D24–D33 *Nucleic Acids Research, 2020, Vol. 48, Database issue* *Published online 8 November 2019*
doi: 10.1093/nar/gkz913

Database Resources of the National Genomics Data Center in 2020

National Genomics Data Center Members and Partners[*,†]

Received September 15, 2019; Revised September 30, 2019; Editorial Decision October 01, 2019; Accepted October 02, 2019

ABSTRACT

The National Genomics Data Center (NGDC) provides a suite of database resources to support worldwide research activities in both academia and industry. With the rapid advancements in higher-throughput and lower-cost sequencing technologies and accordingly the huge volume of multi-omics data generated at exponential scales and rates, NGDC is continually expanding, updating and enriching its core database resources through big data integration and value-added curation. In the past year, efforts for update have been mainly devoted to BioProject, BioSample, GSA, GWH, GVM, NONCODE, LncBook, EWAS Atlas and IC4R. Newly released resources include three human genome databases (*PGG*.SNV, *PGG*.Han and CGVD), eLMSG, EWAS Data Hub, GWAS Atlas, iSheep and PADS Arsenal. In addition, four web services, namely, eGPS Cloud, BIG Search, BIG Submission and BIG SSO, have been significantly improved and enhanced. All of these resources along with their services are publicly accessible at https://bigd.big.ac.cn.

INTRODUCTION

The National Genomics Data Center (NGDC), officially approved by the Ministry of Science & Technology and the Ministry of Finance of the People's Republic of China in June 2019, is a national-level center dedicated to advancing life and health sciences by archiving, managing and processing a wide range of genomics related data. NGDC is established based on the BIG Data Center (1–3) at Beijing Institute of Genomics (BIG) of Chinese Academy of Sciences (CAS), jointly in close collaboration with two CAS institutions, namely, Institute of Biophysics (IBP) and Shanghai Institute of Nutrition and Health (SINH). Considering the rapid advancements in higher-throughput and lower-cost sequencing technologies, huge amounts of multi-omics data are generated at ever-growing rates and scales. Therefore, the primary mission of NGDC is to build archive platforms and information systems, develop advanced algorithms and tools to translate big data into big discovery, and provide open access to a suite of database resources in support of research activities of global users from both academia and industry.

During the past year, NGDC has expanded, updated and enriched the amount and type of data through big data integration and value-added curation, particularly by close collaboration with IBP and SINH, with significant improvements and advances over the previous release. In terms of data attribute and curation intensity, database resources in NGDC can be generally divided into three categories: Data—raw sequence data and metadata, Information—value-added standardized information, and Knowledge—curated knowledge and knowledge graphs. Here, we provide a brief summary of new developments and recent updates, and describe the core resources and services of NGDC (Figure 1). All resources, along with their services, are publicly accessible through the home page of NGDC at https://bigd.big.ac.cn.

NEW DEVELOPMENTS

Human genome resources

***PGG*.SNV** (http://www.pggsnv.org) (4) is a human genome database, which gives much higher weight to previously under-investigated indigenous populations in Asia, as these genomes harbor an enormous number of variants that have not been observed in the extensively studied populations of European ancestry. In the current version, *PGG*.SNV archives 265 million single nucleotide variants (SNVs) across 220 147 present-day human genomes and 1018 ancient genomes and estimates their frequencies in 977 diverse populations, including 1009 newly sequenced genomes rep-

[*]To whom correspondence should be addressed: Zhang Zhang. Tel: +86 10 84097261; Fax: +86 10 84097720; Email: zhangzhang@big.ac.cn
Correspondence may also be addressed to Wenming Zhao. Email: zhaowm@big.ac.cn
Correspondence may also be addressed to Jingfa Xiao. Email: xiaojingfa@big.ac.cn
Correspondence may also be addressed to Yiming Bao. Email: baoym@big.ac.cn
Correspondence may also be addressed to Shunmin He. Email: heshunmin@ibp.ac.cn
Correspondence may also be addressed to Guoqing Zhang. Email: gqzhang@picb.ac.cn
Correspondence may also be addressed to Yixue Li. Email: yxli@sibs.ac.cn
Correspondence may also be addressed to Guoping Zhao. Email: gpzhao@sibs.ac.cn
Correspondence may also be addressed to Runsheng Chen. Email: crs@sun5.ibp.ac.cn
[†]Full list provided in the Appendix.

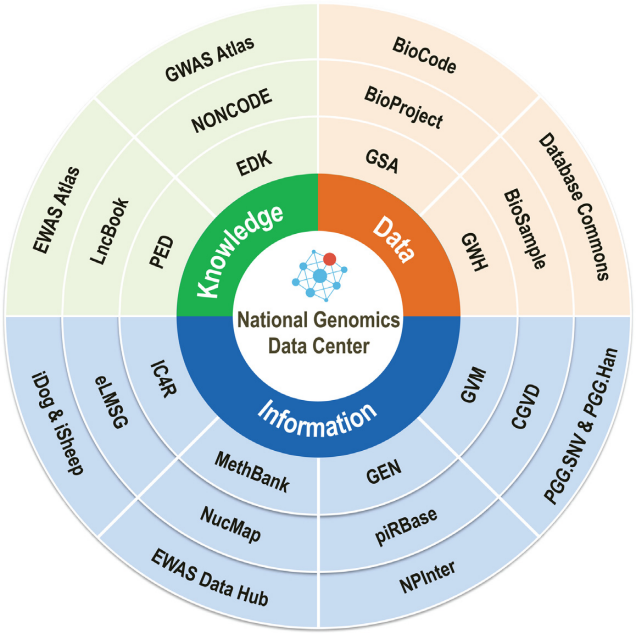

Figure 1. The National Genomics Data Center's core data resources. Three categories, namely, data, information and knowledge, are adopted to represent resources that are typically to deposit raw data/metadata (archives), house value-added information (databases) and integrate validated knowledge through literature curation (knowledgebases), respectively. It is noted that there are several databases that are not introduced in this report, namely, BioCode—Biological Tool Codes, GEN—Gene Expression Nebulas, iDog—Integrated Resource for Dog. A full list of data resources, which contains links to each resource, is available at https://bigd.big.ac.cn/databases.

resenting 16 indigenous populations living in unusual environments (e.g. tropical forests and highlands) in East Asia and Southeast Asia. For each variant, *PGG*.SNV provides various approaches to query SNV information and nine types of annotations. In addition, *PGG*.SNV offers users-friendly interfaces for data browsing and search and is equipped with an online tool for estimation of population genetic diversity and evolutionary parameters.

***PGG*.Han** (http://www.pgghan.org) (detailed in (5) in this issue) is a population genome database, which serves as the central repository of genomic data of the Han Chinese Genomes Initiative (Phase I). *PGG*.Han archives whole-genome sequencing or high-density genome-wide SNVs of 114 783 Han Chinese individuals (a.k.a. the Han100K), representing geographical sub-populations covering 33 of the 34 administrative divisions of China, as well as Singapore. *PGG*.Han provides: (i) an interactive interface for visualization of the fine-scale genetic structure of the Han Chinese population; (ii) genome-wide allele frequency of hierarchical sub-populations; (iii) ancestry inference for individual samples and controlling population stratification based on nested ancestry informative marker panels; (iv) a population-structure-aware shared control for genotype–phenotype association studies and (v) a Han-Chinese-specific reference panel for genotype imputation. Computational tools are implemented in *PGG*.Han and an online user-friendly interface is provided for data analysis and visualization.

