

RESEARCH

Open Access



SynBioStrainFinder: A microbial strain database of manually curated CRISPR/Cas genetic manipulation system information for biomanufacturing

Pengli Cai^{1†}, Mengying Han^{1†}, Rui Zhang¹, Shaozhen Ding², Dachuan Zhang¹, Dongliang Liu¹, Sheng Liu¹ and Qian-Nan Hu^{1*}

Abstract

Background: Microbial strain information databases provide valuable data for microbial basic research and applications. However, they rarely contain information on the genetic operating system of microbial strains.

Results: We established a comprehensive microbial strain database, SynBioStrainFinder, by integrating CRISPR/Cas gene-editing system information with cultivation methods, genome sequence data, and compound-related information. It is presented through three modules, Strain2Gms/PredStrain2Gms, Strain2BasicInfo, and Strain2Compd, which combine to form a rapid strain information query system conveniently curated, integrated, and accessible on a single platform. To date, 1426 CRISPR/Cas gene-editing records of 157 microbial strains have been manually extracted from the literature in the Strain2Gms module. For strains without established CRISPR/Cas systems, the PredStrain2Gms module recommends the system of the most closely related strain as a reference to facilitate the construction of a new CRISPR/Cas gene-editing system. The database contains 139,499 records of strain cultivation and genome sequences, and 773,298 records of strain-related compounds. To facilitate simple and intuitive data application, all microbial strains are also labeled with stars based on the order and availability of strain information. SynBioStrainFinder provides a user-friendly interface for querying, browsing, and visualizing detailed information on microbial strains, and it is publicly available at <http://design.rxnfinder.org/biosynstrain/>.

Conclusion: SynBioStrainFinder is the first microbial strain database with manually curated information on the strain CRISPR/Cas system as well as other microbial strain information. It also provides reference information for the construction of new CRISPR/Cas systems. SynBioStrainFinder will serve as a useful resource to extend microbial strain research and application for biomanufacturing.

Keywords: CRISPR/Cas system, Gene-editing method, Microorganism, Database, Strain cultivation, Genome sequence, Strain-related compound

Background

The development of genome sequencing and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) technology has allowed an increasing number of microbial strains with excellent or unique characteristics to be studied and

[†]Pengli Cai and Mengying Han contributed equally to this work

*Correspondence: qnhu@sibs.ac.cn

¹ 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
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

exploited [1–3]. These strains also provide potential chassis options for biological manufacturing [4]. Meanwhile, several databases with information of microorganisms have been published, besides the National Center for Biotechnology Information (NCBI) and strain culture collections. The Global Catalog of Microorganisms gathers strain catalog information [5]. KOMODO collects the culture medium for all bacterial strains and provides possible formulations for strains without established culture media [6]. Cell2Chem includes 40,370 species and 125,212 compounds with microbial strain information [7]. These databases provide valuable information for microbiological research [8, 9]. However, genetic manipulation systems for microbial strains are not frequently reported in these databases.

Genetic manipulation systems are indispensable for basic research and biomanufacturing applications. CRISPR/Cas has become the most well-known and widely used method for gene editing compared with the Cre/loxP recombination system, zinc-finger nucleases (ZFNs), and transcription activator-like effector nucleases (TALENs) [10–13]. The CRISPR/Cas system originates from the adaptive immune system against invading foreign nucleic acids and is widely found in bacteria and archaea [14]. It forms the basis of powerful gene-editing tools [15]. It consists of an endonuclease and tracrRNA/crRNA, which is further simplified as a single sgRNA [15]. Cas9 and Cpf1 (Cas12a) are the most extensively studied endonucleases for gene editing [1]. In addition, via modification and fusion, the derived mutant Cas9 nickase and nuclease-deficient Cas9 (with only DNA binding activity but no cleavage activity), which can be used for gene regulation independently or by fusion with other elements, further improve and expand the application of this system in model and non-model microorganisms, enabling CRISPR-mediated epigenome editing, genome/chromatin imaging, and manipulation of chromatin topology [1, 2, 12, 16]. The development of simple, rapid, powerful, and economical CRISPR/Cas9 technologies has introduced a new era for genome editing, and they have been applied in a wide range of fields, including clinical, pharmaceutical, agricultural, food, and energy fields [17–21]. Computational tools and resources supporting CRISPR-Cas experiments have also been developed, including a variety of sgRNA design tools [22–26]. Addgene (<https://www.addgene.org/>) [27], an international nonprofit plasmid and data resource, can retrieve the CRISPR/Cas plasmids of some but not all microbial strains, and it does not include additional microbial strain information. Comprehensive and convenient methods for retrieving microbial species-related information are lacking. Therefore, the development of a comprehensive database of strains with information on the CRISPR/Cas

genetic manipulation system, as well as other relevant information, is needed.

Here, we established SynBioStrainFinder (<http://design.rxnfinder.org/biosynstrain/>), a knowledge database containing CRISPR/Cas genetic manipulation system information of microbial strains. The cultivation, genome sequences, and compound-related information of all microbial strains were also integrated to facilitate rapid queries for strain-related information in one place. Information can be retrieved using the following three modules: Strain2BasicInfo for cultivation and genome sequence information integrated from several databases, Strain2Gms for CRISPR/Cas genetic manipulation system information manually curated from the literature, and Strain2Compd for strain-related compounds calculated using the term frequency-inverse document frequency (TF-IDF) method. It also provides CRISPR/Cas system information for the most closely related strains as a reference to facilitate new construction in the PredStrain2Gms module. SynBioStrainFinder provides a useful and comprehensive one-stop microbial strain data resource with a user-friendly interface for querying, browsing, and visualizing detailed information about the CRISPR/Cas system, as well as other microbial strain properties. To facilitate the application of strain information in a simple and intuitive way, all microbial strains are labeled with stars based on the order and availability of data for culture media, genome sequencing, genetic manipulation systems, and strain-related compounds. We expect this database to serve as an important resource and extend the utilization of microbial strains by microbiologists and synthetic biologists.

