中国科学院北京基因组研究所
BEIJING INSTITUTE OF GENOMICS CHINESE ACADEMY OF SCIENCES

国家基因组科学数据中心
National Genomics Data Center

Curation Model and Curation rules of Database Commons

Lina Ma Ph.D

Project Leader

2022-10-10

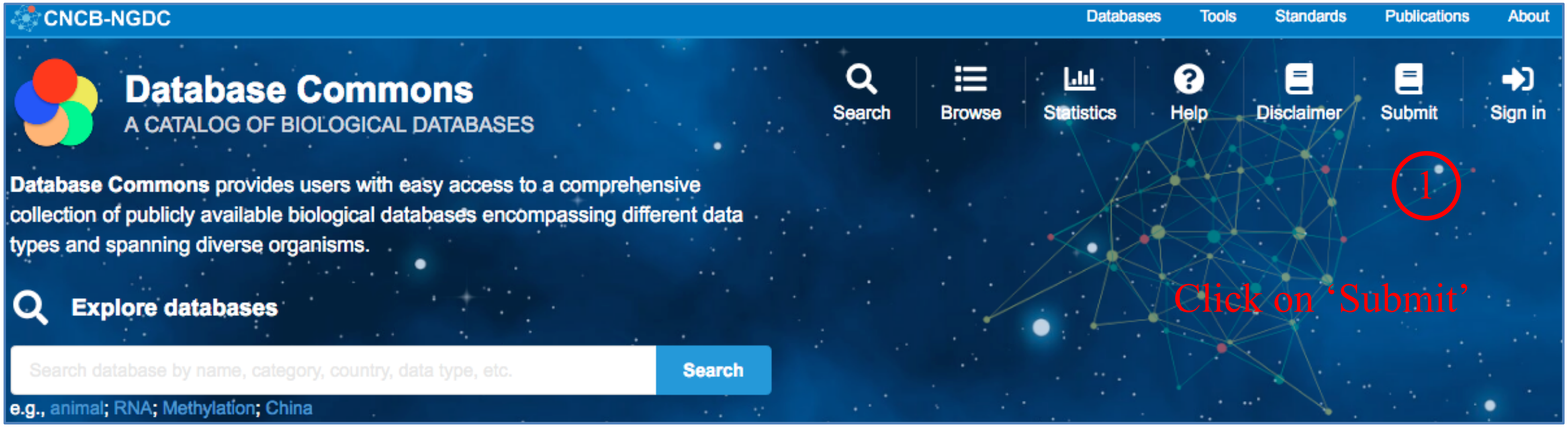
Database Commons, Launched in 2015
https://ngdc.cncb.ac.cn/databasecommons

国家生物信息中心
China National Center for Bioinformation

About Database Commons

- Database Commons is a manually curated catalogue of worldwide biological databases. It aims to make users easily access to a comprehensive collection of publicly available biological databases, conveniently capture a specific collection of databases of interest, and also efficiently retrieve the popular or high-quality databases.
- Based on literature curation, Database Commons integrates relevant information for all collected databases and catalogues each database in terms of data type, organism, subject, hosted institution and location, etc., accordingly enabling people to easily find a specific collection of databases of interest. Moreover, all collected databases are ranked by z-index as well as total citations. Meanwhile, it allows users to rate any database by considering different aspects, facilitating efficient location of appropriate databases of interest.
- Database Commons can be served as both a valuable resource and a comprehensive search engine involving all biological field. It benefits all users in selecting from all publicly available databases for their effective and efficient exploitation.

Submit a new database in Database Commons



Database Commons

A CATALOG OF BIOLOGICAL DATABASES

Database Commons provides users with easy access to a comprehensive collection of publicly available biological databases encompassing different data types and spanning diverse organisms.

