

DATABASE

Open Access



# SynBioTools: a one-stop facility for searching and selecting synthetic biology tools

Pengli Cai<sup>1†</sup>, Sheng Liu<sup>1†</sup>, Dachuan Zhang<sup>2</sup>, Huadong Xing<sup>1</sup>, Mengying Han<sup>1</sup>, Dongliang Liu<sup>1</sup>, Linlin Gong<sup>1</sup> and Qian-Nan Hu<sup>1\*</sup>

<sup>†</sup>Pengli Cai and Sheng Liu have contributed equally to this work

\*Correspondence: qnhu@sibs.ac.cn

<sup>1</sup> CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China

<sup>2</sup> Ecological Systems Design, Institute of Environmental Engineering, ETH Zurich, 8093 Zurich, Switzerland

## Abstract

**Background:** The rapid development of synthetic biology relies heavily on the use of databases and computational tools, which are also developing rapidly. While many tool registries have been created to facilitate tool retrieval, sharing, and reuse, no relatively comprehensive tool registry or catalog addresses all aspects of synthetic biology.

**Results:** We constructed SynBioTools, a comprehensive collection of synthetic biology databases, computational tools, and experimental methods, as a one-stop facility for searching and selecting synthetic biology tools. SynBioTools includes databases, computational tools, and methods extracted from reviews via SScientific Table Extraction, a scientific table-extraction tool that we built. Approximately 57% of the resources that we located and included in SynBioTools are not mentioned in bio.tools, the dominant tool registry. To improve users' understanding of the tools and to enable them to make better choices, the tools are grouped into nine modules (each with subdivisions) based on their potential biosynthetic applications. Detailed comparisons of similar tools in every classification are included. The URLs, descriptions, source references, and the number of citations of the tools are also integrated into the system.

**Conclusions:** SynBioTools is freely available at <https://synbiotools.lifesynter.com/>. It provides end-users and developers with a useful resource of categorized synthetic biology databases, tools, and methods to facilitate tool retrieval and selection.

**Keywords:** Synthetic biology, Computational tool, Database, Tool retrieval, Tool registry, Table extraction

## Background

In synthetic biology research, data processing, computational modeling, and artificial intelligence play important roles in the design and analysis of laboratory experiments [1–3]. For instance, the big data generated by high-throughput sequencing depends on computational data processing. This has promoted the rapid development of databases and computational tools, with large numbers of them being produced in recent decades.

To better manage these resources, various tool registries of different sizes and on different topics have been created, improving convenience for users and developers. These include BioMOBY [4], Bioconductor [5], BioCatalogue [6], SEQanswers (a wiki database



of tools for high-throughput sequencing analysis) [7], BioJavaScript (BioJS) for bioinformatics visualization tools [8, 9], the BioContainers Registry [10], OMICTools (a directory of tools for various kinds of omics analyses) [11], Bio-TDS [12], bio.tools [13], JIB.tools 2.0 [14], Expasy [15], and GSAREfDB (providing tools for gene set analysis) [16]. Among these registries, bio.tools and BioContainers are currently the largest. The bio.tools registry, based on community-driven curation, lists 25,299 tools [13]. BioContainers stores, creates, and distributes bioinformatics tools, containers, and packages [10]. The various existing tool registries make it easier to find tools during experimental design and analysis or tool development. Nonetheless, there is currently no comprehensive tool registry for synthetic biology. While the existing registries list some useful design and analysis tools for synthetic biology research, some of these tool registries, such as OMICTools [11] and SEQwiki [7] for omics analysis, are no longer available. The Secondary Metabolite Bioinformatics Portal (SMBP) provides the computational tools to facilitate synthetic biology research involving secondary metabolite production [17], but does not offer researchers a one-stop search for finding other tools. Furthermore, comparative information on similar tools is lacking in the large tool registries. From the end user's perspective, it is often challenging to choose the right tool for each research task from the many similar tools that have been developed over the years.

At the same time, the development of a large number of tools has been accompanied by the publication of reviews describing them. These reviews have efficiently categorized and compared similar tools or databases for different topics or categories, addressing some of the problems related to the tool registries mentioned. These reviews are, therefore, extremely valuable resources for tool users and developers. Nonetheless, information about the tools is scattered among different reviews, and the information provided by these reviews cannot be explored interactively, as is possible with tool registries.

To address these issues, we constructed SynBioTools, a registry dedicated to synthetic biology tools, with relevant databases, computational tools, and methods. Some relevant experimental methods and tools, such as DNA assembly tools, were integrated for coherence and convenience. These resources were collected from review articles dealing with tools and databases in synthetic biology. To better extract information from reviews, we built SCIENTIFIC Table Extraction (SCITE), a tool for extracting tabular data from articles. We extracted information on tool classification, features, and comparisons, and reorganized it into biosynthetic tool categories. SynBioTools combines the advantages of the reviews' categorical summaries and human-computer interactions via a web-server database. We further integrated other tool-related information to help users to select the appropriate tools to match their needs.