Results

Database summary

SynBioStrainFinder is the first database of microbial strain information with manually curated data for the CRISPR/Cas gene-editing method, as well as strain cultivation, genome sequencing, and strain-related compounds. It consists of three modules, namely, Strain2BasicInfo for genome sequence data and basic information, Strain2Gms/PredStrain2Gms containing information on the CRISPR/Cas genetic manipulation system for each strain or providing a reference for strains without an established CRISPR/Cas gene-editing system, and Strain2Compd providing strain-related compounds. To date, SynBioStrainFinder contains information for 32,320 species, including 16,404 fungi, 14,072 bacteria, 483 archaea, and 1361 algae. There are 139,499 records of strain growth, 1426 records of CRISPR/Cas systems, and 773,298 records of 1768 microbial strains with compound information in Strain2BasicInfo, Strain2Gms, and Strain2Compd, respectively. In SynBioStrainFinder,

11.4% of fungi, 62.8% of bacteria, 69.0% of archaea, and 9.6% of algae have sequenced genomes. Up to June 2020, 157 microbial species had an established CRISPR/Cas gene-editing system. In addition, 4.9% of fungi (78), 0.9% of bacteria (75), 0.3% of archaea (1), and 3.0% of algae (3) among taxa with sequenced genomes had a CRISPR/Cas genome editing system (Fig. 1a).

The CRISPR/Cas gene-editing method is the most popular genetic manipulation method (Fig. 1b). In the Strain2Gms module, we evaluated the delivery type, editing type, promoters used for Cas9 and sgRNA

expression, homologous arm length of donor DNA for Homology directed repair (HDR), and commonly used sgRNA design tools. (1) Based on the current database statistics, plasmid (74%) is the main construction and delivery mode, followed by genomic integration (13%), transient expression (7%), and ribonucleoproteins (RNPs) (4%) (Fig. 1c). (2) The most frequently used promoters for Cas9 expression are the *tef* [28–32] and *teto* promoters [33–36] in fungi and bacteria, respectively. For sgRNA expression, the commonly used promoters are the SNR52 [30, 37] and U6 promoters [28, 38]

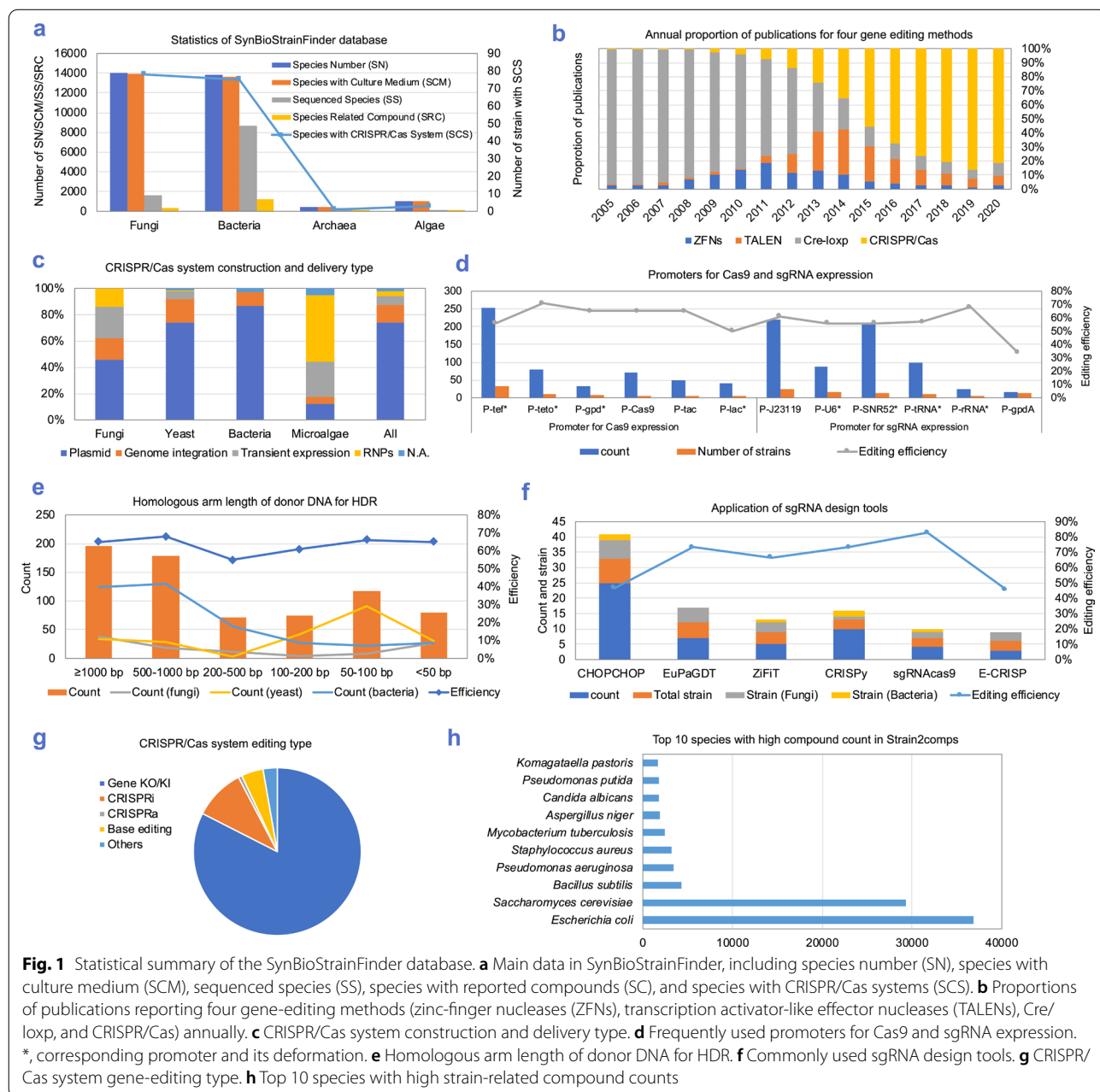


Fig. 1 Statistical summary of the SynBioStrainFinder database. **a** Main data in SynBioStrainFinder, including species number (SN), species with culture medium (SCM), sequenced species (SS), species with reported compounds (SC), and species with CRISPR/Cas systems (SCS). **b** Proportions of publications reporting four gene-editing methods (zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), Cre/ loxp, and CRISPR/Cas) annually. **c** CRISPR/Cas system construction and delivery type. **d** Frequently used promoters for Cas9 and sgRNA expression. *, corresponding promoter and its deformation. **e** Homologous arm length of donor DNA for HDR. **f** Commonly used sgRNA design tools. **g** CRISPR/ Cas system gene-editing type. **h** Top 10 species with high strain-related compound counts

in fungi and the J23119 promoter [39, 40] in bacteria (Fig. 1d). (3) For precise editing by HDR, the average lengths of repair templates for fungi, yeast, and bacteria are 567, 276, and 612 bp, respectively. Templates in the length range of 200–500 bp are the least used and have relatively low editing efficiency. Templates of 50–100 in length bp are the most commonly used for yeast. Bacteria usually use longer homologous arms (Fig. 1e). The length selection of the homologous arm of the donor DNA usually depends on the intrinsic DSB repair mechanism of the strain. (4) CHOPCHOP [41–43] is the most commonly used sgRNA design tool (Fig. 1f). To facilitate sgRNA design, we selected and updated the webserver tools from WeReview [24] as CRISPR tools on the home page of SynBioStrainFinder. (5) In the current database, the main editing types are CRISPR-based gene knockout/knockin (83%) and CRISPR interference (CRISPRi) (10%) (Fig. 1g). In addition to the factors mentioned above, other information related to the CRISPR/Cas system, including edited genes, selection markers, and transformation methods, can be retrieved from the database if this information is included in the literature.