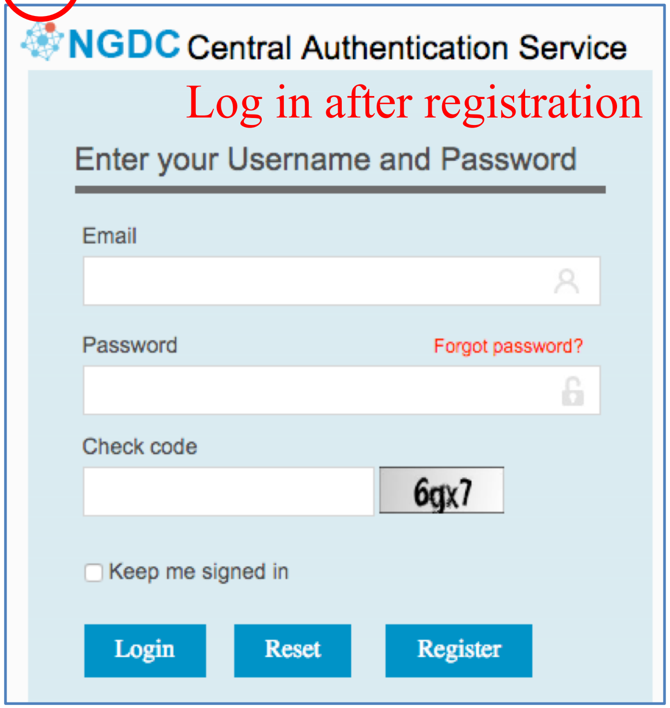
Explore databases

e.g., animal; RNA; Methylation; China

1

Click on ‘Submit’

2



NGDC Central Authentication Service

Log in after registration

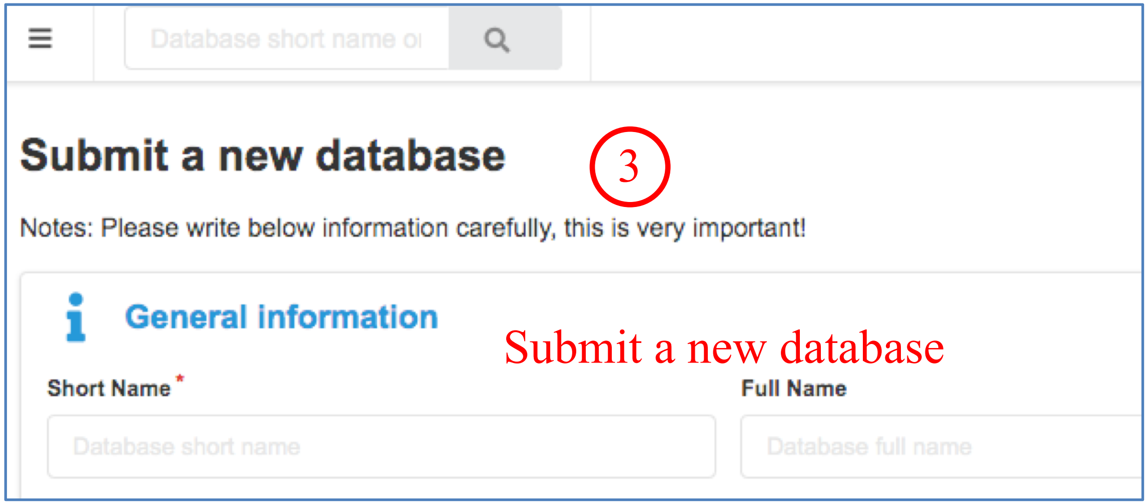
Enter your Username and Password

Email

Password

Check code

☐ Keep me signed in



Submit a new database

3

Notes: Please write below information carefully, this is very important!

