information and experimental data from functional genomics studies. These rich resources deliver most of the essentials for the *E. coli* community. In addition, several more specialized databases are dedicated to specific topics in *E. coli* research. RegulonDB is an authoritative resource for the *E. coli* transcriptional regulatory network, which is built on experimental and computationally predicted data sets (56). The Bacteriome database (57) has a goal of defining all known and predicting novel protein-protein interactions in *E. coli*, and STEPdb (58) catalogs subcellular localization and topology of all its proteins. The comparative proteomics database EcoProDB has experimental information on *E. coli* proteins and experimental and theoretical 2D maps (59). ECMDB contains comprehensive annotation and detailed information about the *E. coli* metabolome (60).

Bacillus subtilis, the best-studied Gram-positive bacterium, serves as a model organism for studying cell differentiation and chromosome replication, and it is widely used in biotechnology. Dedicated databases for *B. subtilis* include SubtiWiki, a collaborative resource of the *Bacillus* community, which links pathway, interaction, and expression information (61). The SporeWeb interactive knowledge platform captures relevant information about the *B. subtilis* sporulation cycle (62), whereas the Bacillus-RegNet database contains the known regulatory network of this organism as well as predicted interactions for other *Bacillus* species (63). Transcriptional regulation in *B. subtilis* and information about upstream intergenic conservation is also captured in the DBTBS database (64).

Cyanobacteria are widely studied due to their ability to derive energy through photosynthesis. ProPortal, which contains genome, metagenome, transcriptomics, and population dynamics information, serves as the main resource for the model cyanobacterium *Prochlorococcus* and its close relatives (65). CyanoBase (66) is one of the Kazusa genomic resources (Japan) that serve researchers studying cyanobacteria, including *Synechocystis*, *Prochlorococcus*, *Synechococcus*, *Nostoc*, and other model photosynthetic bacteria. Its sister database, RhizoBase (66), plays the same role for a large community of microbiologists interested in nitrogen-fixing bacteria, including various rhizobia as well as *Azoarcus*, *Azospirillum*, *Klebsiella*, and *Frankia*. Kazusa resources also include the *Streptomyces* database devoted to the organisms that produce more than two-thirds of clinically useful antibiotics of natural origin. A more specialized StreptomeDB database (67) explicitly focuses on natural bioactive compounds from *Streptomyces* that can potentially be used as pharmaceuticals—antibiotics and antitumor and immunosuppressant drugs. As the endosymbiont of aphids, *Buchnera* is the subject of investigation in the field of host-microbe interactions. The BuchneraBase database was constructed to facilitate the postgenomic analysis of several *Buchnera* strains and closely related insect endosymbionts (68). Due to its extremely small genome, *Mycoplasma genitalium* became a model for synthetic genomics. The first whole-genome genotype-to-phenotype model was constructed using this organism, and the corresponding database WholeCellKB (69) offers a platform for whole-cell modeling in other organisms.

The *Saccharomyces* Genome Database is the key resource for a model eukaryotic microorganism (70), which provides comprehensive biological information for the budding yeast *Saccharomyces cerevisiae*, including ontologies, biochemical pathways, expression data, and phenotypes in addition to the genome browser and similarity search tools. The yeast stress expression database, yStreX (71), is an online repository of analyzed gene expression data related to responses to diverse environmental transitions.

Diversity and metagenomics. Research in microbial diversity has flourished owing to metagenomics approaches utilizing NGS technologies to characterize microbial communities in different ecosystems. Consequently, databases that assist researchers in environmental microbiology, microbial ecology, taxonomy, and phylogenetics become increasingly important. The Ribosomal Database Project (RDP) (72) maintains the largest collection of aligned and annotated rRNA gene sequences from bacteria, archaea, and fungi. It enables researchers to analyze their rRNA sequences in the RDP framework and provides tools to facilitate analysis of high-throughput data. Similar capabilities are provided by the SILVA taxonomic framework (73), which offers a set of rRNA gene sequence databases for bacteria, archaea, and eukaryota based on representative phylogenetic trees, and by the GREENGENES project (74). To help monitor microbial communities within complex environments, the PhylOPDb database offers a large collection of regular and explorative rRNA-targeted probes (75), and the rRNA operon copy number (rrnDB) database provides a means to interpret rRNA gene abundance in bacteria and archaea (76).

The Bacterial Diversity Metadatabase (BacDive) contains detailed information on various aspects of more than 20,000 strains of bacteria and archaea, which includes taxonomy, physiology, sampling, and environmental conditions (77). The Global Catalog of Microorganisms (GCM) is a database for retrieval and analysis of relevant information for hundreds of thousands of microbial strains from different sources (78). Bacterial Isolate Genome Sequence Database (BIGSdb) from the PubMLST (collection of databases for molecular typing and microbial genome diversity) resource stores sequence data for bacterial isolates and enables analysis of genome variation at the population level (79). The List of Prokaryotic Names with Standing in Nomenclature (LPSN) (80) is an important resource that provides up-to-date classification of prokaryotes by listing the names of bacteria and archaea that have been validated and published in the *International Journal of Systematic and Evolutionary Microbiology*.

The Human Microbiome Project (HMP) and other large-scale projects created the need to submit, store, analyze, and share numerous metagenomic data sets. In response to this challenge, platforms dedicated to metagenomics are now offered by several bioinformatics centers. EBI established a dedicated metagenomics resource (15), which allows users to submit raw nucleotide reads for taxonomic and functional analysEs by an automated pipeline. Similar capabilities are provided by IMG in its dedicated IMG/M resource (81), which also enables expert review of metagenome annotations (IMG/M ER). The metaMicrobesOnline (82) and MetaRef (83) databases offer comparative analysis for microbial genomes and metagenomes with emphasis on gene trees, gene family conservation, and genome context comparisons. To promote inference of metagenomic functional networks, the MetaProx database (84) contains information on candidate operons in metagenomic data sets. The MetaBioME database (85) positions itself as a comprehensive metagenomic biomining engine by providing the opportunity to find novel homologs of known commercially useful enzymes in metagenomic data sets. The FOAM (Functional Ontology Assignments for Metagenomes) database offers classification of gene functions in environmental metagenomes based on ontology and orthologous relationships (86).