There are 773,298 records of compound information corresponding to 1768 microbial strains in the Strain2Compd module. Out of the top 10 species with the highest counts of the corresponding compounds, six strains are traditional industrial microorganisms used in biomanufacturing, namely *Escherichia coli* [44], *Saccharomyces cerevisiae* [45], *Bacillus subtilis* [46, 47], *Aspergillus niger* [48], *Pseudomonas putida* [49–51], and *Komagataella pastoris* [52, 53]. *Escherichia coli* and *S. cerevisiae* are the most extensively studied strains, which is also reflected by a higher number of related compounds for these strains than for other strains. The other four species are pathogens, namely *Pseudomonas aeruginosa* [54], *Staphylococcus aureus* [55], *Mycobacterium tuberculosis* [56], and *Candida albicans* [57] (Fig. 1h).

SynBioStrainFinder provides a user-friendly interface for querying, browsing, and visualizing detailed information on microbial strains, especially the CRISPR/Cas gene-editing system. Furthermore, to obtain information of these strains in a simple and intuitive manner, all microbial strains were marked with stars according to the order and extent of information available on strain cultivation, genome sequence, CRISPR/Cas genetic operating system, and related compounds. Star-level information can be used for microbial strain selection. For example, it can be applied to the CF-targeter [58], which is a web server for host organism selection of biosynthetic pathway design. In the search interface for host organism selection for astaxanthin biosynthesis, by adding a strain star-tag next to the strain list, we can clearly and

intuitively obtain information about the corresponding strain, thereby assisting strain selection (Fig. 2).

User interface

Users can quickly browse the database using the Latin name of the species on the first page. First, stars indicating the availability of information about the strain are shown at the top of the retrieved page immediately below the strain name. The following are the three modules: Strain2BasicInfo, Strain2Gms/PredStrain2Gms, and Strain2Compd (Fig. 3a–c). The top of the browse-results page shows statistical information for the Strain2Gms and Strain2Compd modules and the phylogenetic tree for the Strain2BasicInfo module. The bottom of the browse-results page contains detailed information (Fig. 3a–c). SynBioStrainFinder also supports retrieval using the strain's generic name. All strains in this genus will be displayed below the search box on the first browse page, from which specific strains can be selected.

On the Strain2BasicInfo browse page, the strain genome sequence (Fig. 3d), as well as culture medium and conditions, can be obtained by clicking “detail” to enter the detailed information page (Fig. 3e, f). On the Strain2Gms browse page, the CRISPR/Cas construction type, publication time, and numbers are counted and shown at the top of the page. Clicking on each statistical color block or dot displays the details in the detailed data entry at the bottom of the page (Fig. 3b). The detailed CRISPR/Cas information includes the Cas9 and sgRNA expression plasmids, promoters, and terminators (Fig. 3g). In addition, the most commonly used CRISPR/Cas tools are shown in the navigation bar on the first page (Fig. 3j). For strains without an established CRISPR/Cas system, the PredStrain2Gms module appears instead of the Strain2Gms module to show the reference information (Fig. 3h). On the Strain2Compd browse page, the upper left shows the statistics for retrieving compounds associated with the strain. The upper right shows the statistical information for strains associated with the selected compound. An example of compound information is shown in Fig. 3c, i.

Discussion

Genetic operating systems play an important role in the research and application of microorganisms. With the rapid development and application of CRISPR/Cas systems, a growing number of distinctive microbial strains could be developed into microbial cell factories. Nevertheless, most of the current microbial strain databases do not include information about genetic operating systems. Therefore, we built the SynBioStrainFinder database, which includes information on the most common CRISPR/Cas gene-editing system corresponding to all