General information

Submit a new database

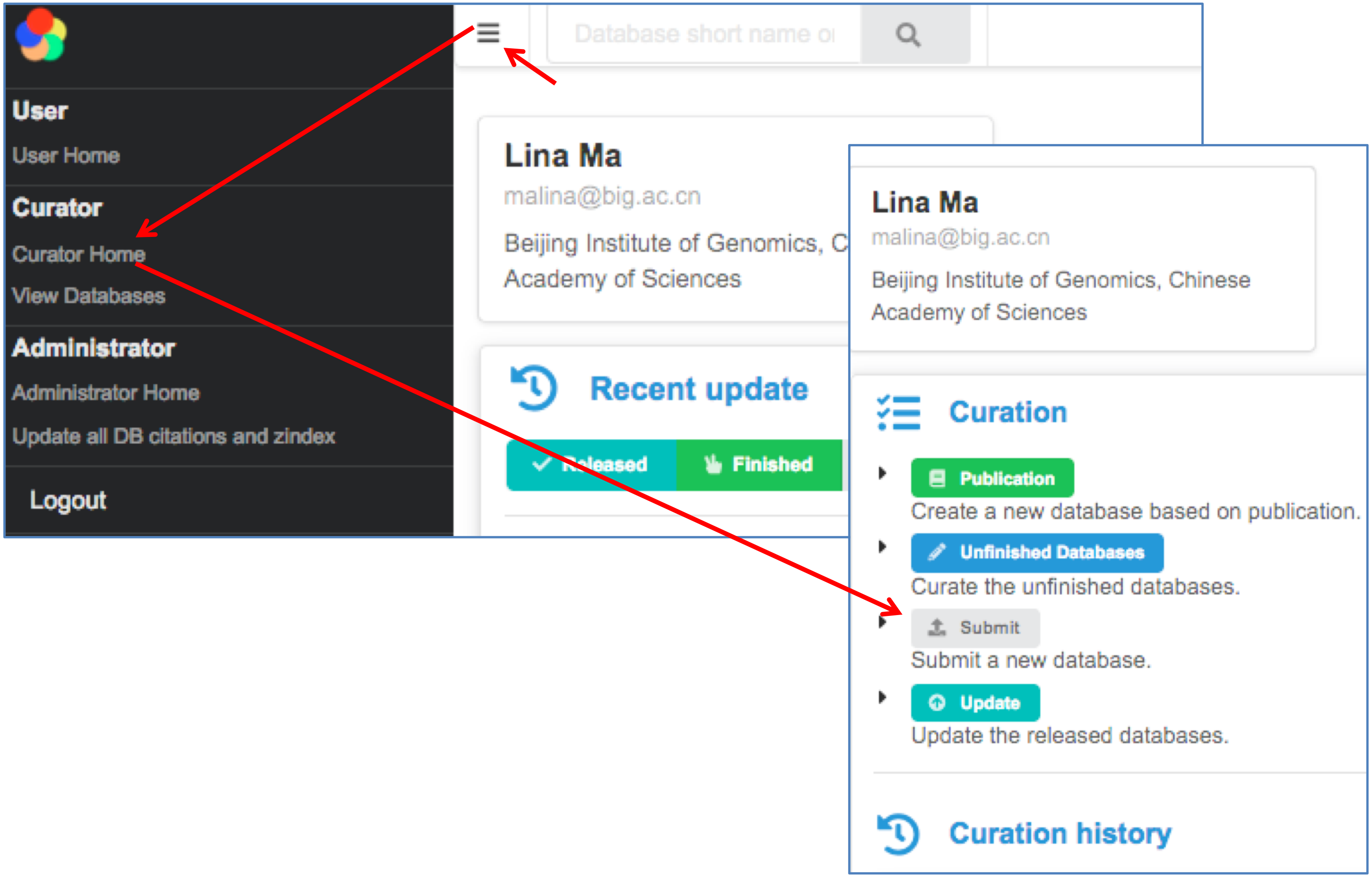
Short Name*

Database short name

Full Name

Database full name

Alternatively, you can submit your database through “Curator Home”.



Lina Ma
malina@big.ac.cn
Beijing Institute of Genomics, Chinese Academy of Sciences

Create a new database based on publication.