The Chinese Genomic Variation Database (CGVD; https://bigd.big.ac.cn/cgvd) (detailed in (6) in this issue) is a genomic variation database for Chinese populations. CGVD is a sub-project of the CAS Precision Medicine Initiative project (CASPMI) (7), with the aim to establish the CAS professional cohort with whole-genome deep sequencing (25–30×) and build precise reference genomes for different Chinese sub-populations. In comparison with *PGG*.Han, CGVD features high-coverage sequencing data of 991 individuals of the CASPMI cohort and 301 Chinese individuals from the 1000 Genome Project (1KGP). Accordingly, it houses genomic variations of 48.30 million SNVs and 5.77 million small indels; in contrast to dbSNP (8), 28.49 million (46.67%) SNVs and 2.25 million (31.88%) indels are novel, indicating the advantage of deeper whole-genome sequencing coverage or/and the heterogeneity of genetic background in Chinese populations. Moreover, CGVD provides star-allele frequencies of drug metabolism related genes that are essential for pharmacogenomics studies in CASPMI and 1KGP related populations. It also integrates curated knowledge of genomic variation impacts on drug absorption, distribution, metabolism, excretion and toxicity.

GWAS Atlas

GWAS Atlas (https://bigd.big.ac.cn/gwas) (detailed in (9) in this issue) is a manually curated resource of genome-wide variant-trait associations in plants and animals. In the current version, GWAS Atlas contains 75 467 variant-trait associations for 614 traits across seven cultivated plants (cotton, Japanese apricot, maize, rapeseed, rice, sorghum and soybean) and two domesticated animals (goat and pig), which were manually extracted and curated from 254 publications. More importantly, associations and traits are annotated and presented based on a set of ontologies (Plant Trait Ontology, Animal Trait Ontology for Livestock, etc.). Taken together, GWAS Atlas integrates high-quality curated GWAS associations for animals and plants and accordingly serves as a valuable resource for genetic research of important traits and breeding application.

EWAS Data Hub

Over the past decade, a large amount of epigenetic data, especially those sourced from DNA methylation array, has been accumulated as a result of numerous EWAS (epigenome-wide association study) projects. Hence, we present EWAS Data Hub (https://bigd.big.ac.cn/ewas/datahub) (detailed in (10) in this issue), a data hub for collecting and normalizing DNA methylation array data as well as archiving associated metadata. The current release of EWAS Data Hub integrates a comprehensive collection of DNA methylation array data from 75 344 samples. Based on an effective normalization method to remove batch effects among different datasets, EWAS Data Hub provides high-quality reference DNA methylation profiles in terms of different contexts, involving 81 tissues/cell types (that contain 25 brain parts and 25 blood cell types), six ancestry categories, and 67 diseases (including 39 cancers).

iSheep

iSheep (https://bigd.big.ac.cn/isheep) is a specialized genomics resource for sheep (*Ovis aries*), providing a wealth of information on genotype and phenotype association, domestication and climatic adaptation of domestic sheep as well as their wild relatives. The current version of iSheep houses 70 390 968 unique SNPs and 12 318 530 indels obtained from 2777 samples (including 355 samples with whole-genome sequences, 1512 samples with 50K-BeadChip and 911 samples with 600K-BeadChip) and provides comprehensive phenotypic information of 1459 worldwide sheep breeds. Meanwhile, iSheep offers an online tool to investigate the variations between individuals or among populations. Collectively, iSheep is a valuable genomics resource for the sheep research community, helpful to promote molecular breeding and farming industry for improved production traits.

eLMSG

eLMSG (eLibrary of Microbial Systematics and Genomics; http://www.biosino.org/elmsg) is a web microbial library that integrates not only taxonomic information, but also genomic information and phenotypic information (including morphology, physiology, biochemistry and enzymology). The taxonomic system of eLMSG is manually curated and composed of all validly and some effectively published taxa. For each taxon, the Latin name, taxon ID (NCBI taxonomy), etymology, rank, lineage, the dates of effective and/or valid publication, feature descriptions, nomenclature type and references for the proposal and emendations during the history of the taxon are presented. Besides these data, the species taxa contain information about 16S rRNA gene and/or genome sequences. All publicly available genome data of each type species including both type and non-type strains were collected, and if needed, re-annotated using the standardized analysis pipeline. Furthermore, pan-genomic data analyses were conducted for species with $\geq$5 genome sequences available. Finally, for all type species, taxonomically relevant phenotypic data were extracted and curated from literatures, which were further indexed into eLMSG as searchable and analyzable data records. Taken together, eLMSG is a comprehensive web platform for studying microbial systematics and genomics, potentially useful for better understanding microbial taxonomy, natural evolutionary processes and ecological relationships.

PADS Arsenal

PADS Arsenal (https://bigd.big.ac.cn/padsarsenal) (detailed in this issue) is a comprehensive public database of prokaryotic defense systems related genes (PADS). To address the challenges of ever-increasing prokaryotic genomic data and the progressive discovery of novel defense systems, we develop PADS Arsenal for browsing, searching, and analyzing various defense system genes. In the current version, PADS Arsenal integrates 6 600 264 defense systems genes, which belong to 18 defense systems, 63 701 genomes and 33 390 species of archaea and bacteria. In addition, it supports defense system gene analysis by equipping with an interactive online pipeline that includes sequence homology search, multiple sequence alignment and phylogenetic analysis. Meanwhile, PADS Arsenal provides a presence-absence variation (PAV) analysis function to visualize the dynamic variation of defense system genes. Collectively, PADS Arsenal integrates a comprehensive collection of defense system genes in archaea and bacteria and thus provides valuable resources to facilitate development of novel genome editing, engineering and regulation tools.

RECENT UPDATES

BioProject and BioSample

BioProject (https://bigd.big.ac.cn/bioproject) and BioSample (https://bigd.big.ac.cn/biosample), designed in compliance with INSDC (International Nucleotide Sequence Database Collaboration; a joint initiative by DDBJ, EMBL-EBI and NCBI) standards, are two public repositories of biological projects and biological samples, respectively. They collect and store descriptive metadata and information about biological projects and biological materials used for experiments. By providing a centralized access to all public projects and reciprocal links to their related data, BioProject supports various projects in terms of data types, ranging from genomic, transcriptomic, epigenomic and metagenomic sequencing projects to genome-wide association studies (GWAS) and variation analyses. Similarly, BioSample serves as a centralized access to all public samples and reciprocal links to BioProject as well as other relevant database resources. In the past year, BioSample has been significantly upgraded by adding the batch submission functionality and allowing users to submit information of multiple samples in a single table, which consequently had greatly improved the efficiency of data submission. As of August 2019, BioProject houses a total of 1248 biological projects submitted by 734 users from 219 organizations and BioSample includes a total of 87 107 samples from 482 species, presenting a dramatic increase in data submission (Figure 2).