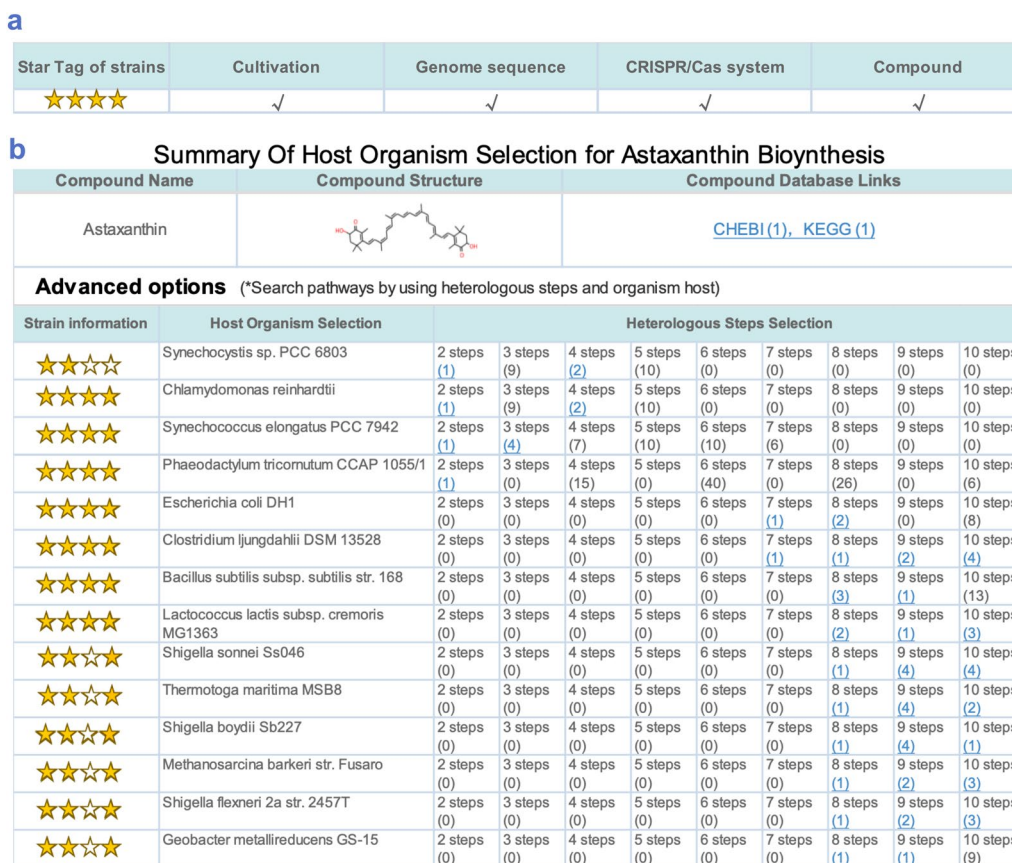


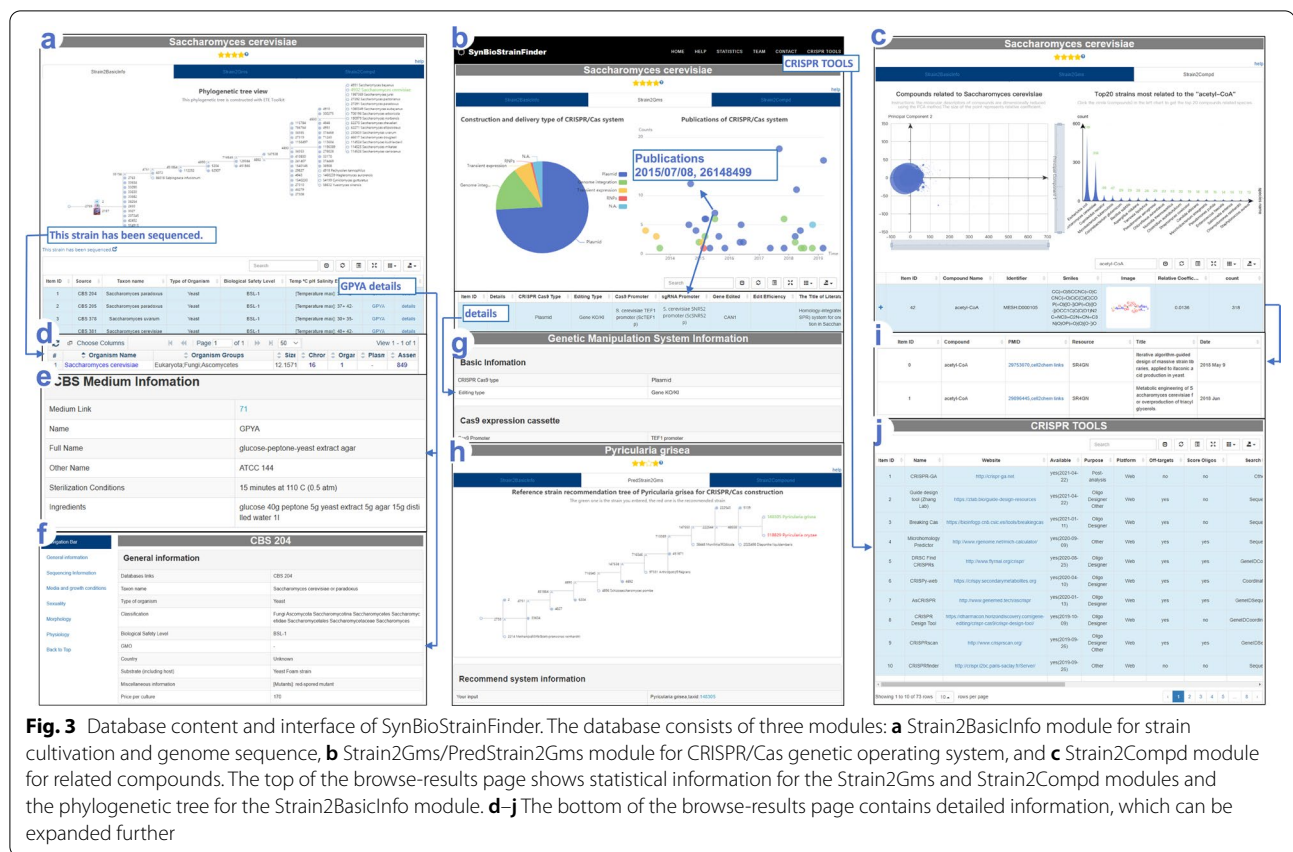
Fig. 2 Application of SynBioStrainFinder data. **a** Example of star-tags for strains in SynBioStrainFinder. All microbial strains were marked with stars according to the order and extent of information available on strain cultivation, genome sequence, CRISPR/Cas genetic operating system, and related compounds. **b** Application of star-tags to host organism selection with CF-targeter. The strain star-tag can be added next to the strain list

microbial strains. Furthermore, it provides a reference strain with an established CRISPR/Cas system for the construction of new systems. We also integrated cultivation, genome sequence, and compound information simultaneously to facilitate the rapid retrieval of microbial strain information in one place.

Although only 157 microbial strains have reported CRISPR/Cas gene-editing systems through the statistics of SynBioStrainFinder, the range of strains with genetic operating systems has expanded substantially. Among the 157 strains with established CRISPR/Cas gene-editing systems, only 29% utilize the Cre/loxP method, illustrating this point. Shared features of the CRISPR/Cas gene-editing method are among different species in the same genus, to a certain extent, although the method of some species requires slight modifications. Accordingly, we developed the PredStrain2Gms module based on the evolutionary relationships among strains with and without genome editing systems, providing a basis for the construction of a new system. Owing to the relatively small number of strains building CRISPR/Cas systems,

PredStrain2Gms may only provide a useful reference for a limited number of strains. The statistical analysis of data in the Strain2Gms module might also provide some suggestions for new system construction, which includes the delivery method, Cas9 and sgRNA expression promoter, homologous arm of HDR, and sgRNA design tool. Although we provide the editing efficiency, it is affected by many factors, including the characteristics of the target genes in addition to the CRISPR/Cas system itself. Additionally, the size of the dataset affects the statistical analysis. Therefore, when building a new system, it is necessary to make reasonable attempts based on the characteristics of microbial strains.

