Synthetic and Systems Biotechnology 8 (2023) 647–653

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: www.keaipublishing.com/en/journals/synthetic-and-systems-biotechnology

Original Research Article

DSEMR: A database for special environment microorganisms resource and associating them with synthetic biological parts

Yuzhou Wang [a], Jinyi Qian [a], Fang Yan [a], Yuetong Wang [a], Tianqiong Shi [a,**], Zhidong Zhang [b,***], Chao Ye [a,*], He Huang [a,****]

[a] School of Food Science and Pharmaceutical Engineering, Nanjing Normal University, Nanjing, 210023, China
[b] Institute of Microbiology, Xinjiang Academy of Agricultural Sciences, Urumqi, 830091, China

ARTICLE INFO

Keywords:
Special environment microorganism
Physiological functions
Synthetic biological parts

ABSTRACT

Special environmental microorganisms are considered to be of great industrial application value because of their special genotypes, physiological functions and metabolites. The research and development of special environmental microorganisms will certainly bring about some innovations in biotechnology processes and change the face of bioengineering. The Special Environmental Microbial Database (DSEMR) is a comprehensive database that provides information on special environmental microbial resources and correlates them with synthetic biological parts. DSEMR aggregates information on specific environmental microbial genomes, physiological properties, culture media, biological parts, and metabolic pathways, and provides online tool analysis data, including 5268 strains from 620 genera, 31 media, and 42,126 biological parts. In short, DSEMR will become an important resource for the study of microorganisms in special environments and actively promote the development of synthetic biology.

1. Introduction

Special environment microorganisms thrive in environments where one or more physical or chemical parameters deviate from the typical range suitable for most organisms. As a result, they develop unique physiological characteristics, such as heat resistance, cold resistance, acid resistance, alkali resistance, salt resistance, radiation resistance, and drought resistance, among others [1–7]. Special living conditions give rise to unique genetic backgrounds and metabolic pathways in special environment microorganisms, enabling them to produce distinct functional enzymes and active substances [8]. For instance, *Herbinix hemicellulosilytica*, a thermophilic bacterium, has demonstrated the capability to produce six thermophilic xylanases [9]. These enzymes have significant potential in high-temperature processing applications involving cellulose and wood products. The acid-stabilized α-amylase derived from *acicyclobacillus acidocaldarius* can be adapted to the low pH environment in the starch industry, enabling low-cost production [10]. In recent years, significant advancements have been made in the field of next-generation industrial biotechnology (NGIB) by leveraging the development and optimization of synthetic biology and fermentation technologies. This progress has been particularly evident in the utilization of special environment microorganisms for industrial fermentation processes [11]. Notably, halophilic microorganisms, such as *Halomonas*, have demonstrated the ability to conduct open, continuous fermentation and achieve large-scale production of bioplastic polyhydroxyalkanoate (PHA) and other diverse products in high salt and alkaline environments [7]. Special environment microorganisms have increasingly garnered attention in industrial biotechnology research due to their remarkable tolerance to harsh conditions. They serve as promising platforms for the advancement of next-generation industrial biotechnology.

Currently, there is a scarcity of reports on the database of special environmental microorganisms. This can be attributed to the widespread distribution of microorganisms in such environments, coupled with the challenging sampling conditions they pose. Isolating strains of

Peer review under responsibility of KeAi Communications Co., Ltd.
* Corresponding authors.
** Corresponding author.
*** Corresponding authors.
**** Corresponding author.
E-mail addresses: tqshi@njnu.edu.cn (T. Shi), zhangzheedong@xaas.ac.cn (Z. Zhang), chaoye09@njnu.edu.cn (C. Ye), huangh@njnu.edu.cn (H. Huang).

https://doi.org/10.1016/j.synbio.2023.09.006
Received 28 June 2023; Received in revised form 11 September 2023; Accepted 20 September 2023
Available online 2 October 2023

special environmental microorganisms and establishing their cultivation conditions can be a challenging task, requiring specialized techniques and conditions. Consequently, information pertaining to microorganisms in special environments is widely dispersed, and there exists a diverse array of experimental data within the literature. This situation presents a challenge for researchers, as it can be time-consuming and arduous to gather comprehensive information about special environment microorganisms. For instance, a quick search using the keyword 'Special environment microorganism' in PubMed yields approximately 6667 results, while a similar search in the Web of Science produces around 3931 results. These numbers indicate the vast amount of research available on this topic. Moreover, researchers often need to navigate through multiple databases to extract various types of information. For example, they may consult NCBI Genome and GOLD for genome information, KEGG for pathway information, CDD for protein sequences, RCSB-PDB for structural information, and BRENDA for enzyme-related information [12–17]. Each database serves a specific purpose and provides valuable insights for studying special environment microorganisms. Therefore, systematic organization and integration of data, along with user-friendly web interfaces, can greatly assist researchers in analyzing the data according to their specific requirements. Such efforts facilitate efficient exploration and extraction of relevant information from diverse sources, ultimately enhancing research on special environment microorganisms.

To overcome these limitations, we have developed DSEMR (www.dsemr.cn), the first comprehensive database dedicated to special environment microorganisms. DSEMR integrates information on discovered special environment microorganisms, including their biological characteristics, colony morphology, culture medium, 16S rDNA sequences, and synthetic biological parts. The website offers advanced search capabilities, online functions, and analytical tools to actively facilitate and advance research on special environment microorganisms.