Genome Sequence Archive

As a public data repository for archiving raw sequence reads, the Genome Sequence Archive (GSA; https://bigd.big.ac.cn/gsa) (11) accepts data submissions from all over the world and provides free access to all publicly available data for global scientific communities. Over the past year, GSA has been significantly enhanced by upgrading the metadata submission functionality to enable batch submission of experiments and runs in a single table. Till August 2019, GSA has archived a total of 55 057 Experiments and 59 566 Runs and housed >1200 Terabytes of submitted raw sequence data (Figure 2), showing the doubled volume by comparison with the previous release last August (namely, ~580 TB). According to the statistics (https://bigd.big.ac.cn/gsa/statistics), data housed in GSA were submitted from 150 organizations and reported in >100 scientific journals, including Cell, Genome Research, Genomics Proteomics Bioinformatics, Nature, Plant Cell and PNAS. More importantly, GSA has been designated as supported repository for genes and gene expression data by Elsevier. All released

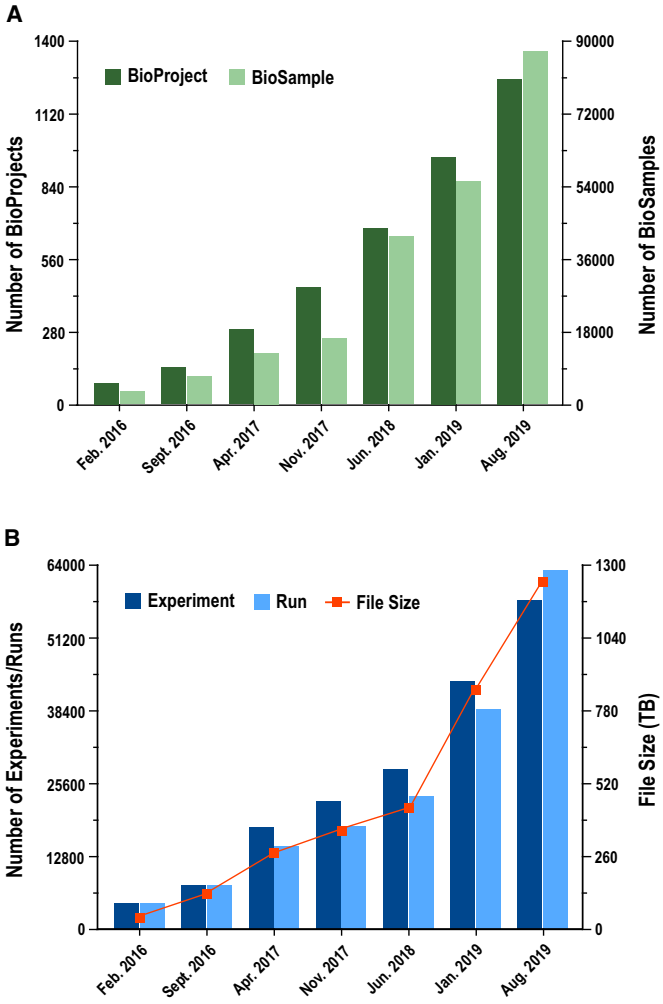

Figure 2. Statistics of data submissions to BioProject, BioSample, and GSA. (**A**) Data statistics of BioProject and BioSample. (**B**) Data statistics of Experiments and Runs as well as submitted files' size in GSA. All statistics are frequently updated and publicly available at https://bigd.big.ac.cn/bioproject, https://bigd.big.ac.cn/biosample and https://bigd.big.ac.cn/gsa.

data in GSA are publicly accessible and downloadable at ftp://download.big.ac.cn/gsa/.

Genome Warehouse

The Genome Warehouse (GWH; https://bigd.big.ac.cn/gwh) is a public archival resource housing genome-scale data for a wide range of species. For each collected genome assembly, GWH incorporates detailed descriptive information, including metadata of biological sample, genome assembly, sequence data and genome annotation, and offers standardized quality control for genome sequence and genome annotation. Notably, in this version, the sequences of the northern Han reference genome (NH1.0; GWHAAAS00000000) has been deposited in GWH, which was *de novo* assembled with a contig N50 size of 3.6 Mb and a scaffold N50 size of 46.63 Mb (see (7) for details). In addition, GWH has been significantly upgraded by accepting updated submissions (including both genome sequence and updates of genome annotation) and improving web services for data submission, release and sharing. In particular, GWH provides data visualization for both genome sequence and genome annotation powered by JBrowse (12) and offers statistics and charts in light of assembly, genome, sequencing platform, assembly method, organization and download. Till September 2019, GWH has accepted 649 data submissions from organizations both nationally and internationally and covered a broad diversity of species, e.g. animals, plants, fungi, bacteria, archaea and viruses. Among them, 133 genome assemblies have been publicly released and reported in 19 international journals.

Genome Variation Map

The Genome Variation Map (GVM; https://bigd.big.ac.cn/gvm) (13) is a public database of genome variations, including single nucleotide polymorphisms (SNP) and small insertions and deletions (indel). Different from dbSNP that only accepts human data submissions, GVM collects genome variations for a wide range of species and accepts submissions of different types of genome variations from all over the world. In the current version, GVM incorporates a total of ∼8.4 billion variants for 13 animals and 19 plants, including 7.2 billion SNPs and 1.2 billion indels. By comparison with the previous version, it has been updated by integrating 47 million variants from two newly added species (diploid wheat and cat). In addition, GVM has accepted 24 genome variation data submissions involving 23 056 samples from 10 species.

Non-coding RNA Resources

NONCODE (http://www.noncode.org) (14) is an integrated knowledgebase dedicated to the complete collection and annotation of non-coding RNAs (ncRNA). Almost all the types of ncRNAs (excluding tRNAs and rRNAs) were filtered automatically from literatures and other public databases and were later manually curated. The ncRNA sequences and their related information (such as chromosomal information, conservation, function, etc.) were collected and recorded. BLAST alignment search service and access through our custom UCSC Genome Browser were also incorporated. In the current version (v5.0), 17 species are included in NONCODE (human, mouse, cow, rat, chicken, fruit fly, zebrafish, nematode, yeast, Arabidopsis, chimpanzee, gorilla, orangutan, rhesus macaque, opossum platypus and pig). Consequently, NONCODE collects a total of 548,640 long ncRNAs (lncRNA), coupled with their expression profiles identified based on RNA-seq data for human and mouse as well as their predicted functions. Moreover, it also includes human lncRNA–disease relationships and SNP–lncRNA–disease relationships, human exosome lncRNA expression profiles and predicted RNA secondary structures of human transcripts.

NPInter (http://bigdata.ibp.ac.cn/npinter) (15) is a database that documents experimentally identified functional interactions between ncRNAs (except tRNAs and rRNAs), especially lncRNAs, and protein related biomacromolecules (proteins, mRNAs or genomic DNAs). NPInter provides the scientific community with a comprehensive and integrated tool for efficient browsing

and extraction of information on interactions between ncRNAs and biomolecules. With the development of high-throughput biotechnology, such as cross-linking immunoprecipitation (CLIP-seq) and Chromatin Isolation by RNA purification (ChIRP-seq), the number of known ncRNA interactions, has grown rapidly in recent years. In the current release, NPInter houses 609 020 RNA-RNA interactions, 488 315 RNA–protein interactions and 892 737 RNA–DNA interactions, and provides more user-friendly interfaces and functional modules.

piRBase (http://www.regulatoryrna.org/database/piRNA/) (16) is a comprehensive database of piRNA sequences, which are a class of small RNAs that is mainly expressed in animal germ line. piRBase integrates various piRNA-related high-throughput data in multiple species, leading to the largest collection of piRNAs and their annotations. Since its launch in 2014, piRBase has incorporated 264 datasets from 21 organisms and accordingly housed a total of ∼173 million piRNAs up to now. Furthermore, piRBase provides comprehensive annotations of piRNA sequences and genomic loci as well as piRNA targets and disease-related piRNAs. In addition, epigenetic and post-transcriptional regulation data were systematically integrated to support piRNA functional study.