Statistical results show that plasmids are the most commonly used method among the four CRISPR/Cas delivery methods, each of which has distinct advantages and disadvantages [2, 16]. For plasmid delivery, Cas9 and sgRNA are constructed in one or two plasmids, which are usually designed to enable curing for subsequent gene editing, such as by using the temperature-sensitive replicons repA101, pSG5, repE, and pBL1^{ts}



[40, 59–64]. Plasmid delivery, transient expression, and genomic integration require the effective expression of Cas9 and/or sgRNA, although sgRNA can also be transcribed or synthesized in vitro. The search and optimization of suitable expression regulatory elements can be time-consuming and expensive. The RNPs method is more suited for the CRISPR/Cas system development of relatively new hosts with limited genetic manipulation tools or without a foundation. Owing to the structural complexity of fungi, fairly diverse construction and delivery methods are used relative to bacteria during CRISPR/Cas system construction, and the availability of plasmids, selection of the nuclear localization sequence, and identification of type III promoter and/or promoter effectiveness should be considered [2, 3]. Therefore, in the Strain2Gms module, we ensured that at least one detailed CRISPR/Cas record was provided for each species and relatively more information for fungal CRISPR/Cas systems was provided. Bacteria are considerably simpler than fungi because they mostly contain plasmids and one type of RNA polymerase promoter [1]. For the selection of sgRNA design tools, those offering more than one scoring algorithm to accurately assess gRNA activity are preferable, such as

CRISPOR [65]. The usage range should also be considered, as some sgRNA software models are not universally applicable. For example, the Moreno-Mateos score is best suited for experiments with gRNAs expressed in vitro [23]. In this version of the database, we collected basic gene-editing types, including gene knock-outs, knock-ins, base editing, CRISPRi, and CRISPRa, which are the most used initially.

Owing to the limitations of the CRISPR/Cas data collection method and the limited strains with constructed CRISPR/Cas gene-editing systems, the volume of data in Strain2Gms is relatively small at present. It will be updated continuously for improvement. Furthermore, to further facilitate information acquisition and, thus, promote the research and utilization of microbial strains, other information on microbial strains will be added, such as the plasmids, promoters, metabolic network models, various omics data, and product information. Nevertheless, SynBioStrainFinder is a useful database to facilitate new CRISPR/Cas system construction, providing abundant and concentrated microbial strain information for chassis construction and basic research, as well as a variety of chassis recommendations for biomanufacturing. In addition, labeling strains with stars allows a

simple, intuitive, and convenient-for-use display of strain information in strain selection tools.

Conclusions

SynBioStrainFinder is the first database with manually curated information on the microbial strain CRISPR/Cas system, the most widely used genetic manipulation method. It provides a reference strain with an established CRISPR/Cas system for the construction of new CRISPR/Cas systems. The database also comprises other microbial strain information (cultivation methods, genome sequence data, and strain-related compound information) to facilitate rapid strain information queries. Tagging stars to indicate strain information provides a simple, intuitive basis for microbial strain selection. SynBioStrainFinder will continue to expand, aiming to serve as an important resource to extend microbial strain research and application for biomanufacturing by microbiologists and synthetic biologists.

Methods

Data collection and database content

The CRISPR/Cas system information in the Strain2Gms module was manually curated from the literature. Other information was processed and compiled from public resources, including NCBI [66], DSMZ (<https://www.dsmz.de/>), CBS (<https://wi.knaw.nl/>), UTEX collection (<https://utex.org>), Global Catalog of Microorganisms [5, 67], and Cell2Chem [7]. The basic strain information in Strain2BasicInfo includes the strain name, taxon, safety level, and culture medium and conditions from the CBS, DSMZ, and UTEX databases, as well as genome sequence information from NCBI [66]. Links to external resources were also provided, including PubMed ID, genome sequencing in NCBI, sgRNA design tools, and chemical ID in PubMed.

Strain2Gms/PredStrain2Gms module construction

For CRISPR/Cas genetic manipulation information, all publications up to June 2020 matching the keyword “CRISPR*” and generic names for taxa in the microbial strain list of SynBioStrainFinder were first retrieved from PubMed [66]. A total of 1326 titles and/or abstracts of publications were reviewed to obtain microbial-related CRISPR/Cas systems. After further filtering, 472 publications related to the construction of a CRISPR/Cas tool for microbial strains were retained for a detailed review of the full text to extract CRISPR/Cas construction-related information. The manually extracted information included the species name, PubMed ID for the publication, CRISPR/Cas editing types (Gene KO/KI, CRISPRi, CRISPRa, base editing, and others), CRISPR/Cas construction and delivery types (plasmid, RNPs, transient expression, and genome

integration), CRISPR/Cas editing targets (DNA, RNA, chromosome, and others), CRISPR/Cas system details (Cas9 and sgRNA expression plasmid or expression cassette, and related promoter, terminator, selection marker, and donor DNA for homologous recombination repair (HDR)), CRISPR/Cas editing targets and editing efficiency, and sgRNA design tools.

If a CRISPR/Cas gene-editing system is not available for a retrieved strain, the PredStrain2Gms module replaces Strain2Gms. In this module, the CRISPR/Cas gene-editing system of the most closely related strain is recommended as a reference. Using the ETE Toolkit (3.0) [68], a phylogenetic tree for all strains according to the strain taxonomy ID was constructed to identify the most closely related strain. If similar relationships to multiple strains are detected, the most suitable strain will be selected by an exhaustive coefficient, which is an index reflecting the completeness of items of the CRISPR/Cas system for strains in our library, such as Cas9 marker, sgRNA marker, and editing efficiency information.

Strain2Compd module construction

In the Strain2Compd module, a weighted statistical method, TF-IDF, was used to find the most relevant compound for the target strain. We first extracted all abstracts of articles obtained in which searches of the strain and its relevant compound co-occurred to form a total abstract text set composed of multiple independent article abstracts (N). The TF value of each compound was then calculated, which refers to the frequency at which each compound appears in the total abstract text. Second, the IDF value for each compound was calculated, which was the total number of article abstracts (N) divided by the number of articles containing the compound. Finally, the product of TF and IDF yielded the relative coefficient for each compound. We calculated the correlation coefficient for all compounds related to the strain of interest using the TF-IDF algorithm. Larger correlation coefficients for compounds indicated higher relevance of the strain of interest.