2. Data collection and processing

2.1. Data collection and curation process

Our data collection process was meticulously planned and executed through multiple steps. Initially, in the stage of sourcing and collection of strains, we selectively targeted diverse specialized environments such as nuclear radiation zones, high-altitude regions, and arid zones to ensure the strains possessed qualities like radiation resistance, cold tolerance, and drought resistance. For drought-resistant strains, we collected samples in regions with annual precipitation below 50 mm. For radiation-resistant strains, our collection focused on the nuclear radiation areas in Xinjiang. And for cold-resistant strains, we gathered samples from areas below 3000 m in altitude with substantial snow cover in the winter, with an average temperature of −8 °C.

Following this, during the isolation and identification phase, we extracted total DNA from the samples. We designed primers targeting conserved regions and added sequencing adapters to them [18,19]. Subsequent PCR amplification was conducted, followed by purification, quantification, and normalization to create sequencing libraries [20,21]. These libraries were subjected to quality checks before being sequenced using Illumina NovaSeq 6000 [22]. The raw image data from high-throughput sequencing was processed through base calling analysis to yield raw sequencing sequences (Sequenced Reads), which were stored in FASTQ file format. After purification and identification, a total of 5591 16S rDNA sequences were collected from the strains.

Subsequently, we performed Blast comparisons in NCBI using specific parameters (outfmt-6, evalue-1e^{-30}, max target seqs-10) to further validate the classification information of these strains. Due to limitations in researching specialized environmental microorganisms, we acquired the full genomic information of only a portion of the strains.

To facilitate subsequent research, we established appropriate culture mediums to support the growth and reproduction of the selected strains. To acquire synthetic biology component information, we conducted Blast comparisons on the collected strains' full genomic information in NCBI and extracted corresponding protein information from the Uniprot website [23,24]. This work successfully obtained 366656 pieces of biological part information from 81 strains of bacteria, covering 42,126 different proteins.

Finally, we compiled the organized information into CSV files and uploaded them to a MySQL database for effective management and utilization of these invaluable research outcomes. The entire process was carried out meticulously, following rigorous steps and detailed operations, providing a robust foundational dataset for our in-depth exploration of specialized environmental microorganisms.

2.2. Data processing

DSEMR encompasses a collection of 5268 strains belonging to 620 genera. Among them, there are 772 radiation-resistant strains, 331 drought-resistant strains, and 4165 cold-resistant strains. The top ten most abundant strains are *Streptomyces*, *Bacillus*, unclassified *Gemmatimonadaceae*, unclassified *Vicinamibacterales*, unclassified *Gaiellales*, *Gemmatimonas*, unclassified *Elsterales*, *Acidothermus*, *Candidatus Solibacter*, and unclassified *Xanthobacteraceae*. The respective counts for these strains are 336, 242, 205, 202, 145, 109, 88, 86, 83, and 68 (Fig. 1A). Taxonomic analysis revealed 77 genera for radiation-tolerant strains, 48 genera for drought-tolerant strains, and 533 genera for cold-tolerant strains (Fig. 1B). There are notable differences in strain abundance among these three categories at the genus level. Comparing the top ten abundant strains within each category, we found low repetition of strain information at the genus level. For radiation-tolerant strains, the top three genera are *Bacillus* (229 strains), *Streptomyces* (163 strains), and *Actinomyces* (23 strains). Among drought-resistant strains, *Streptomyces* (163 strains), *Pseudonocardia* (29 strains), and *Nocardiopsis* (20 strains) rank as the top three. Within the cold-resistant strains, the top three genera are unclassified *Gemmatimonadaceae* (205 strains), unclassified *Vicinamibacterales* (202 strains), and unclassified *Gaiellales* (145 strains) (Fig. 1C).

3. Database implementation

DSEMR (http://www.dsemr.cn/) is built on the framework of LAMP (Linux + Apache + MySQL + PHP). LAMP is a group of Web application platforms commonly used to build dynamic websites or servers. MySQL (https://www.mysql.com/), an open source data management system, is used to plan metadata storage and access. HTML and PHP (https://www.php.net/) are used to code the website front end. Python (https://www.python.org/) and PhyML (http://www.atgc-montpellier.fr/phyml) are used to code the website back end. The website is hosted on an Apache server (https://www.apache.org/). The Apache server and MySQL database are integrated in the site construction software XAMPP (https://www.apachefriends.org/) (Table 1).

The web interfaces of the DSEMR database have undergone extensive testing on various browsers, including Google Chrome, Internet Explorer, Mozilla Firefox, Opera, and Safari on both Windows and Linux platforms. While there may be slight differences in appearance, all tools and features have performed reliably across all tested browsers and platforms.

4. Database objectives and functional features

DSEMR offers three primary types of content: (1) Information on special environmental microbes, (2) Information on special environmental microbes associated with synthetic biology, and (3) Online tools. The data on 'Special Environment Microbial Strain Information' serves to investigate the physiological and biochemical characteristics of these microbes, establishing a foundation for their large-scale application and facilitating the transition of special environment microorganisms from