## Methods

### Data acquisition

We retrieved references for bioinformatics tools from bio.tools, which provides a comprehensive registry of tools and databases. Additionally, the Semantic Scholar Open Research Corpus (S2ORC) dataset (<https://allenai.org/data/s2orc>) and PubMed data (<https://pubmed.ncbi.nlm.nih.gov/download/>) were downloaded as data sources for all literature. The S2ORC and PubMed data were used to obtain citations and review labels. To obtain reviews describing bioinformatics tools, we extracted citations for all tools

from the S2ORC dataset, filtered them for review articles, and then selected reviews citing more than 100 tools that were published between 2010 and 2022. Synthetic biology-related reviews were chosen manually for further tool information extraction. Finally, 37 review articles were used for tool extraction. We used our custom-developed tool, SCITE, to extract information from the tables in the reviews. Based on their characteristics and biosynthetic process application [18], we manually grouped the tools and databases into nine modules: compounds, biocomponents, protein, pathway, gene-editing, metabolic modeling, omics, strains, and others.

#### **Tabular information extraction**

To extract information from the tables in the reviews, we developed a literature-table-extraction tool, SCITE, based on the optical character recognition (OCR) toolkit PaddleOCR (<https://github.com/PaddlePaddle/PaddleOCR>) and the R package tidypmc (<https://github.com/ropensci/tidypmc>). SCITE implements two methods to extract tables from articles. For general articles in PDF format, we built a table extraction tool based on an OCR strategy (Additional file 1). This tool first converts the pages of a PDF document into image format, then identifies and extracts the table information from the images based on PaddleOCR, which is an ultra-light deep learning OCR model. For papers from PubMed Central, we obtained tables by parsing the full-text XML file directly using tidypmc (Additional file 2). We further deployed SCITE as an API using FastAPI and Celery. Finally, the tabular information from review articles was automatically extracted using SCITE.

#### **Data curation and integration**

Data management and integration included table extraction, manual curation, data supplement, and data integration. As most of the tables were formatted differently between papers and the automatically extracted data were not 100% reliable, manual curation was performed after table extraction by SCITE. During the curation process, we corrected some mistakes and formatted each row to one tool. Based on the reference columns in the review tables, we obtained and supplemented direct references for each tool using either programming or manual means. They were subsequently used to obtain information on reference-related common fields. The data integrated into SynBioTools is divided into common and unique fields. Common fields, such as name, module, citation, and other information common to all tools, are displayed on the SynBioTools Browse page, while unique field information from the review table is displayed on the tool Details page.

#### **System design and implementation**

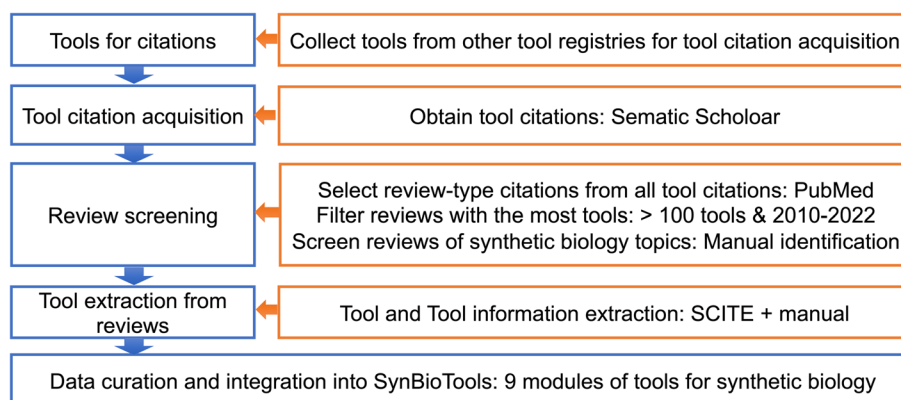
The SynBioTools web server, deployed in the Ubuntu 18.04.2 environment, uses the Python Web framework FastAPI 0.73.0 and the front-end framework Bootstrap 5.2. The project data are stored in the NoSQL database MongoDB 5.0.4. We used the JavaScript libraries Echarts 5.3.3 and Tabulator 5.4.2 for graph and table rendering, respectively. We further developed various search methods using Elasticsearch 7.16.2. SynBioTools is freely available at <https://synbiotools.lifesynther.com>. Users can access it in Google Chrome or Safari for the best experience.