LncBook (17) (https://bigd.big.ac.cn/lncbook) and **LncRNAWiki** (18) (https://bigd.big.ac.cn/lncrnawiki), are two dedicated resources of human lncRNAs, through expert curation and community curation, respectively. In the past year, LncBook has been updated by removing 1196 redundant lncRNA transcripts and updating genomic annotations of 1046 lncRNA transcripts. As a result, LncBook provides a high-quality collection of 268 848 non-redundant lncRNA transcripts and 140 356 lncRNA genes. Also, LncBook presents tissue-specific lncRNAs (TS lncRNAs) for different tissues; among the 32 tissues, testis has the largest number of TS lncRNAs (9024 lncRNAs) and the following tissue is brain (2297 lncRNAs). In addition, LncBook is equipped with an online tool for coding potential prediction, which is able to accurately identify lncRNAs in a wide range of species (19). On the other side, LncRNAWiki (18), a wiki-based platform for community curation of human lncRNAs, has been updated by curating 291 human lncRNAs with functional experiment evidence, including 149 newly added lncRNAs and 142 existing lncRNAs with updated publications. Also, 65 redundant lncRNAs based on the approved and alias symbols (https://www.genenames.org) were removed. Consequently, in the current release, the number of functionally validated human lncRNAs in LncRNAWiki has grown to 1951. Together, LncBook and LncRNAWiki are of great potential to achieve comprehensive integration of human lncRNAs and their annotations (20).

RNA Editing Resources

Editome Disease Knowledgebase (EDK; https://bigd.big.ac.cn/edk) (21) and Plant Editosome Database (PED; https://bigd.big.ac.cn/ped) (22) are two RNA editing resources for human and plants, respectively. In the updated version, EDK incorporates two new diseases associated with 51 experimentally validated abnormal editing events located in six mRNAs, and 10 aberrant activities involved with two editing enzymes. Furthermore, to provide an easy-to-use and downloadable reference for further functional investigation on individual RNA editing event, EDK incorporates detailed structured annotation information for each editing site, including gene, specific gene region, molecular effect, editing enzyme, associated disease and/or phenotype. As a featured database of RNA editosome in plants (22,23), PED has been updated by integrating two more editing factors, which had been recently verified to be involved in RNA editing processes and related to important phenotypes in Arabidopsis and new maize variety. Collectively, EDK and PED integrate more valuable information of editing enzymes (factors) and/or editing events associated with phenotypes, so as to help users facilitate systematic investigations on RNA editing machinery in both human and plants.

MethBank

The Methylation Bank (MethBank; https://bigd.big.ac.cn/methbank) (24,25) is a databank of genome-wide DNA methylomes across a variety of species, with particular focus on human health and aging, animal embryonic development and plant growth and development. In the current version, MethBank offers 43 consensus reference methylomes (CRM) for human owing to large-scale DNA methylation array data public available, which are sourced from 10 healthy human tissues including 4577 peripheral blood samples, 26 prostate samples, 241 saliva samples, 322 skin samples, 98 breast samples, 38 colon samples, 206 kidney samples, 50 liver samples, 150 lung samples and 56 thyroid samples. In addition to CRMs, MethBank provides single-base resolution methylomes (SRM) based on whole-genome bisulfite sequencing data from human, plants and animals. Up to now, MethBank includes 40 SRMs from 26 healthy human tissues, 336 from different developmental stages in five economical plants and 18 from gametes and early embryos in two model animals. In addition, MethBank provides useful information on methylation data analysis tools, helpful for users to easily find any tool of interest.

EWAS Atlas

EWAS Atlas (https://bigd.big.ac.cn/ewas) (26) is a curated knowledgebase of epigenome-wide association studies. During the past year, it has been enriched by adding a total of 121 156 EWAS associations manually extracted and curated from 191 publications. It is noted that the MethylationEPIC (850K/EPIC) array becomes increasingly popular, so that the number of 850K-based publications in EWAS Atlas has increased accordingly. In addition, the online trait enrichment tool was further enhanced and EWAS knowledge graph (https://bigd.big.ac.cn/ewas/network) was newly developed to visualize and explore trait-gene networks. Till September 2019, EWAS Atlas has integrated 450 328 high-quality EWAS associations derived from 1003 studies in 401 publications, including 135 tissues/cell lines, 409 traits, 2689 cohorts and 409 ontology entities.

Information Commons for Rice

Information Commons for Rice (IC4R; http://ic4r.org) (27,28) is a comprehensive resource dedicated to integrating multi-omics data for rice. To improve the completeness of gene structure and identify novel genes, the current implementation of IC4R incorporates a new gene annotation system IC4R-2.0 that is built based on a large number of 1503 public RNA-seq datasets, accordingly achieving higher integrity and quality by comparison with previous annotation systems. Specifically, IC4R-2.0 contains 56,221 protein-coding gene loci corresponding to 80 039 mRNAs, among which more than 27 000 gene loci are substantially improved with structural modification, 456 novel genes are identified, and 3215 lncRNAs and 4373 circular RNAs are annotated. In addition, although IC4R offers a high-density rice variation map of ∼18 million SNPs, these raw SNPs are not readily usable for population genetics, evolutionary analysis, association studies or genomic breeding in rice. To satisfy various needs of rice researchers on data mining of the integrated genotypic data, a committed module—SnpReady for Rice (SR4R, http://sr4r.ic4r.org), is developed and deployed in IC4R. SR4R features the lowest SNP redundancy and highest genetic diversity of rice populations. Currently, SR4R mainly integrates four reference SNP panels, including 'hapmapSNPs' after data filtration and genotype imputation, 'tagSNPs' selected from linkage disequilibrium (LD)-based redundancy removal, 'fixedSNPs' selected from genes exhibiting selective sweep signatures, and 'barcodeSNPs' selected from DNA fingerprinting simulation. The associated SNPs in these four panels as well as online toolkits are publicly available and downloadable.

LSD

The leaf senescence database (LSD; https://bigd.big.ac.cn/lsd) (29,30) is dedicated to the comprehensive collection of senescence-associated genes (SAGs) and their corresponding mutants through manual curation. In the current version (v3.0; see an update in (31) in this issue), LSD incorporates 5,853 SAGs and 617 mutants from 68 species. Notably, it integrates leaf senescence-associated transcriptome data in *Arabidopsis*, rice, soybean and poplar and identifies senescence-differentially expressed small RNAs (Sen-smRNA) in *Arabidopsis*. Moreover, LSD contains senescence phenotypes of 90 natural accessions (ecotypes) and 42 images of ecotypes in *Arabidopsis* and collects mutant seed information of *SAG*s in rice. Also, interaction pairs between Sen-smRNAs and senescence-associated transcription factors are integrated into LSD. Collectively, the updated LSD has the great potential to continue to provide useful information for the plant research community.