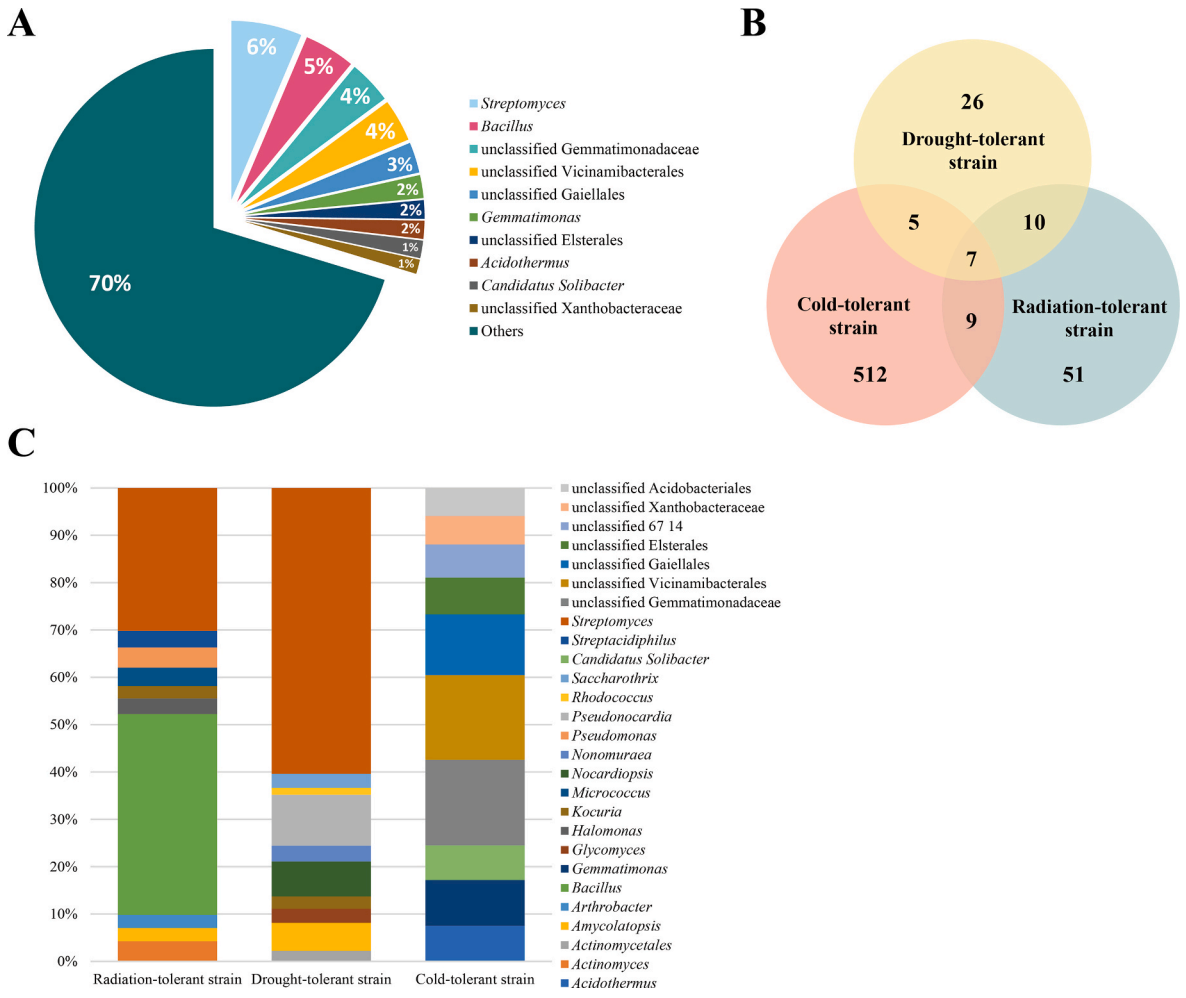


Fig. 1. Distribution analysis of strain species in DSEMR. To make the view best, only the species with the top ten abundance levels are shown, and Others are grouped into others for display in the figure. Unclassified represents species that have not received taxonomic notes. (A) Pie chart of species distribution in DSEMR: color block area represented the proportion of species in relative abundance; (B) Venn diagram drawn according to different characteristics: the overlapping part of the figures in different colors is the number of strains shared between the two characteristics. (C) Histogram of species distribution of strains with different characteristics: color block length represents the proportion of species in relative abundance.

Table 1
The specific versions and download addresses of the relevant software.

Software	Version number	Download link
Linux	CentOS 5.8	https://www.microsoft.com/zh-cn/
XAMPP	3.3.0	https://www.xampps.com/
Apache	2.4.48	https://www.apache.org/
MySQL	8.0.0	https://www.mysql.com/
PHP	8.0.8	https://www.php.net/
phpMyAdmin	5.1.1	http://www.phpmyadmin.net/
Python	3.10.5	https://www.python.org/
Clustal W	2.1	http://www.clustal.org/download/current/
PhyML	3.1	http://www.atgc-montpellier.fr/phyml

unculturable to pure culture, thereby exploring new biological resources. The purpose of providing data on 'Special environmental microbes associated with synthetic biology' is to uncover information about biological elements present in microbial genes.

4.1. Specific environmental microbial information data covers biological characteristics, culture medium, genome information

To present 'special environmental microbial information', DSEMR incorporates two functional modules: 'Strains' and 'Culture'. In the 'Strains' module, users can browse and retrieve essential information about all the microbes included in DSEMR. We have established three search keywords for this module: 'Strain NO', 'Classification', 'Strain type' and 'Biological characteristics'. In this database, 'Strain type' refers to the different adaptability of microorganisms, such as cold resistance, radiation resistance and drought resistance. 'Biological characteristics' refers to the radiation tolerance of the radiation resistant strains in the database. The biological characteristics of the drought- and cold-tolerant strains are still being determined.