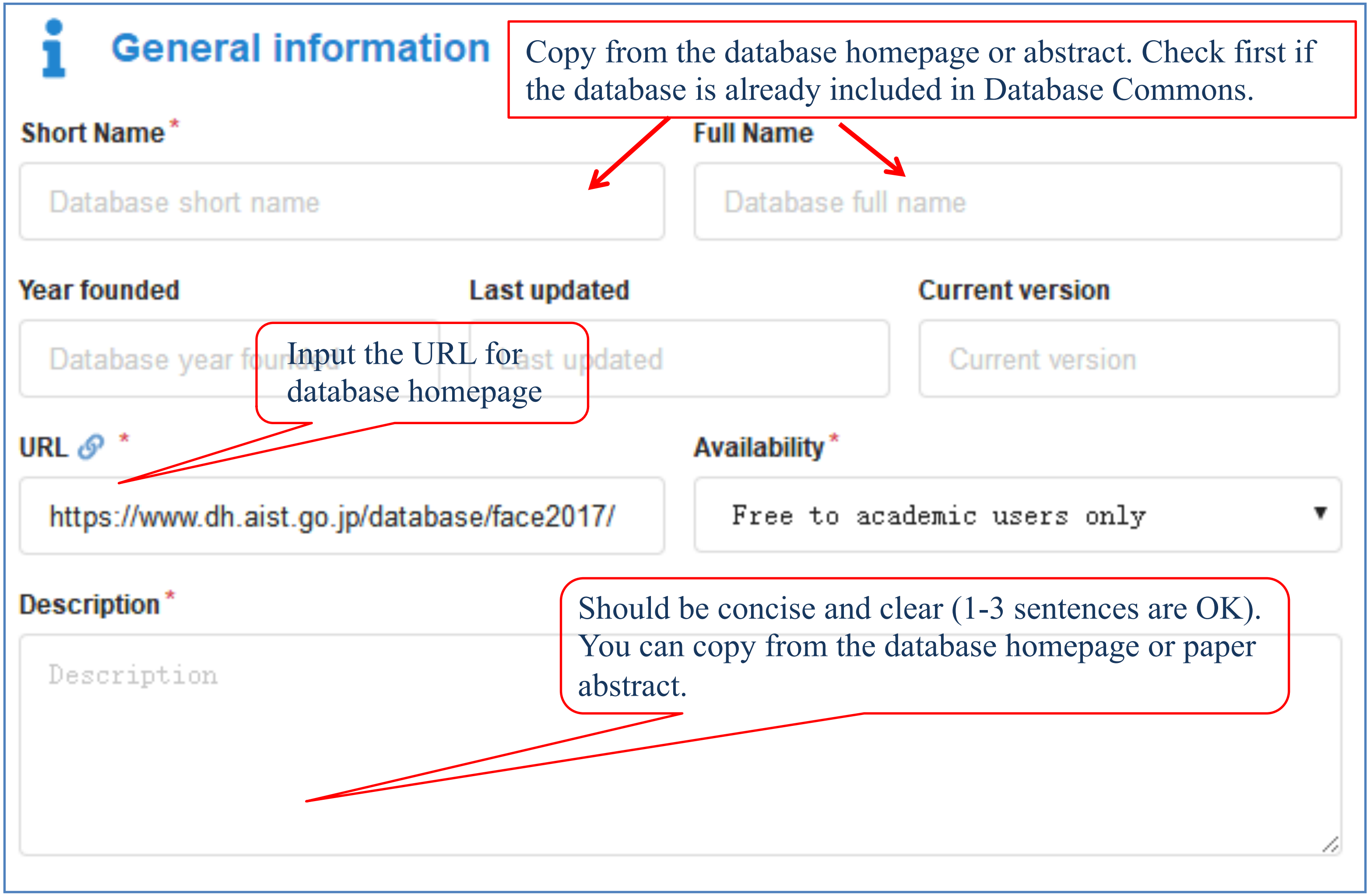
Curate the unfinished databases.

Submit a new database.

Update the released databases.

1. Start from “General Information”

For each database, there are four sections to be curated: general information, data information, contact information, publication.