Pathogenesis. Pathogen Portal (http://pathogenportal.org) is a rich resource on pathogenic microorganisms that includes both bacteria and eukaryotes. The Pathosystems Resource Integration Center (PATRIC) provides the community of microbiologists interested in pathogenic bacteria with access to a variety of data, including genomics, transcriptomics, protein-protein interactions, sequence typing, etc., covering more than 10,000 genomes of genera containing NIAID category A to C/emerging/reemerging pathogens (87). GeneDB (88) is an annotation database for many pathogenic bacteria, eukaryotes, and viruses. Databases that are smaller in scope focus on specific pathogens and their functions. The tuberculosis database (TBDB) integrates genomic sequences and data for *Mycobacterium* species relevant to drug discovery, vaccines, and biomarkers (89), whereas SITVITWEB (90), tbvar (91), and GMTV (92) focus on delivering comprehensive information on genome-wide variation in *Mycobacterium tuberculosis*. Other databases featuring specific pathogenic bacteria include the *Pseudomonas* (93), *Vibrio* (94), *Corynebacterium* (95), and *Helicobacter* (96) genome databases. The HoPaCI-DB resource focuses on host-pathogen interactions of *Pseudomonas aeruginosa* and *Coxiella* spp. and provides thousands of validated interactions between molecules and processes (97). Postanalysis data on the *Staphylococcus aureus* transcriptome can be found in the SATMD database (98). DBSecSys provides comprehensive information about secretion systems in a category B priority pathogen, *Burkholderia mallei* (99). Information about bacterial species with the most relevance to veterinary medicine can be found in the VetBact database (100). Resources on viral pathogens include the molecular and epidemiological ViralZone knowledge base (101), the HBVdb database for the hepatitis B virus (102), the Papillomavirus Episteme (103), and databases of experimentally validated viral small interfering RNA (siRNA)/small hairpin RNA (shRNA) VIRsiRNAdb (104) and microRNA (miRNA) VIRmiRNA (105).

Eukaryotic pathogenic microorganisms are well represented in Eukaryotic Pathogen Database Resources (EuPathDB), a collection of individual databases, each focusing on specific pathogens, accessible through a common portal (106). FungiDB contains information on *Candida*, *Aspergillus*, *Cryptococcus*, and some other fungi (107). The EuPathDB collection includes databases for pathogenic amoeba and microsporidia (108), *Cryptosporidium* (109), *Toxoplasma* (110), *Giardia* and *Trichomonas* (111), *Trypanosoma* (112), and several species of *Plasmodium* malaria parasites (113).

Transport, secretion, and metabolism. Several special-purpose databases are devoted to these functions. Two expert-led databases—the Transporter Classification Database (TCDB) and TransportDB—provide the overview, curated annotations and detailed genomic comparisons of membrane transport proteins across selected genomes of bacteria, archaea, and eucarya (114, 115). More-specialized community databases highlight specific transport systems, including archaeal and bacterial ABC transporters (116) and β-barrel outer membrane proteins (117) as well as type III (118), type IV (119, 120), and type VI (121) secretion systems.

Microme (http://microme.eu) is a European resource for microbial metabolism, and its main goal is to support the large-scale inference of metabolic flux directly from the genome sequence. It includes the well-regarded microbial genome annotation and analysis platform MicroScope (39) and serves as a portal to several analysis and data mining tools. Microorganisms contain many enzymes that assemble, modify, and break down oligo- and polysaccharides. The Carbohydrate-Active Enzymes (CAZy) database provides a classification platform linking the sequences to the specificities and 3D structures of these enzymes (122). Many other community databases are devoted to specific metabolic features and processes in microorganisms. AromaDeg is a database focusing on aerobic bacterial degradation of aromatics (123). DEOP database on osmoprotectants and associated pathways provides curated information for many bacterial species (124). BacMet is a resource for antibacterial biocide and metal resistance genes (125). Information on microbial polyketide and nonribosomal peptide gene clusters can be found in the ClusterMine360 database (126). Detailed information on bacterial carbohydrate structure is provided by the BCSDB database (127); experimentally characterized glycoproteins from bacteria and archaea are listed in the ProGlycProt repository (128). CyanoLyase (129) and mVOC (130) are curated databases of phycobiliproteins and microbial volatile compounds, respectively.

Signal transduction and gene regulation. The Microbial Signal Transduction (MiST) database offers detailed information on receptors, kinases, response regulators, and transcription factors in thousands of bacterial and archaeal genomes (131). A subset of these proteins is also available in the P2CS (prokaryotic two-component systems) database (132). The Quorumpeps database has a collection of specific signaling molecules—quorum-sensing peptides (133). Transcriptional regulation is the main mode of bacterial responses to signals, and several databases are designed to capture known and predicted transcriptional regulators and their targets. The SwissRegulon database (134) provides genome-wide annotation of regulatory sites for model eukaryotes and prokaryotes, including *E. coli*, *B. subtilis*, *S. aureus*, *Vibrio cholerae*, and *M. tuberculosis*. RegTransBase is a database of regulatory sequences and interactions that are based on careful curation of thousands of scientific publications (135), and the RegPrecise database delivers inferred regulatory interactions across hundreds of bacterial genomes with an emphasis on phylogenetic, structural, and functional properties (136). Network Portal (137) provides analysis and visualization tools for selected gene regulatory networks in several bacterial species, including *E. coli*, *B. subtilis*, *M. tuberculosis*, *P. aeruginosa*, and *Campylobacter jejuni* among others. It serves as a modular database for analyzing user-uploaded data and public data. Operons predicted for more than 1,200 genomes of bacteria and archaea can be accessed in the ProOpDB database (138). Collection of experimentally verified transcription factor-binding sites is available in the CollecTF database (139). WebGeSTer DB is the largest database of intrinsic transcription terminators identified in more than a thousand bacterial genomes (140).

CONCLUDING REMARKS

The advent of Internet and high-throughput genome sequencing has opened new horizons for biomedical sciences. Sooner than we think, electronic patient records in hospitals will be merged with patient genomic information to build foundation for truly personalized medicine. In a similar fashion, the wealth of knowledge about microorganisms, which resides in electronic copies of journal articles will be merged with genomic (and other "omic") information to build the foundation for a more vigorous and comprehensive analysis of microbes. While we are at the very beginning of this road, it becomes increasingly important for microbiologists to know how to use genomic resources—databases and computational tools—to enhance their own research and to archive and share obtained knowledge in a robust and widely accessible form. One obvious way is to deposit the results of experimental research, especially high-throughput data, to public repositories. Submission of genome sequencing (141) and microarray (142) data to public repositories has become mandatory. However, there are many other options by which one can enhance the visibility of results and share them more efficiently at the same time. For example, the commentary published in this issue of the *Journal of Bacteriology* by Ivan Erill (143) shows how the CollecTF database (139) can be used by researchers to submit, archive, and share experimentally verified transcription factor-binding sites that are usually only reported in research articles and are hard to mine. Similarly, it is important for authors when publishing their research papers to use database accession numbers that would link genes, proteins, and other data sets described in the paper to genomic data. Because many journal electronic editions now provide hyperlinks to genomic databases, one can then access the relevant data in one click.

While some biological databases will come and go and others may change their substance and appearance, it is clear that in the grand scheme of things, biological databases—from giant repositories and comprehensive information portals to small community databases—will play an increasingly important role in biological discovery.

ACKNOWLEDGMENTS

I thank Michael Galperin, Nikos Kyrpides, and members of my laboratory for constructive criticism and many helpful suggestions.

Work in my laboratory is supported by grants GM072285 and DE024463 from the National Institutes of Health.