By utilizing these search options, users can access relevant information and retrieve corresponding microbial details. For instance, selecting 'Search For Classification' and inputting '*Brevundimonas*' as the keyword will yield 5 records related to *Brevundimonas* (Fig. 2). These records encompass biometric signatures, individual sources, sample descriptions, sample locations, sampling dates, and 16S rDNA sequences. By clicking on 'Details' below the '16S rDNA sequence', users can obtain comprehensive information about the microbial 16S rDNA sequence, along with morphological photos and descriptions of the strain. As research on microorganisms in special environments is limited, only the whole genome information of a few microorganisms has been collected. By clicking on the 16S rDNA sequence of a microorganism, the webpage will redirect to the NCBI database to access specific information.

Within the 'Culture Medium' module, users can explore and access the basic information of all microbial media in DSEMR, comprising a

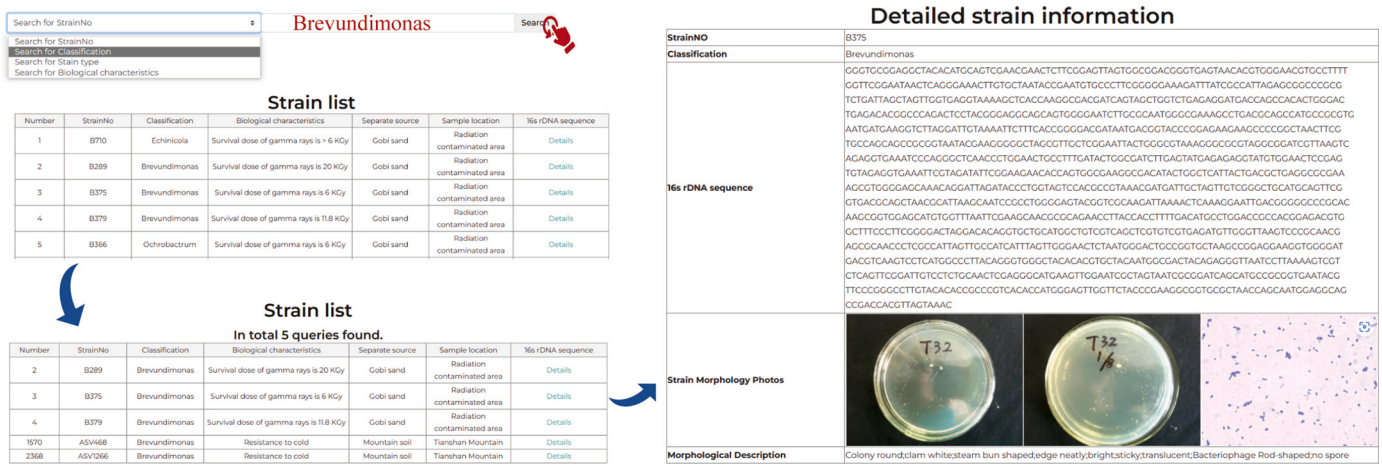


Fig. 2. Example of the use of the 'Strains' module.

total of 31 different media types. We have established two search keywords for this module: 'Strain NO' and 'Classification'. By utilizing these search options, users can retrieve relevant information regarding the composition and preparation method of microbial media, which serves as a valuable foundation for subsequent experiments. For instance, by selecting 'Search For Classification' and entering the keyword '*Ochrobactrum*', users can retrieve 4 records pertaining to *Ochrobactrum* (Fig. 3). These records provide information about the isolation medium and culture conditions associated with *Ochrobactrum* strains.

4.2. *Specific environmental microorganisms are associated with synthetic biology to find bacteria with potential for industrial application*

To establish a connection between specific environmental microorganisms and synthetic biology, DSEMR offers a 'Biological part' module that consolidates information about the biological components found within the genes of the strains. This information includes gene names, EC numbers, metabolic pathways, and more. EC numbers are a set of classification systems developed by the Enzyme Commission, which categorize enzymes based on the specific chemical reactions they catalyze. Each EC number has a specific meaning: EC1 represents oxidoreductases, EC2 represents transferases, EC3 denotes hydrolases, EC4 represents lyases, EC5 indicates isomerases, EC6 stands for ligases, and EC7 refers to translocases. In total, DSEMR contains 42,126 pieces of information regarding biological components, with 16,113 of them lacking EC numbers. The distribution of EC numbers from EC1 to EC7 is displayed in the Figure (note that some biological parts may have multiple EC numbers associated with them) (Fig. 4).