Database Commons

Database Commons (https://bigd.big.ac.cn/databasecommons), a catalog of global biological databases, provides open access to a comprehensive collection of publicly available databases and their descriptive metadata. Currently, it catalogues a total of 4615 databases, involving more than 7000 publications and ∼2000 organizations throughout the world. In the past year, Database Commons has been updated by assigning category tag(s) to each database, linking related databases and providing citation information according to Europe PMC (32). Importantly, to improve the quality of descriptive metadata for each database, we sent invitations to database owners (according to the publications) to call for community curation of their own databases. As a result, a total of 287 database owners have responded and made valuable curations to 345 databases.

eGPS Cloud

eGPS Cloud (http://egpscloud.big.ac.cn) (33) is a multifunctional web portal that integrates comprehensive multi-omics tools and provides online data analysis services for studying evolutionary Genotype-Phenotype Systems (eGPS). In the current release, eGPS Cloud is equipped with 15 tools and 20 visualization scripts, accordingly delivering four modularized web services, that is, genomics data analysis, population data analysis, evolutionary & network data analysis, and multi-omics data visualization. It allows users to configure customized parameters for different tools and perform various data analysis online in a straightforward and friendly manner. Ongoing efforts are linking eGPS Cloud with GSA in order to provide users with seamless services for raw sequence data analysis.

BIG Search

BIG Search (https://bigd.big.ac.cn/search) is a distributed and scalable full-text search engine built based on Elasticsearch (a highly scalable open-source search and analytics engine, https://www.elastic.co/). It features cross-domain search and facilitates users to gain access to a wide range of biological data almost in real-time. In the current version, BIG Search includes data indexes from all NGDC's resources and 25 partner resources (see details at https://bigd.big.ac.cn/partners). Additionally, EBI data resources have also been integrated into BIG Search powered by EBI Search RESTful API (34). In summary, BIG Search has been significantly updated by incorporating more data indexes from internal and external resources and displaying search results in a more user-friendly manner.

BIG Submission

BIG Submission (https://bigd.big.ac.cn/gsub) is a one-stop submission portal that provides submission services for a series of database resources in NGDC, including BioProject, BioSample, GSA, GWH and GVM. During the past year, BIG Submission has been upgraded by optimizing the web interfaces and expanding the storage and computing resources, with the purpose to meet the needs of the rapid growth of data submissions. Importantly, it has been equipped by Aspera, a high-speed transfer tool that can greatly improve the data transfer efficiency and provide users with better submission experiences.

BIG SSO

BIG Single Sign-On (SSO; https://bigd.big.ac.cn/sso) is a user access control system that refers to systems where a sin-

gle authentication provides access to multiple applications by passing the authentication token seamlessly to configured applications. In the past year, HTTPS protocols have been deployed in all web sites for security transfer, so that the BIG SSO system has been updated to be much safer and more reliable. Meanwhile, services for user registration and update have been enhanced and delivered as a micro-service.

CONCLUDING REMARKS

NGDC provides a family of database resources through big data deposition, integration and translation, with the aim to support worldwide research activities in both academia and industry. In the past year, it has been significantly updated by archiving more data submissions, performing value-added curation, and improving web interfaces and services. And most importantly, it has been enhanced as the national center by joint efforts from BIG, IBP and SINH, forming an excellent line-up of field experts from the three institutions. Ongoing and future efforts are standardization of data models and curation processes, unification of web interfaces and SSO authentication across database resources, establishment of cloud infrastructure for big data storage and transfer, and development of a variety of databases and tools to facilitate the translation of big data into big discovery. NGDC is open to worldwide collaborations, particularly seeking the possibility to collaborate with INSDC members in dealing with big data archive. In addition, NGDC promotes big data sharing at a worldwide scale by setting up the Global Biodiversity and Health Big Data Alliance (BHBD; http://bhbd-alliance.org); by July 2019, 20 organizational members from 11 countries have joined the BHBD Alliance, with active collaborations in organizing international meetings/symposia, training courses and joint research projects. With more stable support from the government and CAS, NGDC will continue to grow to deliver a wide range of data resources and services in aid of both domestic and international research activities.

ACKNOWLEDGEMENTS

We thank a number of users for submitting data, sending suggestions, reporting bugs and getting involving in community curation. The National Genomics Data Center is indebted to its funders, including the Ministry of Science & Technology and the Ministry of Finance of the People's Republic of China as well as Chinese Academy of Sciences. We would like to express our sincere thanks to the late Professor Bailin Hao (1934–2018), a leading bioinformatician of his generation, who had first advocated the establishment of national center since the 1990s.

FUNDING

Strategic Priority Research Program of the Chinese Academy of Sciences [XDA19050302, XDB13040500, XDB13040100]; National Key Research & Development Program of China [2018YFD1000505, 2018YFC2000100, 2018YFC1406902, 2018YFC0910400, 2018YFC0310602, 2017YFC1201200, 2017YFC0908405, 2017YFC0908404, 2017YFC0908403, 2017YFC0907505, 2017YFC0907503, 2017YFC0907502, 2016YFE0206600, 2016YFC0906403, 2016YFC0903003, 2016YFC0901904, 2016YFC0901903, 2016YFC0901702, 2016YFC0901604, 2016YFC0901603, 2016YFB0201702]; National Natural Science Foundation of China [91731303, 81670462, 31970565, 31871328, 31871294, 31801104, 31771465, 31771410, 31771388, 31671360, 31571358, 31525014, 1470330, 31961130380, 31711530221]; UK Royal Society-Newton Advanced Fellowship [NAF\R1\191094]; International Partnership Program of the Chinese Academy of Sciences [153F11KYSB20160008, 153D31KYSB20170121]; 13th Five-year Informatization Plan of Chinese Academy of Sciences [XXH13505-05]; Key Program of the Chinese Academy of Sciences [KJZD-EW-L14]; Key Research Program of Frontier Sciences of the Chinese Academy of Sciences [QYZDJ-SSW-SYS009]; Key Technology Talent Program of the Chinese Academy of Sciences; The 100 Talent Program of the Chinese Academy of Sciences; K.C. Wong Education Foundation; The Youth Innovation Promotion Association of the Chinese Academy of Sciences [2019104, 2018134, 2017141]; The Special Project on Precision Medicine under the National Key R&D Program [SQ2017YFSF090210]; The Open Biodiversity and Health Big Data Initiative of IUBS. Funding for open access charge: Strategic Priority Research Program of the Chinese Academy of Sciences.

Conflict of interest statement. None declared.