REFERENCES

1. **Metzker ML.** 2010. Sequencing technologies - the next generation. Nat Rev Genet **11:**31–46. http://dx.doi.org/10.1038/nrg2626.
2. **Howe D, Costanzo M, Fey P, Gojobori T, Hannick L, Hide W, Hill DP, Kania R, Schaeffer M, St Pierre S, Twigger S, White O, Rhee SY.** 2008. Big data: the future of biocuration. Nature **455:**47–50. http://dx.doi.org/10.1038/455047a.
3. **Reddy TB, Thomas AD, Stamatis D, Bertsch J, Isbandi M, Jansson J, Mallajosyula J, Pagani I, Lobos EA, Kyrpides NC.** 2015. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. Nucleic Acids Res **43:**D1099–D1106. http://dx.doi.org/10.1093/nar/gku950.
4. **Galperin MY, Rigden DJ, Fernandez-Suarez XM.** 2015. The 2015 *Nucleic Acids Research* Database Issue and molecular biology database collection. Nucleic Acids Res **43:**D1–D5. http://dx.doi.org/10.1093/nar/gku1241.
5. **NCBI Resource Coordinators.** 2015. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res **43:**D6–D17. http://dx.doi.org/10.1093/nar/gku1130.
6. **Brooksbank C, Bergman MT, Apweiler R, Birney E, Thornton J.** 2014. The European Bioinformatics Institute's data resources 2014. Nucleic Acids Res **42:**D18–D25. http://dx.doi.org/10.1093/nar/gkt1206.
7. **Pruitt KD, Tatusova T, Brown GR, Maglott DR.** 2012. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res **40:**D130–D135. http://dx.doi.org/10.1093/nar/gkr1079.
8. **Kodama Y, Shumway M, Leinonen R, International Nucleotide Sequence Database Collaboration.** 2012. The Sequence Read Archive: explosive growth of sequencing data. Nucleic Acids Res **40:**D54–D56. http://dx.doi.org/10.1093/nar/gkr854.
9. **Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A.** 2013. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res **41:**D991–D995. http://dx.doi.org/10.1093/nar/gks1193.
10. **Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.
11. **Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL.** 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics **13:**134. http://dx.doi.org/10.1186/1471-2105-13-134.
12. **Federhen S.** 2012. The NCBI Taxonomy database. Nucleic Acids Res **40:**D136–D143. http://dx.doi.org/10.1093/nar/gkr1178.
13. **Brister JR, Bao Y, Zhdanov SA, Ostapchuck Y, Chetvernin V, Kiryutin B, Zaslavsky L, Kimelman M, Tatusova TA.** 2014. Virus Variation Resource–recent updates and future directions. Nucleic Acids Res **42:**D660–D665. http://dx.doi.org/10.1093/nar/gkt1268.

14. **UniProt Consortium.** 2014. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res **42:**D191–D198. http://dx.doi.org/10.1093/nar/gkt1140.
15. **Hunter S, Corbett M, Denise H, Fraser M, Gonzalez-Beltran A, Hunter C, Jones P, Leinonen R, McAnulla C, Maguire E, Maslen J, Mitchell A, Nuka G, Oisel A, Pesseat S, Radhakrishnan R, Rocca-Serra P, Scheremetjew M, Sterk P, Vaughan D, Cochrane G, Field D, Sansone SA.** 2014. EBI metagenomics–a new resource for the analysis and archiving of metagenomic data. Nucleic Acids Res **42:**D600–D606. http://dx.doi.org/10.1093/nar/gkt961.
16. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M.** 2014. Pfam: the protein families database. Nucleic Acids Res **42:**D222–D230. http://dx.doi.org/10.1093/nar/gkt1223.
17. **Letunic I, Doerks T, Bork P.** 2015. SMART: recent updates, new developments and status in 2015. Nucleic Acids Res **43:**D257–D260. http://dx.doi.org/10.1093/nar/gku949.
18. **Haft DH, Selengut JD, White O.** 2003. The TIGRFAMs database of protein families. Nucleic Acids Res **31:**371–373. http://dx.doi.org/10.1093/nar/gkg128.
19. **Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA.** 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4:**41. http://dx.doi.org/10.1186/1471-2105-4-41.
20. **Galperin MY, Makarova KS, Wolf YI, Koonin EV.** 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res **43:**D261–D269. http://dx.doi.org/10.1093/nar/gku1223.
21. **Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH.** 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res **43:**D222–D226. http://dx.doi.org/10.1093/nar/gku1221.
22. **Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C, Nuka G, Pesseat S, Sangrador-Vegas A, Scheremetjew M, Rato C, Yong SY, Bateman A, Punta M, Attwood TK, Sigrist CJ, Redaschi N, Rivoire C, Xenarios I, Kahn D, Guyot D, Bork P, Letunic I, Gough J, Oates M, Haft D, Huang H, Natale DA, Wu CH, Orengo C, Sillitoe I, Mi H, Thomas PD, Finn RD.** 2015. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res **43:**D213–D221. http://dx.doi.org/10.1093/nar/gku1243.
23. **Rose PW, Prlic A, Bi C, Bluhm WF, Christie CH, Dutta S, Green RK, Goodsell DS, Westbrook JD, Woo J, Young J, Zardecki C, Berman HM, Bourne PE, Burley SK.** 2015. The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. Nucleic Acids Res **43:**D345–D356. http://dx.doi.org/10.1093/nar/gku1214.
24. **Andreeva A, Howorth D, Chothia C, Kulesha E, Murzin AG.** 2014. SCOP2 prototype: a new approach to protein structure mining. Nucleic Acids Res **42:**D310–D314. http://dx.doi.org/10.1093/nar/gkt1242.
25. **Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL, Furnham N, Laskowski RA, Lee D, Lees JG, Lehtinen S, Studer RA, Thornton J, Orengo CA.** 2015. CATH: comprehensive structural and functional annotations for genome sequences. Nucleic Acids Res **43:**D376–D381. http://dx.doi.org/10.1093/nar/gku947.
26. **Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simao FA, Pozdnyakov IA, Ioannidis P, Zdobnov EM.** 2015. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. Nucleic Acids Res **43:**D250–D256. http://dx.doi.org/10.1093/nar/gku1220.
27. **Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M.** 2014. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res **42:**D199–D205. http://dx.doi.org/10.1093/nar/gkt1076.
28. **Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari A, Kubo A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Subhraveti P, Weaver DS, Weerasinghe D, Zhang P, Karp PD.** 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res **42:**D459–D471. http://dx.doi.org/10.1093/nar/gkt1103.
29. **Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ.** 2013. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res **41:**D808–D815. http://dx.doi.org/10.1093/nar/gks1094.
30. **Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A, Kolas N, O'Donnell L, Reguly T, Nixon J, Ramage L, Winter A, Sellam A, Chang C, Hirschman J, Theesfeld C, Rust J, Livstone MS, Dolinski K, Tyers M.** 2015. The BioGRID interaction database: 2015 update. Nucleic Acids Res **43:**D470–D478. http://dx.doi.org/10.1093/nar/gku1204.
31. **Davidsen T, Beck E, Ganapathy A, Montgomery R, Zafar N, Yang Q, Madupu R, Goetz P, Galinsky K, White O, Sutton G.** 2010. The comprehensive microbial resource. Nucleic Acids Res **38:**D340–D345. http://dx.doi.org/10.1093/nar/gkp912.
32. **Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis K, Pati A, Ivanova NN, Kyrpides NC.** 2014. IMG 4 version of the Integrated Microbial Genomes comparative analysis system. Nucleic Acids Res **42:**D560–D567. http://dx.doi.org/10.1093/nar/gkt963.
33. **Chaudhuri RR, Loman NJ, Snyder LA, Bailey CM, Stekel DJ, Pallen MJ.** 2008. xBASE2: a comprehensive resource for comparative bacterial genomics. Nucleic Acids Res **36:**D543–D546. http://dx.doi.org/10.1093/nar/gkm928.
34. **Uchiyama I, Mihara M, Nishide H, Chiba H.** 2015. MBGD update 2015: microbial genome database for flexible ortholog analysis utilizing a diverse set of genomic data. Nucleic Acids Res **43:**D270–D276. http://dx.doi.org/10.1093/nar/gku1152.
35. **Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, Friedland GD, Huang KH, Keller K, Novichkov PS, Dubchak IL, Alm EJ, Arkin AP.** 2010. MicrobesOnline: an integrated portal for comparative and functional genomics. Nucleic Acids Res **38:**D396–D400. http://dx.doi.org/10.1093/nar/gkp919.
36. **Stein L.** 2001. Genome annotation: from sequence to biology. Nat Rev Genet **2:**493–503. http://dx.doi.org/10.1038/35080529.
37. **Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Ruckert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V.** 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res **33:**5691–5702. http://dx.doi.org/10.1093/nar/gki866.
38. **Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R.** 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res **42:**D206–D214. http://dx.doi.org/10.1093/nar/gkt1226.
39. **Vallenet D, Belda E, Calteau A, Cruveiller S, Engelen S, Lajus A, Le Fevre F, Longin C, Mornico D, Roche D, Rouy Z, Salvignol G, Scarpelli C, Thil Smith AA, Weiman M, Medigue C.** 2013. MicroScope–an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. Nucleic Acids Res **41:**D636–D647. http://dx.doi.org/10.1093/nar/gks1194.
40. **Roberts RJ, Chang YC, Hu Z, Rachlin JN, Anton BP, Pokrzywa RM, Choi HP, Faller LL, Guleria J, Housman G, Klitgord N, Mazumdar V, McGettrick MG, Osmani L, Swaminathan R, Tao KR, Letovsky S, Vitkup D, Segre D, Salzberg SL, Delisi C, Steffen M, Kasif S.** 2011. COMBREX: a project to accelerate the functional annotation of prokaryotic genomes. Nucleic Acids Res **39:**D11–D14. http://dx.doi.org/10.1093/nar/gkq1168.
41. **Gene Ontology Consortium.** 2015. Gene Ontology Consortium: going forward. Nucleic Acids Res **43:**D1049–D1056. http://dx.doi.org/10.1093/nar/gku1179.
42. **Lan Y, Morrison JC, Hershberg R, Rosen GL.** 2014. POGO-DB–a database of pairwise-comparisons of genomes and conserved orthologous genes. Nucleic Acids Res **42:**D625–D632. http://dx.doi.org/10.1093/nar/gkt1094.
43. **Whiteside MD, Winsor GL, Laird MR, Brinkman FS.** 2013. OrtholugeDB: a bacterial and archaeal orthology resource for improved comparative genomic analysis. Nucleic Acids Res **41:**D366–D376. http://dx.doi.org/10.1093/nar/gks1241.