General information

Short Name *: Database short name

Full Name: Database full name

Year founded: Database year founded

Last updated: Last updated

Current version: Current version

URL *: https://www.dh.aist.go.jp/database/face2017/

Availability *: Free to academic users only

Description *: Description

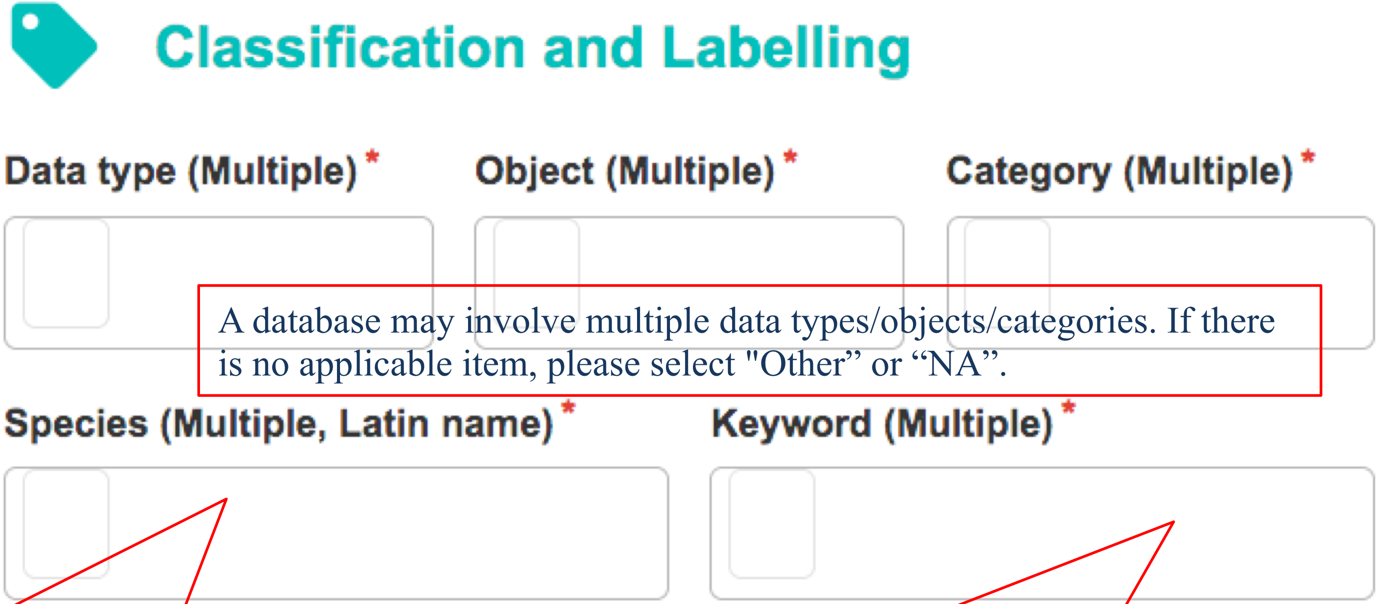
Copy from the database homepage or abstract. Check first if the database is already included in Database Commons.

Input the URL for database homepage

Should be concise and clear (1-3 sentences are OK). You can copy from the database homepage or paper abstract.

2. Classification and Labelling

- All the items are required for this section. “Data type”, “Object”, and “Category” are selected from the drop-down box. Explanation for each database category is attached.
- A database may cover multiple data types, or objects, or categories.



Classification and Labelling

Data type (Multiple) *

Object (Multiple) *

Category (Multiple) *

Species (Multiple, Latin name) *

Keyword (Multiple) *

A database may involve multiple data types/objects/categories. If there is no applicable item, please select "Other” or “NA”.

- Input the Latin name of the major organisms if the database covers quite a large number of organisms.
- If no species information is available, input ‘NA’.

- Input 1-3 keywords which could represent the primary features of the database. Please use the singular form rather than plural form.
- All letters are in lowercase.

Data Type:

- **DNA:** gene/chromosome/genome sequence, DNA mutation/modification, DNA structure, DNA elements including probe, primer, motif, repeat sequence, etc.
- **RNA:** RNA sequence, coding & non-coding transcripts, alternative splicing, RNA editing/modification, RNA probe and primer, RNA motif and structure, RNA expression
- **Protein:** protein sequence, protein motif and domain, protein structure, protein modification, protein-protein interaction, protein expression

Object:

- **Animal**
- **Plant**
- **Fungi**
- **Bacteria**
- **Archaea**
- **Virus**

Category (examples):

Raw bio-data: raw data of nucleic acid/protein sequencing and microarray, and image, digit, video, audio from biological and medical research (SRA, GSA, GEO)