For a term i in document j :

$$W_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$ = number of occurrences of i in j

df_i = number of documents containing i

N = total number of document

System design and implementation

The entire project was conducted using Ubuntu (version 18.04.2). Python (version 3.6.8) and Django (version 1.11.7) were used to build SynBioStrainFinder and the interactive interface. The data for the entire project was stored in MySQL (version 8.0.16). ECharts (version 4.2.0; <http://echarts.baidu.com>) was used as a graphical visualization framework. Bootstrap Table (version 1.15.5) was used for the static and dynamic display of data tables, which relies on Bootstrap (version 3.3.7) and jQuery (version 2.1.1). A modern web browser that supports HTML5, such as Google Chrome, Firefox, Safari, Opera, or IE 9.0+, is recommended. SynBioStrainFinder is freely available to the research community using the web link provided (<http://design.rxnfinder.org/biosynstrain/>). Users are not required to register or login to access the features in the databases.

Abbreviations

Cas: CRISPR-associated; CRISPR: Clustered regularly interspaced short palindromic repeats; CRISPRi: CRISPR interference; HDR: Homology-directed repair; NCBI: National Center for Biotechnology Information; RNPs: Ribonucleoproteins; TALEN: Transcription activator-like effector nuclease; TF-IDF: Term frequency-inverse document frequency; ZFN: Zinc-finger nuclease.

Acknowledgements

Not applicable.

Author contributions

PC, MH, and QH designed the project. PC and MH performed the project. RZ, SD, DL, DZ, and SL validated the database. QH supervised the project. PC and MH wrote the manuscript. SD, DL, DZ, and SL reviewed and edited the manuscript. All authors have read and agreed to the published version of 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 number: 2019YFA0904300], the National Natural Science Foundation of China [Grant numbers: 31700081 and 31570092], the CAS STS program [Grant number: QYZDB-SSW-SMC012], and the International Partnership Program of Chinese Academy of Sciences of China [Grant number: 153D31KYSB20170121].

Availability of data and materials

Not applicable.

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.

Author details

¹CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China. ²Chemical.AI Wuhan, Wuhan 430062, China.

Received: 8 March 2022 Accepted: 2 May 2022

Published online: 14 May 2022

References

1. Yao R, Liu D, Jia X, Zheng Y, Liu W, Xiao Y. CRISPR-Cas9/Cas12a biotechnology and application in bacteria. *Synth Syst Biotechnol*. 2018;3:135–49. <https://doi.org/10.1016/j.synbio.2018.09.004>.
2. Wang Q, Coleman JJ. Progress and challenges: development and implementation of CRISPR/Cas9 technology in filamentous fungi. *Comput Struct Biotechnol J*. 2019;17:761–9. <https://doi.org/10.1016/j.csbj.2019.06.007>.
3. Song R, Zhai Q, Sun L, Huang E, Zhang Y, Zhu Y, Guo Q, Tian Y, Zhao B, Lu H. CRISPR/Cas9 genome editing technology in filamentous fungi: progress and perspective. *Appl Microbiol Biotechnol*. 2019;103:6919–32. <https://doi.org/10.1007/s00253-019-10007-w>.
4. Nora LC, Westmann CA, Guazzaroni ME, Siddaiah C, Gupta VK, Silva-Rocha R. Recent advances in plasmid-based tools for establishing novel microbial chassis. *Biotechnol Adv*. 2019;37: 107433. <https://doi.org/10.1016/j.biotechadv.2019.107433>.
5. Wu L, Sun Q, Sugawara H, Yang S, Zhou Y, McCluskey K, Vasilenko A, Suzuki K, Ohkuma M, Lee Y, et al. Global catalogue of microorganisms (gcm): a comprehensive database and information retrieval, analysis, and visualization system for microbial resources. *BMC Genomics*. 2013;14:933. <https://doi.org/10.1186/1471-2164-14-933>.
6. Oberhardt MA, Zarecki R, Gronow S, Lang E, Klenk HP, Gophna U, Ruppin E. Harnessing the landscape of microbial culture media to predict new organism-media pairings. *Nat Commun*. 2015;6:8493. <https://doi.org/10.1038/ncomms9493>.
7. Liu D, Han M, Tian Y, Gong L, Jia C, Cai P, Tu W, Chen J, Hu QN. Cell 2Chem: mining explored and unexplored biosynthetic chemical spaces. *Bioinformatics*. 2021;36:5269–70. <https://doi.org/10.1093/bioinformatics/btaa660>.
8. Zhulin IB. Databases for microbiologists. *J Bacteriol*. 2015;197:2458–67. <https://doi.org/10.1128/jb.00330-15>.
9. Sun Q, Liu L, Wu L, Li W, Liu Q, Zhang J, Liu D, Ma J. Web resources for microbial data. *Genom Proteom Bioinform*. 2015;13:69–72. <https://doi.org/10.1016/j.gpb.2015.01.008>.
10. Gaj T, Gersbach CA, Barbas CF 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*. 2013;31:397–405. <https://doi.org/10.1016/j.tibtech.2013.04.004>.
11. Mei YZ, Zhu YL, Huang PW, Yang Q, Dai CC. Strategies for gene disruption and expression in filamentous fungi. *Appl Microbiol Biotechnol*. 2019;103:6041–59. <https://doi.org/10.1007/s00253-019-09953-2>.
12. Adli M. The CRISPR tool kit for genome editing and beyond. *Nat Commun*. 2018;9:1911. <https://doi.org/10.1038/s41467-018-04252-2>.
13. Fraczek MG, Naseeb S, Delneri D. History of genome editing in yeast. *Yeast*. 2018;35:361–8. <https://doi.org/10.1002/yea.3308>.
14. Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet*. 2010;11:181–90. <https://doi.org/10.1038/nrg2749>.
15. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337:816–21. <https://doi.org/10.1126/science.1225829>.
16. Wang H, La Russa M, Qi LS. CRISPR/Cas9 in genome editing and beyond. *Annu Rev Biochem*. 2016;85:227–64. <https://doi.org/10.1146/annurev-biochem-060815-014607>.
17. Strich JR, Chertow DS. CRISPR-Cas biology and its application to infectious diseases. *J Clin Microbiol*. 2019;57:e01307-18. <https://doi.org/10.1128/jcm.01307-18>.
18. Chen K, Wang Y, Zhang R, Zhang H, Gao C. CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu Rev Plant Biol*. 2019;70:667–97. <https://doi.org/10.1146/annurev-arplant-050718-100049>.
19. Heidenreich M, Zhang F. Applications of CRISPR-Cas systems in neuroscience. *Nat Rev Neurosci*. 2016;17:36–44. <https://doi.org/10.1038/nrn.2015.2>.
20. Zhang YT, Jiang JY, Shi TQ, Sun XM, Zhao QY, Huang H, Ren LJ. Application of the CRISPR/Cas system for genome editing in microalgae. *Appl Microbiol Biotechnol*. 2019;103:3239–48. <https://doi.org/10.1007/s00253-019-09726-x>.