REFERENCES

1. BIG Data Center Members (2017) The BIG Data Center: from deposition to integration to translation. *Nucleic Acids Res.*, **45**, D18–D24.
2. BIG Data Center Members (2018) Database resources of the BIG data center in 2018. *Nucleic Acids Res.*, **46**, D14–D20.
3. BIG Data Center Members (2019) Database resources of the BIG data center in 2019. *Nucleic Acids Res.*, **47**, D8–D14.
4. Zhang,C., Gao,Y., Ning,Z., Lu,Y., Zhang,X., Liu,J., Xie,B., Xue,Z., Wang,X., Yuan,K. *et al.* (2019) PGG.SNV: Understanding the evolutionary and medical implications of human single nucleotide variations in diverse populations. *Genome Biol.*, doi:10.1186/s13059-019-1838-5.
5. Gao,Y., Zhang,C., Yuan,L., Ling,Y., Wang,X., Liu,C., Pan,Y., Zhang,X., Ma,X., Wang,Y. *et al.* (2020) PGG.Han: The Han Chinese Genome Database and analysis platform. *Nucleic Acids Res.*, doi:10.1093/nar/gkz829.
6. Zeng,J., Yuan,N., Zhu,J., Pan,M., Zhang,H., Wang,Q., Shi,S., Du,Z. and Xiao,J. (2019) CGVD: a genomic variation database for Chinese populations. *Nucleic Acids Res.*, doi:10.1093/nar/gkz952.
7. Du,Z., Ma,L., Qu,H., Chen,W., Zhang,B., Lu,X., Zhai,W., Sheng,X., Sun,Y., Li,W. *et al.* (2019) Whole genome analyses of chinese population and De Novo assembly of a northern han genome. *Genomics Proteomics Bioinform.*, **17**, 229–247.
8. Sherry,S.T., Ward,M.H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M. and Sirotkin,K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
9. Tian,D., Wang,P., Tang,B.-X., Teng,X., Li,C., Liu,X., Zou,D., Song,S. and Zhang,Z. (2019) GWAS Atlas: a curated resource of genome-wide variant-trait associations in plants and animals. *Nucleic Acids Res.*, doi:10.1093/nar/gkz828.
10. Xiong,Z., Li,M., Yang,F., Ma,Y., Sang,J., Li,R., Li,Z., Zhang,Z. and Bao,Y.-M. (2019) EWAS Data Hub: a resource of DNA methylation array data and metadata. *Nucleic Acids Res.*, doi:10.1093/nar/gkz840.
11. Wang,Y., Song,F., Zhu,J., Zhang,S., Yang,Y., Chen,T., Tang,B., Dong,L., Ding,N., Zhang,Q. *et al.* (2017) GSA: Genome Sequence Archive. *Genomics Proteomics Bioinform.*, **15**, 14–18.

12. Buels,R., Yao,E., Diesh,C.M., Hayes,R.D., Munoz-Torres,M., Helt,G., Goodstein,D.M., Elsik,C.G., Lewis,S.E., Stein,L. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.
13. Song,S., Tian,D., Li,C., Tang,B., Dong,L., Xiao,J., Bao,Y., Zhao,W., He,H. and Zhang,Z. (2018) Genome Variation Map: a data repository of genome variations in BIG Data Center. *Nucleic Acids Res.*, **46**, D944–D949.
14. Fang,S., Zhang,L., Guo,J., Niu,Y., Wu,Y., Li,H., Zhao,L., Li,X., Teng,X., Sun,X. *et al.* (2018) NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.*, **46**, D308–D314.
15. Hao,Y., Wu,W., Li,H., Yuan,J., Luo,J., Zhao,Y. and Chen,R. (2016) NPInter v3.0: an upgraded database of noncoding RNA-associated interactions. *Database (Oxford)*, **2016**, baw057.
16. Wang,J., Zhang,P., Lu,Y., Li,Y., Zheng,Y., Kan,Y., Chen,R. and He,S. (2019) piRBase: a comprehensive database of piRNA sequences. *Nucleic Acids Res.*, **47**, D175–D180.
17. Ma,L., Cao,J., Liu,L., Du,Q., Li,Z., Zou,D., Bajic,V.B. and Zhang,Z. (2019) LncBook: a curated knowledgebase of human long non-coding RNAs. *Nucleic Acids Res.*, **47**, D128–D134.
18. Ma,L., Li,A., Zou,D., Xu,X., Xia,L., Yu,J., Bajic,V.B. and Zhang,Z. (2015) LncRNAWiki: harnessing community knowledge in collaborative curation of human long non-coding RNAs. *Nucleic Acids Res.*, **43**, D187–D192.
19. Wang,G., Yin,H., Li,B., Yu,C., Wang,F., Xu,X., Cao,J., Bao,Y., Wang,L., Abbasi,A.A. *et al.* (2019) Characterization and identification of long non-coding RNAs based on feature relationship. *Bioinformatics*, **35**, 2949–2956.
20. Ma,L., Cao,J., Liu,L., Li,Z., Shireen,H., Pervaiz,N., Batool,F., Raza,R.Z., Zou,D., Bao,Y. *et al.* (2019) Community curation and expert curation of human long noncoding RNAs with LncRNAWiki and LncBook. *Curr. Protoc. Bioinform.*, **67**, e82.
21. Niu,G., Zou,D., Li,M., Zhang,Y., Sang,J., Xia,L., Li,M., Liu,L., Cao,J., Zhang,Y. *et al.* (2019) Editome Disease Knowledgebase (EDK): a curated knowledgebase of editome-disease associations in human. *Nucleic Acids Res.*, **47**, D78–D83.
22. Li,M., Xia,L., Zhang,Y., Niu,G., Li,M., Wang,P., Zhang,Y., Sang,J., Zou,D., Hu,S. *et al.* (2019) Plant editosome database: a curated database of RNA editosome in plants. *Nucleic Acids Res.*, **47**, D170–D174.
23. Lo Giudice,C., Hernandez,I., Ceci,L.R., Pesole,G. and Picardi,E. (2019) RNA editing in plants: A comprehensive survey of bioinformatics tools and databases. *Plant Physiol. Biochem.*, **137**, 53–61.
24. Li,R., Liang,F., Li,M., Zou,D., Sun,S., Zhao,Y., Zhao,W., Bao,Y., Xiao,J. and Zhang,Z. (2018) MethBank 3.0: a database of DNA methylomes across a variety of species. *Nucleic Acids Res.*, **46**, D288–D295.
25. Zou,D., Sun,S., Li,R., Liu,J., Zhang,J. and Zhang,Z. (2015) MethBank: a database integrating next-generation sequencing single-base-resolution DNA methylation programming data. *Nucleic Acids Res.*, **43**, D54–D58.
26. Li,M., Zou,D., Li,Z., Gao,R., Sang,J., Zhang,Y., Li,R., Xia,L., Zhang,T., Niu,G. *et al.* (2019) EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res.*, **47**, D983–D988.
27. IC4R Project Consortium. (2016) Information Commons for Rice (IC4R). *Nucleic Acids Res.*, **44**, D1172–D1180.
28. Xia,L., Zou,D., Sang,J., Xu,X., Yin,H., Li,M., Wu,S., Hu,S., Hao,L. and Zhang,Z. (2017) Rice Expression Database (RED): an integrated RNA-Seq-derived gene expression database for rice. *J. Genet. Genomics*, **44**, 235–241.
29. Li,Z., Zhao,Y., Liu,X., Peng,J., Guo,H. and Luo,J. (2014) LSD 2.0: an update of the leaf senescence database. *Nucleic Acids Res.*, **42**, D1200–D1205.
30. Liu,X., Li,Z., Jiang,Z., Zhao,Y., Peng,J., Jin,J., Guo,H. and Luo,J. (2011) LSD: a leaf senescence database. *Nucleic Acids Res.*, **39**, D1103–D1107.
31. Li,Z., Zhang,Y., Zou,D., Zhao,Y., Wang,H.-L., Zhang,Y., Xia,X., Luo,J., Guo,H. and Zhang,Z. (2019) LSD 3.0: a comprehensive resource for the leaf senescence research community. *Nucleic Acids Res.*, doi:10.1093/nar/gkz898.
32. Levchenko,M., Gou,Y., Graef,F., Hamelers,A., Huang,Z., Ide-Smith,M., Iyer,A., Kilian,O., Katuri,J., Kim,J.H. *et al.* (2018) Europe PMC in 2017. *Nucleic Acids Res.*, **46**, D1254–D1260.
33. Yu,D., Dong,L., Yan,F., Mu,H., Tang,B., Yang,X., Zeng,T., Zhou,Q., Gao,F., Wang,Z. *et al.* (2019) eGPS 1.0: comprehensive software for multi-omic and evolutionary analyses. *Natl. Sci. Rev.*, doi:10.1093/nsr/nwz079.
34. Madeira,F., Park,Y.M., Lee,J., Buso,N., Gur,T., Madhusoodanan,N., Basutkar,P., Tivey,A.R.N., Potter,S.C., Finn,R.D. *et al.* (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.*, **47**, W636–W641.