Gene, genome and annotation: gene/genetic element annotation, gene structure/family/motif/domain annotation, genome annotation, comparative genome (metagenome, pan-genome) analysis and annotation **(**GenBank, RefSeq, Uniprot, Pfam, InterPro, miRBase, TRANSFAC, FlyBase)

Genotype, phenotype and variation: genotypes, phenotypes, multiple-scale variations (including SNP, INDEL, CNV, chromosomal rearrangement and other structural variation), genotype-phenotype associations (dbSNP, GWAS-catalog, ClinVar, OMIM, HGMD, HPO)

Phylogeny and homology: phylogeny reconstruction of genes/species, evolutionary history/process/event among individuals/organisms, homology identification (COG, CDD, InParanoid)

Expression: RNA/protein expression, expression abundance and pattern, RNA probe or primer used for gene expression detection, differential expression analysis (GEO, microRNA.org, FlyAtlas, GXD)

Modification: DNA modification, post-transcriptional modification of mRNA and non-coding RNA, post-translational modification of protein, modification type/technology/function (RNAMDB, Phospho.ELM)

Structure: secondary, tertiary and quaternary structure of DNA/RNA/protein, chromatin structure (CATH)

Interaction: direct (physical) and indirect (functional) associations, including protein-protein interaction, RNA-protein interaction, DNA-protein interaction, gene regulatory interaction, biochemical reaction, antigen and antibody, and genetic interaction (STRING, TRANSFAC, miRTarBase)

Pathway: biological pathways for metabolic, signaling, gene regulatory analysis (KEGG, TRANSPATH)

Health and medicine: disease variation/genotype-phenotype association, immune reaction, disease model, clinical biomarker, therapeutic target, drug & chemical compound, pharmacogenomics and pharmacodynamics, electronic health record (OMIM, HGMD, DrugBank, ChEMBL)

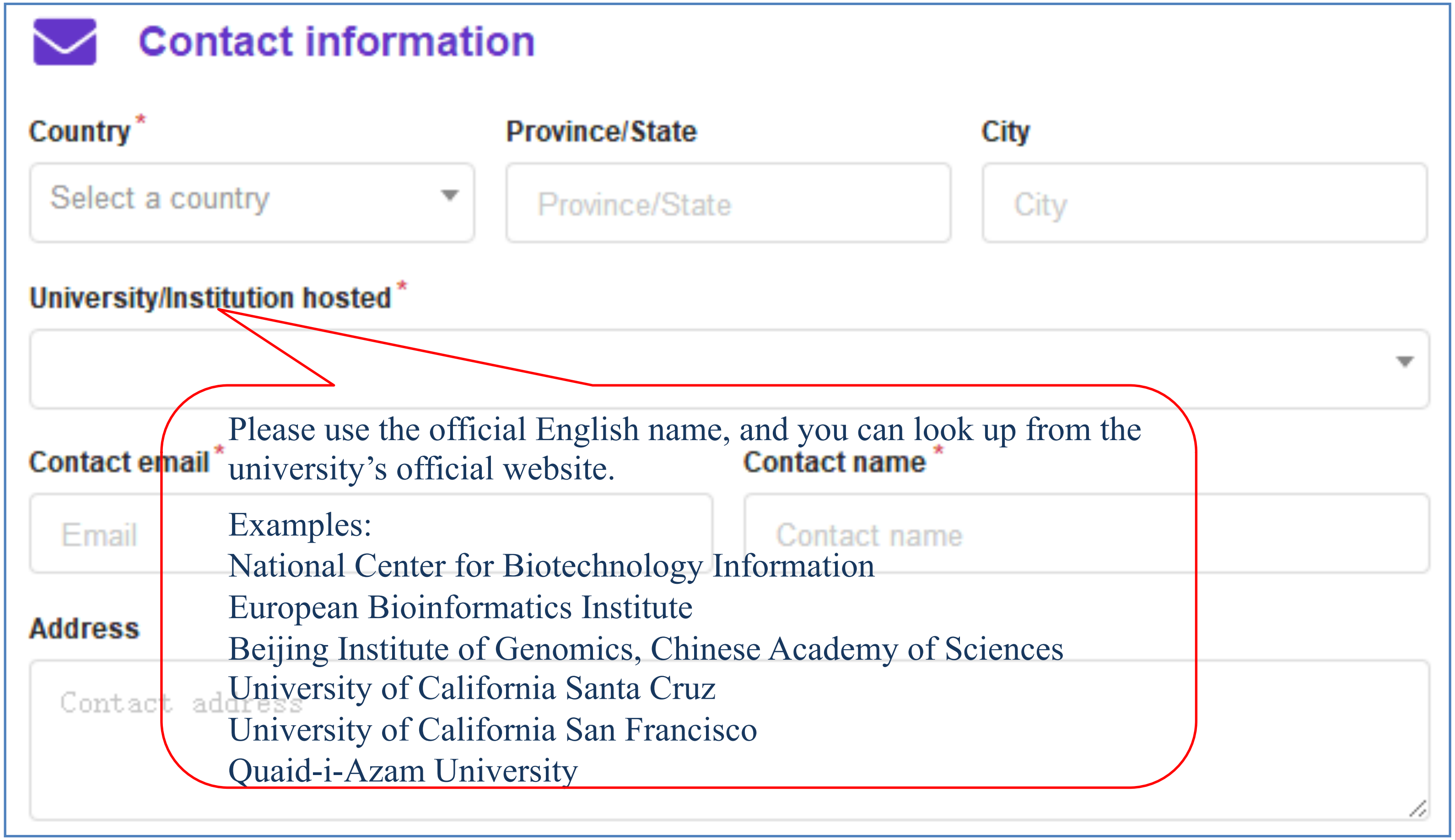
Standard, ontology and nomenclature: standard, ontology and nomenclature for biological entities (GO)