21. Morio F, Lombardi L, Butler G. The CRISPR toolbox in medical mycology: state of the art and perspectives. *PLoS Pathog.* 2020;16:e1008201. <https://doi.org/10.1371/journal.ppat.1008201>.
22. Alkhnbashi OS, Meier T, Mitrofanov A, Backofen R, Voß B. CRISPR-Cas bioinformatics. *Methods.* 2020;172:3–11. <https://doi.org/10.1016/j.jmeth.2019.07.013>.
23. Sledzinski P, Nowaczyk M, Olejniczak M. Computational tools and resources supporting CRISPR-Cas experiments. *Cells.* 2020;9:1288. <https://doi.org/10.3390/cells9051288>.
24. Torres-Perez R, Garcia-Martin JA, Montoliu L, Oliveros JC, Pazos F. WeReview: CRISPR tools—live repository of computational tools for assisting CRISPR/Cas experiments. *Bioengineering.* 2019;6:63. <https://doi.org/10.3390/bioengineering6030063>.
25. Cui Y, Xu J, Cheng M, Liao X, Peng S. Review of CRISPR/Cas9 sgRNA design tools. *Interdiscip Sci.* 2018;10:455–65. <https://doi.org/10.1007/s12539-018-0298-z>.
26. Tong Y, Weber T, Lee SY. CRISPR/Cas-based genome engineering in natural product discovery. *Nat Prod Rep.* 2019;36:1262–80. <https://doi.org/10.1039/c8np00089a>.
27. Kamens J. The Addgene repository: an international nonprofit plasmid and data resource. *Nucleic Acids Res.* 2015;43:D1152–1157. <https://doi.org/10.1093/nar/gku893>.
28. Huang L, Dong H, Zheng J, Wang B, Pan L. Highly efficient single base editing in *Aspergillus niger* with CRISPR/Cas9 cytidine deaminase fusion. *Microbiol Res.* 2019;223–225:44–50. <https://doi.org/10.1016/j.micres.2019.03.007>.
29. Liu JJ, Kong II, Zhang GC, Jayakody LN, Kim H, Xia PF, Kwak S, Sung BH, Sohn JH, Walukiewicz HE, et al. Metabolic engineering of probiotic *Saccharomyces boulardii*. *Appl Environ Microbiol.* 2016;82:2280–7. <https://doi.org/10.1128/aem.00057-16>.
30. Wang L, Deng A, Zhang Y, Liu S, Liang Y, Bai H, Cui D, Qiu Q, Shang X, Yang Z, et al. Efficient CRISPR-Cas9 mediated multiplex genome editing in yeasts. *Biotechnol Biofuels.* 2018;11:277. <https://doi.org/10.1186/s13068-018-1271-0>.
31. Schwartz CM, Hussain MS, Blenner M, Wheelodon I. Synthetic RNA polymerase III promoters facilitate high-efficiency CRISPR-Cas9-mediated genome editing in *Yarrowia lipolytica*. *ACS Synth Biol.* 2016;5:356–9. <https://doi.org/10.1021/acssynbio.5b00162>.
32. Zhang JL, Peng YZ, Liu D, Liu H, Cao YX, Li BZ, Li C, Yuan YJ. Gene repression via multiplex gRNA strategy in *Y. lipolytica*. *Microb Cell Fact.* 2018;17:62. <https://doi.org/10.1186/s12934-018-0909-8>.
33. Wu J, Cheng ZH, Min D, Cheng L, He RL, Liu DF, Li WW. CRISPRi system as an efficient, simple platform for rapid identification of genes involved in pollutant transformation by *Aeromonas hydrophila*. *Environ Sci Technol.* 2020;54:3306–15. <https://doi.org/10.1021/acs.est.9b07191>.
34. Mo XH, Zhang H, Wang TM, Zhang C, Zhang C, Xing XH, Yang S. Establishment of CRISPR interference in *Methyloburbum extorquens* and application of rapidly mining a new phytoene desaturase involved in carotenoid biosynthesis. *Appl Microbiol Biotechnol.* 2020;104:4515–32. <https://doi.org/10.1007/s00253-020-10543-w>.
35. Shin J, Kang S, Song Y, Jin S, Lee JS, Lee JK, Kim DR, Kim SC, Cho S, Cho BK. Genome engineering of *Eubacterium limosum* using expanded genetic tools and the CRISPR-Cas9 System. *ACS Synth Biol.* 2019;8:2059–68. <https://doi.org/10.1021/acssynbio.9b00150>.
36. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell.* 2013;152:1173–83. <https://doi.org/10.1016/j.cell.2013.02.022>.
37. Swiat MA, Dashko S, den Ridder M, Wijsman M, van der Oost J, Daran JM, Daran-Lapujade P. FnCpf1: a novel and efficient genome editing tool for *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 2017;45:12585–98. <https://doi.org/10.1093/nar/gkx1007>.
38. Liu Q, Gao R, Li J, Lin L, Zhao J, Sun W, Tian C. Development of a genome-editing CRISPR/Cas9 system in thermophilic fungal *Myceliophthora* species and its application to hyper-cellulase production strain engineering. *Biotechnol Biofuels.* 2017;10:1. <https://doi.org/10.1186/s13068-016-0693-9>.
39. Qin Q, Ling C, Zhao Y, Yang T, Yin J, Guo Y, Chen GQ. CRISPR/Cas9 editing genome of extremophile *Halomonas* spp. *Metab Eng.* 2018;47:219–29. <https://doi.org/10.1016/j.ymben.2018.03.018>.
40. Liu Y, Wei WP, Ye BC. High GC content Cas9-mediated genome-editing and biosynthetic gene cluster activation in *Saccharopolyspora erythraea*. *ACS Synth Biol.* 2018;7:1338–48. <https://doi.org/10.1021/acssynbio.7b00448>.
41. Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* 2014;42:W401–407. <https://doi.org/10.1093/nar/gku410>.
42. Labun K, Montague TG, Gagnon JA, Thyme SB, Valen E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* 2016;44:W272–276. <https://doi.org/10.1093/nar/gkw398>.
43. Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* 2019;47:W171–w174. <https://doi.org/10.1093/nar/gkz365>.
44. Pontrelli S, Chiu TY, Lan EI, Chen FY, Chang P, Liao JC. *Escherichia coli* as a host for metabolic engineering. *Metab Eng.* 2018;50:16–46. <https://doi.org/10.1016/j.ymben.2018.04.008>.
45. Belda I, Ruiz J, Santos A, Van Wyk N, Pretorius IS. *Saccharomyces cerevisiae*. *Trends Genet.* 2019;35:956–7. <https://doi.org/10.1016/j.tig.2019.08.009>.
46. Liu Y, Liu L, Li J, Du G, Chen J. Synthetic biology toolbox and chassis development in *Bacillus subtilis*. *Trends Biotechnol.* 2019;37:548–62. <https://doi.org/10.1016/j.tibtech.2018.10.005>.
47. Kovacs AT. *Bacillus subtilis*. *Trends Microbiol.* 2019;27:724–5. <https://doi.org/10.1016/j.tim.2019.03.008>.
48. Cairns TC, Nai C, Meyer V. How a fungus shapes biotechnology: 100 years of *Aspergillus niger* research. *Fungal Biol Biotechnol.* 2018;5:13. <https://doi.org/10.1186/s40694-018-0054-5>.
49. Nickel PI, de Lorenzo V. *Pseudomonas putida* as a functional chassis for industrial biocatalysis: from native biochemistry to trans-metabolism. *Metab Eng.* 2018;50:142–55. <https://doi.org/10.1016/j.ymben.2018.05.005>.
50. Weimer A, Kohlstedt M, Volke DC, Nickel PI, Wittmann C. Industrial biotechnology of *Pseudomonas putida*: advances and prospects. *Appl Microbiol Biotechnol.* 2020;104:7745–66. <https://doi.org/10.1007/s00253-020-10811-9>.
51. Loeschcke A, Thies S. Engineering of natural product biosynthesis in *Pseudomonas putida*. *Curr Opin Biotechnol.* 2020;65:213–24. <https://doi.org/10.1016/j.copbio.2020.03.007>.
52. Peña DA, Gasser B, Zanghellini J, Steiger MG, Mattanovich D. Metabolic engineering of *Pichia pastoris*. *Metab Eng.* 2018;50:2–15. <https://doi.org/10.1016/j.ymben.2018.04.017>.
53. Zhu T, Sun H, Wang M, Li Y. *Pichia pastoris* as a versatile cell factory for the production of industrial enzymes and chemicals: current status and future perspectives. *Biotechnol J.* 2019;14:e1800694. <https://doi.org/10.1002/biot.201800694>.
54. Gellatly SL, Hancock RE. *Pseudomonas aeruginosa*: new insights into pathogenesis and host defenses. *Pathog Dis.* 2013;67:159–73. <https://doi.org/10.1111/2049-632x.12033>.
55. Tam K, Torres VJ. *Staphylococcus aureus* secreted toxins and extracellular enzymes. *Microbiol Spectr.* 2019;7:7–2. <https://doi.org/10.1128/microbiolspec.GPP3-0039-2018>.
56. Ehrst S, Schnappinger D, Rhee KY. Metabolic principles of persistence and pathogenicity in *Mycobacterium tuberculosis*. *Nat Rev Microbiol.* 2018;16:496–507. <https://doi.org/10.1038/s41579-018-0013-4>.
57. Dadar M, Tiwari R, Karthik K, Chakraborty S, Shahali Y, Dhama K. *Candida albicans*—biology, molecular characterization, pathogenicity, and advances in diagnosis and control—an update. *Microb Pathog.* 2018;117:128–38. <https://doi.org/10.1016/j.micpath.2018.02.028>.
58. Ding S, Cai P, Yuan L, Tian Y, Tu W, Zhang D, Cheng X, Sun D, Chen J, Hu QN. CF-Targeter: a rational biological cell factory targeting platform for biosynthetic target chemicals. *ACS Synth Biol.* 2019;8:2280–6. <https://doi.org/10.1021/acssynbio.9b00070>.
59. Wang Y, Wang S, Chen W, Song L, Zhang Y, Shen Z, Yu F, Li M, Ji Q. CRISPR-Cas9 and CRISPR-assisted cytidine deaminase enable precise and efficient genome editing in *Klebsiella pneumoniae*. *Appl Environ Microbiol.* 2018;84:e01834-18. <https://doi.org/10.1128/aem.01834-18>.