## Results

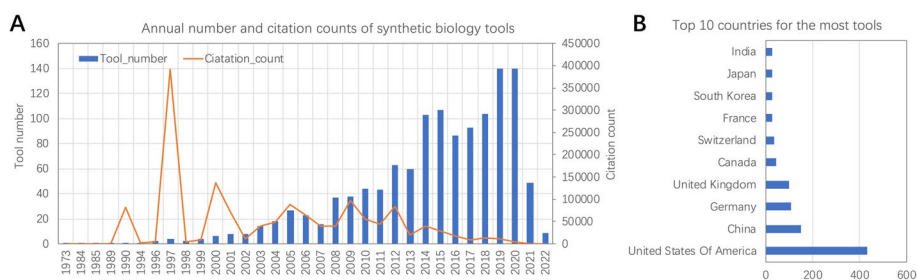
### SynBioTools summary

SynBioTools is a one-stop solution for searching and selecting synthetic biology tools. Here, synthetic biology tools refer to the tools, methods, and databases used for synthetic biology research. All the tools in SynBioTools were extracted from review articles [1, 18–53] (Additional file 3) via SCITE, our custom-built article-table-extraction tool. The method and process of the construction of SynBioTools are summarized in Fig. 1. Based on the tool characteristics and potential biosynthesis application, we manually grouped them into nine modules (compounds, biocomponents, protein, pathway, gene-editing, metabolic modeling, omics, strains, and others), related to compound selection, pathway mining and design, element selection, protein selection and design, gene editing, metabolic network modeling, omics analysis, and strain modification, respectively. Additional parameters were integrated, including tool descriptions, source references, URLs linking to the tools, and hints toward tool availability on the Browse page. The probability that a tool's web server is accessible is positively correlated with the number of citations of that tool [54], and article citation counts are used to estimate tool popularity [16]. Therefore, for each tool, we provided the total numbers of all citations, review citations, citations used for tool development, and citations reflecting the experimental application of the tool (i.e., not including the previously mentioned review and tool-development articles). This grouping and the parameters included will improve users' understanding and selection of tools.

Most of the tools and databases included in SynBioTools were developed within the last 20 years (Fig. 2A). In the past 10 years, the number of tools has increased rapidly, while the number of citations has declined. Familiar and frequently used tools, such as BLAST, KEGG, GO, STRING, NCBI, MAFFT, Reactome, PRIDE, Fastree, and Bowtie, have numerous citations (Fig. 2A). The top three countries developing the tools or databases listed in SynBioTools are the United States of America, China, and Germany



**Fig. 1** Schematic of the process of constructing SynBioTools. Tools and tool information were extracted via the following steps: collect tools from other tool registries to obtain tool citations; select only citations for reviews; filter for reviews published from 2010 to 2022 with more than 100 tool citations; manually select review articles on synthetic biology. After data curation and integrating information such as the tool's URL, description, reference, and the number of citations, the tools were grouped into nine modules, i.e., compounds, biocomponents, protein, pathway, gene-editing, metabolic modeling, omics, strains, and others, based on their potential biosynthetic applications



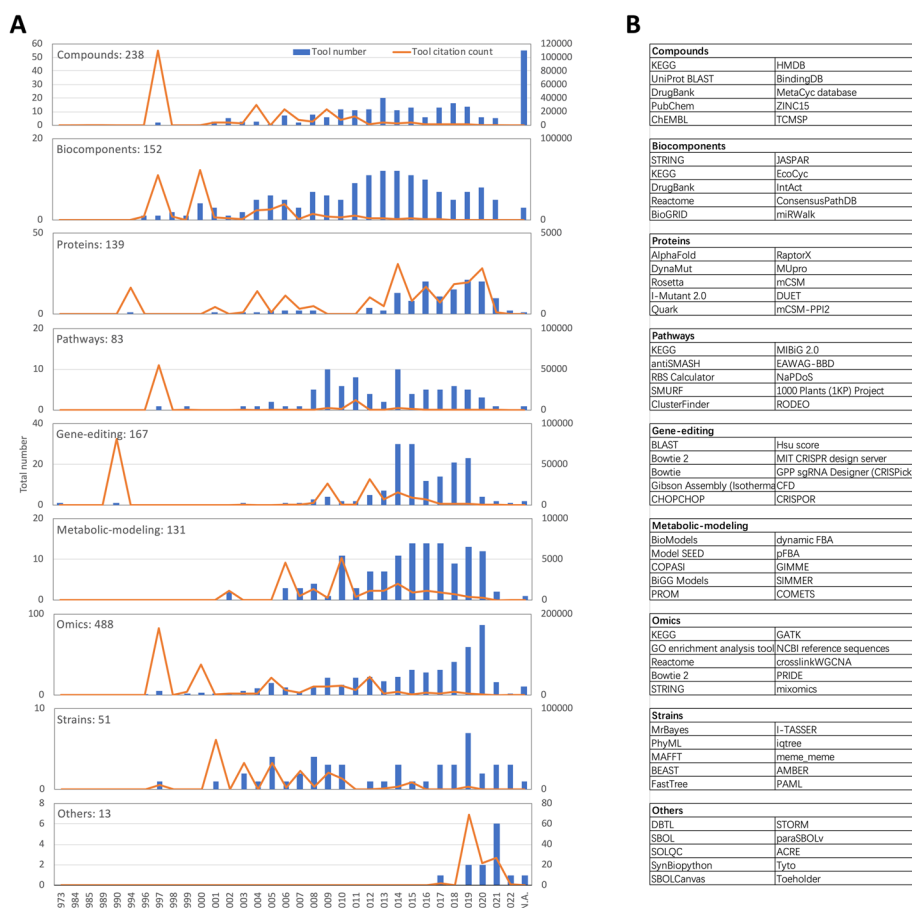
**Fig. 2** Summary of the data in SynBioTools. **A** Annual numbers of tools published and tool citations in synthetic biology. The citation count refers to the annual number of citations of all tools. Most of the tools and databases included in SynBioTools were developed within the last 20 years. The tools and databases with the most citations (and year of origin) are BLAST (1990), KEGG (1997), GO (2000), STRING (2000), NCBI (2000), MAFFT (2005), Reactome (2005), PRIDE (2005), Bowtie (2009), Fastree (2009), and Bowtie 2 (2012), with the years corresponding to five highest citation-count peaks. **B** Top 10 countries contributing the most tools to SynBioTools