44. **Novichkov PS, Ratnere I, Wolf YI, Koonin EV, Dubchak I.** 2009. ATGC: a database of orthologous genes from closely related prokaryotic genomes and a research platform for microevolution of prokaryotes. Nucleic Acids Res **37:**D448–D454. http://dx.doi.org/10.1093/nar/gkn684.

45. **Stothard P, Van Domselaar G, Shrivastava S, Guo A, O'Neill B, Cruz J, Ellison M, Wishart DS.** 2005. BacMap: an interactive picture atlas of annotated bacterial genomes. Nucleic Acids Res **33:**D317–D320. http://dx.doi.org/10.1093/nar/gki075.

46. **Yu NY, Laird MR, Spencer C, Brinkman FS.** 2011. PSORTdb–an expanded, auto-updated, user-friendly protein subcellular localization database for Bacteria and Archaea. Nucleic Acids Res **39:**D241–D244. http://dx.doi.org/10.1093/nar/gkq1093.

47. **Gao F, Luo H, Zhang CT.** 2013. DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes. Nucleic Acids Res **41:**D90–D93. http://dx.doi.org/10.1093/nar/gks990.

48. **Bi D, Xu Z, Harrison EM, Tai C, Wei Y, He X, Jia S, Deng Z, Rajakumar K, Ou HY.** 2012. ICEberg: a web-based resource for integrative and conjugative elements found in Bacteria. Nucleic Acids Res **40:**D621–D626. http://dx.doi.org/10.1093/nar/gkr846.

49. **Mudunuri SB, Patnana S, Nagarajaram HA.** 2014. MICdb3.0: a comprehensive resource of microsatellite repeats from prokaryotic genomes. Database (Oxford) **2014:**bau005. http://dx.doi.org/10.1093/database/bau005.

50. **Seemann T.** 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30:**2068–2069. http://dx.doi.org/10.1093/bioinformatics/btu153.

51. **Zhou J, Rudd KE.** 2013. EcoGene 3.0. Nucleic Acids Res **41:**D613–D624. http://dx.doi.org/10.1093/nar/gks1235.

52. **Otsuka Y, Muto A, Takeuchi R, Okada C, Ishikawa M, Nakamura K, Yamamoto N, Dose H, Nakahigashi K, Tanishima S, Suharnan S, Nomura W, Nakayashiki T, Aref WG, Bochner BR, Conway T, Gribskov M, Kihara D, Rudd KE, Tohsato Y, Wanner BL, Mori H.** 2015. GenoBase: comprehensive resource database of Escherichia coli K-12. Nucleic Acids Res **43:**D606–D617. http://dx.doi.org/10.1093/nar/gku1164.

53. **Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martinez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, Latendresse M, Muniz-Rascado L, Ong Q, Paley S, Schroder I, Shearer AG, Subhraveti P, Travers M, Weerasinghe D, Weiss V, Collado-Vides J, Gunsalus RP, Paulsen I, Karp PD.** 2013. EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Res **41:**D605–D612. http://dx.doi.org/10.1093/nar/gks1027.

54. **Hu JC, Sherlock G, Siegele DA, Aleksander SA, Ball CA, Demeter J, Gouni S, Holland TA, Karp PD, Lewis JE, Liles NM, McIntosh BK, Mi H, Muruganujan A, Wymore F, Thomas PD, Altman T.** 2014. PortEco: a resource for exploring bacterial biology through high-throughput data and analysis tools. Nucleic Acids Res **42:**D677–D684. http://dx.doi.org/10.1093/nar/gkt1203.