60. Yue X, Xia T, Wang S, Dong H, Li Y. Highly efficient genome editing in *N. gorenzanensis* using an inducible CRISPR/Cas9-RecA system. *Biotechnol Lett.* 2020;42:1699–706. <https://doi.org/10.1007/s10529-020-02893-2>.
61. Li L, Wei K, Zheng G, Liu X, Chen S, Jiang W, Lu Y. CRISPR-Cpf1-assisted multiplex genome editing and transcriptional repression in *Streptomyces*. *Appl Environ Microbiol.* 2018;84:e00827-18. <https://doi.org/10.1128/aem.00827-18>.
62. Huang H, Zheng G, Jiang W, Hu H, Lu Y. One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta Biochim Biophys Sin.* 2015;47:231–43. <https://doi.org/10.1093/abbs/gmv007>.
63. Chen W, Zhang Y, Yeo WS, Bae T, Ji Q. Rapid and efficient genome editing in *Staphylococcus aureus* by using an engineered CRISPR/Cas9 system. *J Am Chem Soc.* 2017;139:3790–5. <https://doi.org/10.1021/jacs.6b13317>.
64. Jiang Y, Qian F, Yang J, Liu Y, Dong F, Xu C, Sun B, Chen B, Xu X, Li Y, et al. CRISPR-Cpf1 assisted genome editing of *Corynebacterium glutamicum*. *Nat Commun.* 2017;8:15179. <https://doi.org/10.1038/ncomms15179>.
65. Concordet J-P, Haeussler M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* 2018;46:W242–5. <https://doi.org/10.1093/nar/gky354>.
66. Sayers EW, Beck J, Brister JR, Bolton EE, Canese K, Comeau DC, Funk K, Ketter A, Kim S, Kimchi A, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2020;48:D9–d16. <https://doi.org/10.1093/nar/gkz899>.
67. Wu L, Sun Q, Desmeth P, Sugawara H, Xu Z, McCluskey K, Smith D, Alexander V, Lima N, Ohkuma M, et al. World data centre for microorganisms: an information infrastructure to explore and utilize preserved microbial strains worldwide. *Nucleic Acids Res.* 2017;45:D611–d618. <https://doi.org/10.1093/nar/gkw903>.
68. Huerta-Cepas J, Serra F, Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol.* 2016;33:1635–8. <https://doi.org/10.1093/molbev/msw046>.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