(Fig. 2B). Based on the annual numbers of tools and citations for each module, most of the tools in most of the modules were developed within the past 20 years. Most of the tools in the protein, gene editing, metabolic modeling, and omics modules were developed within the past 10 years (Fig. 3A). SynBioTools lists 1321 de-duplicated tools and 1462 tool records, because some comprehensive tools or databases, such as KEGG, were grouped into more than one module (Fig. 3B). The top 10 tools in terms of citation counts are BLAST, MrBayes, KEGG, GO enrichment analysis tool, PhyML, Bowtie 2, STRING, UniProt BLAST, MAFFT, and BEAST. According to the published sources for each tool, the top 10 databases and tools that are continually updated include KEGG, UniProt BLAST, CTD, NCBI reference sequences, PubChem, EcoCyc, RegulonDB, Reactome, the MetaCyc database, and STRING. SynBioTools shares 564 tools with bio.tools, which is the primary tool registry; of the 757 not shared with bio.tools, 62 are for laboratory experiments, providing cloning strategies and DNA-assembly methods that are critical in synthetic biology. Including these tools provides a one-stop search solution for synthetic biology tools.

### User interface

On the Home Search page, SynBioTools offers two retrieval methods: simple and advanced search (Fig. 4A). In the simple search, possible tools will be displayed while the search term is being typed. For an advanced search, the search term can be the tool name, module, keyword, EDAM term, MeSH term, author, country, institution, or any other term, and search terms can be combined. On the Search Results page, the retrieved tools are shown on the right, with the sorting methods and filtering criteria on the left (Fig. 4B). The tools can be sorted by relevance, recency, and citation count, and filtered by journals, conferences, authors, institutions, and countries. Clicking on the tool name in the search result will load the Tool Details page (Fig. 4C), which includes general information, classifications, labels, credits, publications, and external links, lists other tools in the same category, and provides comparisons with these tools.

The Browse page displays the tool name, module, category, type, publication date, homepage availability, citation, source reference, and review source, allowing tool



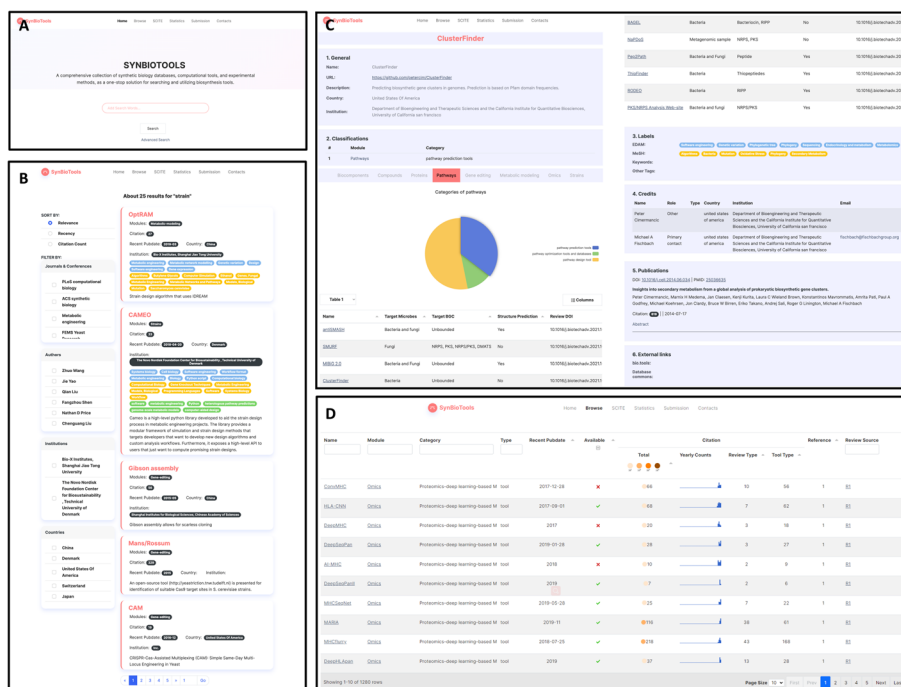
**Fig. 3** Tools in each SynBioTools module. **A** Annual tool number and annual citation count for each of the nine modules. The number of tools in each module is listed next to the module name. **B** Top 10 most-cited tools in each of the nine modules

information retrieval and sorting (Fig. 4D). The Tool Details page can also be accessed by clicking on the tool name on the Browse page.