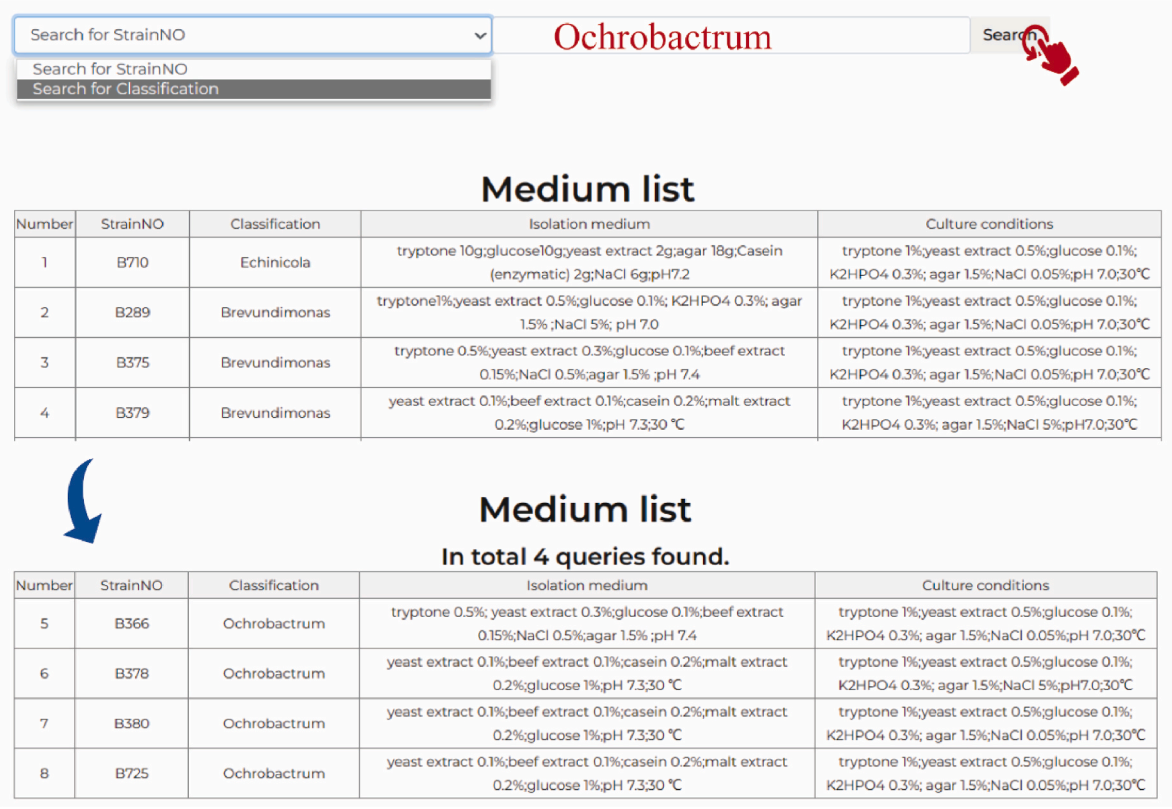


Fig. 3. Example of the use of the 'Culture' module.

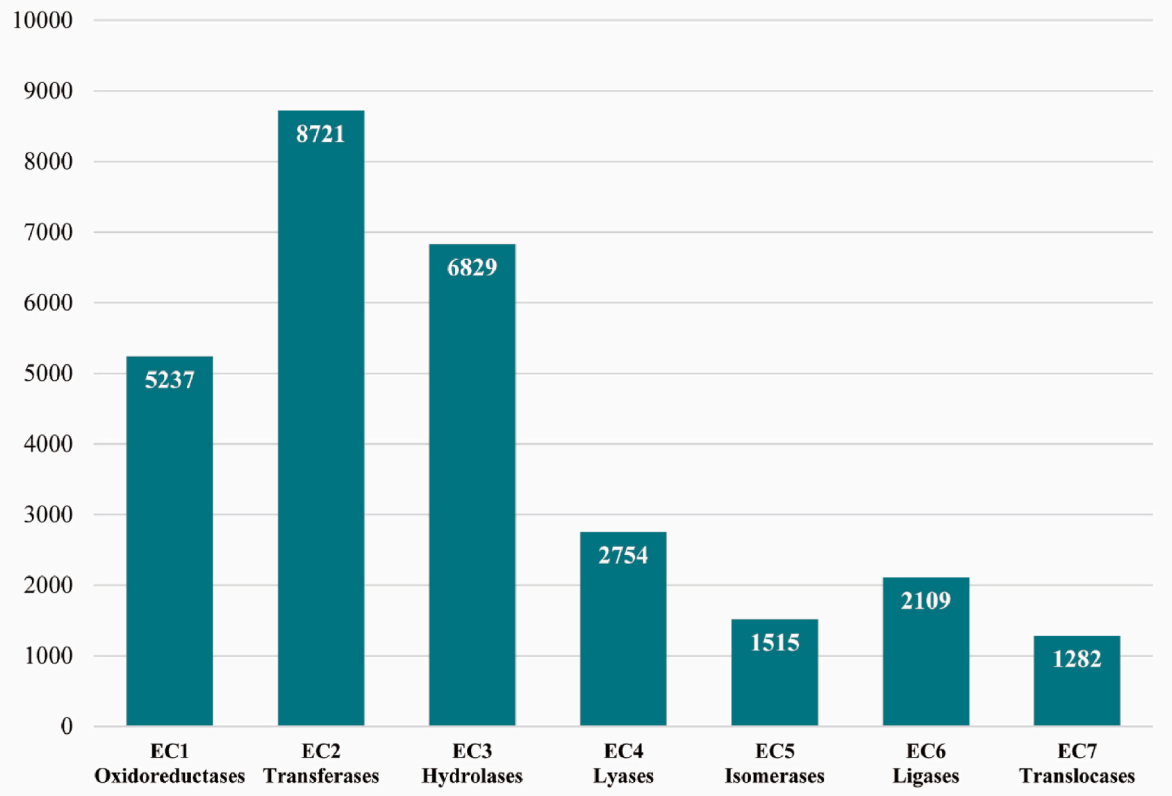


Fig. 4. Comparison of number of EC1-EC7 biological parts in DSEMR.