55. **Glasner JD, Rusch M, Liss P, Plunkett G, III, Cabot EL, Darling A, Anderson BD, Infield-Harm P, Gilson MC, Perna NT.** 2006. ASAP: a resource for annotating, curating, comparing, and disseminating genomic data. Nucleic Acids Res **34:**D41–D45. http://dx.doi.org/10.1093/nar/gkj164.

56. **Salgado H, Peralta-Gil M, Gama-Castro S, Santos-Zavaleta A, Muniz-Rascado L, Garcia-Sotelo JS, Weiss V, Solano-Lira H, Martinez-Flores I, Medina-Rivera A, Salgado-Osorio G, Alquicira-Hernandez S, Alquicira-Hernandez K, Lopez-Fuentes A, Porron-Sotelo L, Huerta AM, Bonavides-Martinez C, Balderas-Martinez YI, Pannier L, Olvera M, Labastida A, Jimenez-Jacinto V, Vega-Alvarado L, Del Moral-Chavez V, Hernandez-Alvarez A, Morett E, Collado-Vides J.** 2013. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. Nucleic Acids Res **41:**D203–D213. http://dx.doi.org/10.1093/nar/gks1201.

57. **Su C, Peregrin-Alvarez JM, Butland G, Phanse S, Fong V, Emili A, Parkinson J.** 2008. Bacteriome.org—an integrated protein interaction database for E. coli. Nucleic Acids Res **36:**D632–D636. http://dx.doi.org/10.1093/nar/gkm807.

58. **Orfanoudaki G, Economou A.** 2014. Proteome-wide subcellular topologies of E. coli polypeptides database (STEPdb). Mol Cell Proteomics **13:**3674–3687. http://dx.doi.org/10.1074/mcp.O114.041137.

59. **Yun H, Lee JW, Jeong J, Chung J, Park JM, Myoung HN, Lee SY.** 2007. EcoProDB: the Escherichia coli protein database. Bioinformatics **23:**2501–2503. http://dx.doi.org/10.1093/bioinformatics/btm351.

60. **Guo AC, Jewison T, Wilson M, Liu Y, Knox C, Djoumbou Y, Lo P, Mandal R, Krishnamurthy R, Wishart DS.** 2013. ECMDB: the E. coli Metabolome Database. Nucleic Acids Res **41:**D625–D630. http://dx.doi.org/10.1093/nar/gks992.

61. **Michna RH, Commichau FM, Todter D, Zschiedrich CP, Stulke J.** 2014. SubtiWiki-a database for the model organism Bacillus subtilis that links pathway, interaction and expression information. Nucleic Acids Res **42:**D692–D698. http://dx.doi.org/10.1093/nar/gkt1002.

62. **Eijlander RT, de Jong A, Krawczyk AO, Holsappel S, Kuipers OP.** 2014. SporeWeb: an interactive journey through the complete sporulation cycle of Bacillus subtilis. Nucleic Acids Res **42:**D685–D691. http://dx.doi.org/10.1093/nar/gkt1007.

63. **Misirli G, Hallinan J, Rottger R, Baumbach J, Wipat A.** 2014. BacillusRegNet: a transcriptional regulation database and analysis platform for Bacillus species. J Integr Bioinform **11:**244. http://dx.doi.org/10.2390/biecoll-jib-2014-244.

64. **Sierro N, Makita Y, de Hoon M, Nakai K.** 2008. DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information. Nucleic Acids Res **36:**D93–D96. http://dx.doi.org/10.1093/nar/gkm910.

65. **Kelly L, Huang KH, Ding H, Chisholm SW.** 2012. ProPortal: a resource for integrated systems biology of Prochlorococcus and its phage. Nucleic Acids Res **40:**D632–D640. http://dx.doi.org/10.1093/nar/gkr1022.

66. **Fujisawa T, Okamoto S, Katayama T, Nakao M, Yoshimura H, Kajiya-Kanegae H, Yamamoto S, Yano C, Yanaka Y, Maita H, Kaneko T, Tabata S, Nakamura Y.** 2014. CyanoBase and RhizoBase: databases of manually curated annotations for cyanobacterial and rhizobial genomes. Nucleic Acids Res **42:**D666–D670. http://dx.doi.org/10.1093/nar/gkt1145.

67. **Lucas X, Senger C, Erxleben A, Gruning BA, Doring K, Mosch J, Flemming S, Gunther S.** 2013. StreptomeDB: a resource for natural compounds isolated from Streptomyces species. Nucleic Acids Res **41:**D1130–D1136. http://dx.doi.org/10.1093/nar/gks1253.

68. **Prickett MD, Page M, Douglas AE, Thomas GH.** 2006. BuchneraBASE: a post-genomic resource for Buchnera sp. APS Bioinformatics **22:**641–642. http://dx.doi.org/10.1093/bioinformatics/btk024.

69. **Karr JR, Sanghvi JC, Macklin DN, Arora A, Covert MW.** 2013. WholeCellKB: model organism databases for comprehensive whole-cell models. Nucleic Acids Res **41:**D787–D792. http://dx.doi.org/10.1093/nar/gks1108.

70. **Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED.** 2012. Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res **40:**D700–D705. http://dx.doi.org/10.1093/nar/gkr1029.

71. **Wanichthanarak K, Nookaew I, Petranovic D.** 2014. yStreX: yeast stress expression database. Database (Oxford) **2014:**bau068. http://dx.doi.org/10.1093/database/bau068.

72. **Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM.** 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res **42:**D633–D642. http://dx.doi.org/10.1093/nar/gkt1244.

73. **Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glockner FO.** 2014. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res **42:**D643–D648. http://dx.doi.org/10.1093/nar/gkt1209.

74. **McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P.** 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J **6:**610–618. http://dx.doi.org/10.1038/ismej.2011.139.

75. **Jaziri F, Parisot N, Abid A, Denonfoux J, Ribiere C, Gasc C, Boucher D, Brugere JF, Mahul A, Hill DR, Peyretaillade E, Peyret P.** 2014. PhylOPDb: a 16S rRNA oligonucleotide probe database for prokaryotic identification. Database (Oxford) **2014:**bau036. http://dx.doi.org/10.1093/database/bau036.

76. **Stoddard SF, Smith BJ, Hein R, Roller BR, Schmidt TM.** 2015. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Res **43:**D593–D598. http://dx.doi.org/10.1093/nar/gku1201.

77. **Sohngen C, Bunk B, Podstawka A, Gleim D, Overmann J.** 2014. BacDive–the Bacterial Diversity Metadatabase. Nucleic Acids Res **42:**D592–D599. http://dx.doi.org/10.1093/nar/gkt1058.

78. **Wu L, Sun Q, Sugawara H, Yang S, Zhou Y, McCluskey K, Vasilenko A, Suzuki K, Ohkuma M, Lee Y, Robert V, Ingsriswang S, Guissart F, Philippe D, Ma J.** 2013. Global catalogue of microorganisms (GCM): a comprehensive database and information retrieval, analysis, and visual-

ization system for microbial resources. BMC Genomics **14:**933. http://dx.doi.org/10.1186/1471-2164-14-933.