Our article-table-extraction tool, SCITE, has been integrated into SynBioTools as an online server application. SCITE provides two ways to extract tabular data from scientific papers, and users can choose the mode based on the file type. If a PDF file of an article is uploaded, SCITE will automatically convert the uploaded file into pictures, and identify tables via artificial intelligence. If the user provides an article’s PMID from PubMed Central, SCITE will extract the table information by parsing the full-text XML document, providing more accurate table retrieval. SCITE can be accessed freely at <https://synbiotools.lifesynter.com/scite.html>.

**Discussion**

Synthetic biology research involves the utilization of many databases and computational tools. We constructed SynBioTools, comprehensively listing categorized synthetic biology tools, to make it easier to search and select biosynthetic tools and conduct synthetic biology research. SynBioTools lists computational tools, databases, and methods grouped into nine modules based on their potential biosynthetic applications. Unlike



**Fig. 4** SynBioTools content and web interface. **A** Home Search page, including simple and advanced search. **B** Search Results page. The results can be sorted by relevance, recency, or citation count, and filtered by journals, authors, institutions, and countries. **C** Tool Details page, including general information, classification, labels, credits, publications, and external links. **D** Browse page, showing the tool’s name, module, category, data type, homepage availability, publication date, citation, tool reference, and review source

existing registries, SynBioTools lists tools, databases, and methods related to most bio-synthesis processes in order to facilitate tool discovery, sharing, and reutilization across the field of synthetic biology. SynBioTools also includes experimental laboratory methods, such as DNA assembly and cloning strategies, to allow researchers to locate and retrieve all methods in one place. Approximately 57% of the tools listed in SynBioTools are not found in the most comprehensive tool registry, bio.tools. Although OMICtools lists a larger number of omics analysis tools and has a good classification system, it is currently not available [14]. Additionally, while SMBP provides computational tools for secondary metabolite production, it does not offer researchers a one-stop search facility for other tools [17].

As well as enabling tool retrieval, SynBioTools provides a comprehensive overview of synthetic biology tools and includes a wealth of tools and database resources for constructing workflows and large comprehensive databases. It reveals that the number of synthetic biology tools has grown rapidly in the past 20 years, especially in the fields of omics and gene editing; this growth is closely related to the emergence and rapid development of sequencing and CRISPR/Cas technologies. Omics and gene editing are driving rapid technological developments in synthetic biology [37]. Genome editing, via programmable nucleases, is revolutionizing the life sciences and medicine; currently available CRISPR/Cas-related tools facilitate convenient and reliable genome-editing experiments at every step, from designing guide RNA to analyzing gene editing outcomes [31]. In recent years, the enormous progress in developing protein design tools

has promoted rapid development in the field of protein design. Protein design is no longer restricted to fundamentals and the analysis of protein folding. Our ability to generate and manipulate synthetic proteins has advanced to the point where they provide realistic alternatives to the functions of natural proteins for both in vitro and intracellular applications. Furthermore, computer-based protein design is becoming increasingly accepted by non-specialists [55]. The collation and classification that SynBioTools provides are conducive to the integration and construction of larger and more comprehensive databases, such as COCONUT, an aggregated open-source dataset of known and predicted natural products [56], as well as integration and interoperability between databases [57]. Workflows can integrate multiple tools to handle analyses that are too complex to be addressed using a single tool [58]. SynBioTools is conducive to the construction of workflows for complex, multi-task data analyses, integrating tools for every step, from chemical selection to pathway design, enzyme selection, gene editing, and omics analysis.

When constructing SynBioTools, we encountered various difficulties, including those related to tabular information extraction and data de-duplication. Data acquisition was a critical step in constructing our tool registry. The current commonly used PDF table batch-extraction tools for extracting structured data from the literature are Tabula (<https://github.com/tabulapdf/tabula>) and Camelot (<https://github.com/camelot-dev/camelot>), which have been used for table extraction [59, 60]. However, for some PDF documents, these tools do not perform very well. Therefore, to improve performance and generality, we developed SCITE, which can better extract tabular data from reviews and other types of scientific papers. Further, SynBioTools provides a new strategy for data extraction: find reviews that cite the tool from the identified tools, filter the reviews for the topics of interest, then acquiring additional tools and information from the screened reviews. This makes it possible to rapidly locate topic-specific tools and tool information.

Duplicate removal and tool updates presented difficulties in terms of data curation during our construction of SynBioTools. For example, the same tool may be referred to in different source papers, requiring the merging of records. However, tool disambiguation is difficult because tools do not have a unique identification number. Therefore, we identified unique tools based on the tool name, reference, link, and other factors. Further, some tool updates are described in published articles, while others are provided as ongoing updates. If each tool could be assigned a unique ID number through a system or platform upon tool release, and all updates are linked to the same ID, this would provide a potential solution. However, this would depend on consensus among all tool publishers and publication journals, as well as ID registration and maintenance platforms.