The database contains a total of 42,126 different biological parts, which encompass 623 metabolic pathways. Metabolism refers to the chemical processes occurring within living organisms, encompassing both anabolism and catabolism. For instance, the metabolic pathway of methionine synthase (EC 2.1.1.13) from *Arthrobacter* sp. PGP41 is associated with amino-acid biosynthesis. The metabolic pathway of short-chain dehydrogenase/reductase prx1 (EC 1.1.99.-) from *Microbacterium oxydans* is involved in sesquiterpene biosynthesis. Additionally, the metabolic pathway of Beta-ketoacyl-[acyl-carrier-protein] synthase III (beta-ketoacyl-ACP synthase III) (KAS III) (EC 2.3.1.180) from *Streptomyces* sp. SGAir0924 is related to lipid metabolism. These examples illustrate the diverse roles and functions of biological parts within specific metabolic pathways. Through the analysis of metabolic pathway enzymes, researchers can gain insights into the molecular basis and regulatory mechanisms underlying specific metabolic functions in important industrial microorganisms. Additionally, they can uncover the stress resistance and adaptation mechanisms of these microorganisms to the industrial environment. By utilizing metabolic engineering techniques, it becomes possible to transform and optimize the physiological and metabolic functions of these microorganisms. This optimization can lead to significant improvements in the production levels of specific metabolites, thereby promoting the advancement of cost-effective biological manufacturing.

Within the 'Biological part' module of DSEMR, users can conduct searches using three keywords: Organism, EC number, and Pathway. For example, by selecting Organism search option and entering '*Arthrobacter* sp. PGP41', users can retrieve 3758 records of biological parts associated with this strain (Fig. 5).

4.3. Online tools to map evolutionary trees and analyze phylogenetic relationships of strains

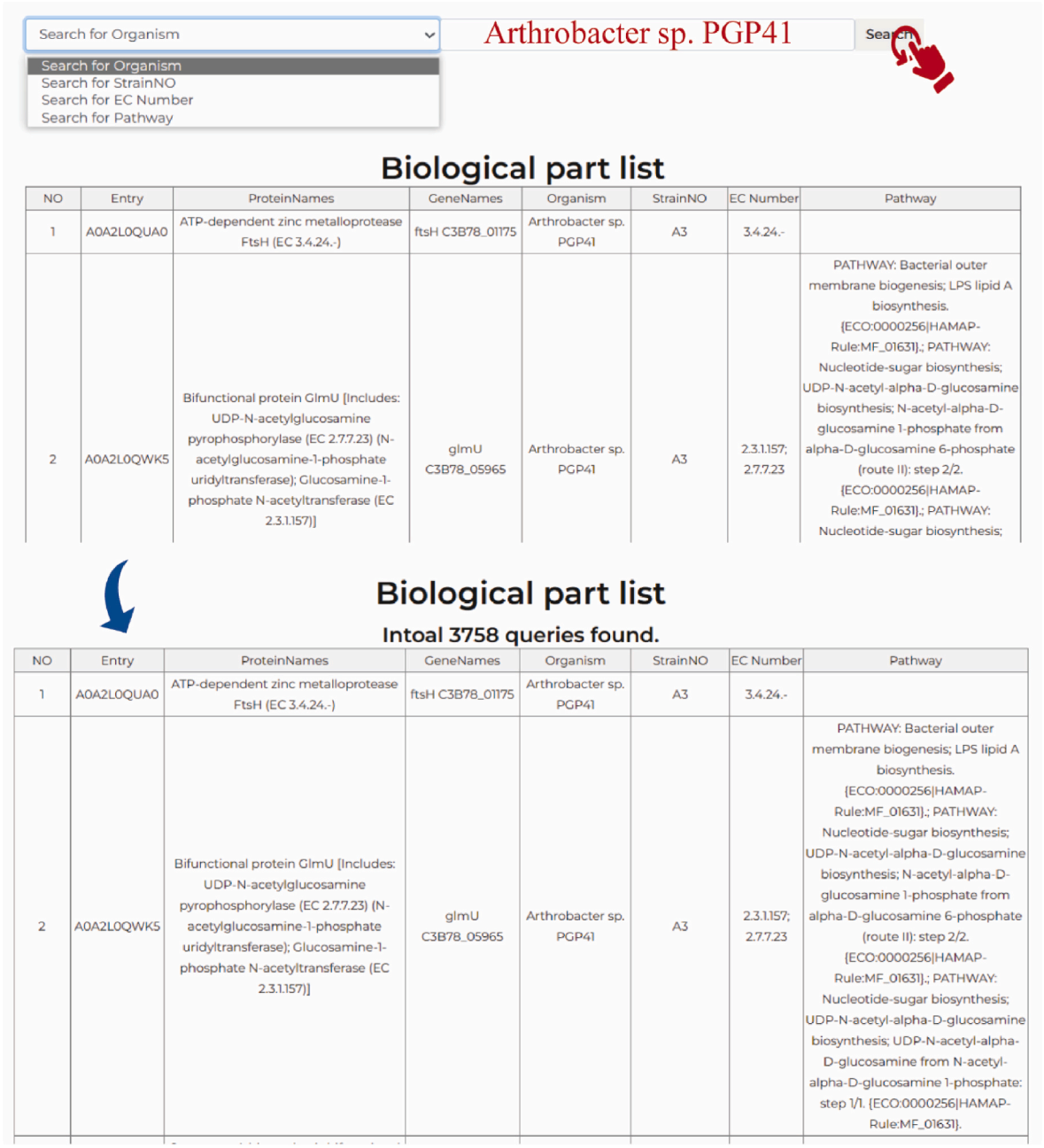
In phylogenetic analysis, the construction of evolutionary tree is an essential link. Evolutionary tree can use tree-like branch graph to represent the relationship between species or genes. An evolutionary tree usually contains nodes, clades, outgroups, and branching lengths. Each node represents a taxon, usually genus, species, and so on. Clade represents an evolutionary relationship consisting of two or more organisms or sequences. Outgroups are biological sequences that are related to the analytical sequence, but are more distantly related. Generally, the length of evolutionary clades is marked on the branching line, representing the degree of clade change. The shorter it is, the smaller the difference is and the closer the evolutionary distance is.