APPENDIX

Corresponding author: Zhang Zhang[1,2,3,10,11,*]

Co-corresponding authors: Wenming Zhao[1,2,3,10,*], Jingfa Xiao[1,2,3,10,*], Yiming Bao[1,2,3,10,11,*], Shunmin He[1,4,10,*], Guoqing Zhang[1,5,*], Yixue Li[1,5,*], Guoping Zhao[1,5,6,7,*], Runsheng Chen[1,4,10,*]

NGDC MEMBERS (Arranged by project role and then by contribution except for Team Leader (TL), as indicated)

PGG.Han: Yang Gao[5,#], Chao Zhang[5,#], Liyun Yuan[5,#], Guoqing Zhang[1,5,*] (TL), Shuhua Xu[5,14,15,16] (TL)

PGG.SNV: Chao Zhang[5,#], Yang Gao[5,#], Zhilin Ning[5,#], Yan Lu[5,#], Shuhua Xu[5,14,15,16] (TL)

CGVD: Jingyao Zeng[1,2,3,#], Na Yuan[1,2,#], Junwei Zhu[1,2], Mengyu Pan[1,2], Hao Zhang[1,2,3,10], Qi Wang[1,2,3,10], Shuo Shi[1,2,3,10], Meiye Jiang[1,2,3,10], Mingming Lu[1,2,3,10], Qiheng Qian[1,2,3,10], Qianwen Gao[1,2,3,10], Yunfei Shang[1,2,3,10], Jinyue Wang[1,2,3,10], Zhenglin Du[1,2,#] (TL), Jingfa Xiao [1,2,3,10,*] (TL)

GWAS Atlas: Dongmei Tian[1,2,#], Pei Wang[1,2,3,10,#], Bixia Tang[1,2,#], Cuiping Li[1,2,#], Xufei Teng[1,2,3,10], Xiaonan Liu[1,2,3,10], Dong Zou[1,2,3], Shuhui Song[1,2,3,#] (TL)

EWAS Data Hub: Zhuang Xiong[1,2,3,10,#], Mengwei Li[1,2,3,10,#], Fei Yang[1,2,3,10,#], Yingke Ma[1,2,3], Jian Sang[1,2,3,10], Zhaohua Li [1,2,3,10,11], Rujiao Li[1,2,3,#] (TL)

iSheep: Zhonghuang Wang[1,2,10,#], Qianghui Zhu[9,10,#], Junwei Zhu[1,2], Xin Li[9], Sisi Zhang[1,2], Dongmei Tian[1,2], Hailong Kang[1,2,10], Cuiping Li[1,2], Lili Dong[1,2], Cui Ying[1,2,10], Guangya Duan[1,2,10], Shuhui Song[1,2,3], Menghua Li[9,10] (TL), Wenming Zhao[1,2,3,10,*] (TL)

eLMSG: Xiaoyang Zhi[12,#] (TL), Yunchao Ling[5,#], Ruifang Cao[5,#], Zhao Jiang[12], Haokui Zhou[7], Daqing Lv[5], Wan Liu[5], Hans-Peter Klenk[13], Guoping Zhao[1,5,6,7,*], Guoqing Zhang[1,5,*] (TL)

PADS: Yadong Zhang[1,2,3,10,#], Zhewen Zhang[1,2,3,#], Hao Zhang[1,2,3,10], Jingfa Xiao[1,2,3,10,*] (TL)

BioProject & BioSample & GSA & BIG Submission: Tingting Chen[1,2,#], Sisi Zhang[1,2,#], Xu Chen[1,2,#], Junwei Zhu[1,2,#], Zhonghuang Wang[1,2,3,10], Hailong Kang[1,2,3,10], Lili Dong[1,2], Yanqing Wang[1,2,#] (TL)

GWH: Yingke Ma[1,2,3,#], Song Wu[1,2,3,10], Zhaohua Li[1,2,3,10,11], Zheng Gong[1,2,3,10], Meili Chen[1,2,3,#] (TL)

GVM: Cuiping Li[1,2,#], Dongmei Tian[1,2,#], Xufei Teng[1,2,3,10,#], Pei Wang[1,2,3,10,#], Bixia Tang[1,2,#], Xiaonan Liu[1,2,3,10], Dong Zou[1,2,3], Shuhui Song[1,2,3,#] (TL)

NONCODE: Shuangsang Fang[8], Lili Zhang[4,10], Jincheng Guo[8], Yiwei Niu[4,10], Yang Wu[8], Hui Li[8], Lianhe Zhao[8], Xiyuan Li[8], Xueyi Teng[4,10], Xianhui Sun[4,10], Liang Sun[8], Runsheng Chen[1,4,10,*], Yi Zhao[8] (TL)