All of the tools in SynBioTools were extracted from reviews. However, due to the publication lag for review articles, the list includes little to no tools that have appeared within the past two years. To address this, we added a small number of synthetic biology tools that are not derived from the review literature. Additionally, we provide a channel for users to manually submit tool information. In the future, given the constant publication of synthetic biology reviews, we will regularly update the data in SynBioTools. This includes updating changes to existing tools and adding new tools to SynBioTools. Concretely, we will perform the data process steps shown in



**Fig. 1.** The only difference is to remove reviews that have been previously processed. In addition, due to the lagging nature of the review literature, we will periodically add synthetic biology tools that are not derived from the review to provide basic tool search, although these tools lack information like detailed comparisons of similar tools extracted from reviews. At the same time, new natural language processing techniques will be applied to optimize the entire data processing pipeline to minimize the reliance on expert curation. SynBioTools focuses on synthetic biology, rather than attempting to address all aspects of computational biology. Nevertheless, it presents a useful catalog of synthetic biology tools for researchers and tool developers.

## Conclusions

We constructed SynBioTools, which includes computational tools, databases, and methods, to improve the ease of locating tools used in synthetic biology. SynBioTools combines the advantages of data collation and comparison of review articles with the ease of interaction of databases. It extracts biosynthesis-related tools from published reviews of synthetic biology tools, classifies them according to their characteristics and potential biosynthetic applications, and integrates extra information, such as tool URLs, source references, and the number of citations, to assist users and developers in tool retrieval and selection. SynBioTools provides researchers with an efficient, one-stop search and selection facility for finding synthetic biology tools, as well as a source of tools for further workflow construction.

## Abbreviations

BioJS	BioJavaScript
SMBP	Secondary Metabolite Bioinformatics Portal
SCITE	SCientific Table Extraction
S2ORC	Semantic Scholar Open Research Corpus
OCR	Optical Character Recognition

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-023-05281-5>.

**Additional file 1.** The code zip file on extracting tabular information from papers in PDF format based on OCR.

**Additional file 2.** The code zip file on extracting tabular information from papers by parsing the full-text XML file.

**Additional file 3.** The list of reviews used for the tool and tool information extraction.

## Acknowledgements

Not applicable.

## Author contributions

PC, SL, and QH designed the project. PC and SL conducted the project. DZ, HX, DL, MH, and LG validated the database. QH supervised the project. PC and SL wrote the manuscript. DZ, HX, DL, MH, and LG reviewed and edited the manuscript. All authors read and approved the final manuscript.

## Funding

This work was financially supported by the National Key Research and Development Program of China [Grant Numbers: 2021YFC2103001, 2019YFA0904300] and the International Partnership Program of the Chinese Academy of Sciences of China [Grant Number: 153D31KYSB20170121].

## Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 9 January 2023 Accepted: 11 April 2023