79. **Jolley KA, Maiden MC.** 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics **11:**595. http://dx.doi.org/10.1186/1471-2105-11-595.
80. **Parte AC.** 2014. LPSN–list of prokaryotic names with standing in nomenclature. Nucleic Acids Res **42:**D613–D616. http://dx.doi.org/10.1093/nar/gkt1111.
81. **Markowitz VM, Chen IM, Chu K, Szeto E, Palaniappan K, Pillay M, Ratner A, Huang J, Pagani I, Tringe S, Huntemann M, Billis K, Varghese N, Tennessen K, Mavromatis K, Pati A, Ivanova NN, Kyrpides NC.** 2014. IMG/M 4 version of the integrated metagenome comparative analysis system. Nucleic Acids Res **42:**D568–D573. http://dx.doi.org/10.1093/nar/gkt919.
82. **Chivian D, Dehal PS, Keller K, Arkin AP.** 2013. MetaMicrobesOnline: phylogenomic analysis of microbial communities. Nucleic Acids Res **41:** D648–D654. http://dx.doi.org/10.1093/nar/gks1202.
83. **Huang K, Brady A, Mahurkar A, White O, Gevers D, Huttenhower C, Segata N.** 2014. MetaRef: a pan-genomic database for comparative and community microbial genomics. Nucleic Acids Res **42:**D617–D624. http://dx.doi.org/10.1093/nar/gkt1078.
84. **Vey G, Charles TC.** 2014. MetaProx: the database of metagenomic proximons. Database (Oxford) **2014:**bau097. http://dx.doi.org/10.1093/database/bau097.
85. **Sharma VK, Kumar N, Prakash T, Taylor TD.** 2010. MetaBioME: a database to explore commercially useful enzymes in metagenomic datasets. Nucleic Acids Res **38:**D468–D472. http://dx.doi.org/10.1093/nar/gkp1001.
86. **Prestat E, David MM, Hultman J, Tas N, Lamendella R, Dvornik J, Mackelprang R, Myrold DD, Jumpponen A, Tringe SG, Holman E, Mavromatis K, Jansson JK.** 2014. FOAM (Functional Ontology Assignments for Metagenomes): a Hidden Markov Model (HMM) database with environmental focus. Nucleic Acids Res **42:**e145. http://dx.doi.org/10.1093/nar/gku702.
87. **Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, Machi D, Mao C, Nordberg EK, Olson R, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MJ, Yoo HS, Zhang C, Zhang Y, Sobral BW.** 2014. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res **42:**D581–D591. http://dx.doi.org/10.1093/nar/gkt1099.
88. **Logan-Klumpler FJ, De Silva N, Boehme U, Rogers MB, Velarde G, McQuillan JA, Carver T, Aslett M, Olsen C, Subramanian S, Phan I, Farris C, Mitra S, Ramasamy G, Wang H, Tivey A, Jackson A, Houston R, Parkhill J, Holden M, Harb OS, Brunk BP, Myler PJ, Roos D, Carrington M, Smith DF, Hertz-Fowler C, Berriman M.** 2012. GeneDB–an annotation database for pathogens. Nucleic Acids Res **40:** D98–D108. http://dx.doi.org/10.1093/nar/gkr1032.
89. **Reddy TB, Riley R, Wymore F, Montgomery P, DeCaprio D, Engels R, Gellesch M, Hubble J, Jen D, Jin H, Koehrsen M, Larson L, Mao M, Nitzberg M, Sisk P, Stolte C, Weiner B, White J, Zachariah ZK, Sherlock G, Galagan JE, Ball CA, Schoolnik GK.** 2009. TB database: an integrated platform for tuberculosis research. Nucleic Acids Res **37:** D499–D508. http://dx.doi.org/10.1093/nar/gkn652.
90. **Demay C, Liens B, Burguiere T, Hill V, Couvin D, Millet J, Mokrousov I, Sola C, Zozio T, Rastogi N.** 2012. SITVITWEB–a publicly available international multimarker database for studying Mycobacterium tuberculosis genetic diversity and molecular epidemiology. Infect Genet Evol **12:** 755–766. http://dx.doi.org/10.1016/j.meegid.2012.02.004.
91. **Joshi KR, Dhiman H, Scaria V.** 2014. tbvar: a comprehensive genome variation resource for Mycobacterium tuberculosis. Database (Oxford) **2014:**bat083. http://dx.doi.org/10.1093/database/bat083.
92. **Chernyaeva EN, Shulgina MV, Rotkevich MS, Dobrynin PV, Simonov SA, Shitikov EA, Ischenko DS, Karpova IY, Kostryukova ES, Ilina EN, Govorun VM, Zhuravlev VY, Manicheva OA, Yablonsky PK, Isaeva YD, Nosova EY, Mokrousov IV, Vyazovaya AA, Narvskaya OV, Lapidus AL, O'Brien SJ.** 2014. Genome-wide Mycobacterium tuberculosis variation (GMTV) database: a new tool for integrating sequence variations and epidemiology. BMC Genomics **15:**308. http://dx.doi.org/10.1186/1471-2164-15-308.
93. **Winsor GL, Lam DK, Fleming L, Lo R, Whiteside MD, Yu NY, Hancock RE, Brinkman FS.** 2011. Pseudomonas Genome Database: improved comparative analysis and population genomics capability for Pseudomonas genomes. Nucleic Acids Res **39:**D596–D600. http://dx.doi.org/10.1093/nar/gkq869.
94. **Choo SW, Heydari H, Tan TK, Siow CC, Beh CY, Wee WY, Mutha NV, Wong GJ, Ang MY, Yazdi AH.** 2014. VibrioBase: a model for next-generation genome and annotation database development. ScientificWorldJournal **2014:**569324. http://dx.doi.org/10.1155/2014/569324.
95. **Heydari H, Siow CC, Tan MF, Jakubovics NS, Wee WY, Mutha NV, Wong GJ, Ang MY, Yazdi AH, Choo SW.** 2014. CoryneBase: Corynebacterium genomic resources and analysis tools at your fingertips. PLoS One **9:**e86318. http://dx.doi.org/10.1371/journal.pone.0086318.
96. **Choo SW, Ang MY, Fouladi H, Tan SY, Siow CC, Mutha NV, Heydari H, Wee WY, Vadivelu J, Loke MF, Rehvathy V, Wong GJ.** 2014. HelicoBase: a Helicobacter genomic resource and analysis platform. BMC Genomics **15:**600. http://dx.doi.org/10.1186/1471-2164-15-600.
97. **Bleves S, Dunger I, Walter MC, Frangoulidis D, Kastenmuller G, Voulhoux R, Ruepp A.** 2014. HoPaCI-DB: host-Pseudomonas and Coxiella interaction database. Nucleic Acids Res **42:**D671–D676. http://dx.doi.org/10.1093/nar/gkt925.
98. **Nagarajan V, Elasri MO.** 2007. SAMMD: Staphylococcus aureus microarray meta-database. BMC Genomics **8:**351. http://dx.doi.org/10.1186/1471-2164-8-351.
99. **Memisevic V, Kumar K, Cheng L, Zavaljevski N, DeShazer D, Wallqvist A, Reifman J.** 2014. DBSecSys: a database of Burkholderia mallei secretion systems. BMC Bioinformatics **15:**244. http://dx.doi.org/10.1186/1471-2105-15-244.
100. **Johansson KE.** 2014. VetBact - culturing bacteriological knowledge for veterinarians. Vet Rec **174:**162–164. http://dx.doi.org/10.1136/vr.g162.
101. **Masson P, Hulo C, De Castro E, Bitter H, Gruenbaum L, Essioux L, Bougueleret L, Xenarios I, Le Mercier P.** 2013. ViralZone: recent updates to the virus knowledge resource. Nucleic Acids Res **41:**D579–D583. http://dx.doi.org/10.1093/nar/gks1220.
102. **Hayer J, Jadeau F, Deleage G, Kay A, Zoulim F, Combet C.** 2013. HBVdb: a knowledge database for hepatitis B virus. Nucleic Acids Res **41:**D566–D570. http://dx.doi.org/10.1093/nar/gks1022.
103. **Van Doorslaer K, Tan Q, Xirasagar S, Bandaru S, Gopalan V, Mohamoud Y, Huyen Y, McBride AA.** 2013. The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. Nucleic Acids Res **41:**D571–D578. http://dx.doi.org/10.1093/nar/gks984.
104. **Thakur N, Qureshi A, Kumar M.** 2012. VIRsiRNAdb: a curated database of experimentally validated viral siRNA/shRNA. Nucleic Acids Res **40:**D230–D236. http://dx.doi.org/10.1093/nar/gkr1147.
105. **Qureshi A, Thakur N, Monga I, Thakur A, Kumar M.** 2014. VIRmiRNA: a comprehensive resource for experimentally validated viral miRNAs and their targets. Database (Oxford) **2014:**bau103. http://dx.doi.org/10.1093/database/bau103.
106. **Aurrecoechea C, Brestelli J, Brunk BP, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Innamorato F, Iodice J, Kissinger JC, Kraemer ET, Li W, Miller JA, Nayak V, Pennington C, Pinney DF, Roos DS, Ross C, Srinivasamoorthy G, Stoeckert CJ, Jr, Thibodeau R, Treatman C, Wang H.** 2010. EuPathDB: a portal to eukaryotic pathogen databases. Nucleic Acids Res **38:**D415–D419. http://dx.doi.org/10.1093/nar/gkp941.
107. **Stajich JE, Harris T, Brunk BP, Brestelli J, Fischer S, Harb OS, Kissinger JC, Li W, Nayak V, Pinney DF, Stoeckert CJ, Jr, Roos DS.** 2012. FungiDB: an integrated functional genomics database for fungi. Nucleic Acids Res **40:**D675–D681. http://dx.doi.org/10.1093/nar/gkr918.
108. **Aurrecoechea C, Barreto A, Brestelli J, Brunk BP, Caler EV, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Iodice J, Kissinger JC, Kraemer ET, Li W, Nayak V, Pennington C, Pinney DF, Pitts B, Roos DS, Srinivasamoorthy G, Stoeckert CJ, Jr, Treatman C, Wang H.** 2011. AmoebaDB and MicrosporidiaDB: functional genomic resources for Amoebozoa and Microsporidia species. Nucleic Acids Res **39:**D612–D619. http://dx.doi.org/10.1093/nar/gkq1006.
109. **Heiges M, Wang H, Robinson E, Aurrecoechea C, Gao X, Kaluskar N, Rhodes P, Wang S, He CZ, Su Y, Miller J, Kraemer E, Kissinger JC.** 2006. CryptoDB: a Cryptosporidium bioinformatics resource update. Nucleic Acids Res **34:**D419–D422. http://dx.doi.org/10.1093/nar/gkj078.
110. **Gajria B, Bahl A, Brestelli J, Dommer J, Fischer S, Gao X, Heiges M, Iodice J, Kissinger JC, Mackey AJ, Pinney DF, Roos DS, Stoeckert CJ, Jr, Wang H, Brunk BP.** 2008. ToxoDB: an integrated Toxoplasma gondii database resource. Nucleic Acids Res **36:**D553–D556. http://dx.doi.org/10.1093/nar/gkm981.
111. **Aurrecoechea C, Brestelli J, Brunk BP, Carlton JM, Dommer J, Fischer**