The 'Evolution Tree' module within DSEMR is a Python-based program that we have developed. Initially, we employ Clustalw and Phyml tools for multiple sequence alignment and the subsequent creation of phylogenetic trees. The Phylo module from Biopython aids us in interpreting and visualizing these evolutionary trees. Specifically, the 'Phylo.read' function is employed to parse the evolutionary tree data, followed by the use of the 'Phylo.draw' function to visually render the tree. Ultimately, the 'plt.show' function is employed to showcase the phylogenetic tree graph, which effectively illustrates the evolutionary relationships and branch structures among various samples.

The process of drawing an evolutionary tree online involves four steps: (1) Searching for reference sequences, (2) Selecting up to 10 reference sequences, (3) Uploading the 16s rDNA sequence of the target microorganism (in FASTA format), and (4) Optionally providing an email address to receive the generated evolution tree. Once the task is submitted to the DSEMR database, the evolution tree will automatically appear. If an email address is provided, the result of the evolution tree will be sent to the corresponding email, and users can view and download the results using the provided link in the email (Fig. 6).

5. Future directions

In the coming years, we plan to expand the content of DSEMR and gradually display more types of special environmental microorganisms, such as acid-resistant, alkali-resistant, and salt-resistant strains.



NO	Entry	ProteinNames	GeneNames	Organism	StrainNO	EC Number	Pathway
1	A0A2L0QUA0	ATP-dependent zinc metalloprotease FtsH (EC 3.4.24.-)	ftsH C3B78_01175	Arthrobacter sp. PGP41	A3	3.4.24.-	
2	A0A2L0QWK5	Bifunctional protein GlmU [Includes: UDP-N-acetylglucosamine pyrophosphorylase (EC 2.7.7.23) (N-acetylglucosamine-1-phosphate uridyltransferase); Glucosamine-1-phosphate N-acetyltransferase (EC 2.3.1.157)]	glmU C3B78_05965	Arthrobacter sp. PGP41	A3	2.3.1.157; 2.7.7.23	PATHWAY: Bacterial outer membrane biogenesis; LPS lipid A biosynthesis. {ECO:0000256|HAMAP-Rule:MF_01631}.; PATHWAY: Nucleotide-sugar biosynthesis; UDP-N-acetyl-alpha-D-glucosamine biosynthesis; N-acetyl-alpha-D-glucosamine 1-phosphate from alpha-D-glucosamine 6-phosphate (route II): step 2/2. {ECO:0000256|HAMAP-Rule:MF_01631}.; PATHWAY: Nucleotide-sugar biosynthesis;

Fig. 5. Example of the use of the 'Biological part' module.

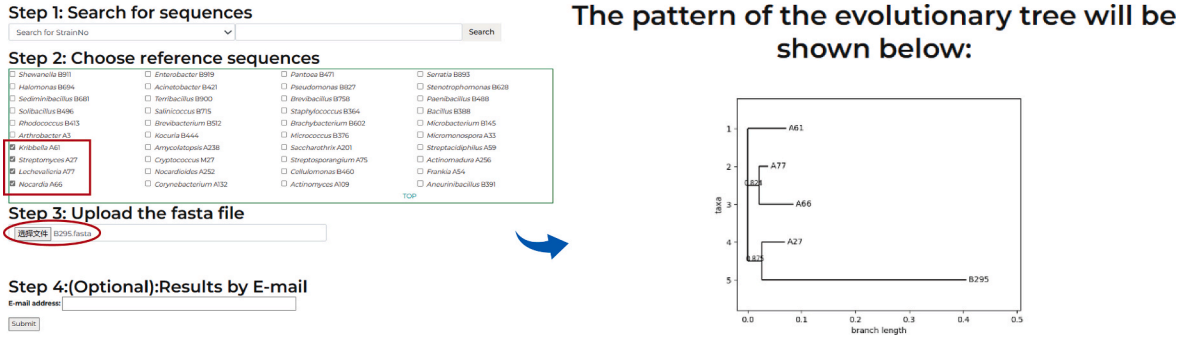


Fig. 6. Example of the use of the 'Evolution Tree' module.

Additionally, we have plans to incorporate new elements into DSEMR, such as 18S rRNA, mRNA, genomic DNA, genomic functional signatures, and metabolic pathway profiles of the collected samples. Moreover, we aim to introduce new features to the platform, including but not limited to biological part searches, functional gene abundance prediction, and rapid identification and classification of CRISPR repeats. These enhancements will enable functional analysis of special environment microorganisms, further advancing the study of these organisms and their applications in industrial biotechnology.

6. Conclusions

DSEMR is a comprehensive and curated database that focuses on special environmental microbes and their relevance to synthetic biology. Currently, the database contains detailed information on 5268 strains from 620 genera, 31 media, and 42,126 biological parts. DSEMR not only provides comprehensive data on special environmental microorganisms but also offers advanced gene analysis functions, thereby establishing a solid foundation and facilitating effective gene analysis for the study of these unique microorganisms. The database also serves as a valuable resource for synthetic biology by collecting and sharing data on biological parts, supporting research and application in this field. The availability of rich biological parts information in the database greatly contributes to the advancement of synthetic biology.