Published online: 17 April 2023

## References

- Otero-Muras I, Carbonell P. Automated engineering of synthetic metabolic pathways for efficient biomanufacturing. *Metab Eng.* 2021;63:61–80.
- Lawson CE, Martí JM, Radivojevic T, Jonnalagadda SVR, Gentz R, Hillson NJ, Peisert S, Kim J, Simmons BA, Petzold CJ, et al. Machine learning for metabolic engineering: a review. *Metab Eng.* 2021;63:34–60.
- Volk MJ, Lourentzou I, Mishra S, Vo LT, Zhai C, Zhao H. Biosystems design by machine learning. *ACS Synth Biol.* 2020;9(7):1514–33.
- Wilkinson MD, Links M. BioMOBY: an open source biological web services proposal. *Brief Bioinform.* 2002;3(4):331–41.
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 2004;5(10):R80.
- Bhagat J, Tanoh F, Nzuobontane E, Laurent T, Orłowski J, Roos M, Wolstencroft K, Aleksejevs S, Stevens R, Pettifer S, et al. BioCatalogue: a universal catalogue of web services for the life sciences. *Nucleic Acids Res.* 2010;38(Web Server issue):W689–94.
- Li JW, Robison K, Martin M, Sjödin A, Usadel B, Young M, Olivares EC, Bolser DM. The SEQanswers wiki: a wiki database of tools for high-throughput sequencing analysis. *Nucleic Acids Res.* 2012;40(Web Server issue):D1313–7.
- Yachdav G, Goldberg T, Wilzbach S, Dao D, Shih I, Choudhary S, Crouch S, Franz M, García A, García LJ, et al. Anatomy of BioJS, an open source community for the life sciences. *Elife.* 2015;4:e07009.
- Corpas M, Jimenez R, Carbon SJ, García A, García L, Goldberg T, Gomez J, Kalderimis A, Lewis SE, Mulvany I, et al. BioJS: an open source standard for biological visualization—its status in 2014. *F1000Res.* 2014;3:55.
- Bai J, Bandla C, Guo J, Vera Alvarez R, Bai M, Vizcaíno JA, Moreno P, Grüning B, Sallou O, Perez-Riverol Y. BioContainers registry: searching bioinformatics and proteomics tools, packages, and containers. *J Proteome Res.* 2021;20(4):2056–61.
- Henry VJ, Bandrowski AE, Pepin AS, Gonzalez BJ, Desfeux A. OMICTools: an informative directory for multi-omic data analysis. *Database (Oxford).* 2014;2014:bau069.
- Gnimpieba EZ, VanDiermen MS, Gustafson SM, Conn B, Lushbough CM. Bio-TDS: bioscience query tool discovery system. *Nucleic Acids Res.* 2017;45(D1):D1117–d1122.
- Ison J, Rapacki K, Ménager H, Kalaš M, Rydza E, Chmura P, Anthon C, Beard N, Berka K, Bolser D, et al. Tools and data services registry: a community effort to document bioinformatics resources. *Nucleic Acids Res.* 2016;44(D1):D38–47.
- Friedrichs M, Shoshi A, Chmura PJ, Ison J, Schwämmle V, Schreiber F, Hofestädt R, Sommer B. JIB.tools 2.0—A Bioinformatics Registry for Journal Published Tools with Interoperability to bio.tools. *J Integr Bioinform.* 2020;16(4):201.
- Duvaud S, Gabella C, Lisacek F, Stockinger H, Ioannidis V, Durinx C. Expaty, the Swiss bioinformatics resource portal, as designed by its users. *Nucleic Acids Res.* 2021;49(W1):W216–w227.
- Xie C, Jauhari S, Mora A. Popularity and performance of bioinformatics software: the case of gene set analysis. *BMC Bioinformatics.* 2021;22(1):191.
- Weber T, Kim HU. The secondary metabolite bioinformatics portal: computational tools to facilitate synthetic biology of secondary metabolite production. *Synth Syst Biotechnol.* 2016;1(2):69–79.
- Zielinski DC, Patel A, Palsson BO. The expanding computational toolbox for engineering microbial phenotypes at the genome scale. *Microorganisms.* 2020;8(12):2050.
- Majewska M, Wysokińska H, Kuźma Ł, Szymczyk P. Eukaryotic and prokaryotic promoter databases as valuable tools in exploring the regulation of gene transcription: a comprehensive overview. *Gene.* 2018;644:38–48.
- Misra BB, Langefeld CD, Olivier M, Cox LA. Integrated omics: tools, advances, and future approaches. *J Mol Endocrinol.* 2018.
- Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. Current status and applications of genome-scale metabolic models. *Genome Biol.* 2019;20(1):121.
- Chen C, Hou J, Tanner JJ, Cheng J. Bioinformatics methods for mass spectrometry-based proteomics data analysis. *Int J Mol Sci.* 2020;21(8):2873.
- Dhingra S, Sowdhamini R, Cadet F, Offmann B. A glance into the evolution of template-free protein structure prediction methodologies. *Biochimie.* 2020;175:85–92.
- Ejigu GF, Jung J. Review on the computational genome annotation of sequences obtained by next-generation sequencing. *Biology (Basel).* 2020;9(9):295.
- Guala D, Ogris C, Müller N, Sonnhammer ELL. Genome-wide functional association networks: background, data & state-of-the-art resources. *Brief Bioinform.* 2020;21(4):1224–37.
- Hanna RE, Doench JG. Design and analysis of CRISPR-Cas experiments. *Nat Biotechnol.* 2020;38(7):813–23.
- Kapli P, Yang Z, Telford MJ. Phylogenetic tree building in the genomic age. *Nat Rev Genet.* 2020;21(7):428–44.
- Makrodimitris S, van Ham R, Reinders MJT. Automatic gene function prediction in the 2020's. *Genes (Basel).* 2020;11(11):1264.