Literature: literature information, literature/text mining, textual annotation based on literature (PubMed, LncRNAWiki, BioCreAtIvE)

Metadata: metadata information for biological entities, e.g., project/sample/experiment/run/database/tool (Database Commons, BioSamples)

3. Contact Information

- Refer to ‘contact’ information of the database or input contact information of the corresponding author.
- This is used to contact people who take charge of database maintenance, and they are encouraged to participate in database curation.



Contact information

Country * : Select a country

Province/State: Province/State

City: City

University/Institution hosted * :

Contact email * : Email

Contact name * : Contact name

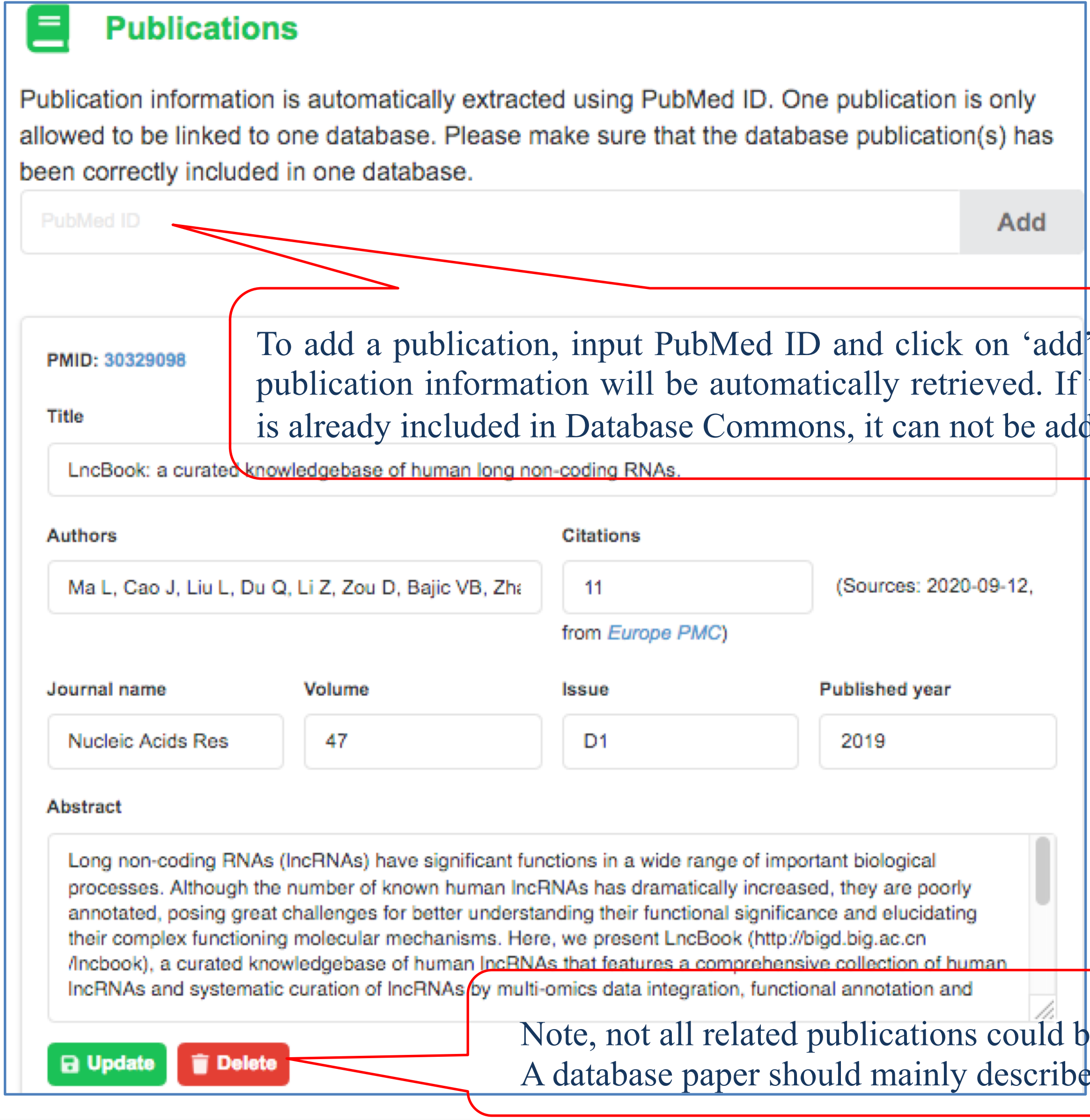
Address: Contact address

Please use the official English name, and you can look up from the university’s official website.

Examples:
National Center for Biotechnology Information
European Bioinformatics Institute
Beijing Institute of Genomics, Chinese Academy of Sciences
University of California Santa Cruz
University of California San Francisco
Quaid-i-Azam University

4. Publication

Make sure that all the publications of the database have been added.
We rank databases based on total citations of all the database papers.



Publications

Publication information is automatically extracted using PubMed ID. One publication is only allowed to be linked to one database. Please make sure that the database publication(s) has been correctly included in one database.

PMID: 30329098

Title