Funding

This work was supported by the National Natural Science Foundation of China (22208167, 32060004), Xinjiang Academy of Agricultural Sciences Science and technology innovation key cultivation project (xjkcpy-2021002), Tianshan Talent Plan (2022TSYCCX0067), Postgraduate Research & Practice Innovation Program of Jiangsu Province (181200003023337).

Declaration of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Yuzhou Wang: Writing – original draft, Conceptualization, Data curation. **Jinyi Qian:** Writing – original draft. **Fang Yan:** Writing – original draft. **Yuetong Wang:** Writing – review. **Tianqiong Shi:** Writing – review. **Zhidong Zhang:** Writing – review, Data collection. **Chao Ye:** Conceptualization, Resources, Supervision, Writing – original draft, Writing – review & editing. **He Huang:** Writing – review.

References

[1] Jiang Y, Jiang W, Xin F, et al. Thermophiles: potential chassis for lignocellulosic biorefinery. Trends Biotechnol 2022;40:643–6.
[2] Jin S, Wang Y, Zhao X. Cold-adaptive mechanism of psychrophilic bacteria in food and its application. Microb Pathog 2022;169:105652.
[3] Razia S, Hadibarata T, Lau SY. Acidophilic microorganisms in remediation of contaminants present in extremely acidic conditions. Bioproc Biosyst Eng 2023;46:341–58.
[4] Mamo G, Mattiasson B. Alkaliphiles: the versatile tools in biotechnology. In: Mamo G, Mattiasson B, editors. Alkaliphiles in biotechnology. Cham: Springer International Publishing; 2020. p. 1–51.
[5] Alpha-Bazin B, Gorlas A, Lagorce A, et al. Lysine-specific acetylated proteome from the archaeon Thermococcus gammatolerans reveals the presence of acetylated histones. J Proteonomics 2021;232:104044.
[6] Fardelli E, D'Arco A, Lupi S, et al. Spectroscopic evidence of the radioresistance of Chroococcidiopsis biosignatures: a combined Raman, FT-IR and THz-TDs spectroscopy study. Spectrochim Acta Mol Biomol Spectrosc 2023;288:122148.
[7] Ye J-W, Chen G-Q. Halomonas as a chassis. Essays Biochem 2021;65:393–403.
[8] Mohammadipanah F, Wink J. Actinobacteria from arid and desert habitats: diversity and biological activity. Front Microbiol 2016;6.
[9] Mechelke M, Koeck DE, Broeker J, et al. Characterization of the arabinoxylan-degrading machinery of the thermophilic bacterium Herbinix hemicellulosilytica—six new xylanases, three arabinofuranosidases and one xylosidase. J Biotechnol 2017;257:122–30.
[10] Shah AR, Shah RK, Madamwar D. Improvement of the quality of whole wheat bread by supplementation of xylanase from Aspergillus foetidus. Bioresour Technol 2006;97:2047–53.
[11] Chen X, Yu L, Qiao G, et al. Reprogramming Halomonas for industrial production of chemicals. J Ind Microbiol Biotechnol 2018;45:545–54.
[12] Rangwala SH, Kuznetsov A, Ananiev V, et al. Accessing NCBI data using the NCBI sequence viewer and genome data viewer (GDV). Genome Res 2021;31:159–69.
[13] Bernal A, Ear U, Kyrpides N. Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. Nucleic Acids Res 2001;29:126–7.
[14] Kanehisa M, Araki M, Goto S, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res 2008;36:D480–4.
[15] Marchler-Bauer A, Anderson JB, Cherukuri PF, et al. CDD: a conserved domain database for protein classification. Nucleic Acids Res 2005;33:D192–6.
[16] Burley SK, Bhikadiya C, Bi C, et al. RCSB Protein Data bank: tools for visualizing and understanding biological macromolecules in 3D. Protein Sci 2022;31:e4482.
[17] Schomburg I, Chang A, Hofmann O, et al. BRENDA: a resource for enzyme data and metabolic information. Trends Biochem Sci 2002;27:54–6.
[18] Clarridge JE. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev 2004;17:840–62.
[19] Srinivasan R, Karaoz U, Volegova M, et al. Use of 16S rRNA gene for identification of a broad range of clinically relevant bacterial pathogens. PLoS One 2015;10.
[20] Drancourt M, Berger P, Raoult D. Systematic 16S rRNA gene sequencing of atypical clinical isolates identified 27 new bacterial species associated with humans. J Clin Microbiol 2004;42:2197–202.
[21] Kolbert CP, Persing DH. Ribosomal DNA sequencing as a tool for identification of bacterial pathogens. Curr Opin Microbiol 1999;2:299–305.
[22] Modi A, Vai S, Caramelli D, et al. The Illumina sequencing protocol and the NovaSeq 6000 system. In: Mengoni A, Bacci G, Fondi M, editors. Bacterial pangenomics: methods and protocols. New York, NY: Springer US; 2021. p. 15–42.
[23] Ye J, McGinnis S, Madden TL. BLAST: improvements for better sequence analysis. Nucleic Acids Res 2006;34:W6–9.
[24] The UniProt C. Reorganizing the protein space at the universal protein resource (UniProt). Nucleic Acids Res 2012;40:D71–5.