piRBase: Jiajia Wang[4,10,#], Peng Zhang[4,#], Yanyan Li[4,10], Yu Zheng[4,10], Runsheng Chen[1,4,10,*], Shunmin He[1,4,10,*] (TL)
NPInter: Xueyi Teng[4,10,#], Xiaomin Chen[4,10,#], Hua Xue[4,10,#], Yiheng Teng[4,10], Peng Zhang[4], Quan Kang[4], Yajing Hao[4], Yi Zhao[8], Runsheng Chen[1,4,10,*], Shunmin He[1,4,10,*] (TL)
LncBook & LncRNAWiki: Jiabao Cao[1,2,3,10,#], Lin Liu[1,2,3,10,#], Zhao Li[1,2,3,10,#], Qianpeng Li[1,2,3,10], Dong Zou[1,2,3], Qiang Du[1,2,3,10], Amir A. Abbasi[25], Huma Shireen[25], Nashaiman Pervaiz[25], Fatima Batool[25], Rabail Z. Raza[25], Lina Ma[1,2,3,#] (TL)
EDK & PED: Guangyi Niu[1,2,3,10,#], Yuansheng Zhang[1,2,3,10,#], Dong Zou[1,2,3,#], Tongtong Zhu[1,2,3,10,11], Jian Sang[1,2,3,10], Mengwei Li[1,2,3,10], Lili Hao[1,2,3,#] (TL)
MethBank: Dong Zou[1,2,3,#], Guoliang Wang[24,#], Mengwei Li[1,2,3,10,#], Rujiao Li[1,2,3,#] (TL)
EWAS Atlas: Mengwei Li[1,2,3,10,#], Rujiao Li[1,2,3], Yiming Bao[1,2,3,10,11,*] (TL)
IC4R: Jun Yan[17,#], Jian Sang[1,2,3,10,#], Dong Zou[1,2,3,#], Chen Li[22], Zhennan Wang[10,23], Yuansheng Zhang[1,2,3,10], Tongtong Zhu[1,2,3,10,11], Shuhui Song[1,2,3] (TL), Xiangfeng Wang[17] (TL), Lili Hao[1,2,3] (TL)
LSD: Zhonghai Li[18,#] (TL), Yang Zhang[1,2,3,10,#], Dong Zou[1,2,3], Yi Zhao[19], Houling Wang[18], Yi Zhang[18], Xinli Xia[18,20], Hongwei Guo[18,21], Zhang Zhang[1,2,3,10,11,*]
Database Commons: Dong Zou[1,2,3,#], Lina Ma[1,2,3,#] (TL)
eGPS Cloud: Lili Dong[1,2,#], Bixia Tang[1,2,#], Junwen Zhu[1,2,#], Qing Zhou[1,2,10], Zhonghuang Wang[1,2,10], Hongen Kang[1,2,10], Xu Chen[1,2], Li Lan[1,2], Yiming Bao[1,2,3,10,11,*] (TL), Wenming Zhao[1,2,3,10,*] (TL)
BIG Search: Dong Zou[1,2,3,#] (TL)
BIG SSO: Junwei Zhu[1,2,#] (TL), Bixia Tang[1,2,#]
BHBD: Yiming Bao[1,2,3,10,11,*], Li Lan[1,2], Xin Zhang[1,2], Yingke Ma[1,2,3], Yongbiao Xue[26] (Project Leader)
Hardware & System Administration: Yubin Sun[1,2], Shuang Zhai[1,2], Lei Yu[1,2], Mingyuan Sun[1,2], Huanxin Chen[1,2] (TL)
Writing Group: Zhang Zhang[1,2,3,10,11,*], Wenming Zhao[1,2,3,10,*], Jingfa Xiao[1,2,3,10,*], Yiming Bao[1,2,3,10,11,*], Lili Hao[1,2,3]
NGDC PARTNERS (Listed in alphabetical order by database names)
AnimalTFDB: Hui Hu[27], An-Yuan Guo[27]
dbPAF & WERAM: Shaofeng Lin[27], Yu Xue[27]
dbPPT: Chenwei Wang[27], Yu Xue[27]
dbPSP: Wanshan Ning[27], Yu Xue[27]
CellMarker: Xinxin Zhang[28], Yun Xiao[28], Xia Li[28]
CGDB: Yiran Tu[27], Yu Xue[27]
circAtlas: Wanying Wu[29], Peifeng Ji[29], Fangqing Zhao[29]
DEG & DoriC: Hao Luo[30,31,32], Feng Gao[30,31,32]
iEKPD: Yaping Guo[27], Yu Xue[27]
GenTree: Hao Yuan[33,34], Yong E. Zhang[10,33,34]
hTFtarget: Qiong Zhang[27], An-yuan Guo[27]
iUUCD: Jiaqi Zhou[27], Yu Xue[27]
LncRNADisease: Zhou Huang[35], Qinghua Cui[35,36]
lncRNASNP: Ya-Ru Miao[27], An-Yuan Guo[27]
MiCroKiTS: Chen Ruan[27], Yu Xue[27]
PceRBase: Chunhui Yuan[37], Ming Chen[37]
PlantTFDB: Jin-Pu Jin[38], Feng Tian[38], Ge Gao[38]
PLMD: Ying Shi[27], Yu Xue[27]
PTMD: Lan Yao[27], Yu Xue[27], Qinghua Cui[35,36]
RhesusBase: Xiangshang Li[39], Chuan-Yun Li[39]
SEGreg: Qing Tang[27], An-Yuan Guo[27]
THANATOS: Di Peng[27], Yu Xue[27]
[1]National Genomics Data Center, Beijing 100101, China
[2]BIG Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China
[3]CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China
[4]Key Laboratory of RNA Biology, Center for Big Data Research in Health, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[5]Bio-Med Big Data Center, Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai 200231, China
[6]CAS Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200231, China
[7]Center for Quantitative Synthetic Biology, Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[8]Key Laboratory of Intelligent Information Processing, Advanced Computer Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[9]CAS Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China
[10]University of Chinese Academy of Sciences, Beijing 100049, China
[11]School of Future Technology, University of Chinese Academy of Sciences, Beijing 100049, China
[12]Yunnan Institute of Microbiology, School of Life Sciences, Yunnan University, Kunming, Yunnan 650091, China
[13]School of Natural and Environmental Sciences, Ridley Building 2, Newcastle University, Newcastle upon Tyne, UK
[14]School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China
[15]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China
[16]Collaborative Innovation Center of Genetics and Development, Shanghai 200438, China
[17]Department of Crop Genomics and Bioinformatics, College of Agronomy and Biotechnology, China Agricultural University, Beijing 100094, China
[18]Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing Forestry University, Beijing 100083, China
[19]College of Life Sciences, Peking University, Beijing 100871, China
[20]College of Biological Sciences and Biotechnology, National Engineering Laboratory for Tree Breeding, Beijing Forestry University, Beijing 100083, China

[21]Institute of Plant and Food Science, Department of Biology, Southern University of Science and Technology (SUSTech), Shenzhen, Guangdong 518055, China
[22]Rice Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China
[23]Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China
[24]College of Plant Protection, Hunan Agricultural University, Hunan 410128, China
[25]National Center for Bioinformatics, Programme of Comparative and Evolutionary Genomics, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan
[26]Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China
[27]Department of Bioinformatics and Systems Biology, Key Laboratory of Molecular Biophysics of the Ministry of Education, Hubei Bioinformatics and Molecular Imaging Key Laboratory, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China
[28]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang 150081, China
[29]Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China
[30]Department of Physics, School of Science, Tianjin University, Tianjin 300072, China
[31]Frontier Science Center of Synthetic Biology, Key Laboratory of Systems Bioengineering, Tianjin University, Tianjin 300072, China
[32]SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), Tianjin 300072, China
[33]Key Laboratory of Zoological Systematics and Evolution and State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China
[34]CAS Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, Yunnan 650223, China
[35]Department of Biomedical Informatics, School of Basic Medical Sciences, MOE Key Lab of Cardiovascular Sciences, Center for Noncoding RNA Medicine, Peking University, Beijing 100190, China
[36]Center of Bioinformatics, Key Laboratory for Neuro-Information of Ministry of Education, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China
[37]Department of Bioinformatics, State Key Laboratory of Plant Physiology and Biochemistry, Institute of Plant Science, College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China
[38]Biomedical Pioneering Innovation Center (BIOPIC), Beijing Advanced Innovation Center for Genomics (ICG), Center for Bioinformatics (CBI), and State Key Laboratory of Protein and Plant Gene Research at School of Life Sciences, Peking University, Beijing 100871, China
[39]Institute of Molecular Medicine, Peking University, Beijing 100871, China
*To whom correspondence should be addressed: Zhang Zhang (zhangzhang@big.ac.cn).
Correspondence may also be addressed to Wenming Zhao (zhaowm@big.ac.cn), Jingfa Xiao (xiaojingfa@big.ac.cn), Yiming Bao (baoym@big.ac.cn), Shunmin He (heshunmin@ibp.ac.cn), Guoqing Zhang (gqzhang@picb.ac.cn), Yixue Li (yxli@sibs.ac.cn), Guoping Zhao (gpzhao@sibs.ac.cn) and Runsheng Chen (crs@sun5.ibp.ac.cn).
#The authors wish it to be known that, in their opinion, these authors should be regarded as Joint First Authors.