LncBook: a curated knowledgebase of human long non-coding RNAs.

Authors

Ma L, Cao J, Liu L, Du Q, Li Z, Zou D, Bajic VB, Zh

Citations

11

(Sources: 2020-09-12, from *Europe PMC*)

Journal name	Volume	Issue	Published year
Nucleic Acids Res	47	D1	2019

Abstract

Long non-coding RNAs (lncRNAs) have significant functions in a wide range of important biological processes. Although the number of known human lncRNAs has dramatically increased, they are poorly annotated, posing great challenges for better understanding their functional significance and elucidating their complex functioning molecular mechanisms. Here, we present LncBook (http://bigd.big.ac.cn/lncbook), a curated knowledgebase of human lncRNAs that features a comprehensive collection of human lncRNAs and systematic curation of lncRNAs by multi-omics data integration, functional annotation and

To add a publication, input PubMed ID and click on 'add', and the publication information will be automatically retrieved. If the paper is already included in Database Commons, it can not be added.

Note, not all related publications could be considered as database papers.
A database paper should mainly describe the database content and data.

Don’t forget to click on “Save” when you have completed all the fields or made any changes.
Click on “Finished” to Submit.

Your database will be publicly released after manual review by the Database Commons Team.

This page https://ngdc.cncb.ac.cn/databasecommons/policies also details the rules for curation.

Any questions, please contact us by emailing to databasecommons@big.ac.cn

THANKS