29. McCarty NS, Graham AE, Studena L, Ledesma-Amaro R. Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nat Commun.* 2020;11(1):1281.
30. Ren H, Shi C, Zhao H. Computational tools for discovering and engineering natural product biosynthetic pathways. *iScience.* 2020;23(1):100795.
31. Sledzinski P, Nowaczyk M, Olejniczak M. Computational tools and resources supporting CRISPR-Cas experiments. *Cells.* 2020;9(5):1288.
32. Sorokina M, Steinbeck C. Review on natural products databases: where to find data in 2020. *J Cheminform.* 2020;12(1):20.
33. Wen B, Zeng WF, Liao Y, Shi Z, Savage SR, Jiang W, Zhang B. Deep learning in proteomics. *Proteomics.* 2020;20(21–22):e1900335.
34. Alam K, Hao J, Zhang Y, Li A. Synthetic biology-inspired strategies and tools for engineering of microbial natural product biosynthetic pathways. *Biotechnol Adv.* 2021;49:107759.
35. Ayres LB, Gomez FJV, Linton JR, Silva MF, Garcia CD. Taking the leap between analytical chemistry and artificial intelligence: a tutorial review. *Anal Chim Acta.* 2021;1161:338403.
36. Baltoumas FA, Zafeiropoulou S, Karatzas E, Koutrouli M, Thanati F, Voutsadaki K, Gkonta M, Hotova J, Kasionis I, Hatzis P, et al. Biomolecule and bioentity interaction databases in systems biology: a comprehensive review. *Biomolecules.* 2021;11(8):1245.
37. Bao XR, Pan Y, Lee CM, Davis TH, Bao G. Tools for experimental and computational analyses of off-target editing by programmable nucleases. *Nat Protoc.* 2021;16(1):10–26.
38. Bin Hafeez A, Jiang X, Bergen PJ, Zhu Y. Antimicrobial peptides: an update on classifications and databases. *Int J Mol Sci.* 2021;22(21):11691.
39. Chung CH, Lin DW, Eames A, Chandrasekaran S. Next-generation genome-scale metabolic modeling through integration of regulatory mechanisms. *Metabolites.* 2021;11(9):606.
40. Jendoubi T. Approaches to integrating metabolomics and multi-omics data: a primer. *Metabolites.* 2021;11(3):184.
41. Luo J, Wei Y, Lyu M, Wu Z, Liu X, Luo H, Yan C. A comprehensive review of scaffolding methods in genome assembly. *Brief Bioinform.* 2021;22(5):bbab033.
42. Marabotti A, Scafuri B, Facchiano A. Predicting the stability of mutant proteins by computational approaches: an overview. *Brief Bioinform.* 2021;22(3):bbaa074.
43. Misra BB. New software tools, databases, and resources in metabolomics: updates from 2020. *Metabolomics.* 2021;17(5):49.
44. Pakhrin SC, Shrestha B, Adhikari B, Kc DB. Deep learning-based advances in protein structure prediction. *Int J Mol Sci.* 2021;22(11):5553.
45. Pereira JM, Vieira M, Santos SM. Step-by-step design of proteins for small molecule interaction: a review on recent milestones. *Protein Sci.* 2021;30(8):1502–20.
46. Santiago-Rodriguez TM, Hollister EB. Multi-omic data integration: a review of concepts, considerations, and approaches. *Semin Perinatol.* 2021;45(6):151456.
47. Sequeiros-Borja CE, Surpeta B, Brezovsky J. Recent advances in user-friendly computational tools to engineer protein function. *Brief Bioinform.* 2021;22(3):bbaa150.
48. Suthers PF, Foster CJ, Sarkar D, Wang L, Maranas CD. Recent advances in constraint and machine learning-based metabolic modeling by leveraging stoichiometric balances, thermodynamic feasibility and kinetic law formalisms. *Metab Eng.* 2021;63:13–33.
49. Worheide MA, Krumsiek J, Kastenmuller G, Arnold M. Multi-omics integration in biomedical research—a metabolomics-centric review. *Anal Chim Acta.* 2021;1141:144–62.
50. Wu M, Yi H, Ma S. Vertical integration methods for gene expression data analysis. *Brief Bioinform.* 2021;22(3):bbaa169.
51. Young R, Haines M, Storch M, Freemont PS. Combinatorial metabolic pathway assembly approaches and toolkits for modular assembly. *Metab Eng.* 2021;63:81–101.
52. Zou Y, Zhu Y, Li Y, Wu FX, Wang J. Parallel computing for genome sequence processing. *Brief Bioinform.* 2021;22(5):bbab070.
53. Luo L, Yang J, Wang C, Wu J, Li Y, Zhang X, Li H, Zhang H, Zhou Y, Lu A, et al. Natural products for infectious microbes and diseases: an overview of sources, compounds, and chemical diversities. *Sci China Life Sci.* 2022;65(6):1123–45.
54. Kern F, Fehlmann T, Keller A. On the lifetime of bioinformatics web services. *Nucleic Acids Res.* 2020;48(22):12523–33.
55. Woolfson DN. A brief history of de Novo protein design: minimal, rational, and computational. *J Mol Biol.* 2021;433(20):167160.
56. Sorokina M, Merseburger P, Rajan K, Yirik MA, Steinbeck C. COCONUT online: collection of open natural products database. *J Cheminform.* 2021;13(1):2.
57. van Santen JA, Kautsar SA, Medema MH, Lington RG. Microbial natural product databases: moving forward in the multi-omics era. *Nat Prod Rep.* 2021;38(1):264–78.
58. Wratten L, Wilm A, Göke J. Reproducible, scalable, and shareable analysis pipelines with bioinformatics workflow managers. *Nat Methods.* 2021;18(10):1161–8.
59. Huang Y, Burgoine T, Essman M, Theis DRZ, Bishop TRP, Adams J. Monitoring the nutrient composition of food prepared out-of-home in the united kingdom: database development and case study. *JMIR Public Health Surveill.* 2022;8(9):e39033.
60. Jaber-Douraki M, Taghian Dinani S, Millagaha Gedara NI, Xu X, Richards E, Maunsell F, Zad N, Tell LA. Large-scale data mining of rapid residue detection assay data from HTML and PDF documents: improving data access and visualization for veterinarians. *Front Vet Sci.* 2021;8:674730.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.