S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Miller JA, Morrison HG, Nayak V, Pennington C, Pinney DF, Roos DS, Ross C, Stoeckert CJ, Jr, Sullivan S, Treatman C, Wang H. 2009. GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens Giardia lamblia and Trichomonas vaginalis. Nucleic Acids Res **37:**D526–D530. http://dx.doi.org/10.1093/nar/gkn631.

112. **Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP, Fischer S, Gajria B, Gao X, Gardner MJ, Gingle A, Grant G, Harb OS, Heiges M, Hertz-Fowler C, Houston R, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Logan FJ, Miller JA, Mitra S, Myler PJ, Nayak V, Pennington C, Phan I, Pinney DF, Ramasamy G, Rogers MB, Roos DS, Ross C, Sivam D, Smith DF, Srinivasamoorthy G, Stoeckert CJ, Jr, Subramanian S, Thibodeau R, Tivey A, Treatman C, Velarde G, Wang H.** 2010. TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res **38:**D457–D462. http://dx.doi.org/10.1093/nar/gkp851.

113. **Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Miller JA, Nayak V, Pennington C, Pinney DF, Roos DS, Ross C, Stoeckert CJ, Jr, Treatman C, Wang H.** 2009. PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res **37:**D539–D543. http://dx.doi.org/10.1093/nar/gkn814.

114. **Ren Q, Chen K, Paulsen IT.** 2007. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. Nucleic Acids Res **35:**D274–D279. http://dx.doi.org/10.1093/nar/gkl925.

115. **Saier MH, Jr, Reddy VS, Tamang DG, Vastermark A.** 2014. The transporter classification database. Nucleic Acids Res **42:**D251–D258. http://dx.doi.org/10.1093/nar/gkt1097.

116. **Fichant G, Basse MJ, Quentin Y.** 2006. ABCdb: an online resource for ABC transporter repertories from sequenced archaeal and bacterial genomes. FEMS Microbiol Lett **256:**333–339. http://dx.doi.org/10.1111/j.1574-6968.2006.00139.x.

117. **Tsirigos KD, Bagos PG, Hamodrakas SJ.** 2011. OMPdb: a database of beta-barrel outer membrane proteins from Gram-negative bacteria. Nucleic Acids Res **39:**D324–D331. http://dx.doi.org/10.1093/nar/gkq863.

118. **Wang Y, Huang H, Sun M, Zhang Q, Guo D.** 2012. T3DB: an integrated database for bacterial type III secretion system. BMC Bioinformatics **13:**66. http://dx.doi.org/10.1186/1471-2105-13-66.

119. **Souza RC, del Rosario Quispe Saji G, Costa MO, Netto DS, Lima NC, Klein CC, Vasconcelos AT, Nicolas MF.** 2012. AtlasT4SS: a curated database for type IV secretion systems. BMC Microbiol **12:**172. http://dx.doi.org/10.1186/1471-2180-12-172.

120. **Bi D, Liu L, Tai C, Deng Z, Rajakumar K, Ou HY.** 2013. SecReT4: a web-based bacterial type IV secretion system resource. Nucleic Acids Res **41:**D660–D665. http://dx.doi.org/10.1093/nar/gks1248.

121. **Li J, Yao Y, Xu HH, Hao L, Deng Z, Rajakumar K, Ou HY.** 30 January 2015. SecReT6: a web-based resource for type VI secretion systems found in bacteria. Environ Microbiol http://dx.doi.org/10.1111/1462-2920.12794.

122. **Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B.** 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res **42:**D490–D495. http://dx.doi.org/10.1093/nar/gkt1178.

123. **Duarte M, Jauregui R, Vilchez-Vargas R, Junca H, Pieper DH.** 2014. AromaDeg, a novel database for phylogenomics of aerobic bacterial degradation of aromatics. Database (Oxford) **2014:**bau118. http://dx.doi.org/10.1093/database/bau118.

124. **Bougouffa S, Radovanovic A, Essack M, Bajic VB.** 2014. DEOP: a database on osmoprotectants and associated pathways. Database (Oxford) **2014:**bau100. http://dx.doi.org/10.1093/database/bau100.

125. **Pal C, Bengtsson-Palme J, Rensing C, Kristiansson E, Larsson DG.** 2014. BacMet: antibacterial biocide and metal resistance genes database. Nucleic Acids Res **42:**D737–D743. http://dx.doi.org/10.1093/nar/gkt1252.

126. **Conway KR, Boddy CN.** 2013. ClusterMine360: a database of microbial PKS/NRPS biosynthesis. Nucleic Acids Res **41:**D402–D407. http://dx.doi.org/10.1093/nar/gks993.

127. **Toukach PV.** 2011. Bacterial carbohydrate structure database 3: principles and realization. J Chem Inf Model **51:**159–170. http://dx.doi.org/10.1021/ci100150d.

128. **Bhat AH, Mondal H, Chauhan JS, Raghava GP, Methi A, Rao A.** 2012. ProGlycProt: a repository of experimentally characterized prokaryotic glycoproteins. Nucleic Acids Res **40:**D388–D393. http://dx.doi.org/10.1093/nar/gkr911.

129. **Bretaudeau A, Coste F, Humily F, Garczarek L, Le Corguille G, Six C, Ratin M, Collin O, Schluchter WM, Partensky F.** 2013. CyanoLyase: a database of phycobilin lyase sequences, motifs and functions. Nucleic Acids Res **41:**D396–D401. http://dx.doi.org/10.1093/nar/gks1091.

130. **Lemfack MC, Nickel J, Dunkel M, Preissner R, Piechulla B.** 2014. mVOC: a database of microbial volatiles. Nucleic Acids Res **42:**D744–D748. http://dx.doi.org/10.1093/nar/gkt1250.

131. **Ulrich LE, Zhulin IB.** 2010. The MiST2 database: a comprehensive genomics resource on microbial signal transduction. Nucleic Acids Res **38:**D401–D407. http://dx.doi.org/10.1093/nar/gkp940.

132. **Ortet P, Whitworth DE, Santaella C, Achouak W, Barakat M.** 2015. P2CS: updates of the prokaryotic two-component systems database. Nucleic Acids Res **43:**D536–D541. http://dx.doi.org/10.1093/nar/gku968.

133. **Wynendaele E, Bronselaer A, Nielandt J, D'Hondt M, Stalmans S, Bracke N, Verbeke F, Van De Wiele C, De Tre G, De Spiegeleer B.** 2013. Quorumpeps database: chemical space, microbial origin and functionality of quorum sensing peptides. Nucleic Acids Res **41:**D655–D659. http://dx.doi.org/10.1093/nar/gks1137.

134. **Pachkov M, Balwierz PJ, Arnold P, Ozonov E, van Nimwegen E.** 2013. SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates. Nucleic Acids Res **41:**D214–D220. http://dx.doi.org/10.1093/nar/gks1145.

135. **Cipriano MJ, Novichkov PN, Kazakov AE, Rodionov DA, Arkin AP, Gelfand MS, Dubchak I.** 2013. RegTransBase–a database of regulatory sequences and interactions based on literature: a resource for investigating transcriptional regulation in prokaryotes. BMC Genomics **14:**213. http://dx.doi.org/10.1186/1471-2164-14-213.

136. **Novichkov PS, Kazakov AE, Ravcheev DA, Leyn SA, Kovaleva GY, Sutormin RA, Kazanov MD, Riehl W, Arkin AP, Dubchak I, Rodionov DA.** 2013. RegPrecise 3.0–a resource for genome-scale exploration of transcriptional regulation in bacteria. BMC Genomics **14:**745. http://dx.doi.org/10.1186/1471-2164-14-745.

137. **Turkarslan S, Wurtmann EJ, Wu WJ, Jiang N, Bare JC, Foley K, Reiss DJ, Novichkov P, Baliga NS.** 2014. Network portal: a database for storage, analysis and visualization of biological networks. Nucleic Acids Res **42:**D184–D190. http://dx.doi.org/10.1093/nar/gkt1190.

138. **Taboada B, Ciria R, Martinez-Guerrero CE, Merino E.** 2012. ProOpDB: Prokaryotic Operon DataBase. Nucleic Acids Res **40:**D627–D631. http://dx.doi.org/10.1093/nar/gkr1020.

139. **Kilic S, White ER, Sagitova DM, Cornish JP, Erill I.** 2014. CollecTF: a database of experimentally validated transcription factor-binding sites in Bacteria. Nucleic Acids Res **42:**D156–D160. http://dx.doi.org/10.1093/nar/gkt1123.

140. **Mitra A, Kesarwani AK, Pal D, Nagaraja V.** 2011. WebGeSTer DB–a transcription terminator database. Nucleic Acids Res **39:**D129–D135.

141. **Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW.** 2013. GenBank. Nucleic Acids Res **41:**D36–D42. http://dx.doi.org/10.1093/nar/gks1195.

142. **Ball CA, Brazma A, Causton H, Chervitz S, Edgar R, Hingamp P, Matese JC, Parkinson H, Quackenbush J, Ringwald M, Sansone SA, Sherlock G, Spellman P, Stoeckert C, Tateno Y, Taylor R, White J, Winegarden N.** 2004. Submission of microarray data to public repositories. PLoS Biol **2:**E317.

143. **Erill I.** 2015. Every site counts: submitting transcription factor-binding site information through the CollecTF portal. J Bacteriol **197:**2454–2457. http://dx.doi.org/10.1128/JB.00031-15.